

# Identifying Causal Effects in Information Provision Experiments

Dylan Balla-Elliott

February 2026

Standard estimators in information provision experiments place more weight on individuals who update their beliefs more in response to new information. This paper shows that, in practice, these individuals who update the most have the weakest causal effects of beliefs on outcomes. Standard estimators therefore understate these causal effects. I propose an alternative local least squares (LLS) estimator that recovers a representative unweighted average effect in a broad class of learning rate models that generalize Bayesian updating. I reanalyze six published studies. In five, estimates of the causal effects of beliefs on outcomes increase; in two, they more than double.

JEL CODES: C26, C9, D83, D9

---

[dbell@uchicago.edu](mailto:dbell@uchicago.edu) University of Chicago, Kenneth C. Griffin Department of Economics.

Thanks especially to Zoë Cullen, Ricardo Perez-Truglia, Alex Torgovitsky, and Max Tabord-Meehan for early feedback and also to Magne Mogstad, Julia Gilman, Santiago Lacouture, Max Maydanchik, Isaac Norwich, Francesco Ruggieri, Sofia Shchukina, Alex Weinberg, Jun Wong, Itzhak Rasooly, Vod Vilfort, Whitney Zhang and many conference and seminar participants at the University of Chicago, UChicago Booth School of Business, and Purdue for helpful comments and suggestions. I am also indebted to Armona, Cantoni, Coibion, Fuster, Kumar, Gorodnichenko, Roth, Settele, Wiswall, Wohlfart, Yang, Yuchtman, Zafar, and Zhang for their useful replication packages. This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1746045. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. A companion R package is available at [dballae Elliott.github.io/lls/](https://github.com/dballae Elliott/lls/).

Information provision experiments have become a standard tool for studying the causal effects of beliefs (Bottan and Perez-Truglia, 2022b; Jensen, 2010; Wiswall and Zafar, 2015). But standard panel and two-stage least squares (TSLS) estimators systematically misrepresent average effects because they overweight individuals who update their beliefs the most. This matters because individuals whose beliefs most strongly affect their choices tend to update their beliefs the least, perhaps because they already sought out information before the experiment began. I propose a local least squares (LLS) estimator that weights all individuals equally. In five of six recent studies I reanalyze, LLS yields substantially larger estimates; in two cases the estimates more than double.

This paper is about experiments that study the causal effects of beliefs: how beliefs affect behavior, policy preferences, and even other beliefs. In these experiments, researchers vary the information (“signal”) shown to participants, then estimate the effect of beliefs on behavior using panel or TSLS regressions. It is well known that such estimators target weighted averages of individual causal effects.<sup>1</sup> In information provision experiments, these weights are proportional to the first-stage effect of information on beliefs.

Strong dependence between belief updating and belief effects makes panel and TSLS estimators substantially misrepresent average effects. When belief updating is negatively correlated with belief effects, standard panel or TSLS estimators can severely understate the average effect. The central empirical finding of this paper is that belief effects and belief updating are systematically negatively correlated: individuals whose beliefs most strongly affect their choices tend to update their beliefs least when provided new information.

I therefore propose a local least squares (LLS) estimator that consistently estimates an unweighted average effect, even when there is strong dependence between belief updating and belief effects. Researchers may prefer targeting an unweighted average as it is a representative summary of heterogeneous effects.<sup>2</sup> This estimator can be applied to

---

<sup>1</sup>The weighted average interpretation of TSLS follows from Imbens and Angrist (1994). Similar results apply to difference-in-differences and other settings (Callaway and Sant’Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2020).

<sup>2</sup>In an early application, Guenther and Nunnari (2025) use results from the working paper version of this

panel, active control, and passive control experiments.<sup>3</sup>

I apply the LLS estimator to six recent information provision studies published in leading economics journals.<sup>4</sup> In five of these six applications, the LLS estimates are meaningfully larger than the panel or TSLS estimates. In two cases the estimates more than double. To study mechanisms, I show how LLS can also be used to estimate effects of beliefs on outcomes conditional on the learning rate. Empirically, belief effects are generally larger for the groups with smaller learning rates. A simple model of endogenous information acquisition can rationalize this pattern. People whose beliefs strongly affect their decisions are incentivized to form precise priors; when researchers provide new information, they update only modestly. When beliefs matter less, people start with noisier priors and update more.<sup>5</sup>

The identification arguments in this paper use results in correlated random coefficients models from Masten and Torgovitsky (2016) and Graham and Powell (2012), generalized here to a nonparametric potential outcomes framework. Vilfort and Zhang (2025) study TSLS in information provision experiments and provide conditions under which TSLS targets *some non-negatively weighted average*. This paper proposes an alternative to TSLS that targets the *equally-weighted average*.

The remainder of this paper is organized as follows. Section 1 develops the conceptual paper (Balla-Elliott, 2025). They use LLS because unequal weights cause “2SLS [to] substantially misrepresent average effects” and so “we adopt [LLS] which identifies the unweighted average effect (p. 18-19).”

<sup>3</sup>The LLS estimator applies immediately in the panel experiment. In experiments with active control groups, the LLS estimator identifies an unweighted average under a learning rate updating assumption. In experiments with passive control groups, the LLS estimator identifies the unweighted average when the variance of the prior is elicited in addition to the mean and the learning rate comes from Bayesian updating. An alternative approach with a passive control imposes the strong assumption that covariates are sufficiently rich to predict the belief update and that there is no residual variation in beliefs that cannot be predicted (i.e. “selection on observables”).

<sup>4</sup>These applications span diverse contexts: college major choice (Wiswall and Zafar, 2015), housing investment (Armona et al., 2019), gender policy preferences (Settele, 2022), household (Roth et al., 2022) and firm (Kumar et al., 2023) responses to macroeconomic uncertainty, and protest participation (Cantoni et al., 2019). These six studies include examples of within-person panel experiments, and between person experiments with both active and passive control groups.

<sup>5</sup>Maćkowiak and Wiederholt (Forthcoming) consider a similar model with rational inattention before and during the experiment. Since they argue that the rational inattention dynamics *before* the experiment dominate, their results are consistent with the model proposed in Appendix B that does not include rational inattention during the experiment.

framework. Section 2 shows that standard panel and TSLS estimators target weighted averages of individual slopes; panel regressions place negative weights on individuals with belief updates between zero and the mean. Section 3 proposes the LLS estimator, which identifies an unweighted average. Section 4 shows that under linearity, the unweighted average slope is a structural parameter: it characterizes the effect of beliefs on outcomes for any change in beliefs, not just those induced by the experiment. Section 5 shows that attenuation is empirically widespread. Section 6 concludes.

## 1. Conceptual Framework and Identifying Assumptions

This paper is about experiments that study how beliefs affect behavior. I analyze three leading experimental designs: panel experiments that compare the same individual before and after information provision, active control experiments that compare individuals receiving different signals, and passive control experiments that compare treated individuals to an untreated control group.<sup>6</sup>

The identification argument follows a simple causal chain: treatment assignment  $Z$  determines the signal  $S$  shown to participants, which affects their beliefs  $X$ , which in turn affects outcomes  $Y$ . This  $Z \rightarrow S \rightarrow X \rightarrow Y$  structure allows us to study how exogenous variation in information provision translates into belief changes and ultimately behavioral responses. I formalize this causal chain in three parts: the outcome equation that links beliefs to behavior, the experimental designs that generate exogenous variation in beliefs, and the identifying assumptions that permit causal inference.

---

<sup>6</sup>In between-subject experiments (with active or passive controls), I will focus on experimental designs where the information treatment is quantitative, for example “12 percent of the US population are immigrants” (Grigorieff et al., 2020; Hopkins et al., 2019) and not treatments that are qualitative, for example “[t]he chances of a poor kid staying poor as an adult are extremely large” (Alesina et al., 2018). The results for within-person (panel) experiments extend to qualitative or other kinds of signals.

## 1.1. Potential Outcomes

The outcome equation allows for arbitrary heterogeneity in how beliefs affect outcomes:

$$Y_i = G_i(X_i) \tag{1}$$

where  $Y_i$  is the outcome or behavior of interest,  $X_i$  is the belief, and  $G_i(\cdot)$  is the individual-specific response function. The function  $G_i(\cdot)$  generates potential outcomes:  $Y_i(x) = G_i(x)$ , where  $Y_i(x)$  is  $i$ 's potential outcome when beliefs are exogenously set to  $x$ . This formulation places no restriction on treatment effect heterogeneity; agents can differ both in their average responsiveness to beliefs and in the shape of their response functions.

We assume that beliefs  $X_i$  are endogenous in the sense that  $\mathbb{E}[Y_i | X_i = x] \neq \mathbb{E}[G_i(x)]$  for at least some  $x$ . This says that the difference in outcomes at two values of  $X$  is not a causal effect.<sup>7</sup> This occurs when unobserved determinants of outcomes also affect beliefs.

## 1.2. Experimental Designs

This paper considers three broad classes of information provision experiments. The first design uses within-person panel variation.

**Panel:** The panel design uses within-individual contrasts before and after the information treatment. The first-stage variation in beliefs induced by treatment is the individual difference between beliefs before and after the information treatment.

The second and third designs use between-person variation, but differ in the construction of the control group.

**Active Control:** The active control design uses contrasts between individuals who see a “high” signal and those who see a “low” signal. The first-stage

---

<sup>7</sup>If  $G_i(x) = cx + U_i$  this is a familiar expression of endogeneity bias  $\mathbb{E}[U_i | X_i] \neq \mathbb{E}[U_i]$ .

variation in beliefs induced by treatment is the individual difference between potential beliefs if shown the “high” signal instead of the “low” signal.

