Revealed Beliefs and the Marriage Market Return to Education[*]

Alison Andrew[†] & Abi Adams[‡]

November 11, 2024

1 Practical Design & Analysis Considerations

This note provides some practical design considerations, and suggestions for robustness checks to perform in analysis, for researchers looking to implement the approach outlined in Andrew and Adams (2022). It is intended as a complement to Section 2 of the main paper, which provides the formal assumptions underlying our approach, and our demonstration simulation and estimation code.[1] This list is not meant to be comprehensive but rather a useful starting point for researchers hoping to tailor a revealed belief measurement approach to their own context.

1.1 Experiment Design

1. **Decision Making Model.** The first thing carefully to consider is whether the decision making process under study can be expressed as a model that is consistent with assumptions A0, A1 and A2. The underlying model must be a finite-horizon, dynamic discrete choice model with features

[*]We are enormously grateful to Abhishek Gautam, Hemlata Verma, Ronak Soni and Amit Kumar for invaluable support in developing and piloting this instrument and to Kuhika Seth for collecting insightful qualitative data. We would like to thank the Centre for Public Policy at the Institute for Fiscal Studies, the Children's Investment Fund Foundation and the John Fell Fund, University of Oxford for generous financial support. Additional research support was provided from the Leverhulme Trust (RPG-2021-391) and the European Research Council (grant nos. 695300 and 948070). Ethical clearance for the project was granted by the Sigma IRB Committee (New Delhi), ICRW IRB (Washington DC), and Oxford University IRB (Oxford, UK).

[†]University of Oxford & Institute for Fiscal Studies, alison.andrew@economics.ox.ac.uk.

[‡]University of Oxford & Institute for Fiscal Studies, abi.adams@economics.ox.ac.uk.

[1]The demonstration code is hosted at `https://github.com/revealedbelief/revealedbelief`

described in A0, with preferences described in A1 and beliefs described in A2. Ensuring that the underlying model fits these assumptions will involve considering: What should the relevant time period be? Is there a terminal action at every decision node? Can preferences be written as an additive sum of utility of the observed path taken through the model and the discounted sum of idiosyncratic preference shocks?

2. **Ex-Post Experiment Design.** This requires designing a discrete choice experiment that asks respondents to make binary choices between all paths through the model that end in a terminal action. In the notation of our paper, we refer to these paths as $\bar{\Psi}_j$. In some cases where researchers want to allow that no terminal action is chosen before the last period (e.g. in our case that the daughter remains unmarried in the last period), it may be useful to also include paths that end in the terminal period without the terminal action is being chosen (e.g. in our case, paths where the daughter is age 23, with education E and still unmarried).

3. **Ex-Ante Experiment Design.** This requires designing a discrete choice experiment that asks respondents to make discrete choices corresponding to choice at given decision nodes of the model. These experiments will specify the time-invariant and deterministic state variables that define the decision node $\bar{\omega}_t$ (in our case, age, education etc), the characteristics feeding into the stochastic variable (in our case characteristics of the groom and match). Respondents should be asked to choose between the terminal action and all other actions that are available at that node.

4. **Visual Aid.** Depending on the complexity of the choice problem and the number of characteristics that researchers want to specify in the vignettes, a visual aid may be appropriate to reduce the cognitive load on respondents.

5. **Scenario Heterogeneity.** Researchers should consider which elements of heterogeneity they want to build into the model and whether these should enter preferences, beliefs or both. Characteristics driving heterogeneity in preferences should be included in both the ex-post and ex-ante experiments whereas characteristics driving heterogeneity in beliefs should only be reflected in the ex-ante experiments. In our case, this involved building in heterogeneity in a daughter's behavior (whether or not they complied with gender norms), in parental wealth, in the need for domestic help at home and in the daughter's attitude towards school.

6. **Anchoring Scenario Characteristics.** Where possible, researchers will find it useful to anchor

the definitions and descriptions of scenario characteristics to those used in observational data from the same (or similar) samples. Having exact analogues of these characteristics in the observational data will help test for consistency between the experimental results and patterns in the observational data. For instance, in our case, we anchored descriptors of the wealth of households to quintiles of the observed asset ownership of households in our sample.

7. **Sampling Choice Scenarios.** The sampling of experimental choice scenarios will affect the precision of estimates. In general, how researchers approach this will depend on the strength of their priors about preferences and beliefs. If a researcher has a strong prior about the direction and magnitude of preference and belief parameters, it can be useful to weight the sampling of choice scenarios to those where the researcher does not predict a huge difference in how the different options will be valued (i.e. where there is not one clearly preferred option). If they have more diffuse priors they may want to simply ensure a good spread of scenarios across the entire state space; heavily weighing towards choice scenarios where the researcher *expects* respondents to be close to indifference may risk harming power if these priors are wrong. If researchers are hoping identify individual-level heterogeneity, or the parameters characterizing the distribution of such heterogeneity, they could consider building an algorithm to optimally select the next choice scenario to maximise new information. This could involve extending the algorithms of Drake et al. (2022) for preference estimation to the revealed belief case. Researchers can simulate the power and precision of the estimators for a given model set-up and given parameterizations of preferences and beliefs by adapting the demonstration code provided with this paper.[2]

8. **Tests for the Importance of Vignette Salience.** Researchers may wish to test whether respondents' choice behavior depends on the salience of a choice scenario, i.e. how close it is to their own situation. In this case, researchers could consider using scenario characteristics based on a respondent's own situation in one round of the experiment. Researchers should randomly vary which round is the "salient" round.

