Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Belief Adjustment: A Double Hurdle Model and Experimental Evidence

---

CAMA Working Paper 1/2018
January 2018

Timo Henckel
Centre for Applied Macroeconomic Analysis, ANU

Gordon D. Menzies
University of Technology Sydney and
Centre for Applied Macroeconomic Analysis, ANU

Peter G. Moffatt
CBESS & School of Economics
University of East Anglia

Daniel J. Zizzo
BENC
Newcastle University Business School and
Centre for Applied Macroeconomic Analysis, ANU

Abstract

We present an experiment where subjects sequentially receive signals about the true state of the world and need to form beliefs about which one is true, with payoffs related to reported beliefs. We control for risk aversion using the Offerman et al. (2009) technique. Against the baseline of Bayesian updating, we test for belief adjustment under-reaction and over-reaction and model the decision making process of the agent as a double hurdle model where agents first decide whether to adjust their beliefs and then, if so, decide by how much. We find evidence for periods of belief inertia interspersed with belief adjustment. This is due to a combination of: random belief adjustment; state-dependent belief adjustment, with many subjects requiring considerable evidence to change their beliefs; and Quasi-Bayesian belief adjustment, with insufficient belief adjustment when a belief change does occur.

Keywords

belief revision, double hurdle model, expectations, overreaction, under re-action

JEL Classification

C34, C91, D03, D84, E03

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Belief Adjustment: A Double Hurdle Model and Experimental Evidence*

Timo Henckel
Australian National University & CAMA

Gordon D. Menzies
University of Technology Sydney & CAMA

Peter G. Moffatt
CBESS & School of Economics, University of East Anglia

Daniel J. Zizzo†
BENC, CAMA & Newcastle University Business School

This Version: 5 January 2018

Abstract

We present an experiment where subjects sequentially receive signals about the true state of the world and need to form beliefs about which one is true, with payoffs related to reported beliefs. We control for risk aversion using the Offerman et al. (2009) technique. Against the baseline of Bayesian updating, we test for belief adjustment under-reaction and over-reaction and model the decision making process of the agent as a double hurdle model where agents first decide whether to adjust their beliefs and then, if so, decide by how much. We find evidence for periods of belief inertia interspersed with belief adjustment. This is due to a combination of: random belief adjustment; state-dependent belief adjustment, with many subjects requiring considerable evidence to change their beliefs; and Quasi-Bayesian belief adjustment, with insufficient belief adjustment when a belief change does occur.

Keywords: belief revision; double hurdle model; expectations; overreaction; underreaction.

JEL classification codes: C34; C91; D03; D84 ; E03.

---

*We acknowledge the support of the ESRC (Network for Integrated Behavioural Science, ES/K002201/1) and the ANU College of Business and Economics. We thank Rohan Alexander and Ailko van der Veen for research assistance, participants at the University of Technology Sydney seminar, the BENC/CEAR workshop on "Recent advances in modelling decision making under risk and uncertainty", Durham University, the Society for Experimental Finance workshop, University of Arizona, Tucson, and the 2017 BEAM Workshop, Kyoto, as well as Han Bleichrodt, Chris Carroll and John Rust for useful feedback. The usual disclaimer applies. The experimental instructions are in an online appendix.

†Daniel John Zizzo (corresponding author), Newcastle University Business School, Newcastle University, Newcastle upon Tyne, Tyne and Wear NE1 4SE, United Kingdom. E-mail address: daniel.zizzo@ncl.ac.uk. Phone: +44-191-20858544.

1 Introduction

Agents form and update their beliefs when they receive new information. The assumptions about how they do this are fundamental to a plethora of theoretical and empirical models, both in micro- and macroeconomics. In the presence of rational expectations, new information leads to smooth and continuous belief updating according to Bayes' rule. In reality many agents systematically misunderstand basic statistics, and complexity and inattention may contribute to deviations from Bayesian predictions (Rabin, 2013). Such violations of rational expectations have been studied in static settings, where all the information is presented at once to subjects who discount priors (Tversky and Kahneman, 1982; Camerer, 1987; El-Gamal and Grether, 1995). In this paper we study a dynamic setting in which new information arrives sequentially and consider the *frequency* as well as the *extent* of belief adjustment, referring to *sticky* belief adjustment when it is insufficient in either domain. Furthermore, we construct a double hurdle econometric model to combine in a single framework different types of belief adjustment we observe in the laboratory: time-dependent (random) belief adjustment and state-dependent (Bayesian, Quasi-Bayesian) belief adjustment. We control for risk aversion using Offerman et al.'s (2009) technique and we also consider how increased task complexity or scope for inattention affect our results.

There exists no widely accepted, fully developed alternative to rational expectations. In microeconomics, Quasi-Bayesian (QB) belief adjustment has been the preferred route to think about boundedly rational belief adjustment. Rabin (2013) distinguishes between warped Bayesian models which encapsulate a false model of how signals are generated, for example by ignoring the law of large numbers (Benjamin et al., 2015), and information-misreading Bayesian models that misinterpret signals as supporting agents' hypotheses, thus giving rise to confirmation bias (Rabin and Schrag, 1999) which underweights information. While various anomalies have been considered within this framework, one simple way of modeling QB adjustment is that the agent adjusts beliefs continuously in response to new information—in the sense that it takes place whenever there is new information—but this adjustment is either too big or too small (Massey and Wu, 2005; Ambuehl and Li, 2014). See also Schmalensee (1976).

In macroeconomics, too, there have been many attempts to model departures from rational expectations. For example, there is a large literature examining time-dependent versus state-dependent price adjustment, with mixed empirical findings (e.g., Costain and Nakov, 2011; Aucremann and Dhyne, 2005; Stahl, 2005; Dias et al., 2007; Klenow and Kryvtsov, 2008; Midrigan, 2010). State-dependence implies a dependence of belief adjustment on the economic state, which in turn depends on new information arriving. Time-dependence is often viewed stochastically (as for example in Caballero, 1989) and therefore yields random belief adjustment.[1] Insuf-

[1] We do not consider non-stochastic time-dependence as in Fischer (1977) and Taylor (1980).

ficient belief adjustment can be seen as a possible microfoundation of sticky price adjustment, for example as a result of inattention and observation costs (Alvarez et al., 2016), information costs (Abel et al., 2013), cognitive costs (Magnani et al., 2016) and the costly consultation of experts by inattentive agents (Carroll, 2003).

In this paper we analyze the results of an experiment using an econometric model based on the inferential expectations (IE) framework of Menzies and Zizzo (2009). This provides a stylized way of modeling infrequent belief adjustment, and can be modified to allow for partial or excessive belief adjustment. We assume subjects hold a belief until enough evidence has accumulated to pass a threshold of statistical significance, at which point beliefs are updated. Each subject is assumed to draw a test size α from a distribution, which we derive from the results of our econometric model. Furthermore, once the decision is made to update, subjects can under- or overreact to data according to Quasi-Bayesian updating. Sticky belief adjustment is then fully specified by an agent's α distribution and a Quasi-Bayesian adjustment parameter, both of which are generated by a double hurdle model.

One source of deviation from Bayesian updating could be task complexity (Caplin et al., 2011, Charness and Levin, 2005), providing our rationale for including a complexity treatment. For instance, within consumer markets, complexity has often been blamed for suboptimality of consumer choices (e.g., Joskow, 2008; Ofgem, 2011; Independent Commission on Banking, 2011); the evidence from consumer experiments is less clear but consistent with at least some effect of complexity on consumer choice (Kalayci and Potters, 2011; Sitzia and Zizzo, 2011; Sitzia et al., 2015).

Another possible source of deviation from Bayesian updating is inattention (Alvarez et al., 2016; Magnani et al., 2016; Carroll, 2003). We thus also include an inattention treatment, which consists of an alternative task being available to the main task. The setup is closest to Corgnet et al. (2014), who find an effect on team effort in a work experiment, and Sitzia and Zizzo (2015), who find an effect on consumption choices.

We are not aware of research on complexity and inattention that has identified their effect on belief updating with sequential information flow. In brief, our results are as follows. Subjects change their beliefs about half the time, which is consistent with random belief adjustment, but they also consider the amount of evidence available, which is consistent with state-dependent belief adjustment. When subjects do change beliefs, they do so by around 35 per cent of the full Bayesian update, which is consistent with our version of Quasi-Bayesian belief adjustment. There is substantial heterogeneity in our results and, importantly, the frequency and extent of belief adjustment are positively correlated: agents who update with low frequency do so by even less than 35 percent of the full Bayesian update. Our results are perhaps surprisingly robust to either the task complexity or inattention treatment, but

---

In macroeconomics these types of models have been largely superceded by stochastic models such as Calvo (1983).

we find evidence that inattention reduces the propensity to update and complexity reduces the extent of updating.

Section 2 presents our experimental design and treatments, section 3 our expectation models, Sections 4 and 5 our results. Section 5 provides a discussion and concludes.

2 Experimental Design and Treatments

Our experiment was fully computerized and run in the experimental laboratory of the University of East Anglia with $n = 245$ subjects who were separated by partitions. The experiment was divided in two parts, labelled the *practice part* and the *main part*. Experimental instructions were provided at the beginning of each part for the tasks in that part (see the online appendix for a copy of the instructions). A questionnaire was administered to ensure understanding after each batch of instructions.

Main part of the experiment. After the practice part described below, in the main part of the experiment subjects played 7 stages, each with 8 rounds, thus generating $T = 56$ observations. At the beginning of each stage the computer randomly chose one of two urns (Urn 1 or Urn 2), with Urn 1 being selected at a known probability of 0.6. Each urn represents a different state of the world. While this prior probability was known and it was known that the urn would remain the same throughout the stage, the chosen urn was not known to subjects. It was known that Urn 1 had seven white balls and three orange balls, and Urn 2 had three white balls and seven orange balls. At the beginning of each of the 8 rounds (round = t), there was a draw from the chosen urn (with replacement) and subjects were told the color of the drawn ball. These were therefore signals that could be used by subjects to update their beliefs. It was made clear to the subjects that the probability an urn was chosen in each of the seven stages was entirely independent of the choices of urns in previous stages.

Once they saw the draw for the round, subjects were asked to make a probability guess between 0% and 100%, on how likely it was that the chosen urn was Urn 1. The corresponding variable for analysis is their probability guess expressed as a proportion, denoted g. Once a round was completed, the following round started with a new ball draw, up to the end of the 8^{th} round.

Payment for the main part of the experiment was based on the guess made in a randomly chosen stage 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 penalize incorrect answers. The payoff for each subject was equal to 18 GBP minus 18 GBP $\times$ $(\textit{guess} - \textit{correct probability})^2$. Therefore, for the randomly chosen stage and round, subjects could earn between 0 and 18 GBP depending on the accuracy of their guesses.