**Passive Control:** The passive control design uses contrasts between individuals who receive a signal and those who do not. The first-stage variation in beliefs induced by treatment is the individual difference between potential beliefs if shown the signal instead of not being shown the signal.

Within and between person designs use different kinds of identifying variation and rely on qualitatively different kinds of identifying assumptions. The within person design uses the panel structure on the outcome and does not rely on any assumption on how people update beliefs in response to new information. In contrast, between person designs use assumptions on belief updating to match treatment units to the appropriate control units.<sup>8</sup>

### 1.3. Panel Identifying Assumption

The identifying assumption is that outcomes follow a “panel” form. Let time  $t$  have two periods, denoting pre ( $t = 0$ ) and post ( $t = 1$ ) information provision. Then, let

$$Y_{it} = G_i(X_{it}) + \gamma_t \quad (2)$$

The response function  $G_i(\cdot)$  is time-invariant but arbitrarily heterogeneous across individuals; the time effects  $\gamma_t$  are additively separable. This is a nonparametric generalization of the standard panel model used in the literature (e.g. Armona et al., 2019; Wiswall and Zafar, 2015). The special case  $G_i(x) = \tau_i x + U_i$  generates the classic linear panel model  $Y_{it} = \tau_i X_i + U_i + \gamma_t$  with heterogeneous treatment effects.

The identifying assumption is that different changes in outcomes are due only to differ-

---

<sup>8</sup>In principle, panel and active control designs could be combined by eliciting pre-treatment outcomes in an active control experiment. Exploring the identification implications of such hybrid designs is beyond the scope of this paper but is an interesting direction for future research.

ent changes in beliefs.<sup>9</sup> There are no assumptions on how beliefs are updated; researchers who do not wish to place structure on belief updating may find the panel design particularly appealing.

#### 1.4. Active and Passive Control Identifying Assumption

In active and passive control experiments, the relationship between beliefs and outcomes is completely flexible. The identifying assumption is that belief updating follows a simple learning rate structure. This includes the workhorse linear updating or “signal averaging” models like Bayesian updating. Randomization to a particular signal generates variation in posterior beliefs through this learning rate updating.

##### 1.4.1. Learning Rate Belief Updating

For exposition, the main text uses the familiar linear form throughout; potential beliefs are a linear function of the prior  $X_i^0$  and an experimental signal  $s$ :

$$X_i(s) = \alpha_i (s - X_i^0) + X_i^0 \quad (3)$$

The heterogeneous coefficient on the signal  $\alpha_i$  is often called the learning rate. In this model, posterior beliefs are a weighted average of the prior and the signal, with weight  $\alpha_i$  on the signal. This updating rule is widely used in applied work and fits observed belief changes well in information provision experiments.<sup>10</sup>

This linear updating rule is often microfounded in a normal-normal Bayesian updating, but it also arises in several other behavioral models. This class of linear updating models

---

<sup>9</sup>The time trend  $\gamma_t$  is commonplace in empirical practice (Armona et al., 2019; Wiswall and Zafar, 2015). This allows for all respondents to, for example, respond with a higher number when the outcome is re-elicited, perhaps because of salience or other behavioral factors. The time trend  $\gamma_t$  can be interacted with observables  $W_i$  to allow for these time trends to vary across observables, like the prior belief. Without a time trend, the model implies that outcomes should not change when beliefs do not change:  $\mathbb{E}[\Delta Y_i | \Delta X_i = 0] = 0$ . This restriction is testable in the data.

<sup>10</sup>See for example Cavallo et al. (2017), Cullen et al. (2023), Cullen and Perez-Truglia (2022), Fuster et al. (2022), and Giacobasso et al. (2022).

includes rational inattention (Fuster et al., 2022), base-rate neglect, over-reaction, under-reaction (Grether, 1980) and anchoring on the prior or signal (Gabaix, 2019).<sup>11</sup> See Appendix A for further discussion.

#### 1.4.2. Randomization and Potential Beliefs

Denote treatment arms by  $Z_i$ . In the active and passive control designs, assume that the researcher randomizes over two arms  $Z_i \in \{A, B\}$ . In the active design, arm  $A$  will be the treatment arm that receives the “high” signal and arm  $B$  will be the treatment arm that receives the “low” signal. In the passive design, arm  $A$  will be the treatment arm that receives a signal and arm  $B$  will be the control arm that does not receive a signal. Finally,  $S_i(z)$  is the signal that is shown to individual  $i$  in treatment arm  $z$ .<sup>12</sup>

Treatment is assigned randomly in the sense that  $Z_i$  is independent of the potential outcomes: the outcome function  $G_i(\cdot)$ , the prior  $X_i^0$ , the potential signals  $S_i(\cdot)$ , and the learning rate  $\alpha_i$ .<sup>13</sup> In passive designs, treatment arm  $B$  does not receive any signal. For the sake of completeness, define  $S_i(B) \equiv X_i^0$  in passive designs. It will be convenient to work with the following shorthand where potential beliefs are directly a function of the treatment assignment  $z$ . In a slight abuse of notation, we redefine

$$X_i(z) \equiv X_i(S_i(z)) = \alpha_i (S_i(z) - X_i^0) + X_i^0 \quad (4)$$

<sup>11</sup>Linearity in belief updating can be relaxed as long as differences in updating are still driven only by the learning rate. Nonlinear learning rate models take the form  $X_i(s) = \alpha_i f(s, X_i^0) + X_i^0$ , where  $f(\cdot, X_i^0)$  is any function monotonic in the signal with  $f(X_i^0, X_i^0) = 0$ . For example,  $f$  could be a nonlinear “dampener” that discounts signals further away from the prior. Or, it could be asymmetric around zero so that people respond more to signals of a particular sign. The remainder of the paper uses the linear updating rule with  $f(s, X_i^0) = s - X_i^0$  due to its overwhelming popularity in practice and because it can be microfounded in many popular models of belief updating.

<sup>12</sup>In the panel design, the researcher may randomly assign  $Z_i$  in the same way, or may choose to show the information to all participants. If the panel design includes a treatment arm that receives no information, denote that arm with  $B$ . Since the panel design uses within-person contrasts, identification does not come from randomization across people. Thus it is sufficient to work with the realized signal  $S_i$ .

<sup>13</sup>While the treatment  $Z_i$  will be randomly assigned, it is important to note that the realized signal  $S_i(Z_i)$  can generally vary across individuals endogenously. In Bottan and Perez-Truglia (2022a),  $S_i(A)$  and  $S_i(B)$  are high and low estimates of the home value and thus the realized signal is only randomly assigned conditional on the potential signals.

We will use this equation for potential beliefs along with the potential outcome equations (1) and (2) to study common empirical specifications.

## 2. Standard Panel and TSLS Estimators

The following three sections compare standard estimators to a local least squares (LLS) alternative. Standard estimators weight individuals by their belief updating; LLS weights all individuals equally. When belief updates are negatively correlated with causal effects, standard estimators understate the average effect. The current section begins by introducing the individual slopes, which are the causal building block of all the estimators considered in this paper, and then shows that standard panel and TSLS estimators recover weighted averages of these slopes.

### 2.1. Individual Slopes: The Causal Building Block

Define the individual slope as the ratio of outcome change to belief change induced by the experiment:

$$\beta_i \equiv \frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \quad (5)$$

This is the average rate of change in individual  $i$ 's outcome as beliefs move from  $X_i(B)$  to  $X_i(A)$ , which are the individual-specific beliefs in treatment arms  $A$  and  $B$ . Equivalently, this is the individual-specific average partial effect  $G'_i(x)$  over the individual-specific interval of beliefs induced by the experiment. These individual slopes  $\beta_i$  thus depend both on the individual response function  $G_i(\cdot)$  and the variation in beliefs induced by the experiment  $\{X_i(B), X_i(A)\}$ . In the panel design, define  $X_i(A) \equiv X_{i1}$  and  $X_i(B) \equiv X_{i0}$ . The standard estimators used in the literature and the new LLS estimator aggregate these individual slopes differently. The differences between the parameters targeted by LLS and TSLS or panel estimators come *entirely* from differences in aggregation. The remainder of this section characterizes standard panel and TSLS estimators.

## 2.2. Standard Panel and TSLS Specifications

Standard estimators in information provision experiments yield weighted averages of individual effects  $\beta_i$ , with weights proportional to belief updating. In panels, individuals with belief updates between zero and the mean have negative weights.

$$\beta^{design} \equiv \mathbb{E} [\beta_i \times \omega_i(design)] \quad (6)$$

The precise form of these weights varies, but in all three cases, standard specifications weight individual effects  $\beta_i$  in proportion to the first-stage belief updating. In all specifications, these weights integrate to one. Online Appendix C.1 contains derivations for all expressions in this section and Online Appendix F provides a more general discussion of TSLS in information experiments.

We now examine three representative specifications and derive the implicit weights each places on different individuals.

### 2.2.1. A Representative Panel Specification

Armona et al. (2019) use a regression in first-differences. Since there are only two time periods, this is equivalent to a panel regression with individual and time fixed effects. Let  $\Delta X_i$  denote the difference between the post- and pre-treatment observations,  $X_{i1} - X_{i0}$ . The regression specification is simply

$$\beta^{Panel} \equiv \frac{\text{Cov} [\Delta Y_i, \Delta X_i]}{\text{Var} [\Delta X_i]} \quad (7)$$

which has implied weights

$$\omega_i(Panel) \propto \Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i]) \quad (8)$$

The regression of  $\Delta Y_i$  on  $\Delta X_i$  and a constant can assign negative weights to observations

with  $\Delta X_i$  between zero and the mean  $\mathbb{E}[\Delta X_i]$ .