9. **Sample Size (Respondents and Rounds).** The sample size will clearly affect the power and precision of estimates. Researchers can explore the implications of varying the number of respondents and number of rounds using simulations described in the "sampling choice scenarios" point above. In general, having many rounds per respondent may permit allowing for some individual-

[2] See `https://github.com/revealedbelief/revealedbelief`

level heterogeneity in preferences and/or beliefs. Even if this is not precisely identified at the individual level, its distribution across the sample may be.

10. **Piloting.** We strongly recommend multiple rounds of thorough piloting of these instruments. During this process researchers should check whether the most critical scenario characteristics (as perceived by respondents) are indeed adequately captured in the choice experiments. Open-ended feedback from respondents should be used following trials of the instruments to assess whether respondents interpreted the scenarios in the manner intended, where any misunderstandings arose, whether respondents paid attention to all aspects of the choice scenarios and whether respondents' reported reasoning in making choices was consistent with the underlying structure of the model. We would recommend documenting this stage carefully, to allow researchers to report on instrument development in any later published work.

11. **Measures of Complexity and Understanding Checks.** Researchers could consider following the choice scenarios with measures of how complex the respondent found the task and checks of understanding. These could take the form of a subjective enumerator assessment, a likert-scale based measure directly asking the respondent or a more formal measure of subjective complexity based on binary comparisons as described in Gabaix and Graeber (2023).

1.2 Analysis Robustness

1. **Functional Form.** In many applications, researchers will have to impose a specific functional form on preferences and beliefs for estimation. Researchers should conduct robustness checks to assess whether or not conclusions are sensitive to assumptions made in estimation. First, it is useful to assess robustness to using different functional forms for both preferences and beliefs (we showed such checks in Tables 2 and A4). Second, researchers should explore whether results are sensitive to varying the assumed discount factor (we showed this check in Table A4). Third, researchers may want to explore the impacts of removing the possibility that respondents may be inattentive (we showed this check in Table A4).

2. **Consistency of Taste Shock Scale.** The dynamic choice model underlying the revealed belief approach implies a specific relationship between the scale of the unobserved drivers of choice behavior across the ex-post and ex-ante experiments. In the ex-post experiment, the aggregated taste shocks comprise the discounted sum of per-period taste shocks. In our case with

10 periods and setting $\beta = 0.95$ this is $\nu_{i,j} \equiv \sum_{t=1}^{10} 0.95^t \varepsilon_{i,t,j}(d_{t,j})$.[3] This implies that $\sigma_\varepsilon = \sqrt{\sigma_\nu^2 / \sum_{t=1}^{10} 0.95^{2t}}$. Plugging in our estimate of $\hat{\sigma}_\nu = 1.202$ from the ex-post experiment of gives an implied σ_ε of 0.450 which is well within the confidence interval for our direct estimate of σ_ε from the ex-ante experiment (Table 3).

3. **Patterns of Heterogeneity.** Researchers may have strong priors on the heterogeneity in preference and belief parameters based on observable characteristics of respondents. In this case, it is natural to test whether these priors are borne out in the empirical results. In cases (like our own) where respondents' are asked about how a fictional or typical decision maker would behave, we would typically have weaker priors on the relationship between a respondent's own characteristics and their choice behavior. In this instance, heterogeneity by observable characteristics could reflect different respondents having different reference points for the preferences and beliefs of a typical decision maker. Similarly, heterogeneity by proxies of respondents' own exposure to the type of situations being asked about (e.g. whether the respondent has gone through the marriage process for one of their own children) may be indicative of learning from experience and could suggest that respondents without such experience may have beliefs based on more limited information.

4. **Impact of Vignette Salience .** If researchers built in "salient" choice scenarios into their design then a check of whether choices vary with vignette salience is a natural check of whether respondents were able to position themselves in and consider choice scenarios that were more removed from their own situation. Even when researchers did not explicitly build in such salient scenarios, randomly occurring variation in scenario characteristics leading to some choice scenarios being more similar to a respondents own situation than others, could also be used for this purpose.

5. **Inattention.** We follow Mas and Pallais (2017) in suggesting that researchers estimate the inattention rate directly from the discrete choice experiments. Researchers can then check whether the inattention rate estimated appears plausible and in line with other studies.[4] To assess whether excess cognitive load is a major concern, researchers may wish to check whether the inattention rate varies with respondent characteristics that we might expect to be correlated with cognitive load (such as education) or by round of the experiment to assess whether growing fatigue caused

[3] See Section 2.3.