Practice part of the experiment. The practice part was similar to the main part but simpler and therefore genuinely useful as practice. It was modelled after Offerman et al. (2009) to enable us to infer people's risk attitude, as detailed in section 3.

It consisted of 10 stages with one round each. In each stage a new urn was drawn (with probabilities 0.05, 0.1, 0.15, 0.2, 0.25, 0.75, 0.8, 0.85, 0.9, 0.95). Subjects were told the prior probability of Urn 1 being chosen but did not receive any futher information. In particular, no balls were drawn. The guessing task (single round) was to nominate a probability that Urn 1 was chosen; this is non-trivial because subjects should take account of the payoff structure rather than repeat the announced probabilities. Payment for the practice part of the experiment was based on the guess made in a randomly chosen round picked at the end of the experiment. A quadratic scoring rule was applied as in the main part, but this time this was equal to 3 GBP minus 3 GBP $\times$ $(\textit{guess} - \textit{correct probability})^2$.[2]

Experimental treatments. There were three treatments. The practice parts were identical across all treatments, and the main part of the *Baseline* treatment was as described.

In the main part (only) of the *Complexity* treatment, the information on the ball drawn from the chosen urn at the beginning of each round was presented as follows: it was presented as a statement about whether the sum of three numbers (of three digits each) is true or false. If true (e.g., $731 + 443 + 927 = 2101$), this meant that a white ball draw was drawn. If false (e.g., $731 + 443 + 927 = 2121$), this meant that an orange ball draw was drawn.

In the main part (only) of the *Inattention* treatment, subjects were given a non-incentivized alternative counting task which they could do instead of working on the probability. The counting exercise was a standard one from the real effort experimental literature (see Abeler et al., 2011, for an example) and consisted in counting the number of 1s in matrices of 0s and 1s. Subjects were told that they could do this exercise for as little or as long as they liked within 60 seconds for each round, and that we were not asking them in any way to engage in this exercise at all unless they wanted to.[3]

[2]This ensured similar marginal incentives for each round in the practice part (3 GBP prize picked up from 1 out of 10 rounds) and the main part (18 GBP prize picked up from 1 out of 56 rounds).

[3]They were also told that, if they did not make a guess in the guessing task within 60 seconds, they would automatically keep the guess from the previous round and move to the next round (or to the next stage). The length of 60 seconds was chosen based on piloting, in such a way that this would not be a binding constraint if subjects focused on the guessing task.

3 Model Variables, Risk and Expectation Process

3.1 Model Variables and Risk Attitude Correction

Table 1 lays out the main model variables and the interrelationships between them, when the event is described in terms of the chosen urn (row 1) and when it is described in terms of the probability of a white ball being drawn (row 2). The two descriptions are equivalent since the subject's subjective guess of the probability that Urn 1 was chosen generates an implied subjective probability that a white ball is drawn.[4] In our modelling, we sometimes use the former probability—that Urn 1 was chosen—and it will be useful to transform this probability guess using the inverse cumulative Normal distribution, so that the support has the same dimensionality as a classic z-statistic. Alternatively, we sometimes describe agents' guesses in terms of the probability that a white ball is drawn.

Time is measured by t, the draw (round) number for the ball draws in each stage. We define the value of t for which subjects last moved their guess (viz. updated their beliefs) to be m (for 'last Move'). Thus, for any sequence of ball draws at time t, the time that has elapsed since the last change in the guess is always $t - m$.

Event	Estimator of event probability	Estimator symbol	P(Event) Subject guesses	Transformed guess	Strength of Evidence z_t against earlier choice at m
Urn 1 drawn	Bayes rule	P_t	g_t : optimal guess (observed) g_t^* : inferred guess	$r_t^* = \Phi^{-1}(g_t^*)$	$z_t = \frac{\Phi^{-1}(P_t)-(\Phi_m)}{\sqrt{t}}$
White drawn	Prop. white balls out of t	P_t^w	Not guessed	Not guessed	$z_t \approx \frac{1}{\sqrt{3}}\left(\frac{P_t^w - P_m^w}{\sqrt{0.5^2/t}}\right)$

Table 1: Model variables

Along the top row the theoretical estimator for the probability that Urn 1 was drawn is provided by Bayes rule, which we denote by P_t after t ball draws. Many subjects do not use Bayes rule when they are guessing the probability that Urn 1 is chosen, though some guesses are closer to it than others.

As derived in Offerman et al. (2009), the elicited guess g_t in the fourth column is the result of maximizing utility based on a Constant Relative Risk Aversion (CRRA) utility function, $U\{\text{Payoff}\}$:

$$U\{\text{Payoff}\} = \frac{\text{Payoff}^{1-\theta} - 1}{1-\theta},$$

[4] For example, at the start of the experiment, before any ball is drawn, subjects know that the chance that Urn 1 was drawn is 0.6. It therefore follows that the chance of a white ball being drawn for the very first time is 0.7*0.6+0.3*(1-0.6) = 0.54.

and a true guess g_t^* in $E\left[U\{\text{Payoff}\}\right] = g_t^* U\{1-(1-g_t)^2\} + (1-g_t^*)U\{1-g_t^2\}$ where the payoffs for Urn 1 and Urn 2 are proportional to $1-(1-g_t)^2$ and $1-g_t^2$, according to the quadratic scoring rule, as explained in section 2. Expected utility is assumed to be maximized with respect to g_t and yields the following relationship between g_t^* and g_t:

$$\ln\left(\frac{g_t^*\left(1-g_t\right)}{g_t\left(1-g_t^*\right)}\right) = \theta \ln\left(\frac{g_t\left(2-g_t\right)}{\left(1+g_t\right)\left(1-g_t\right)}\right). \tag{1}$$

In the practice part the prior probabilities given to the subjects (by way of reminder, 0.05, 0.1, 0.15, 0.2, 0.25, 0.75, 0.8, 0.85, 0.9, 0.95 for 10 separate stages/rounds) are in fact the correct probabilities P_t. We see no reason not to credit subjects with realizing this, and they possess no other information anyway, so we define their true guess to be $g_t^* = \Phi^{-1}\left(P_t\right)$. Offerman et al. (2009) then interpret the deviations of g_t from g_t^* as being due to the subjects' risk preferences, and so do we. We use the ten datapoints (g_t^*, g_t) for each subject to estimate θ in a version of (1) appended with a regression error.[5] Armed with a subject-specific value of θ from the practice part, all the observable g_t values in the main experiment can be transformed to a set of inferred g_t^*. This transformation is accomplished by exponentiating both sides of (1), and solving for g_t^*. By taking the inverse cumulative Normal function, Φ^{-1}, of g_t^* we move it outside the $[0, 1]$ interval and give it the same dimensionality as a test statistic, namely $(-\infty, \infty)$. The variable in the penultimate column, $r_t^* = \Phi^{-1}\left(g_t^*\right)$, thus becomes the basis for all subsequent analysis.

In the final column of Table 1, we provide a measure z_t of the strength of evidence against the probability guess at the time of the last change. Agents change their guesses from time to time, and z_t tells us if the value of P at the last change, denoted P_m, seems mistaken in the light of subsequent evidence.

As shown in Appendix A, z_t is a good approximation for the standard test statistic for a proportion, using the maximal value of the variance of the sampling distribution (namely $(\frac{1}{2})^2$):

$$z_t \approx \frac{P_t^w - P_m^w}{(0.5^2/t)^{1/2}}. \tag{2}$$

When we analyze the inferential expectations of each agent we use (2) to recover from subject behaviour the entire distribution of the test size as a key component in our description of sticky belief adjustment.

3.2 Expectation Processes

Using the notation of Table 1, we define three processes of expectation formation that will be relevant for our double hurdle model in section 4.

[5] If an agent ever declared zero or unity in this preliminary stage, a regression version of (1) cannot be run, and so we set $\theta = -1$ in those cases.

Rational Expectations

The rational expectations (RE) solution predicts straightforward Bayesian updating. The (conditional) probability that the subject is being asked to guess is the rational expectation (RE) which is given by P_t. Calling $P_{initial}$ the initial prior probability and noting that the number of white balls is tP_t^w we can write down P_t in a number of ways:

$$P_t = \left(\frac{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}}{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}P_{initial} + (0.3)^{tP_t^w}(0.7)^{t-tP_t^w}(1-P_{initial})} \right) P_{initial}$$
$$= \frac{1}{1 + \frac{(0.3)^{tP_t^w}(0.7)^{t-tP_t^w}(1-P_{initial})}{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}P_{initial}}}. \quad (3)$$

The second line is a useful simplification (which we use in appendix A) whereas the bracketed fraction in the first line is the probability of obtaining the tP_t^w white balls when Urn 1 is drawn versus the total probability of obtaining this number of white balls.

Quasi-Bayesian Updating

In our version of Quasi-Bayesian updating (QB), agents use Bayesian updating as each new draw is received, but they incorrectly weight this bracketed probability fraction:

$$P_t^{QB} = \left(\frac{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}}{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}P_{initial} + (0.3)^{tP_t^w}(0.7)^{t-tP_t^w}(1-P_{initial})} \right)^{\beta} P_{initial} \quad (4)$$

The parameter β may be thought of as the QB parameter: if $\beta = 1$, agents are straightforward Bayesians; if $\beta > 1$ they overuse information and under-weight priors; if $0 \leq \beta < 1$ they underuse information and over-weight priors and if $\beta < 0$ they respond the wrong way to information—raising the conditional probability when they should be lowering it, and vice versa.

Agents' attitude towards the extent of belief change in the light of evidence can be summarized by the distribution $f(\beta)$ across subjects. If $f(\beta)$ has most probability mass between 0 and 1, most agents only partially adjust, and subjects converge to full adjustment at $\beta = 1$ to the extent that the probability mass in $f(\beta)$ converges towards unity.

Inferential Expectations

In our version of state dependent belief adjustment, agents form a belief and do not depart from that belief until the weight of evidence against the belief is sufficiently strong, as measured by $|z_t|$ from equation (2).

Under inferential expectations (IE), each agent is assumed to start with a belief about the probability of U (that is, $P_0 = 0.6$) and an implied probability of a white ball ($P_0^w = 0.54$) and conducts a hypothesis test that the latter is true after drawing a test size from his or her own distribution of α, namely $f_i(\alpha)$. Agents are assumed to draw this every round during the experiment. The p-value is then derived using (2) as the test statistic. We assume for simplicity that z_t is distributed as a standard Normal. The purpose of modeling the full distribution of α is to let the data adjudicate whether our subjects act as "classical statisticians" with a high probability mass for α over $5 - 10\%$.