*Heterogeneity Bias Causes Negative Weights in Panel Regressions.* This negative weights result restates Chamberlain’s classic (1982) “heterogeneity bias” as negative weights in a weighted average of individual effects. A closely related expression appears in Theorem 3.4c of Callaway et al. (2025), who show that units with below-mean treatment intensity receive negative weights in difference-in-differences with continuous treatment. The panel regression here is analogous: it compares outcomes for big changers to small changers. Small changers act as the control group and their outcomes are subtracted from outcomes for big changers. Increasing the treatment effects of small changers thus *decreases* the slope estimate. This is what it means for them to have negative weights. Heterogeneity bias arises because these cross-update comparisons are contaminated by differences in treatment effects.

### 2.2.2. A Representative Active Control Specification

Settele (2022) uses an IV specification where assignment to the “high” signal  $T_i \equiv \mathbb{1}\{Z_i = A\}$  is a binary instrument for beliefs. The estimand takes the canonical Wald form:

$$\beta^{Active} \equiv \frac{\mathbb{E}[Y | Z = A] - \mathbb{E}[Y | Z = B]}{\mathbb{E}[X | Z = A] - \mathbb{E}[X | Z = B]} \quad (9)$$

$$\omega_i(Active) \propto X_i(A) - X_i(B) \quad (10)$$

which under learning rate updating simplifies further to

$$\omega_i(Active) \propto \alpha_i(S_i(A) - S_i(B)) \quad (11)$$

These weights are non-negative under learning rate updating with  $\alpha_i \geq 0$  and in a general class of updating models when a monotonicity assumption holds such that  $(X_i(A) -$

$X_i(B))$  has the same sign for everyone.

### 2.2.3. A Representative Passive Control Specification

Cullen et al. (2023) use an IV specification where the instrument is an indicator for assignment to the information treatment interacted with the initial gap in beliefs.<sup>14</sup>

$$T_i^{ex} \equiv T_i(S_i(A) - X_i^0) \quad (12)$$

Since these specifications control for the exposure  $S_i(A) - X_i^0$ , the residual variation in the instrument is simply a re-centered version of the instrument.<sup>15</sup>

$$\tilde{T}_i^{ex} \equiv (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \quad (13)$$

The TSLS coefficient is then given by

$$\beta^{Passive} \equiv \frac{\text{Cov}[\tilde{T}_i^{ex}, Y_i]}{\text{Cov}[\tilde{T}_i^{ex}, X_i]} \quad (14)$$

$$\omega_i(Passive) \propto (X_i(A) - X_i(B))(S_i(A) - X_i^0) \quad (15)$$

which under learning rate updating simplifies further to

$$\omega_i(Passive) \propto \alpha_i(S_i(A) - X_i^0)^2 \quad (16)$$

These weights are non-negative under learning rate updating with  $\alpha_i \geq 0$  and in a general class of updating models when monotonicity holds:  $\text{sign}(X_i(A) - X_i(B)) = \text{sign}(S_i(A) - X_i^0)$ .

---

<sup>14</sup>Vilfort and Zhang (2025) point out that similar specifications that also include the treatment indicator as an excluded instrument have negative weights.

<sup>15</sup>To see this, notice that random assignment implies that  $\mathbb{E}[T_i^{ex} | S_i(A) - X_i^0] = \mathbb{E}[T_i](S_i(A) - X_i^0) = \mathbb{E}[T_i^{ex} | S_i(A) - X_i^0]$ . By FWL  $\tilde{T}_i^{ex} \equiv T_i^{ex} - \mathbb{E}[T_i^{ex} | S_i(A) - X_i^0] = (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0)$ .

### 2.3. Discussion

The key takeaway from these expressions is that these standard specifications weight individual effects by the strength of belief updating. In the active and passive controls, weights are non-negative and thus the resulting estimands are “weakly causal” (Blandhol et al., 2025).

## 3. The Local Least Squares Estimator

This section presents a local least squares (LLS) estimator that recovers an equally weighted average of individual belief effects. With a linear outcome equation, these individual slopes have a structural interpretation as partial derivatives of the outcome with respect to beliefs and so the equally weighted average is the (structural) average partial effect (APE).

LLS is a control function estimator. It works by constructing a vector of controls that isolates the experimental variation in beliefs. In this setting, learning rate updating means that people who have the same prior, the same potential signals, *and the same learning rate* have the same potential beliefs; the only variation in their actual beliefs comes from the random assignment to the actual signal. The LLS approach aggregates many “local” regressions that use only this exogenous (i.e. experimental) variation in beliefs.<sup>16</sup>

### 3.1. Intuition: Conditioning on Potential Beliefs

The LLS estimator recovers equally weighted averages of individual slopes  $\mathbb{E}[\beta_i]$  by constructing local regressions that isolate purely experimental variation in beliefs. The ideal regression conditions on the potential beliefs  $X_i(A)$  and  $X_i(B)$ , which isolates only the remaining variation in beliefs that comes from being assigned randomly to treatment  $A$  or  $B$ . This ideal regression is:

---

<sup>16</sup>Graham and Powell (2012) and Masten and Torgovitsky (2016) show how to construct these “local” regressions in panel and IV settings more generally. I generalize their results from the linear random coefficients model to a more general nonparametric potential outcome model.

$$\frac{\text{Cov}[Y_i, X_i \mid X_i(A) = x_A, X_i(B) = x_B]}{\text{Var}[X_i \mid X_i(A) = x_A, X_i(B) = x_B]} = \mathbb{E} \left[ \underbrace{\frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)}}_{\equiv \beta_i} \mid X_i(A) = x_A, X_i(B) = x_B \right]$$

This regression recovers a conditional average  $\mathbb{E}[\beta_i \mid X_i(A) = x_A, X_i(B) = x_B]$ . Iterating expectations thus recovers the average individual slope  $\mathbb{E}[\beta_i]$ .<sup>17</sup> This is an easily interpretable causal parameter: it answers the question, “On average, how much do outcomes change per unit change in beliefs, over the range of beliefs induced by the experiment?”.

The LLS estimation strategy also produces intermediate estimates  $\mathbb{E}[\beta_i \mid \alpha_i]$  that reveal how causal effects vary with belief updating. Many behavioral models make strong predictions about the relationship between belief updating and belief effects (Enke et al., 2024; Fuster et al., 2022; Maćkowiak and Wiederholt, Forthcoming; Yang, 2024). Section 5 presents estimates of these conditional average slopes to document strong negative correlation between belief updates and causal effects across a range of settings.

The identification strategy in practice is then to condition on a set of controls that is as good as conditioning on the potential beliefs directly. The following sections show how to construct feasible local regressions.

### 3.2. Constructing Feasible Local Regressions

The following sections show how to construct feasible local regressions in the three experimental designs. Appendix C.2 provides proofs for the results in this section.

---

<sup>17</sup>When all individuals respond to the signal ( $\alpha_i > 0$  for all  $i$ ), LLS identifies an average over the full experimental sample. When some individuals ignore the signal entirely ( $\alpha_i = 0$ ), LLS identifies the equally weighted average among those who update:  $\mathbb{E}[\beta_i \mid \alpha_i \neq 0]$ . Notice that when  $\alpha_i = 0$  the ideal regression is infeasible since  $X_i(A) = X_i(B)$  which means there is no experimental variation in beliefs. Bayesian updating is a sufficient condition to imply that  $\alpha_i > 0$  for all  $i$ , since under Bayes’ rule everyone updates at least slightly toward the signal (Appendix A.1.4).

### 3.2.1. Local Regressions in Panel Experiments

The panel approach works with any information treatment (including qualitative treatments or bundles of signals) because identification relies only on the panel structure, not on the content of the signal.

For any belief change  $x \neq 0$ :

$$\mathbb{E}[\beta_i | \Delta X_i = x] = \frac{\text{Cov}[\Delta Y_i, \Delta X_i | \Delta X_i \in \{0, x\}]}{\text{Var}[\Delta X_i | \Delta X_i \in \{0, x\}]} \quad (17)$$

The right hand side is a feasible local regression using only observations with  $\Delta X_i = x$  or  $\Delta X_i = 0$ . Iterating over  $x$  and averaging yields  $\mathbb{E}[\beta_i]$ . This requires that some individuals have (close to) zero change in beliefs.<sup>18</sup>

### 3.2.2. Local Regressions in Active Control Experiments

Active designs rely on the Bayesian updating assumption (4) and identify learning rates directly from observed belief updates:  $\alpha_i = (X_i - X_i^0) / (S_i - X_i^0)$ . Under Bayesian updating, people with the same learning rate, prior, and potential signals have the same potential beliefs; the only remaining variation comes from random assignment.

The control vector is  $C_i \equiv [\alpha_i \ X_i^0 \ S_i(A) \ S_i(B)]$ . Conditional on  $C_i = c$ :

$$\mathbb{E}[\beta_i | C_i = c] = \frac{\text{Cov}[Y_i, X_i | C_i = c]}{\text{Var}[X_i | C_i = c]} \quad (18)$$

Iterating over  $c$  and averaging yields  $\mathbb{E}[\beta_i]$ . The regression is feasible when  $(S_i - X_i^0) \neq 0$  and  $\text{Var}[X_i | C_i = c] > 0$ , which excludes cases with no learning ( $\alpha_i = 0$ ) or identical signals ( $S_i(A) = S_i(B)$ ).