[4] For instance, Mas and Pallais (2017) estimate an inattention rate of 30.16% when estimating the rate internally using maximum likelihood estimation.

some respondents to become inattentive.

6. **Evidence of Cognitive Overload.** Estimated inattention rate and heterogeneity therein will be informative about whether cognitive overload was a problem. In addition, supplementary measures of cognitive load, complexity and fatigue (either self reported by the respondent or assessed by the interviewer) can be analyzed and correlated with estimated preferences and beliefs to assess whether such problems may have led to attenuation of the patterns found.

7. **Consistency with Other Measurements.** In many cases, it may be useful to check the consistency of the preference and belief estimates with estimates of other objects from the same sample. Exactly what these other objects are and how they are best measured will of course depend on the exact application. In our marriage market context, we collected simpler measures of stated expectations about likely marriage offers at different points in time and preferences on the groom's side of the marriage market as checks of internal consistency. These are described in Section 6.4. Whether these internal consistency checks are qualitative or quantitative in nature will depend on the nature of the measures used and the appropriateness of any further assumptions that would need to be made in order to make the measures quantitatively comparable. For instance, in our case, for our estimates of preferences on the groom's side of the marriage market to be comparable with beliefs on the bride's side, we would have had to make very strong assumptions about the nature of frictions in the marital search process.

8. **Consistency with Observational Data.** If observational data on choice behaviour with scenario characteristics is available, one can check how the trajectories implied by the estimated preference and belief parameters compare to the patterns in observational data. This is most useful when the observational data is from the same sample of respondents but a comparison to similar samples could be used when this is not possible. In Section 6.4, we compared the implied vs. predicted fraction of daughters who were in school and married at each age. Where researchers have built in heterogeneity into the model (in preferences, beliefs, or both) along dimensions that have analogues in the observational data, the magnitude of this heterogeneity can also be compared between the estimated predictions and the observed patterns.

   While we think of this exercise as highly informative about the plausibility of the experimental estimates, we note that researchers shouldn't necessarily expect observed and predicted trajectories to be completely aligned. The predicted trajectories will capture the *expected* trajectories based upon respondents' beliefs at the time when the data was collected. Given that the observed trajectories necessarily result from choices made over a prolonged period of time, any unanticipated aggregated shocks to the distribution of the stochastic variable of interest may cause a divergence that is perfectly compatible with economic theory.

1.3 Possible Extensions Checklist

1. **Preference Shifting Instruments.** In Appendix B4, we show how the use of instruments that shift preferences but are excluded from beliefs allows for the identification of beliefs when the stochastic state variable q has more than two dimensions. i.e. rather than impose that q has two dimensions (e.g. low versus high), one might want to allow for q to be high-dimensional (e.g. low versus medium versus high). In our application, we use whether or not the fictional daughter "likes school" as an instrument to allow us to identify the beliefs over the probability of not receiving a marriage offer at all in a given period.

2. **Own Choice *vs.* Choice of a Fictional or Typical Decision Maker.** In our application, we always asked respondents about the choices they thought that a fictional couple we described to them would make for their fictional daughter. We made this choice to bypass concerns about social desirability bias and to bypass concerns about unobserved heterogeneity driving choice behavior. In other applications, researchers may want to ask respondents directly about what they themselves would choose (in response to different hypothetical scenarios) which would allow researchers to interpret estimates, and heterogeneity by observables, as indicative of respondents own preferences and beliefs rather than their beliefs about typical preferences and beliefs. Researchers could usefully compare the impact of these two framings by randomizing whether respondents are asked about their own choice vs. the choice of a fictional decision maker and comparing the resulting estimates.

3. **Aggregation of Preferences & Beliefs into Collective Choice.** In our application, were interested in understanding the *collective* choice of married couples over their daughter's education and marriage. We thus did not make a distinction between the preferences of beliefs of the husband or wife or seek to understand how these might be combined to form collective preferences and beliefs. In other settings, researchers might find it productive to expand our approach to assess separately the preferences and beliefs of different individuals involved in collective choice

(perhaps in addition to also measuring collective preferences and beliefs). Researchers could consider, for instance, asking the same respondents about their perceptions of what both wives and husbands would separately choose in addition to what their collective choice would be. Researchers could also compare the choices of male and female respondents to such questions to assess whether they have different insights or information about preferences of different groups.

References

Andrew, A. and A. Adams (2022). Revealed beliefs and the marriage market return to education. Technical report, IFS Working Papers.

Drake, M., F. Payró, N. Thakral, and L. T. Tô (2022). Bayesian adaptive choice experiments. Technical report, Mimeo.

Gabaix, X. and T. Graeber (2023). The complexity of economic decisions. *Available at SSRN*.

Mas, A. and A. Pallais (2017). Valuing alternative work arrangements. *American Economic Review 107*(12), 3722–3759.