To flag one of our main results, we need to develop a more nuanced account of agents' attitude towards evidence. If $f_i(\alpha)$ has most probability mass near zero, agent i is sluggish to adjust. Probability mass in $f_i(\alpha)$ near unity implies a very strong willingness to use evidence, and probability mass at unity implies (stochastic) time dependent updating. That is, if the probability mass at unity in $f_i(\alpha)$ is, say, 0.3, it implies that there is a thirty per cent chance that agent i will update regardless of what the evidence says. This is because the decision rule in a hypothesis test is to reject H_0, the status quo, if the p-value $\leq \alpha$. A value for α of unity implies the status quo will be rejected, which is the same as updating in this context, for any p-value whatsoever.

Relationship between Expectations Benchmarks

When agent i rejects H_0 within the IE framework we assume she updates her probability guess. This agent can then either be fully Bayesian ($\beta_i = 1$), or she can be quasi-Bayesian. If $0 < \beta_i < 1$, she moves by a fraction β_i of the distance she should move; if $\beta_i > 1$, she over-reacts, and if $\beta_i < 0$, she misinterprets the information.

Since each agent has a full distribution of α, namely $f_i(\alpha)$, we need a representative α_i to summarize the extent of sticky belief adjustment for agent i and to relate to her β_i. There are a number of possibilities, but a natural choice which permits analytic solutions is the median α_i from their $f_i(\alpha)$. For the purposes of our empirical analysis a fully rational (Bayesian) agent is one who has (median) $\alpha_i = \beta_i = 1$, whereas any other sort of agent does not have RE.

We now parameterize all three expectation processes in a double hurdle model. We find evidence for all of them in our data, and importantly we find that the IE representation of $f_i(\alpha)$ has non-zero measure at unity. As discussed above, this is the fraction of agents who undertake random belief adjustment.

4 Experimental Results

4.1 Preliminary Analysis of "No-change"

The baseline and complex treatments each had 82 subjects, and the inattention treatment had 81 subjects. In this sub-section, we consider the number of times our subjects executed a "no-change", meaning a guessed probability equal to that of the previous period. This is interesting because, given the nature of the information and comparatively small number of draws, incidences of "no-change" are not predicted by either Bayesian or Quasi-Bayesian updating, and so, if such observations are widespread in the data, this is the first piece of evidence that these standard models are incomplete.

The maximum of the number of "no-changes" for each subject is 49: seven opportunities for no change out of eight draws, times the seven stages. The distributions over subjects separately by treatment are shown in Figure 1. The baseline distribution shows a concentration at low values; for both Complex and Inattention, there appears to be a shift in the distribution towards higher values, as one might expect. The means for each treatment are represented by the vertical lines on the right hand side. The vertical lines on the left hand side show the mean number of no-changes due to subjects rounding the Bayesian probabilities to two decimal places. Clearly, rounding cannot account for the prevalence of no-changes found in the empirical distribution.

The mean is higher under C (22.79) than under B (18.74) (Mann-Whitney test gives $p = 0.007$); and higher under I (26.97) than under B ($p < 0.001$).[6] This is expected: complexity and inattention are both expected to increase the tendency to leave guesses unchanged. When C and I are compared, the p-value is 0.06, indicating mild evidence of a difference between the two treatments.

In Figure 1 it is clear from the nonparametric evidence of widespread incidence of "no-changes" that any successful model of our data will have to deal with the phenomenon of whether to adjust, before considering how much to adjust. This in turn can imply that the waiting process is stochastic and is unrelated to the actual information arriving (time-dependent belief adjustment) or, that the information that arrives influences the timing (state-dependent belief adjustment). Our double hurdle model enables us to consider both together.

4.2 A Double Hurdle Model of Belief Adjustment

In this section, we develop a parametric double hurdle model which simultaneously considers the decision to update beliefs and the extent to which beliefs are changed when updates occur. The purpose of the model is to act as a testing tool for state-dependent belief adjustment, namely Bayesian belief adjustment and Quasi Bayesian belief adjustment in the simple version previously defined, as well as (stochastic) time-dependent belief adjustment.

Our econometric task is to model the transformed implied belief $r_t^* = \Phi^{-1}\left(g_t^*\left(\theta_i\right)\right)$, which in turn requires an estimate for risk aversion. We estimate this at the individual level using the technique by Offerman et al. (2009). Appendix B contains the subject-level details surrounding the estimation of θ_i.

[6] All p-values in the paper are two tailed. All bivariate tests use subject level means as the independent observations to avoid the problem of dependence of within-subject choices.

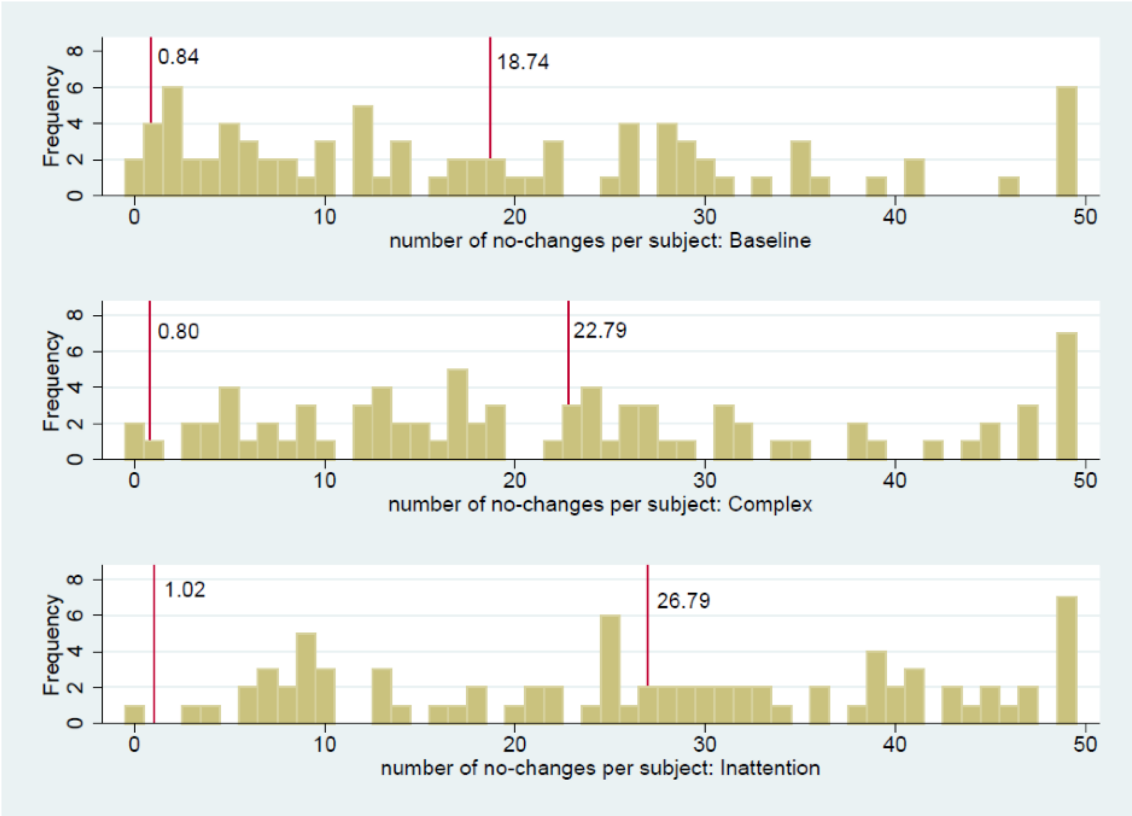

Figure 1: Distributions of number of no-changes over subjects separately by treatment. In each panel, the vertical line on the right represents the mean number of no-changes, while the vertical line on the left represents the mean number of no-changes that would arise if subjects rounded Bayesian probabilities to two decimal places.

We will refer to r^*_{it} as subject i's "belief" in period t, as shorthand for 'transformed implied belief'. We will treat r^*_{it} as the focus of the analysis, because r^*_{it} has the same dimensionality as z_{it}, the test statistic defined in (2). That is, both have support $(-\infty, \infty)$. Sometimes r^*_{it} changes between $t-1$ and t; other times, it remains the same. Let Δr^*_{it} be the change in belief of subject i between $t-1$ and t. That is, $\Delta r^*_{it} = r^*_{it} - r^*_{it-1}$.

In the following estimation we exploit the near equivalence between (2) and the scaled difference since the last update $(\Phi^{-1}(P_t) - \Phi^{-1}(P_m))/\sqrt{t}$. In round 1 P_m equals the prior 0.6 and the movement of the guess for a given subject is $\Delta r^*_{it} = r^*_1 - \Phi^{-1}(0.6)$. That is, both the objective measure of the information change and the subjective guess of the agent are assumed to anchor onto the prior probability that Urn 1 is chosen, 0.6, in the first period.

First Hurdle: The probability that a belief is updated (in either direction) in period t is given by:

$$P\left(\Delta r_{it}^{*}\neq 0\right)=\Phi\left[\delta_{i}+x_{i}'\theta_{1}+\gamma\left|z_{it}\right|\right], \tag{5}$$

where $\Phi[\cdot]$ is the standard Normal cdf and δ_i represents subject i's idiosyncratic propensity to update beliefs, and therefore models random probabilistic belief adjustment (time-dependent belief adjustment). The probability of an update is assumed to depend (positively) on the absolute value of z_{it}, the test statistic. The vector x_i contains treatment and gender dummy variables together with an age variable, all of which are time invariant and can be expected to affect the propensity to update.

One econometric problem that arises is the endogeneity of the variable $|z_{it}|$: subjects who are averse to updating tend to generate large values of $|z_{it}|$ while subjects who update regularly do not allow it to grow beyond small values. This is likely to create a severe downward bias in the estimate of the parameter γ in the first hurdle. To address this problem we use an instrumental variables (IV) estimator which uses the variable $\widehat{|z_{it}|}$ in place of $|z_{it}|$, where $\widehat{|z_{it}|}$ comprises the fitted values from a regression of $|z_{it}|$ on a set of suitable instruments. This IV procedure is explained in detail in Appendix C.

Second hurdle: Conditional on subject i choosing to update beliefs in draw t, the next question relates to how much they do so. This is given by:

$$\Delta r_{it}^{*}=\left(\beta_{i}+x_{i}'\theta_{2}\right)\sqrt{t}z_{it}+\varepsilon_{it,}\qquad \varepsilon_{it}\sim N\left(0,\sigma^{2}\right). \tag{6}$$

As a reminder, the Quasi-Bayesian belief adjustment parameter β_i represents subject i's idiosyncratic responsiveness to the accumulation of new information: if $\beta_i = 1$, subject i responds fully; if $\beta_i = 0$, subject i does not respond at all. Remember that β_i is not constrained to $[0, 1]$. In particular, a value of β_i greater than one would indicate the plausible phenomenon of over-reaction. Again, treatment variables are included: the elements of the vector θ_2 tell us how responsiveness differs by treatment.