---

<sup>18</sup>This is an easily verifiable condition. It is satisfied if  $P[\Delta X_i = 0] > 0$ , or more generally if  $\Delta X_i$  has positive mass in any neighborhood around zero. See Graham and Powell (2012) for detailed discussion of technical considerations with continuous  $\Delta X_i$ .

### 3.2.3. Local Regressions in Passive Control Experiments

Passive designs also rely on the Bayesian updating assumption (4), but require additional assumptions because learning rates for the control group are unobserved. Consider two possible approaches to infer learning rates in the control group:

*Case 1: Observed Prior Variance.* In normal-normal Bayesian updating,  $\alpha_i = \sigma_{X_i}^2 / (\sigma_{X_i}^2 + \sigma_S^2)$ . If signal precision  $\sigma_S^2$  is common across individuals, then conditioning on the rank of prior variance  $\sigma_{X_i}^2$  is equivalent to conditioning on  $\alpha_i$ . The control vector becomes  $C_i \equiv [\text{rank}(\sigma_{X_i}^2) \ X_i^0 \ S_i(A)]$ .

*Case 2: Rich Observables.* When researchers can predict beliefs from observables (Ballar-Elliott et al., 2022; Cantoni et al., 2019), they can use *predicted updates* instead of observed updates. The implied predicted learning rate  $\tilde{\alpha}_i$  replaces the observed rate. The control vector becomes  $C_i \equiv [\tilde{\alpha}_i \ X_i^0 \ S_i(A)]$ .

In either case, under the linear outcome equation (1) and Bayesian updating (4):

$$\mathbb{E}[\beta_i \mid C_i = c] \equiv \frac{\text{Cov}[Y_i, X_i \mid C_i = c]}{\text{Var}[X_i \mid C_i = c]} \quad (19)$$

Online Appendix C.2.3 formally states the assumptions in both of these cases.

### 3.2.4. Comparing Assumptions Across Designs

The three experimental designs require progressively stronger assumptions to implement LLS. Panel designs impose no new behavioral assumptions. Active designs require Bayesian updating. Passive designs require Bayesian updating and also require either elicited prior variances or rich observables to infer unobserved learning rates.

The assumptions in the active case are weaker than in the passive case because in the active case researchers observe all participants update beliefs in response to new

information. The experiment reveals heterogeneity in belief updating. In contrast, in a passive design, researchers need to use observables to infer heterogeneity in belief updating for a control group that the researcher never sees update their beliefs.<sup>19</sup> This suggests that researchers interested in implementing an LLS estimator may find active designs more attractive since they reveal more information about belief updating.<sup>20</sup>

### 3.3. Practical Implementation

Conditioning on high-dimensional control vectors is often impractical in experimental samples. When belief updating is linear in the signal and prior, it is sufficient to control for  $C_i$  semi-parametrically. The local regressions in between-person designs need only condition on the learning rate and can simply control linearly for the prior and signals in each local regression. In passive designs, or designs with person-specific high and low signals (i.e. Roth et al. (2022)), it is also necessary to reweight by the inverse of the exposure. This weighted local regression recovers  $\mathbb{E}[\beta_i | \alpha_i]$ . Online Appendix C.3 shows that this modified local regression is sufficient and Online Appendix D provides general implementation guidance.

## 4. Comparing Estimators and Interpreting Individual Slopes

The estimators in Sections 2 and 3 target parameters that can be written as  $\mathbb{E}[\beta_i \times \omega_i]$  for some weights  $\omega_i$ . The interpretation of these parameters depends on the interpretation of the individual slopes  $\beta_i$ , but the difference between estimators comes only from the weights  $\omega_i$ . LLS assigns equal weights. Under linearity, these equal weights deliver a structural APE that characterizes the average effect of beliefs on outcomes for any change

<sup>19</sup>Recall that the learning rate is identified from the observed update  $\alpha_i = (X_i - X_i^0) / (S_i(Z_i) - X_i^0)$ , which is undefined for the passive control group that receives no information. Randomization is enough to ensure that the learning rates have the same distribution in both groups, but the individual learning rates are not directly identified in the passive control group.

<sup>20</sup>There are many design considerations beyond the scope of this paper. Haaland et al. (2023) discuss implementation considerations of active and passive control designs. List (2025) discusses within- and between-subject experimental designs more generally.

in beliefs, not just those induced by the experiment. Online Appendix G discusses the nonlinear case in greater detail.

#### 4.1. Equal Weights Deliver a Representative Average

With treatment effect heterogeneity, researchers must decide how to summarize heterogeneous effects. LLS recovers a simple average  $\mathbb{E}[\beta_i]$ . This equally weighted average  $\mathbb{E}[\beta_i]$  answers the question: “On average, how much did outcomes change per unit change in beliefs?” Like the non-parametric ATE  $\mathbb{E}[G'_i(X_i)]$ , this parameter is local to the variation the experiment actually induced (Heckman and Vytlacil, 2007). A TSLS-weighted average may be policy-relevant when the intervention under consideration is information provision, since it captures effects among those whose beliefs would actually change.<sup>21</sup> However, the attenuation documented in Section 5 suggests that relying on TSLS outside this narrow case is risky: researchers may conclude that belief effects are generally unimportant on the basis of an unrepresentative average.

#### 4.2. Under Linearity, the Average Slope Permits Extrapolation

In the linear outcome equation  $G_i(x) = \tau_i x + U_i$ , the slope  $\tau_i$  does not depend on the level of beliefs: it fully characterizes  $i$ ’s response to any hypothetical belief shift, not just those induced by the experiment. The average slope  $\mathbb{E}[\tau_i]$  inherits this property, permitting extrapolation; predictions for interventions that shift beliefs by any amount can be formed by scaling  $\mathbb{E}[\tau_i]$  appropriately.<sup>22</sup>

---

<sup>21</sup>If the policy question is whether to implement an information campaign, the reduced form (the effect of treatment assignment on outcomes) answers this directly.

<sup>22</sup>Since  $\tau_i$  is the individual partial effect, the average  $\mathbb{E}[\tau_i]$  is also called the average partial effect (APE).

## 5. Empirical Applications

This section demonstrates that attenuation due to dependence between belief updating and belief effect is empirically relevant. I compare standard panel and TSLS specifications to LLS estimates in six recent studies from leading economics journals.<sup>23</sup> See Online Appendix D for estimation details.

Table 1 contrasts LLS estimate with estimates recovered by the standard specification in each study. In five of the six studies, standard estimators are substantially attenuated. Figure 1 plots an estimate of conditional slopes for each study:  $\mathbb{E}[\beta_i | |\Delta X_i|]$  in panel experiments (Panel A) and  $\mathbb{E}[\beta_i | \text{rank}(\alpha_i)]$  in active and passive control experiments (Panels B and C). These curves directly show that people with the strongest causal effects tend to have smaller belief updates.

### 5.1. Results from Panel Experiments

Wiswall and Zafar (2015) study how beliefs about field-specific earnings affect college students' major choices. The LLS estimate of 0.721 (s.e. 0.29) is roughly 125% larger than the panel estimate of 0.32 (s.e. 0.086). Armona et al. (2019) study how beliefs about home prices affect investment decisions. The LLS estimate of 1.8 (s.e. 0.387) is roughly 50% larger than the panel estimate of 1.15 (s.e. 0.234).

### 5.2. Results from Active Control Experiments

Settele (2022) studies how beliefs about the gender wage gap affect support for gender equality policies. The LLS estimate of 0.16 (s.e. 0.042) is roughly 66% larger than the TSLS estimate of 0.096 (s.e. 0.033). Roth et al. (2022) study how recession expectations affect

---

<sup>23</sup>I searched the Web of Science database for papers in the top five economics journals, ReStat, AER: Insights, and all AEJs containing “beliefs,” “information,” or “perception” together with “experiment” or “treatment.” This yielded 116 potentially eligible experiments. I replicated the two most highly cited studies of each experimental design. To standardize the presentation of the results, I flip the sign of the outcome variable when necessary to ensure that mean effects are always positive. I also omit additional demographic controls and probability weights from all estimates for simplicity.

subjective personal unemployment risk. The LLS estimate of 0.882 (s.e. 0.365) is slightly larger than the TSLS estimate of 0.755 (s.e. 0.435).

### 5.3. Results from Passive Control Experiments

Kumar et al. (2023) study how firms' beliefs about GDP growth affect their employment decisions. The LLS estimate of 1.787 (s.e. 0.465) is nearly four times the TSLS estimate of 0.466 (s.e. 0.19). Cantoni et al. (2019) study how beliefs about others' protest participation affect one's own willingness to participate. Here the pattern reverses: the TSLS estimate of 0.68 (s.e. 0.253) is larger than the LLS estimate of 0.18 (s.e. 0.133). Since TSLS overweights people with larger belief updates, this difference implies that people with larger belief effects also had *larger* belief updates. Admittedly, the CAPE curve in Panel C.ii of Figure 1 reveals only modest variation across learning rate ranks, with wide confidence intervals, so these results should be interpreted suggestively.

### 5.4. Discussion

The conditional average partial effects in Figure 1 reveal that individuals who update their beliefs the least tend to have the strongest causal effects across a range of contexts. This provides direct empirical support for models of endogenous information acquisition where people with decision-relevant beliefs invest in forming precise priors (Appendix B; see also Cavallo et al. (2017), Fuster et al. (2022), and Maćkowiak and Wiederholt (Forthcoming)).