Considering the complete model, there are two idiosyncratic parameters, δ_i and β_i. These are assumed to be distributed over the population of subjects as follows:

$$\left(\begin{array}{c}\delta_{i}\\ \beta_{i}\end{array}\right)\sim N\left[\left(\begin{array}{c}\mu_{1}\\ \mu_{2}\end{array}\right),\left(\begin{array}{cc}\eta_{1}^{2} & \rho\eta_{1}\eta_{2}\\ \rho\eta_{1}\eta_{2} & \eta_{2}^{2}\end{array}\right)\right]. \tag{7}$$

In total, there are fifteen parameters to estimate: μ_1, η_1, μ_2, η_2, ρ, γ, σ, four treatment effects (two in each hurdle); two gender effects (one in each hurdle); and two age effects (one in each hurdle).

The results are presented in Table 2, for four different models. The last column shows the preferred model.

Model 1 estimates the QB benchmark, in which it is assumed that the "first hurdle" is crossed for every observation—that is updates always occur. Zero updates

	QB	IE	IE-QB (ρ=0)	IE-QB Preferred model
	(1)	(2)	(3)	(4)
Propensity to update:				
μ_1	$+\infty$	$0.065^{**}(0.222)$	$-0.007(0.175)$	$-0.165(0.161)$
[Baseline time-dependent update probability $\Phi(\mu_1)$]	[1.00]	[0.53]	[0.50]	[0.43]
η_1	0	$0.979^{**}(0.044)$	$0.987^{**}(0.064)$	$1.014^{**}(0.045)$
γ extent of state-dependence	0	$1.250^{**}(0.055)$	$1.242^{**}(0.056)$	$1.239^{**}(0.056)$
Complex	0	$-0.044(0.112)$	$-0.155(0.134)$	$-0.036(0.114)$
Inattention	0	$-0.249^{*}(0.115)$	$-0.270^{*}(0.121)$	$-0.263^{**}(0.101)$
Male	0	$-0.118(0.091)$	$-0.148(0.127)$	$-0.022(0.089)$
Age-22	0	$-0.003(0.009)$	$-0.009(0.007)$	$-0.014^{*}(0.006)$
Extent of update:				
μ_2	$0.506^{**}(0.134)$	1	$0.467^{**}(0.084)$	$0.350^{**}(0.078)$
η_2	$0.516^{**}(0.030)$	0	$0.315^{**}(0.021)$	$0.337^{**}(0.022)$
Complex	$-0.242^{**}(0.055)$	0	$-0.217^{**}(0.047)$	$-0.172 * (0.057)$
Inattention	$-0.235^{**}(0.064)$	0	$-0.102^{*}(0.050)$	$-0.028(0.049)$
Male	$0.014(0.045)$	0	$0.058(0.038)$	$0.080^{*}(0.042)$
Age-22	$0.017^{**}(0.005)$	0	$0.009^{**}(0.003)$	$0.007^{*}(0.003)$
σ	$0.513^{**}(0.003)$	$0.760^{**}(0.006)$	$0.676^{**}(0.006)$	$0.675^{**}(0.006)$
ρ propensity to update vs. extent of update correlation				$0.454^{**}(0.080)$
LogL	-9944.9997	-16,277.909	-14740.158	-14723.67
AIC (=2k-2LogL)	N/A	32,571.8	29,508.3	29,477.3
Wald test (df, p-value)	58,367(8, 0.000)	1,237(7, 0.000)	24.3(1, 0.000)	N/A
Number of subjects (n)	245	245	245	245
Observations per subject (T)	56	56	56	56

Note: LogL for QB cannot be compared with that of other columns.

Table 2: Results of hurdle model with risk adjustment

are treated as zero realizations of the update variable in the second hurdle, and their likelihood contribution is a density instead of a probability. Because of this difference in the way the likelihood function is computed, the log-likelihoods and AICs cannot be used to compare the performance of QB to that of the other models.

Model 2 estimates the IE benchmark, in which the update parameter (β_i) is fixed at 1 for all subjects. Consequently the extra residual variation in updates is reflected in the higher estimate of σ. The parameters in the first hurdle are free.

Model 3 combines IE and QB, but constrains the correlation (ρ) between δ and β to be zero. Model 4 is the same model with ρ unconstrained.

The overall performance of a model is best judged using the AIC; the preferred model is the one with the lowest AIC. Among the models that can be compared, the best model is the most general model 4: IE-QB with ρ unrestricted, whose results are presented in the final column of Table 2.

To confirm the superiority of the general model over the restricted models, we conduct Wald tests of the restrictions implied by the three less general models. We see that, in all three cases, the implied restrictions are strongly rejected, implying that the general model is superior. Note in particular that this establishes the superiority of the general model 4 (IE-QB with ρ unrestricted) over the QB model 1 (a comparison that was not possible on the basis of AIC).

We interpret the results from the preferred model as follows. Consider the first hurdle (propensity to update). The intercept parameter in the first hurdle (μ_1) tells us that a typical subject has a predicted probability of $\Phi(-0.165) = 0.43$ of updating in any task, in the absence of any evidence (i.e. when $|z_{it}| = 0$). We note that this estimate is not significantly different from zero, which would imply a 50% probability of updating:

Result 1 *There is evidence of time-dependent (random) belief adjustment. In every period subjects update their beliefs idiosyncratically around half the time.*

The Inattention treatment effect is significant, suggesting that Result 1 is more pronounced when subjects are not paying attention. The large estimate of η_1 tells us that there is however considerable heterogeneity in this propensity to update (see Figures 2 and 3 below), something we will explore further in section 4.3. The parameter γ is estimated to be significantly positive, and this tells us, as expected, that the more cumulative evidence there is, in either direction, the greater the probability of an update:

Result 2 *There is evidence of state-dependent belief adjustment. Subjects are more likely to adjust if there is more evidence to suggest that an update is appropriate (thus making it costlier not to update).*

In the second hurdle, the intercept (μ_2) is estimated to be 0.35 in our preferred model 4: when a typical (baseline) subject does update, she updates by a proportion 0.35 of the difference from the Bayes probability. The large estimate of η_2 tells us that there is considerable heterogeneity in this proportion also (see Figure 2 below). However, in all models where the second hurdle is meaningful (models 1, 3 and 4), many of the β_i's are between 0 and 1. If we take model 4, only 18 out of 245 subjects have $\beta < 0$, which indicates noise or confused subjects who adjusted in the wrong direction. More interestingly, only 4 out of 245 subjects ($< 2\%$) display overreaction to the evidence in model 4. We summarize this in the following result:

Result 3 *There is evidence of Quasi-Bayesian partial belief adjustment. On average, subjects who adjust do so by around 35%. There is no evidence of prior information under-weighting: virtually none of the subjects overreact to evidence once they decide to adjust.*

The estimate of ρ is strongly positive, indicating that subjects who have a higher propensity to update, also tend to update by a higher proportion of the difference from the Bayes probability. This positive correlation is seen in the Model 4 plot in Figure 2.

The treatment effects in the second hurdle are of the expected sign but the Inattention effect is insignificant in the second hurdle while being significant in the first hurdle. There is instead evidence that further complexity in the decision

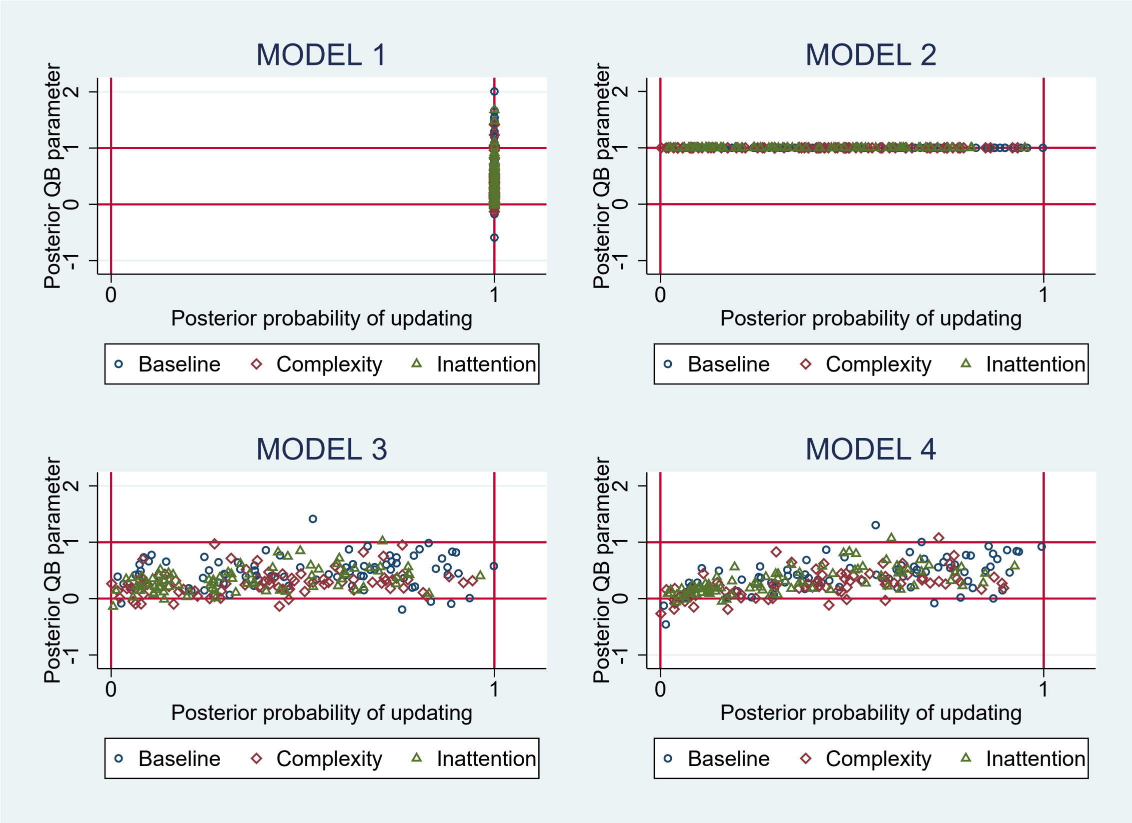

Figure 2: Posterior probability of updating

problem reduces the extent of update by around 15 percentage points (from 35% to approx. 20%):

Result 4 *Complexity does not affect whether subjects decide to update or not, but, if they do, they partially adjust by around 15 percentage points less on average. Inattention reduces the propensity to update but does not affect the extent of updating.*

4.3 The Empirical Distribution of α and β

To get a better sense of the population heterogeneity in belief adjustment, this subsection maps out the empirical distribution of the IE α_i and QB β_i parameters across subjects against each other. Estimating $f(\beta)$ is easy enough to see from our double hurdle model since β_i is directly estimated and we have done so in Figure 2. We next use the first hurdle information to generate $f_i(\alpha)$, the empirical distribution of α_i.

As we flagged earlier each agent has a full distribution of α and so we need a representative α_i to summarize the extent of sticky belief adjustment for agent i, to then relate to their β_i. As will be clear below, the choice that permits analytic solutions is the median α_i from $f_i(\alpha)$.

The econometric equation for the first hurdle is equivalent to the probability of rejecting the null under IE. We omit the dummy variables. We begin by re-writing the first hurdle, namely (5) without dummies:

$$\Pr(\text{reject } H_0)_{it} = \Phi(\delta_i + \gamma |z_{it}|), \tag{8}$$

where δ_i and γ are estimated parameters and $|z_{it}|$ is the test statistic based on the proportion of white balls.

$$z_{it} = \frac{P_{it}^w - P_{im}^w}{\sqrt{0.5^2/t}}. \tag{9}$$

For any $|z_{it}|$ it is possible to work out an implied p-value and we do so by assuming that (9) is approximately distributed $N(0,1)$. This in turn allows us to work out $f_i(\alpha)$ from the econometric equation for the first hurdle. When $|z_{it}| = 0$, the p-value for a hypothesis test is unity, and so the equation says that a fraction of agents will reject H_0 if the p-value is unity. Since the criteria for rejecting H_0 in a hypothesis test is always $\alpha \geq p$-value it implies that there must be a non-zero probability mass on $f_i(\alpha)$ at the value of α *exactly* equal to 1. The pdf of α_i will thus have a discrete 'spike' at unity and be continuous elsewhere. We know what that spike is from equation (8) with $|z_{it}| = 0$, namely $\Phi(\delta_i)$.

The probability of rejecting H_0 depends on the probability that the test size is greater than the p-value, but this is also equal to the econometric equation for the first hurdle.

$$\Pr(\text{reject } H_0)_{it} = \int_{p\text{-value}_{it}}^{1} f_i(\alpha)\, d\alpha_i = 1 - F_i(p\text{-value}_{it}) = \Phi(\delta_i + \gamma |z_{it}|) \tag{10}$$

Upper case F in the last equality is the anti-derivative of the density. We define $F_i(1)$ to be unity since 1 is the upper end of the support of α but we also note that there is a discontinuity such that F jumps from $1 - \Phi(\delta_i)$ to 1 at $\alpha = 1$, as a consequence of the non-zero probability mass on $f_i(\alpha)$ at unity. To solve the equation we use an expression for the p-value of $|z_{it}|$ on a two-sided Normal test.

$$p\text{-value}_{it} = 2(1 - \Phi(|z_{it}|)). \tag{11}$$

We use a 'single parameter' approximation to the cumulative Normal (see Bowling et al. 2009). For our purposes $\sqrt{3}$ is sufficient for the single parameter.

$$\Phi(|z_{it}|) = \frac{1}{1 + \exp\left(-\sqrt{3}\,|z_{it}|\right)}. \tag{12}$$

We can now write down $|z_{it}|$ as a function of the p-value using (11) and (12).

$$|z_{it}| = \frac{1}{\sqrt{3}} \ln\left(\frac{2 - p\text{-value}_{it}}{p\text{-value}_{it}}\right). \tag{13}$$

Intuitively, a p-value of zero implies an infinite $|z_{it}|$ and p-value of unity implies $|z_{it}|$ is zero. We can now use the relationship between $F_i\left(p\text{-value}_{it}\right)$ and our estimated first hurdle to generate $F_i(\alpha)$.

$$\begin{aligned}
1 - F_i\left(p\text{-value}_{it}\right) &= \Phi\left(\delta_i + \gamma\left|z_{it}\right|\right) \\
\therefore F_i\left(p\text{-value}_{it}\right) &= 1 - \Phi\left(\delta_i + \gamma\left|z_{it}\right|\right) \\
&= 1 - \Phi\left(\delta_i + \gamma\left\{\frac{1}{\sqrt{3}}\ln\left(\frac{2 - p\text{-value}_{it}}{p\text{-value}_{it}}\right)\right\}\right). \qquad (14)
\end{aligned}$$

In the above expression the variable 'p-value' is just a place-holder and can be replaced by anything with the same support leaving the meaning of (14) unchanged. Thus, it can be replaced by α giving the cumulative density of α.

$$\begin{aligned}
F_i(\alpha) &= 1 - \Phi\left(\delta_i + \gamma\left\{\frac{1}{\sqrt{3}}\ln\left(\frac{2-\alpha}{\alpha}\right)\right\}\right) \\
&= 1 - \frac{1}{1 + \left[\frac{\alpha}{2-\alpha}\right]^{\gamma}\exp\left(-\sqrt{3}\delta_i\right)}. \qquad (15)
\end{aligned}$$

Substitution of $\alpha = 1$ does not give unity, which is what we earlier assumed for the value of $F_i(1)$. However, it does give $1 - \Phi(\delta_i)$, which of course concurs with the econometric equation for the first hurdle when $|z_{it}| = 0$. This discontinuity in F_i is consistent with a discrete probability mass in $f_i(\alpha)$ at unity, as we noted earlier. It now just remains to differentiate F_i to obtain the continuous density $f_i(\alpha)$ for α strictly less than unity. The description of the function at the upper end of the support (unity) is completed with a discrete mass at unity of $\Phi(\delta_i)$.

$$\left.\begin{aligned}
f_i(\alpha) &= \frac{2\gamma\alpha^{\gamma-1}\exp\left(-\sqrt{3}\delta_i\right)}{(2-\alpha)^{1+\gamma}\left\{1+\left[\frac{\alpha}{2-\alpha}\right]^{\gamma}\exp\left(-\sqrt{3}\delta_i\right)\right\}^2}, & \alpha < 1 \\
\Pr_i(\alpha = 1) &= \frac{1}{1+\exp\left(-\sqrt{3}\delta_i\right)} = \Phi(\delta_i), & \alpha = 1
\end{aligned}\right\} \qquad (16)$$

Figure 3 illustrates the distribution $f_i(\alpha)$ for $\delta_i = -0.17$ and $\gamma = 1.24$ together with the distributions one standard deviation either side of δ_i. The former is the mean of δ across subjects, from our estimation (from the last column of Table 2, rounded to two decimal places. On the right-most of the chart is the probability mass when $\alpha = 1$. As discussed earlier, this corresponds to the proportion of agents who update without any evidence at all ($|z_{it}| = 0$). There is clearly a great deal of interesting heterogeneity. One distribution has a near-zero probability of a random update (11%) and when the agent uses information they are very conservative, with α close to zero. Another distribution has a virtually certain probability of a random update (81%) and agents who look at information are not conservative at all (α is likely close to unity).

Since there are idiosyncratic values of δ_i there will be a separate distribution for every subject varying over δ_i. So we must use a summary statistic for $f_i(\alpha)$, and the

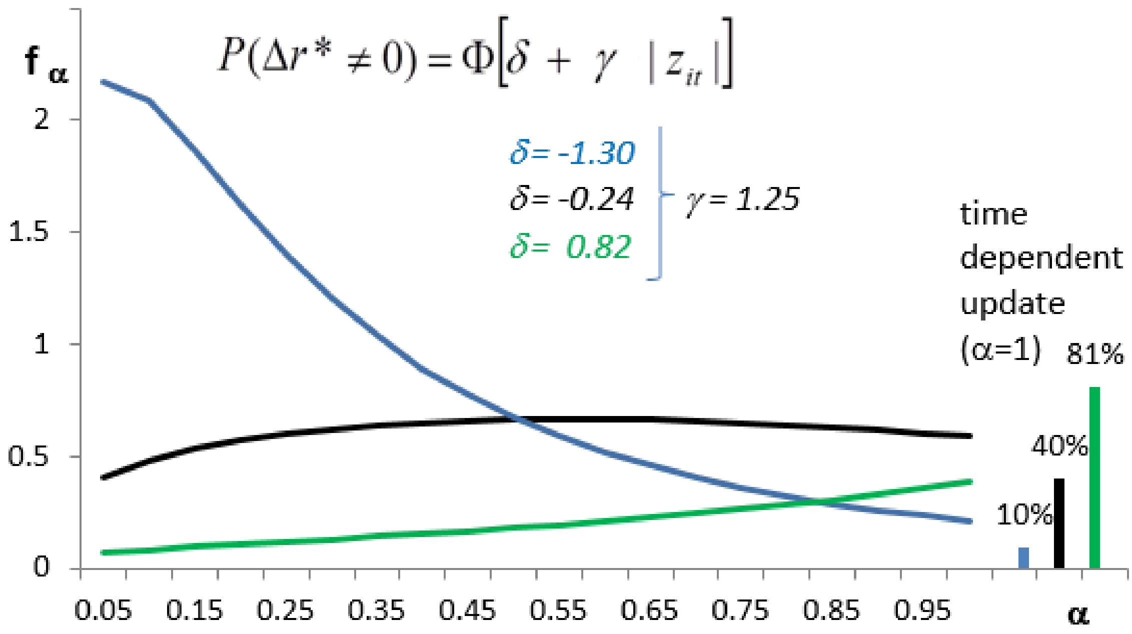

Figure 3: Distribution of $f_i(\alpha)$ for $\delta_i = -0.17$ and $\gamma = 1.24$ together with the distributions one standard deviation either side of δ_i

one which comes to hand is the median α value, obtained by solving $F_i(\alpha) = 0.5$ in equation (15). In Figure 4, we plot the collection of subject i's (median α, β) duples for model 4, our preferred equation, along with a regression line. Table 3 lists the percentage of subjects in each (median α_i, β_i) 0.2 bracket.

Fully rational agents, whose α_i are always identically equal to unity, are hard to come by since they would require a modelled probability mass of unity at $\alpha = 1$ in the distribution of α_i (16), which in turn would require an infinite δ_i in (5). So, our procedure in columns of Table 3 is to describe agents as rational on the α-dimension (the rows of Table 3) if they have a median in the top range (0.8 to 1.0).

With that in mind, we can now comment on the subjects' use of information. Roughly half of the subjects update regardless of evidence, so the median α's cluster at unity along the bottom axis with over half of them (52%) in the range at or above 0.8. Only 5% of agents could be described as classical statisticians with median α's around the 5-10% level. However, nearly one quarter of median α's point to conservative belief adjustment, with α-values no more than 0.2.