This negative correlation attenuates existing estimates; in five of the six applications, LLS estimates exceed the standard estimates. The exception is Cantoni et al. (2019), where the point estimates suggest a *positive* correlation between belief effects and belief updating. This is plausible in a setting where it is difficult for anyone to form precise beliefs about their classmates' protest intentions, so that the relevant heterogeneity comes from within-experiment attention rather than pre-experimental information acquisition. People whose decisions depend on their beliefs pay more attention to the signal than people whose

decisions do not depend on these beliefs.<sup>24</sup> The LLS estimator recovers the unweighted average effect regardless of the sign of this dependence. Estimating the CAPE is also valuable in either case; the dependence between belief updating and belief effects helps distinguish competing models of how people form beliefs and make decisions.

## **6. Conclusion**

Standard empirical specifications in information provision experiments systematically understate the causal effects of beliefs on behavior. This paper demonstrates that in five of six high-profile studies in leading economics journals, ranging from college major choice to macroeconomic expectations, LLS estimates average effects of beliefs that are larger than estimates from standard specifications.

---

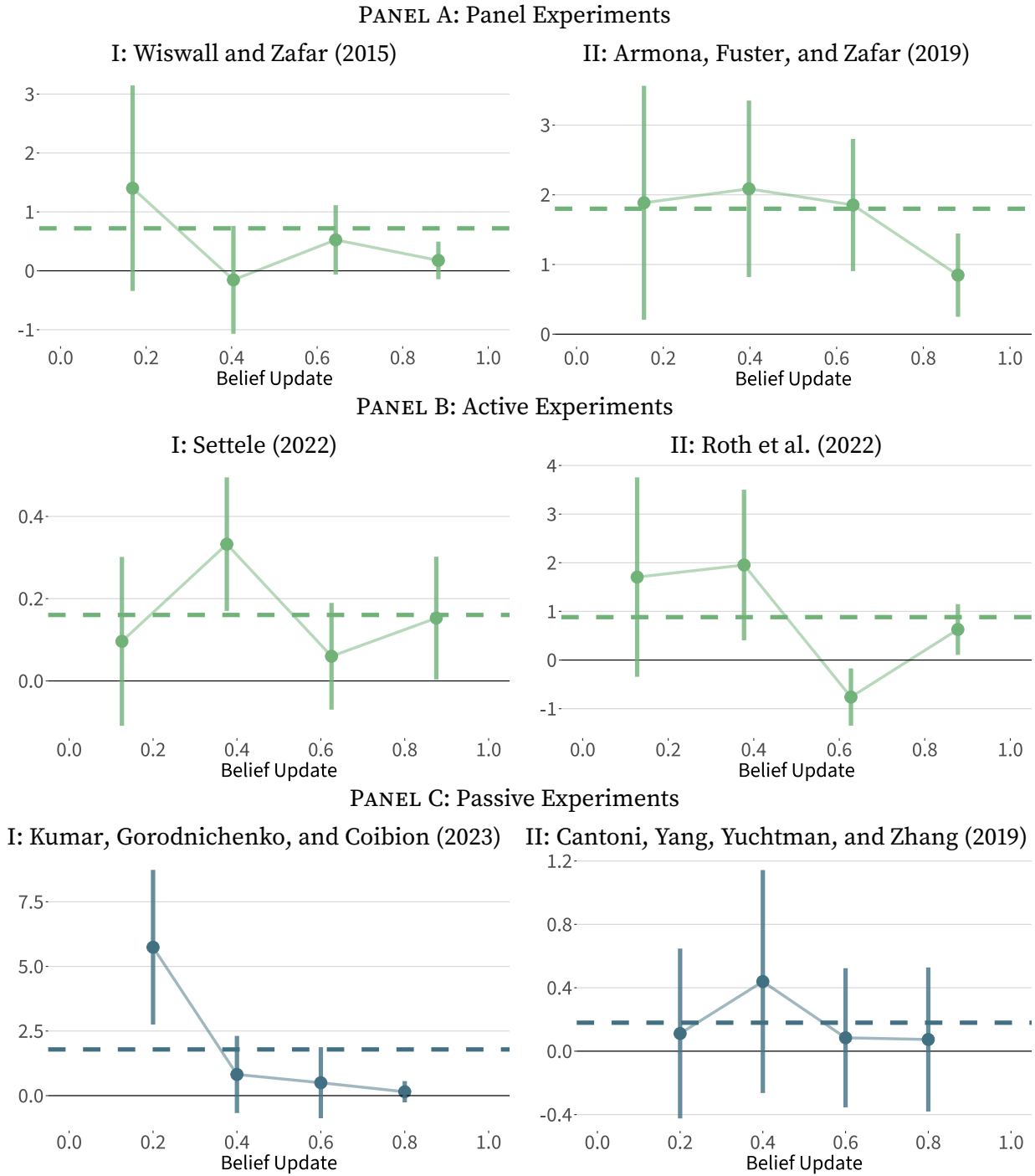
<sup>24</sup>In principle, either pre-experimental information acquisition or within-experiment attention could dominate. When prior information is readily available, the first channel generates a negative correlation; when prior information is scarce, the second channel could generate a positive correlation. Future work could use the tools developed here to help distinguish these mechanisms in a broader range of settings.

TABLE 1. LLS and Standard Specifications in Six Studies

<b>Panel A: Panel Experiments</b>	LLS	FD Regression
Wiswall and Zafar (2015)	0.721 (0.290)	0.320 (0.086)
Armona, Fuster, and Zafar (2019)	1.800 (0.387)	1.147 (0.234)
<b>Panel B: Active Experiments</b>	LLS	TSLS Regression
Settele (2022)	0.160 (0.042)	0.096 (0.033)
Roth, Settele, and Wohlfart (2022)	0.882 (0.365)	0.755 (0.435)
<b>Panel C: Passive Experiments</b>	LLS	TSLS Regression
Kumar et al. (2023) - Employment	1.787 (0.465)	0.466 (0.199)
Cantoni, Yang, Yuchtman, and Zhang (2019)	0.180 (0.133)	0.680 (0.253)

*Notes:* This table compares local least squares (LLS) estimates of the unweighted average effect to standard first-difference (FD) or two-stage least squares (TSLS) estimates across all six replication studies. Bootstrap standard errors are reported in parentheses. Online Appendix D discusses implementation details and reports results for alternative choices of bandwidth.

FIGURE 1. Dependence between Belief Updating and Belief Effects in Six Studies



*Notes:* Each panel plots conditional estimates of the effects of beliefs on outcomes  $\mathbb{E}[\beta_i | \cdot]$ . Panel A (panel experiments) conditions on the absolute value of observed belief changes  $|\Delta X_i|$ . Panels B and C (active and passive control experiments) condition on the rank of the estimated learning rate  $\alpha_i$ , which measures responsiveness to experimental information. In all panels, smaller values on the horizontal axis correspond to individuals who update their beliefs less. Confidence intervals are twice the bootstrap standard error.

## References

- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). “Intergenerational Mobility and Preferences for Redistribution”. *American Economic Review* 108.2, pp. 521–554. DOI: 10.1257/aer.20162015 (p. 3).
- Anderson, Michael L. (2008). “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. *Journal of the American Statistical Association* 103.484, pp. 1481–1495. DOI: 10.1198/016214508000000841 (p. A.31).
- Angrist, Joshua D. and Guido W. Imbens (1995). “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity”. *Journal of the American Statistical Association* 90.430, pp. 431–442. DOI: 10.1080/01621459.1995.10476535 (p. A.40).
- Armona, Luis, Andreas Fuster, and Basit Zafar (2019). “Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment”. *The Review of Economic Studies* 86.4 (309), pp. 1371–1410 (p. 2, 5, 6, 9, 18, 22, i, A.19, A.21, A.29, A.30).
- Balla-Elliott, Dylan (2025). *Identifying Causal Effects in Information Provision Experiments*. DOI: 10.48550/arXiv.2309.11387 (p. 2).
- Balla-Elliott, Dylan, Zoë B. Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton (2022). “Determinants of Small Business Reopening Decisions After Covid Restrictions Were Lifted”. *Journal of Policy Analysis and Management* 41.1, pp. 278–317. DOI: 10.1002/pam.22355 (p. 15, 28, A.11).
- Bhuller, Manudeep and Henrik Sigstad (2024). *2SLS with Multiple Treatments*. DOI: 10.48550/arXiv.2205.07836 (p. A.35).
- Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (2025). *When Is TSLS Actually LATE?* Working Paper 29709. National Bureau of Economic Research. DOI: 10.3386/w29709 (p. 12).
- Bottan, Nicolas L. and Ricardo Perez-Truglia (2022a). “Betting on the House: Subjective Expectations and Market Choices”, p. 53. DOI: 10.3386/w27412 (p. 7).
- (2022b). “Choosing Your Pond: Location Choices and Relative Income”. *The Review of Economics and Statistics* 104.5, pp. 1010–1027. DOI: 10.1162/rest\_a\_00991 (p. 1).
- Brinch, Christian N., Magne Mogstad, and Matthew Wiswall (2017). “Beyond LATE with a Discrete Instrument”. *Journal of Political Economy* 125.4, pp. 985–1039. DOI: 10.1086/692712 (p. A.46).
- Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant’Anna (2025). *Difference-in-Differences with a Continuous Treatment*. DOI: 10.48550/arXiv.2107.02637 (p. 10).
- Callaway, Brantly and Pedro H.C. Sant’Anna (2021). “Difference-in-Differences with Multiple Time Periods”. *Journal of Econometrics* 225.2, pp. 200–230. DOI: 10.1016/j.jeconom.2020.12.001 (p. 1).
- Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang (2019). “Protests as Strategic Games: Experimental Evidence from Hong Kong’s Antiauthoritarian Movement\*”. *The Quarterly Journal of Economics* 134.2, pp. 1021–1077. DOI: 10.1093/qje/qjz002 (p. 2, 15, 19, 22, ii, A.11, A.19, A.25, A.36, A.37).
- Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia (2017). “Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments”. *American Economic Journal: Macroeconomics* 9.3, pp. 1–35. DOI: 10.1257/mac.20150147 (p. 6, 19).
- Chamberlain, Gary (1982). “Multivariate Regression Models for Panel Data”. *Journal of Econometrics* 18.1, pp. 5–46. DOI: 10.1016/0304-4076(82)90094-X (p. 10).