Regarding the size of updating, we already know from Result 3 that it is far from complete. In Table 3 just under one third of subjects (29 per cent) only update between 20 and 40 per cent of what they should, and we have already noted from model 4 of Table 2 that the average amount of updating over all subjects is within this range (35%). Figure 4 and Table 3 show that those agents who are relatively likely to update ($\alpha \to 1$) are likely to accomplish relatively more complete updating

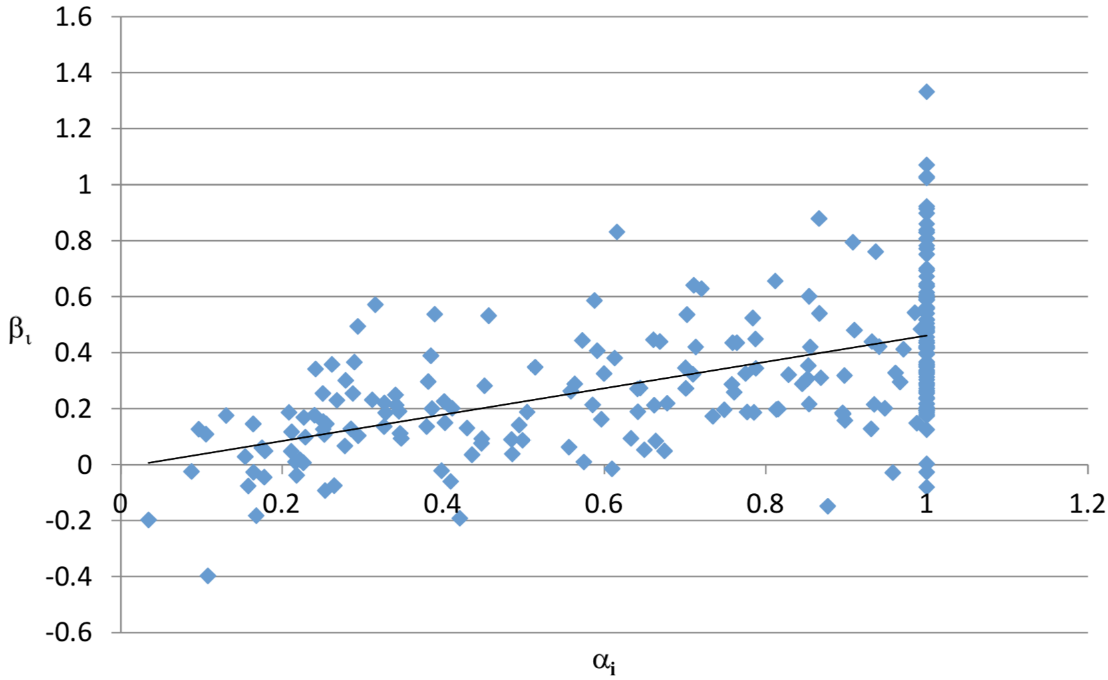

Figure 4: The collection of subject i's (median α, β) duples for model 4.

than those who do not.

Result 5 *Estimated test sizes spread over the whole support* $[0, 1]$*. There is a positive correlation (0.45) between the median* α_i *and the extent of belief adjustment when it occurs.*

This positive correlation between α_i and β_i suggest the existence of a small group (less than 5%) of rational agents. They inhabit the bottom RHS of Table 3, where α_i and β_i both exceed 0.8.

5 Discussion and Conclusion

Our paper reframes the debate about time- versus state-dependent behavior and finds clear evidence for both in subject play. To date each of these has been considered separately, and this has clearly been a good starting point to observe them in the laboratory. Sticky belief adjustment—inadequate frequency and extent of belief updating—is not a novel idea, and initial experimental evidence for it in settings with evidence presented all at once were discussed as long ago as Phillips and Edwards (1966) and Edwards (1968). A pilot study described in Menzies and Zizzo (2005) found evidence for sticky belief adjustment in an experiment with dynamically provided information, but, apart from the small nature of the study, it neither

	α_i					
β_i	0 - 0.2	0.2 - 0.4	0.4 - 0.6	0.6 - 0.8	0.8 - 1	Total
-0.4 - -0.2	1	0	0	0	0	0.4%
-0.2 - 0	6	4	2	1	4	6.9%
0 - 0.2	7	23	13	9	16	27.8%
0.2 - 0.4	0	13	8	12	39	29.4%
0.4 - 0.6	0	3	4	8	39	22.0%
0.6 - 0.8	0	0	0	2	16	7.3%
0.8 - 1	0	0	0	1	10	4.5%
1 - 1.2	0	0	0	0	3	1.2%
1.2 - 1.4	0	0	0	0	1	0.4%
Total	5.7%	17.6%	11.0%	13.5%	52.2%	100.0%

Table 3: Percentage of subjects in each (α_i, β_i) bracket in model 4

controlled for risk aversion nor did it account for different forms of sticky belief adjustment. Massey and Wu (2005) contains a related but different experiment with dynamically provided information where the goal of the subjects is to identify whether a regime shift has taken place, but they are allowed to change their mind only once; they identify conditions for which, in a decision problem of this kind, their subjects display underweighting or overweighting of priors.

In the context of an experiment in which there is only one piece of information provided at the beginning of trading, Camerer (1987) argues that probability updating anomalies wash away in the light of market discipline. Conversely, again in a setting where information is provided all at once, Menzies and Zizzo (2012) find greater evidence of stickiness in market prices in a Walrasian auction market setting intended to model an exchange rate market, than in the corresponding individual beliefs as revealed by the market choices of traders. There is a range of empirical applications where belief stickiness appears plausible in natural economic environments, including markets. Applications of sticky belief adjustment include, among others, optimal principal agent contracts (Rabin and Schrag, 1999), individual responses to market signals (Sims, 2003), a micro-foundation for the New Keynesian Phillips curve (Mankiw and Reis, 2002), consumer and producer behavior (Reis, 2006a, 2006b), and pricing under information costs (Woodford, 2009). Inferential expectations modeling has been applied to explain the uncovered interest rate parity failure (Menzies and Zizzo, 2009, 2012), central bank credibility (Henckel et al., 2011, 2013) and merger decisions by competition regulators (Lyons et al., 2012).

The double hurdle model we have developed in this paper allows us to integrate both time- and state-dependent belief adjustment in a unified econometric framework. Our experiment uses a quadratic scoring rule with monetary payoffs to incentivize subjects, and we operationalize Offerman et al. (2009) in order to generate risk-adjusted beliefs for our regression analysis. To our knowledge this is one of the first such applications.

Our econometric model found evidence for considerable heterogeneity in both the propensity and extent of updating, with only a relatively small subset of subjects displaying rational expectations. We observe random belief adjustment around half the time, which is consistent with stochastic time-dependent belief adjustment. Deviations from Bayesian updating are systematically in the direction of under-adjustment, with around 5% in the neighborhood of full adjustment ($0.8 < \beta < 1.2$). We find that, when beliefs change, they do so by only around 35% of the amount required by Bayesian updating, well short of what full rationality requires. The likelihood of a belief change increases as the amount of evidence against the null hypothesis increases, which is consistent with state-dependent belief adjustment.

Our quadratic scoring incentive mechanism implies that the greater the amount of evidence against the currently held belief, the greater the expected costs of maintaining this belief. Thus, threshold-based rational inattention models of belief adjustment (Sims, 2003) are consistent with agents holding inferential expectations (Menzies and Zizzo, 2009). That is, agents hold onto their status quo belief until the cumulated evidence makes this too costly, modeled as passing a threshold determined by the test size α. We estimate that roughly one quarter of agents are belief conservative with $\alpha \leq 0.4$. Deriving the full α distribution, as we do in Figure 3, summarizes a great deal about subject behavior. First, the probability mass at $\alpha = 1$ measures the extent of time-dependent adjustment. Second, where agents instead adopt state-dependent adjustment, the density over the support $\alpha = [0, 1)$, as shown in Figure 3, informs about the extent of belief conservatism.

As evidenced by a significant positive correlation between α and β, subjects who are less likely to adjust their beliefs (low α's) are also subjects who adjust them less when they do (low β's). This departs from common statistical practice which, while adopting low α's (1%, 5%, 10%), recommends full adjustment when beliefs do change. Our results are robust to treatments effect, but we do find that, plausibly, our inattention manipulation makes subjects less likely to adjust their stated beliefs, whereas additional complexity makes it harder to adjust fully. Finally, older subjects tend to have a slightly lower propensity to update but adjust by more when they do update.

The evidence presented for the co-existence of time- and state-dependent behavior has important implications for the modelling of expectations. Having shown their co-existence we leave to future research the task of incorporating them into theoretical models and refining their properties in specific economic contexts.

References

[1] Abeler, Johannes, Armin Falk, Lorenz Goette and David Huffman (2011). "Reference Points and Effort Provision". *American Economic Review* 101(2), 470-492.

[2] Alvarez, Fernando, Francesco Lippi and Juan Passadore (2016). "Are State and Time dependent models really different?" *NBER Macroeconomics Annual* 2016, Vol. 31.

[3] Ambuehl, Sandro and Shengwu Li (2014). "Belief Updating and the Demand for Information", SSRN Discussion Paper, June.

[4] Aucremanne, Luc, and Emmanuel Dhyne (2005). "Time-Dependent versus State-Dependent Pricing: A Panel Data Approach to the Determinants of Belgian Consumer Price Changes," ECB Working Paper No. 462, March.

[5] Benjamin, Daniel J., Matthew Rabin, and Collin Raymond (2015). "A Model of Non-Belief in the Law of Large Numbers." *Journal of the European Economics Association* 14(2), 515-544.

[6] Bowling, S. Khasawnch, M., Kaewkuekool, S., and B. Cho, (2009). "A logistic approximation to the cumulative normal distribution", *Journal of Industrial Engineering and Management* 2(1), 114-127.

[7] Caballero, R.J. (1989). "Time dependent rules, aggregate stickiness and information externalities." Columbia University Department of Economics Discussion Paper 428.

[8] Calvo, Guillermo A. (1983). "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics* 12(3), 383-398.

[9] Camerer, Colin F. (1987). "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence," *American Economic Review* 77(5), 981-997.

[10] Caplin, Andrew, Mark Dean, and Daniel Martin (2011). "Search and Satisficing," *American Economic Review* 101(7), 2899-2922.

[11] Carroll, Christopher D. (2003), "Macroeconomic Expectations of Households and Professional Forecasters" *Quarterly Journal of Economics*, 118, 269-298.

[12] Charness, Gary, and Dan Levin (2005). "When Optimal Choices Feel Wrong: A laboratory Study of Bayesian Updating, Complexity and Affect," *American Economic Review* 95, 1300-1309.

[13] Corgnet, Brice, Roberto Hernan-Gonzalez and Eric Schniter (2015). “Why real leisure really matters: incentive effects on real effort in the laboratory,” *Experimental Economics* 18, 284-301.

[14] Costain, James, and Anton Nakov (2011). “Price Adjustments in a General Model of State-Dependent Pricing,” *Journal of Money, Credit and Banking* 43(2-3), 385-406.

[15] Dias, D. A., C. Robalo Marques, and J. M. C. Santos Silva (2007). “Time- or state-dependent price setting rules? Evidence from micro data,” *European Economic Review* 51, 1589-1613.