- Cullen, Zoë, Will Dobbie, and Mitchell Hoffman (2023). “Increasing the Demand for Workers with a Criminal Record”. *The Quarterly Journal of Economics*. DOI: 10.1093/qje/qjac029 (p. 6, 11).
- Cullen, Zoë and Ricardo Perez-Truglia (2022). “How Much Does Your Boss Make? The Effects of Salary Comparisons”. *Journal of Political Economy* 130.3, pp. 766–822. DOI: 10.1086/717891 (p. 6, 28).
- Enke, Benjamin, Thomas Graeber, Ryan Oprea, and Jeffrey Yang (2024). *Behavioral Attenuation*. Tech. rep. w32973. Cambridge, MA: National Bureau of Economic Research, w32973. DOI: 10.3386/w32973 (p. 13).
- Fuster, Andreas, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar (2022). “Expectations with Endogenous Information Acquisition: An Experimental Investigation”. *The Review of Economics and Statistics* 104.5, pp. 1059–1078. DOI: 10.1162/rest\_a\_00994 (p. 6, 7, 13, 19, 29, 37, A.38).
- Gabaix, Xavier (2019). “Behavioral Inattention”. *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2. Elsevier, pp. 261–343. DOI: 10.1016/bs.hesbe.2018.11.001 (p. 7, 30).
- Giacobasso, Matias, Brad C. Nathan, Ricardo Perez-Truglia, and Alejandro Zentner (2022). “Where Do My Tax Dollars Go? Tax Morale Effects of Perceived Government Spending”. Working Paper Series. DOI: 10.3386/w29789 (p. 6, A.41).
- Goodman-Bacon, Andrew (2021). “Difference-in-Differences with Variation in Treatment Timing”. *Journal of Econometrics* 225.2, pp. 254–277. DOI: 10.1016/j.jeconom.2021.03.014 (p. 1).
- Graham, Bryan S. and James L. Powell (2012). “Identification and Estimation of Average Partial Effects in “Irregular” Correlated Random Coefficient Panel Data Models”. *Econometrica* 80.5, pp. 2105–2152. DOI: 10.3982/ECTA8220 (p. 2, 12, 14, A.7, A.16, A.17).
- Grether, David M. (1980). “Bayes Rule as a Descriptive Model: The Representativeness Heuristic”. *The Quarterly Journal of Economics* 95.3, p. 537. DOI: 10.2307/1885092 (p. 7, 29).
- Grigorieff, Alexis, Christopher Roth, and Diego Ubfal (2020). “Does Information Change Attitudes Toward Immigrants?” *Demography* 57.3, pp. 1117–1143. DOI: 10.1007/s13524-020-00882-8 (p. 3).
- Guenther, Laurenz and Salvatore Nunnari (2025). “Do Political Representation Gaps Cause Populism? Evidence from the 2025 German Election”. *Working Paper* (p. 1).
- Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023). “Designing Information Provision Experiments”. *Journal of Economic Literature* 61.1, pp. 3–40. DOI: 10.1257/jel.20211658 (p. 16).
- Hansen, Bruce E. (2022). *Econometrics*. Princeton: Princeton University Press (p. A.18, A.19).
- Heckman, James J. and Edward J. Vytlacil (2001). “Local Instrumental Variables”. *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*. Ed. by Cheng Hsiao, Kimio Morimune, and James L. Powell. 1st ed. Cambridge University Press, pp. 1–46. DOI: 10.1017/CB09781139175203.003 (p. 31).
- (2007). “Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments”. *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 4875–5143 (p. 17, A.47).
- Hoff, Peter D. (2009). *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. New York, NY: Springer New York. DOI: 10.1007/978-0-387-92407-6 (p. 28).

- Hopkins, Daniel J., John Sides, and Jack Citrin (2019). “The Muted Consequences of Correct Information about Immigration”. *The Journal of Politics* 81.1, pp. 315–320. DOI: 10.1086/699914 (p. 3).
- Imbens, Guido and Whitney Newey (2009). “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity”. *Econometrica* 77.5, pp. 1481–1512. DOI: 10.3982/ECTA7108 (p. A.47).
- Imbens, Guido W. and Joshua D. Angrist (1994). “Identification and Estimation of Local Average Treatment Effects”. *Econometrica* 62.2, pp. 467–475. DOI: 10.2307/2951620 (p. 1).
- Jensen, Robert (2010). “The (Perceived) Returns to Education and the Demand for Schooling”. *Quarterly Journal of Economics* 125.2, pp. 515–548. DOI: 10.1162/qjec.2010.125.2.515 (p. 1).
- Kerwin, Jason T and Divya Pandey (2023). “Navigating Ambiguity: Imprecise Probabilities and the Updating of Disease Risk Beliefs”. *Working Paper* (p. 29).
- Kumar, Saten, Yuriy Gorodnichenko, and Olivier Coibion (2023). “The Effect of Macroeconomic Uncertainty on Firm Decisions”. *Econometrica* 91.4, pp. 1297–1332. DOI: 10.3982/ECTA21004 (p. 2, 19, 22, ii, A.11, A.18, A.19, A.24, A.35, A.36).
- List, John (2025). *The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs*. Tech. rep. w33456. Cambridge, MA: National Bureau of Economic Research, w33456. DOI: 10.3386/w33456 (p. 16).
- Maćkowiak, Bartosz and Mirko Wiederholt (Forthcoming). “Rational Inattention during an RCT”. *American Economic Review: Insights*. DOI: 10.1257/aeri.20240509 (p. 2, 13, 19).
- Masten, Matthew and Alexander Torgovitsky (2016). “Identification of Instrumental Variable Correlated Random Coefficients Models”. *The Review of Economics and Statistics* 98.5, pp. 1001–1005. DOI: 10.1162/REST\_a\_00603 (p. 2, 12, A.47).
- Mogstad, Magne and Alexander Torgovitsky (2024). “Instrumental Variables with Unobserved Heterogeneity in Treatment Effects”. *Handbook of Labor Economics*. Vol. 5. Elsevier, pp. 1–114. DOI: 10.1016/bs.heslab.2024.11.003 (p. A.46).
- Pallais, Amanda (2014). “Inefficient Hiring in Entry-Level Labor Markets”. *American Economic Review* 104.11, pp. 3565–3599. DOI: 10.1257/aer.104.11.3565 (p. A.27).
- Robert, Christian P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd ed. Springer Texts in Statistics. New York: Springer. DOI: 10.1007/0-387-71599-1 (p. 28).
- Roth, Christopher, Sonja Settele, and Johannes Wohlfart (2022). “Risk Exposure and Acquisition of Macroeconomic Information”. *American Economic Review: Insights* 4.1, pp. 34–53. DOI: 10.1257/aeri.20200662 (p. 2, 16, 18, 22, 29, ii, A.15, A.16, A.19, A.23, A.33).
- Roth, Christopher and Johannes Wohlfart (2020). “How Do Expectations about the Macroeconomy Affect Personal Expectations and Behavior?” *The Review of Economics and Statistics* 102.4, pp. 731–748. DOI: 10.1162/rest\_a\_00867 (p. 29).
- Settele, Sonja (2022). “How Do Beliefs about the Gender Wage Gap Affect the Demand for Public Policy?” *American Economic Journal: Economic Policy* 14.2, pp. 475–508. DOI: 10.1257/pol.20200559 (p. 2, 10, 18, 22, i, A.16, A.18, A.19, A.22, A.31, A.32, A.34).
- Sun, Liyang and Sarah Abraham (2020). “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects”. *Journal of Econometrics*. DOI: 10.1016/j.jeconom.2020.09.006 (p. 1).
- Vilfort, Vod and Whitney Zhang (2025). “Interpreting TSLS Estimators in Information Provision Experiments”. *American Economic Review: Insights* 7.3, pp. 376–395. DOI: 10.1257/aeri.20240353 (p. 2, 11, A.32, A.35, A.36, A.44, A.46).

Wiswall, Matthew and Basit Zafar (2015). “Determinants of College Major Choice: Identification Using an Information Experiment”. *The Review of Economic Studies* 82.2 (291), pp. 791–824 (p. 1, 2, 5, 6, 18, 22, i, A.19, A.20, A.27, A.29).

Yang, Jeffrey (2024). “On the Decision-Relevance of Subjective Beliefs”. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4425080 (p. 13).

## Appendix

### A. Learning Rate Models

This appendix provides microfoundations for the belief updating model in (3). Section A.1 introduces a general class of updating rules that preserve rank invariance, which is the minimal condition required for LLS to have equal weights. Several behavioral deviations from Bayesian updating fall within this class.

#### A.1. Generalized Updating and Rank Invariance

For LLS to have equal weights across individuals, we need *rank invariance*: the relative magnitude of belief updates must be consistent across signals. Formally, let  $X_i(s, x_0)$  denote individual  $i$ 's posterior given signal  $s$  and prior  $x_0$ . If  $X_i(s_A, x_0) > X_j(s_A, x_0)$  for some signal  $s_A$ , then  $X_i(s_B, x_0) > X_j(s_B, x_0)$  for any other signal  $s_B$  on the same side of the prior.

Intuitively, people who respond more strongly to one signal also respond more strongly to another. Suppose Chris and Dianne have the same prior and receive the same signal above their prior. If Chris's posterior is higher than Dianne's, rank invariance requires that Chris's posterior would also be higher if both received a different signal that was also above their prior.

This section introduces a general class of updating rules that preserve rank invariance. The Bayesian baseline is a special case, but so are several behavioral deviations.

##### A.1.1. General Learning Rate Updating

Consider a general updating rule:

$$X_i(s, x_0) = x_0 + \alpha_i \times f(s, x_0) \tag{20}$$

where  $f(\cdot, X_i^0)$  is any function monotonic in the signal with  $f(X_i^0, X_i^0) = 0$ . The function

$f$  is a link function that allows for nonlinearity in the effects of signals on the posterior belief. The individual parameter  $\alpha_i > 0$  controls differences in updating between people with the same prior and signal.

The leading special case is when  $f(s, x_0)$  is the difference  $s - x_0$ . Then, (20) reduces to:

$$X_i(s, x_0) = \alpha_i s + (1 - \alpha_i)x_0 \quad (21)$$

Which is the simple linear updating rule generated by Bayesian updating, among others.

### A.1.2. Bayesian Learning as a Baseline

The literature often motivates the weighted-average expression in (21) with a Bayesian learning model featuring normally distributed beliefs (Balla-Elliott et al., 2022; Cullen and Perez-Truglia, 2022). Consider individuals with uncertain prior beliefs. The subjective probability that the variable  $X_i$  takes value  $x$  is given by the density of  $\mathcal{N}(X_i^0, \sigma_{iX}^2)$ . We interpret  $X_i^0$  as the mean of the prior distribution and call it the *prior belief*.

People observe a signal  $S_i$  drawn from  $\mathcal{N}(S_i^*, \sigma_{iS}^2)$ . The variances reflect subjective (inverse) precision: people for whom  $\sigma_{iS}^2/\sigma_{iX}^2$  is large think their prior is more precise than the signal, while those with small  $\sigma_{iS}^2/\sigma_{iX}^2$  think the signal is more precise than their prior.

The posterior distribution is:

$$\mathcal{N}\left((1 - \alpha_i) X_i^0 + \alpha_i S_i, \frac{\sigma_{iS}^2 \sigma_{iX}^2}{\sigma_{iS}^2 + \sigma_{iX}^2}\right) \quad (22)$$

$$\text{where } \alpha_i \equiv \frac{\sigma_{iX}^2}{\sigma_{iS}^2 + \sigma_{iX}^2} \quad (23)$$

The mean of the posterior is a weighted average of the prior  $X_i^0$  and signal  $S_i$ , with weights determined by relative precision.<sup>25</sup> We call this mean the *posterior belief*  $X_i$ . The prior  $X_i^0$ ,

---

<sup>25</sup>See Robert (2007) or Hoff (2009) for textbook treatments.

signal  $S_i$ , and posterior  $X_i$  are thus related by:

$$X_i = (1 - \alpha_i)X_i^0 + \alpha_i S_i \quad (24)$$

which generates the potential outcomes for beliefs in (4).

There is direct empirical support for this foundation. Roth et al. (2022) find that belief updating is driven entirely by people who report being “very unsure”, “unsure”, or “somewhat unsure”. Those who are “sure” or “very sure” do not update. Similarly, Roth and Wohlfart (2020) find that people less confident in their priors update roughly twice as much. Kerwin and Pandey (2023) find that people with less precise priors update more in a more general model.

### **A.1.3. Linear Deviations from Bayesian Updating**

Any model where updating takes the linear form (21) satisfies rank invariance, regardless of how  $\alpha_i$  is determined. Three behavioral deviations retain this structure.

*Diagnostic Expectations.* Grether (1980) models deviations from Bayesian updating by raising the likelihood and prior to different powers. Under normal-normal learning, this rescales the effective variances of prior and signal. The learning rate  $\alpha_i$  then depends on “behavioral” variances rather than true variances, but updating remains linear. People can vary in the heuristics they use to update and could either over-update or under-update, as long as these differences are reflected in the learning rate.

*Rational Inattention.* Fuster et al. (2022) develop a rational inattention model where the learning rate  $\alpha_i$  depends on the marginal cost of attention and the value of information. The posterior is still a weighted average of signal and prior, but the weight reflects optimal attention allocation rather than prior precision. Importantly, some people can have a “corner” solution and ignore the information entirely, which allows some people to have

$\alpha_i = 0$ .

*Anchoring on Signal or Prior.* In anchoring models (Gabaix, 2019), people form posteriors as a weighted average of a Bayesian posterior and an anchor:

$$X_i(s, x_0) = \kappa_i A(s, x_0) + (1 - \kappa_i) X_i^B(s, x_0) \quad (25)$$

where  $X_i^B(s, x_0) = \alpha_i s + (1 - \alpha_i) x_0$  is the Bayesian posterior. When the anchor is itself a weighted average of the signal and prior ( $A(s, x_0) = \gamma s + (1 - \gamma) x_0$  for  $\gamma \in [0, 1]$ ) substitution yields:

$$X_i(s, x_0) = \underbrace{[\kappa_i \gamma + (1 - \kappa_i) \alpha_i]}_{\alpha_i^{\text{eff}}} s + [1 - \alpha_i^{\text{eff}}] x_0 \quad (26)$$

The anchored posterior is simply a weighted average with effective learning rate  $\alpha_i^{\text{eff}}$ . Two leading special cases are anchoring on the signal or anchoring on the prior. Both preserve the linear structure.

However, anchoring on a constant  $A(s, x_0) = \bar{x}$  does not have this learning-rate representation and violates rank invariance.

#### A.1.4. Representative Estimates When Everyone Updates

Many of the updating models discussed above have micro-foundations that ensure  $\alpha_i \in (0, 1)$ . This includes standard Bayesian updating, diagnostic expectations, and anchoring on the prior or signal. When learning rates are strictly positive, everyone updates at least somewhat in response to the signal. There are no never-takers who ignore information entirely.

When everyone responds to the signal, LLS identifies an average over the full experimental sample. In contrast, if some individuals have  $\alpha_i = 0$  and completely ignore the

signal, both LLS and TSLS only identify an average over those who update. Individuals who never respond receive zero weight in both estimators. The LLS estimand then downgrades to a local average treatment effect (LATE): an average only among compliers who respond to the instrument.

This is closely related to a more general identification result emphasized by Heckman and Vytlacil (2001) that the ATE is identified when there are values of the instrument  $z, z'$  with propensity scores of zero and one (everyone is a  $z \rightarrow z'$  complier). Targeting an ATE-like parameter requires that the instrument affects the endogenous variable for all individuals. When  $\alpha_i > 0$  for everyone, the signal moves everyone's beliefs (at least slightly) and so everyone is a complier.

*Downgrading to LATE with Rational Inattention.* The rational inattention framework allows  $\alpha_i = 0$  for some individuals. When the cost of attention exceeds the value of information, optimal behavior is to ignore the signal entirely. In this case, the LLS estimand downgrades from a structural APE or ATE-like parameter (representative of all experimental subjects) to a LATE parameter (representative of those who update). The estimand remains interpretable; it simply averages over compliers rather than the full sample.

This downgrade does not change the fundamental difference between LLS and TSLS. Among those who update ( $\alpha_i > 0$ ), LLS continues to place equal weight while TSLS weights by the size of the update. The weighting distinction persists regardless of whether the average is over everyone or only over compliers.

#### **A.1.5. Estimation with Nonlinear Updating**

Under linear updating, we can control for the prior linearly and condition nonparametrically only on the learning rate  $\alpha_i$ . This simplification extends to all linear learning rate models. Under nonlinear updating (20), the estimation approach changes. The prior no longer enters linearly, so we cannot separate conditioning on  $\alpha_i$  from conditioning on  $x_0$ .

Further, the learning rate  $\alpha_i$  is not directly identified from the ratio of the belief update to the difference between the signal and the prior. Instead, the conditional rank of the learning rate is identified from the sign-corrected conditional rank of the posterior given the prior and the signal:

$$R_i \equiv \text{rank}(\alpha_i | X_i^0) = \begin{cases} \text{rank}(X_i | X_i^0, S_i) & \text{if } S_i > X_i^0 \\ 1 - \text{rank}(X_i | X_i^0, S_i) & \text{if } S_i < X_i^0 \end{cases} \quad (27)$$

The local regression must condition on the full vector  $C_i = (R_i, X_i^0, S_i(A), S_i(B))$  non-parametrically, or make alternative simplifying assumptions.

## **B. Endogenous Belief Formation Through Costly Information Acquisition**

This section formalizes a model of endogenous information acquisition. When beliefs strongly affect decisions—think of a homeowner whose refinancing choices depend critically on house price expectations—individuals rationally invest in gathering precise information before any experiment takes place. These well-informed individuals update their beliefs only modestly when researchers provide new information, while those for whom the belief matters less start with noisier priors and update more dramatically. Since standard specifications weight individuals by the strength of their belief updating, they systematically under-weight precisely those people for whom beliefs matter most. I formalize this intuition by modeling how individuals trade off the cost of acquiring information against the risk of making decisions with imprecise beliefs. The resulting negative correlation between causal effects and belief updating leads to attenuated estimates in information provision experiments.