[16] Dias, Daniel A., Carlos Robalo Marques, and Fernando Martins (2012). “Choosing between time and state dependence: Micro evidence on firms' price-reviewing strategies, *mimeo*, University of Illinois at Urbana-Champaign.

[17] Edwards, W. (1968). “Conservatism in human information processing,” in: Kleinmuntz, B. (Ed.). *Formal representation of human judgment.* New York: Wiley, 17-52.

[18] El-Gamal, Mahmoud A., and David M. Grether(1995). “Are People Bayesian? Uncovering Behavioral Strategies”, *Journal of the American Statistical Association* 90, no. 432, 1137-145.

[19] Fischer, Stanley (1977). “Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule,” *Journal of Political Economy* 85(1), 191-205.

[20] Grether, David M. (1980). “Bayes Rule as a Descriptive Model: The Representativeness Heuristic,” *Quarterly Journal of Economics* 95(3), 537-57.

[21] Henckel, Timo, Gordon D. Menzies and Daniel J. Zizzo (2013). “The Great Recession and the Two Dimensions of European Central Bank Credibility” CAMA Working Paper 55/2013, August.

[22] Henckel, Timo, Gordon D. Menzies, Nicholas Prokhovnik and Daniel J. Zizzo (2011). “Barro-Gordon Revisited: Reputational Equilibria with Inferential Expectations”, *Economics Letters* 112(2), 144-147.

[23] Independent Commission on Banking, 2011. Final Report Recommendations. London: U.K. Government.

[24] Joskow, Paul L. (2008). “Lessons Learned From Electricity Market Liberalization”, *Energy Journal* 29 (Special Issue 2), 9-42.

[25] Klenow, Peter J., and Oleksiy Kryvtsov (2008). “State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?” *Quarterly Journal of Economics* 123(3), 863-904.

[26] Lyons, Bruce, Gordon D. Menzies and Daniel J. Zizzo (2012). "Conflicting Evidence and Decisions by Agency Professionals: An Experimental Test in the Context of Merger Regulation", *Theory and Decision* 73(3), 465-499.

[27] Magnani, Jacopo, Aspen Gorry, and Ryan Oprea (2016). "Time and State Dependence in an Ss Decision Experiment," *American Economic Journal: Macroeconomics* 8(1), 285-310.

[28] Mankiw, N. Gregory and Reis, Ricardo (2002). "Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve", *Quarterly Journal of Economics*, 117, 1295–1328

[29] Martins, F. (2005). "The price setting behaviour of Portuguese firms: Evidence from survey data," ECB Working Paper No. 562.

[30] Massey, Cade and George Wu (2005). "Detecting regime shifts: The causes of under- and overreaction", *Management Science* 51(6), 932-947.

[31] Menzies, Gordon D., and Daniel J. Zizzo (2012). "Monetary Policy and Inferential Expectations of Exchange Rates", *Journal of International Financial Markets, Institutions & Money* 22(2), 359-380.

[32] Menzies, Gordon D., and Daniel J. Zizzo (2009). "Inferential Expectations", *B.E. Journal of Macroeconomics* 9(1) (Advances), Article 42.

[33] Menzies, Gordon D., and Daniel J. Zizzo (2005). "Inferential Expectations", CAMA Discussion Paper No. 12, June.

[34] Midrigan, Virgiliu (2010). "Is Firm Pricing State or Time Dependent? Evidence from U.S. Manufacturing," *Review of Economics and Statistics* 92(3), 643-656.

[35] Offerman, Theo, Joep Sonnemans, Gijs Van De Kuilen, and Peter P. Wakker (2009), "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *Review of Economic Studies* 76 (4), 1461-1489.

[36] Ofgem, 2011. Simpler Energy Tariffs. London: U.K. Government.

[37] Phillips, L.D., and W. Edwards (1966). "Conservatism in a simple Probability Inference Task." *Journal of Experimental Psychology* 72, 346-354.

[38] Rabin, Matthew (2013). "Incorporating Limited Rationality into Economics," *Journal of Economic Literature* 51(2), 528-43.

[39] Rabin, Matthew, and Joel L. Schrag (1999). "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics* 114(1), 37-82.

[40] Reis, Ricardo (2006a). "Inattentive Consumers", *Journal of Monetary Economics* 53 (8), 1761-1800.

[41] Reis, Ricardo (2006b). "Inattentive Producers", *Review of Economic Studies* 73 (3), 793-821.

[42] Schmalensee, Richard (1976). "An Experimental Study of Expectation Formation", *Econometrica* 44(1): 17-41.

[43] Sims, Christopher A. (2003). "Implications of Rational Inattention", *Journal of Monetary Economics* 50(3): 665–690.

[44] Sitzia, Stefania, Jiwei Zheng and Daniel J. Zizzo. (2015). "Inattentive Consumers in Markets for Services", *Theory and Decision* 79(2), 307-332.

[45] Stahl, Harald (2005). "Time Dependent of State Dependent Price Setting? Micro-Evidence from German Metal-Working Industries," ECB Working Paper No. 534, October.

[46] Taylor, John B. (1980). "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy* 88(1), 1-23.

[47] Tversky, Amos, and Daniel Kahneman (1982). "Judgments of and by representativeness," in D. Kahneman, P. Slovic, and A. Tversky, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

[48] Woodford, Michael (2009). "Information-constrained state-dependent pricing", *Journal of Monetary Economics* 56(15), S100-S124.

[49] Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

A Closeness of Two Strength-of-Evidence Measures

Our measure of evidence that the guess should change between times t and m is:

$$z_t = \frac{\Phi^{-1}(P_t) - \Phi^{-1}(P_m)}{\sqrt{t}},$$

where the cumulative Normal Φ, and its inverse, are approximated.

$$\Phi(P_t) \approx \frac{1}{1 + \exp\left(-\sqrt{3}P_t\right)} \qquad \Leftrightarrow \qquad \Phi^{-1}(P_t) \approx \frac{1}{\sqrt{3}} \ln\left(\frac{P_t}{1 - P_t}\right).$$

Writing P_t in terms of t, and noting that the number of white balls is tP_t^w, we obtain:

$$\begin{aligned}
P_t &= \frac{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}(0.6)}{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}(0.6) + (0.3)^{tP_t^w}(0.7)^{t-tP_t^w}(1-0.6)} \\
&= \frac{1}{1 + \frac{(0.3)^{tP_t^w}(0.7)^{t-tP_t^w}}{(0.7)^{tP_t^w}(0.3)^{t-tP_t^w}}\frac{2}{3}} \\
\therefore \frac{P_t}{1-P_t} &= \frac{3}{2}\left(\frac{7}{3}\right)^{2tP_t^w - t} \approx \frac{3}{2}e^{2tP_t^w - t} \\
z_t &= \frac{\Phi^{-1}(P_t) - \Phi^{-1}(P_m)}{\sqrt{t}} \\
&\approx \frac{1}{\sqrt{3t}}\left[\ln\left(\frac{3}{2}\right)e^{2tP_t^w - t} - \ln\left(\frac{3}{2}\right)e^{2mP_m^w - m}\right] \\
&= \frac{1}{\sqrt{3t}}\left[(2tP_t^w - t) - (2mP_m^w - m)\right].
\end{aligned}$$

If $t \approx m$, we have

$$z_t \approx \frac{1}{\sqrt{3}}\left(\frac{P_t^w - P_m^w}{\sqrt{\frac{0.5^2}{t}}}\right).$$

Numerical simulations are available from the authors, which confirm the closeness of the first and last lines.

B Method for Estimating CRRA Risk Parameter

The log relationship between transformations of g_t and g_t^* in the text has an i.i.d. error added to it and run with 10 observations as an OLS regression. The variable t in (17) refers to the 10 rounds in the practice part, not to the rounds in the main part. The estimated parameter θ_i is subscripted for subjects, because (17) is run for each subject to provide her own θ.

$$\ln\left(\frac{g_t^*(1-g_t)}{g_t(1-g_t^*)}\right) = \theta_i \ln\left(\frac{g_t(2-g_t)}{(1+g_t)(1-g_t)}\right) + \eta_t, \qquad t = 1, 2, \ldots, 10. \tag{17}$$

We note the following:

1. The regression has no intercept. If an intercept is included the θ_i estimates are inefficient.
2. For some subjects $g_t = 0.5$ in every period. In this case, the RHS variable is $\ln(1)$ in every period. This means that θ approaches $+\infty$. These estimates need to be re-coded to a high positive number, and we use +10.
3. There is a logical requirement that θ cannot be less than -1 in this model. Hence, any estimates less than -1 need to be re-coded to -1.

The above procedure gives rise to the following distribution of θ over the 245 subjects:

Once each subject has her own estimated θ_i, the full set of implied g_t^* values can be generated from the observed guesses g_t in the main experiment. As discussed in the main text, this is accomplished by rewriting (17) without an error, which is identical to equation (1), exponentiating both sides, and then solving for g_t^*. The following figure shows g_t^* against g_t for the full sample. The multiple values of g_t^* for every g_t are due to the different estimated values of θ_i for each subject.

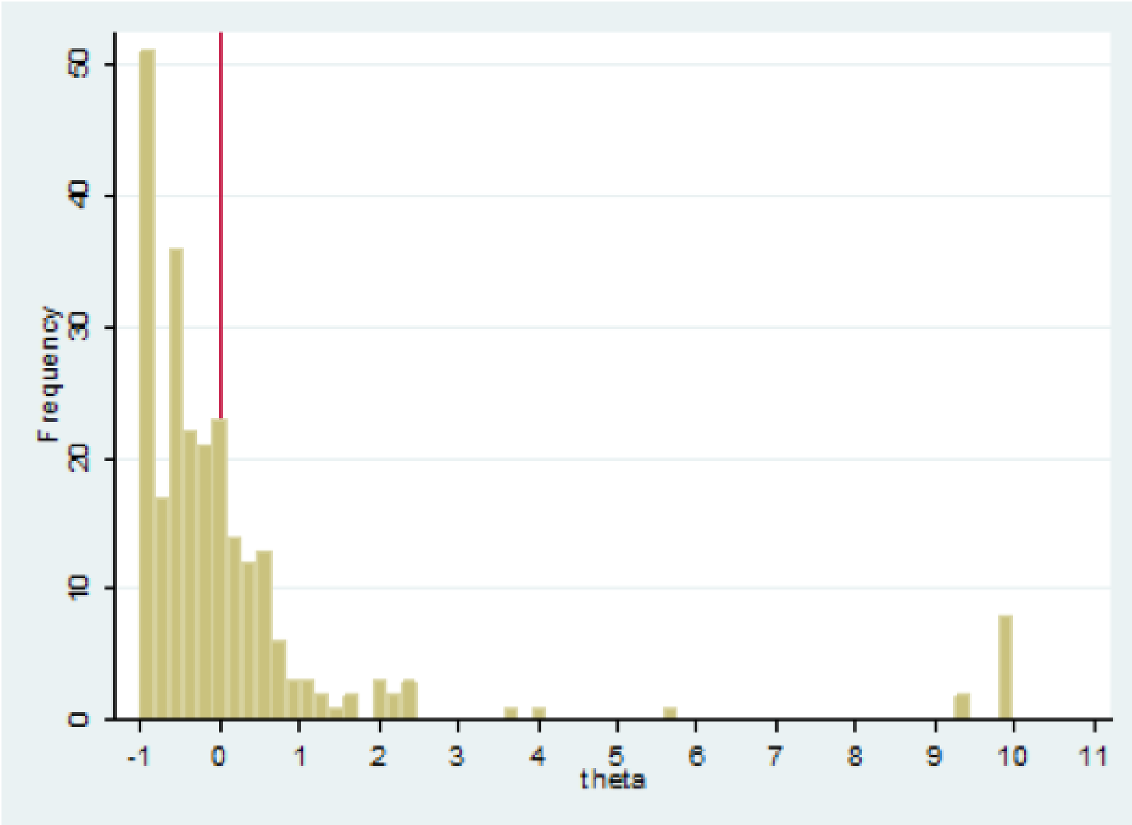

Figure 5: Distribution of θ over the 245 subjects.

C IV Estimator

As mentioned in subsection 4.2, there is an endogeneity problem with using the strength of evidence against the previously chosen value as an explanatory variable in the first hurdle, and in this appendix we resolve this problem. The problem is that the variable is endogenous, because subjects with a low propensity to update are clearly likely to generate large values of $|z_{it}|$ simply by virtue of rarely updating. Hence $|z_{it}|$ always appears to have a perverse negative effect on the propensity to update.

We proceed using an IV estimator. We first create a prediction of the absolute value of z_{it}, $\widehat{|z_{it}|}$, for use in the second stage of a two-stage least squares estimation. The two instruments for $|z_{it}|$ that we use to create this predictor are the round number (t), and the absolute value of the contribution to $|z_{it}|$ in the *current* round $|\Delta z_{it}|$. This is not to be confused with the difference built into z_{it} which spans the current period to period m, namely, $\Phi^{-1}(P_{it}) - \Phi^{-1}(P_{im})$. We note that, because $\Phi^{-1}(P_{im})$ is fixed, a difference operator will eliminate it, leaving Δz_{it} as the change in $\Phi^{-1}(P_{it})$ over the last period.

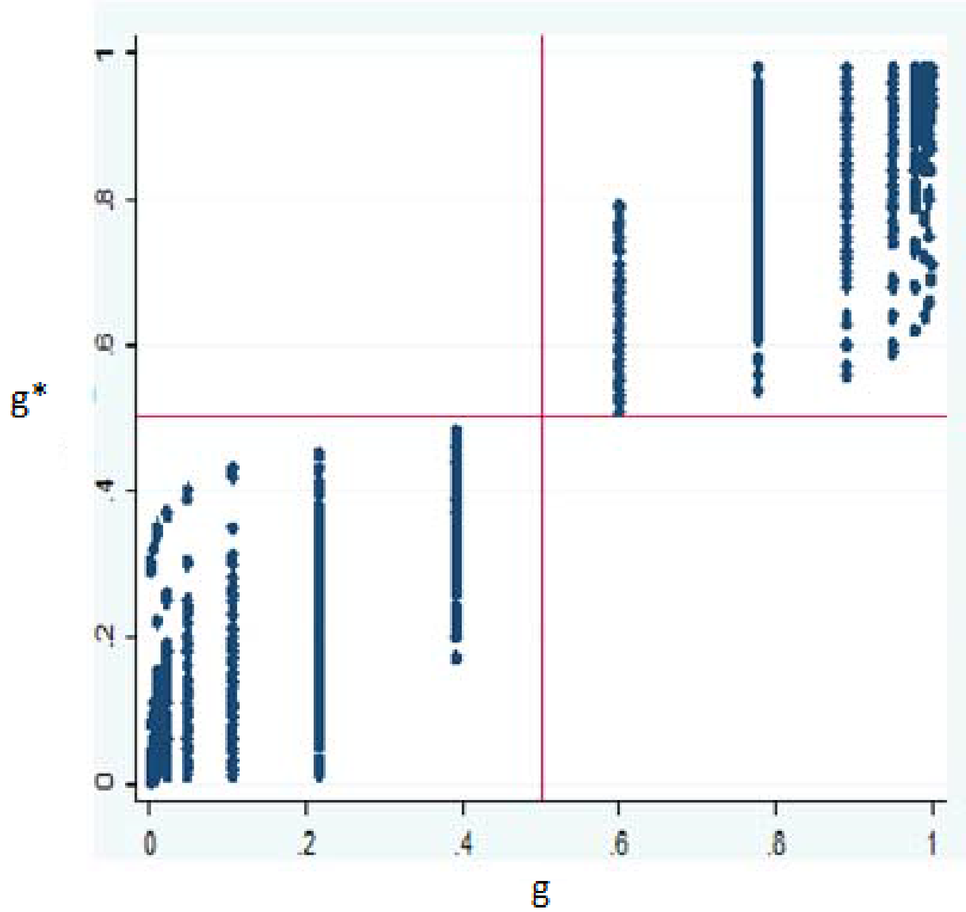

Figure 6: g_t^* against g_t for the full sample.

The stage 1 OLS regression therefore is:

$$|z_{it}| = \pi_0 + \pi_1 t + \pi_3 |\Delta z_{it}| + \varepsilon_{it}. \tag{18}$$

It is important that the dependent variable in (18) is $|z_{it}|$ and not z_{it}. If z_{it} were used as the dependent variable, and consequently the absolute value of the prediction, $\widehat{z}_{it}$, were used in the second stage, we would have what Wooldridge (2010, pp 267-8) refers to as a "forbidden regression". Using $|z_{it}|$ as the dependent variable in (18) avoids this problem.

The results are shown below. Both variables show strong significance in the expected direction, implying that the weak instrument problem is avoided.

Having estimated the stage 1 regression we obtain the predicted values, $\widehat{|z_{it}|}$, and use these in place of $|z_{it}|$ in the first hurdle of the main model. To underline the importance of the instruments, we show two plots below. The left panel of Figure

Source	SS	df	MS		
Model	1121.54757	2	560.773786		
Residual	1145.09497	13, 717	.083479986		
Total	2266.64254	13,719	0.165219225		
cumtbayes_r s	Coef.	Std. Err.	t	P > $\|t\|$	[95% Conf. Interval]
dtbayes_r_abs	.9154853	.0089259	102.57	0.000	.8979894 - .9329812
round	.0131963	.0012816	10.30	0.000	.0106842 - .0157084
_cons	.0502277	.0079752	6.30	0.000	.0345952 - .0658602
Number of obs	13,720				
F(2, 13717)	6717.46				
Prob > F	0.0000				
R-squared	0.4948				
Adj R-squared	0.4947				
Root MSE	.28893				

Table 4: Estimation of (18) in Appendix C

7 shows the estimated probability of updating against $|z_{it}|$, while the right panel Figure 7 displays the estimated probability of updating against $\widehat{|z_{it}|}$.

The left panel makes clear the endogeneity problem identified above: over most of the range of $|z_{it}|$ its effect on the propensity to update is negative. The right panel is of the same binary variable against $\widehat{|z_{it}|}$ (the prediction from the stage 1 regression). It shows completely the opposite pattern: a monotonically increasing effect of $\widehat{|z_{it}|}$ on the propensity to update.

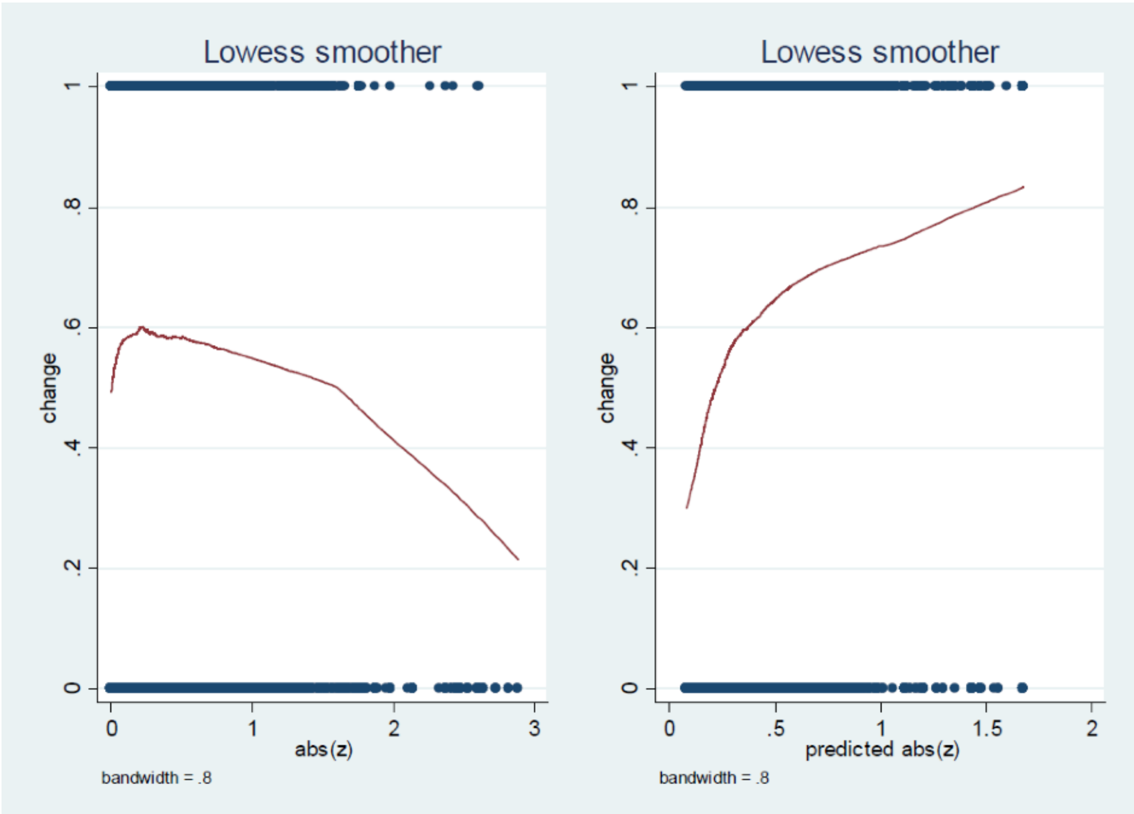

Figure 7: Estimated probability of updating