### B.1. General Model

People have a subjective belief distribution given by  $F_i(\cdot)$ . To make the analysis tractable, focus on belief distributions that can be characterized by their mean  $\mu_i$  and variance  $\sigma_i^2$ , with  $F_i$  belonging to a parametric family (e.g., normal distributions). People are uncertain about their beliefs, and this uncertainty about their beliefs generates uncertainty about the action that they would like to take. Let  $R(\tau_i, \sigma_i^2)$  denote the subjective risk or ex-ante regret (for example, the expected loss) that an individual with causal effect  $\tau_i$  faces when their belief variance is  $\sigma_i^2$ . Note that  $R$  depends on the distribution  $F_i$  only through its variance  $\sigma_i^2$ , as the mean belief affects the level of the action but not the risk from uncertainty.

We make the following assumptions on  $R$ . First, uncertainty is costly:  $\frac{\partial R}{\partial \sigma^2} \geq 0$ , where  $\frac{\partial R}{\partial \sigma^2} = 0$  if and only if  $\tau_i = 0$ . Second, since there is uncertainty in beliefs, it is costly to base behavior on these beliefs:  $\frac{\partial R}{\partial |\tau|} \geq 0$ , where  $\frac{\partial R}{\partial |\tau|} = 0$  if and only if  $\sigma^2 = 0$ . Finally, uncertainty is more costly for people whose beliefs affect actions more:  $\frac{\partial^2 R}{\partial \sigma^2 \partial |\tau|} > 0$ .

People make a decision to pay a cost  $c > 0$  to obtain new information or to do nothing. There is an updating process such that the variance of beliefs after viewing a signal  $\sigma_+^2$  is less than the variance of the initial beliefs  $\sigma^2$ . People then trade off the reduction in risk from the new information against the cost of the signal. Thus, when person  $i$  has beliefs with variance  $\sigma^2$ , her loss can be given recursively by

$$V(\tau_i, \sigma^2) = \min \{R(\tau_i, \sigma^2), V(\tau_i, \sigma_+^2) + c\} \quad (28)$$

Given the assumptions we have made on  $R$ , for any beliefs with  $\sigma^2 > 0$ , there is some threshold value  $\tau^*$  such that people with  $|\tau_i| > \tau^*$  prefer to pay  $c$  to update their beliefs. That such a threshold exists is guaranteed by the fact that  $R(\tau_i, \sigma^2) = R(\tau_i, \sigma_+^2)$  when  $\tau_i = 0$ , which implies that  $R(\tau_i, \sigma^2) < R(\tau_i, \sigma_+^2) + c$  at  $\tau_i = 0$ . However, since  $\frac{\partial^2 R}{\partial \sigma^2 \partial |\tau_i|} > 0$ , we also know that  $\frac{\partial R(\tau_i, \sigma^2)}{\partial |\tau_i|} > \frac{\partial R(\tau_i, \sigma_+^2)}{\partial |\tau_i|}$  since  $\sigma^2 > \sigma_+^2$ .

At  $\tau_i = 0$ ,  $R(\tau_i, \sigma^2)$  is below  $R(\tau_i, \sigma_+^2) + c$ . However,  $R(\tau_i, \sigma^2)$  is increasing faster than  $R(\tau_i, \sigma_+^2)$  in  $|\tau_i|$  such that eventually these curves will cross. And since  $R(\tau_i, \sigma^2)$  is always increasing faster than  $R(\tau_i, \sigma_+^2)$  in  $|\tau_i|$ , they will cross exactly once. Figure B.1 illustrates this graphically. When beliefs are formed through such a process, people with larger causal effects of beliefs will have (weakly) more precise beliefs in equilibrium.

## B.2. A Simple Example with Quadratic Loss and Normal Beliefs

This example illustrates how the general framework applies in an example where beliefs are normally distributed and the risk function takes a particularly tractable form.

Let  $Y$  be the action (e.g., list price of a house) and  $X$  denote beliefs (e.g., about the market value). People start with a prior belief distribution centered around  $\pi_i$  with variance  $\sigma_{X_0}^2$  so that their beliefs are represented by the normal  $\mathcal{N}(\pi_i, \sigma_{X_0}^2)$ . For simplicity,  $\sigma_{X_0}^2$  is common. Signals  $S$  are drawn from a normal distribution  $\mathcal{N}(\mu_S, \sigma_S^2)$ . This is an assumption that people have the same information environment.

People are uncertain about their beliefs, and this uncertainty about their beliefs generates uncertainty about the action that they would like to take. People act to minimize the loss function  $L_i(y, x) = D(y, Y_i(x))$ , for some distance function  $D$ , which is the disutility associated with taking action  $y$  when  $X = x$ . Intuitively, integrating  $L_i(y, x)$  over the distribution of beliefs converts uncertainty about beliefs (i.e., what is the probability that  $X = x$ ) into regret about actions (i.e., how far is the choice  $y$  from  $Y_i(x)$ , which is optimal when  $X = x$ ). In this loss function, beliefs affect utility only through their effect on actions. There is no direct “psychic” cost of imprecise beliefs.

People choose  $Y_i(x)$  following the rule  $Y_i(x) = \tau_i x + U_i$ , where  $\tau_i$  and  $U_i$  vary across individuals, and have quadratic loss  $D(a, b) = (a - b)^2$ . They act to minimize their expected loss, which is simply the expectation of  $L_i(y, x)$  with respect to  $X$  (i.e.  $\int L_i(y, x) dF(x)$ ).

Let  $\bar{X}$  denote the mean of the belief distribution. When beliefs are given by the normal  $\mathcal{N}(\bar{X}, \sigma_X^2)$ , the choice of  $Y$  that minimizes expected loss is simply  $Y^* \equiv Y_i(\bar{X}) = \tau_i \bar{X} + U_i$ .

Use this to further simplify the expression for expected loss and write

$$\int L_i(Y^*, x) dF(x) = \int D(Y_i(\bar{X}), Y_i(x)) dF(x) \quad (29)$$

$$= \int ((\tau_i \bar{X} + U_i) - (\tau_i x + U_i))^2 dF(x) = \tau_i^2 \sigma_X^2 \quad (30)$$

since  $\mathbb{E}[(\tau_i \bar{X} - \tau_i X)^2] = \tau_i^2 \text{Var}(X) = \tau_i^2 \sigma_X^2$ . Notice that with quadratic loss, the risk function takes the form  $R(\tau_i, \sigma_X^2) = \tau_i^2 \sigma_X^2$ , which satisfies the assumptions about  $R$  given in Section B.1.

The disutility generated by uncertainty about  $X$  is increasing in both the variance of the belief distribution and the magnitude of the causal effect of beliefs on the outcome. This expression allows us to study the information acquisition problem.

I endogenize belief formation by allowing people to pay a fixed cost  $C$  to view a signal that is centered around the unknown true value. They then update beliefs following the normal-normal Bayesian learning formula. When a person's beliefs are given by  $\mathcal{N}(\bar{X}, \sigma_X^2)$ , her loss is given recursively by

$$V_i(\bar{X}, \sigma_X^2) = \min \{ \mathbb{E}_X[L_i(Y_i(\bar{X}), x)], \mathbb{E}_S[V_i(X'(s), \sigma_{X'}^2)] + C \} \quad (31)$$

where  $\sigma_{X'}^2 = \frac{\sigma_X^2 \sigma_S^2}{\sigma_X^2 + \sigma_S^2}$  is the posterior variance after observing signal  $S$  and the expectation  $\mathbb{E}[S]$  is with respect to the signal distribution. The benefit of the signal comes from the fact that the posterior variance is less than the prior variance as long as the prior distribution is not already degenerate. Notice that in this example, the value function depends on the belief distribution only through its variance  $\sigma_X^2$ , since the mean  $\bar{X}$  affects the level of the optimal action but not the expected loss from uncertainty.

Solving this recursive problem gives the equilibrium condition

$$\tau_i^2 \sigma_X^2 = \tau_i^2 \sigma_{X'}^2 + C \quad (32)$$

In equilibrium, agents will be indifferent between paying the fixed cost to obtain new information and living with the uncertainty they have.<sup>26</sup> Replacing  $\sigma_{X'}^2$  with its definition, and recalling that  $1 - \frac{\sigma_S^2}{\sigma_S^2 + \sigma_X^2} = \alpha_i$  yields the following equality

$$\alpha_i \tau_i^2 \sigma_X^2 = C \quad (33)$$

Agents for whom the outcome is very sensitive to the beliefs ( $\tau_i^2$  is very large) will update their information until  $\sigma_X^2 \alpha_i$  is small.<sup>27</sup> On the other hand, agents for whom the outcome is not sensitive to beliefs ( $\tau_i^2$  is small) will stop after seeing fewer signals, so that  $\sigma_X^2 \alpha_i$  is relatively large.

This simple model illustrates how the causal relationship of interest affects the formation of beliefs before the experiment takes place. People whose actions depend more on their beliefs will be more willing to pay to obtain new information, and will therefore have more precise beliefs. In a Bayesian updating model, people with more precise beliefs will be less responsive to new information. In this way, the amount of variation in beliefs that can be induced by experimentally providing new information directly depends on the causal effects of interest.

### B.3. Using Models of Belief Formation and Updating to Interpret TSLS Estimates

The class of parameters that are targeted by existing standard specifications depend not only on the causal effects of beliefs on outcomes  $\tau_i$ , but also on heterogeneity in the way that beliefs are updated in response to new information.

In the model proposed in this section, beliefs are formed endogenously through a

---

<sup>26</sup>To ease exposition, I have ignored integer constraints that will, in general, prevent this from holding with equality. People will purchase signals until the next signal reduces their expected loss by less than the cost of the signal and will generally be strictly worse off if they buy another signal, not indifferent. This technicality makes exposition more cumbersome without any conceptual payoff.

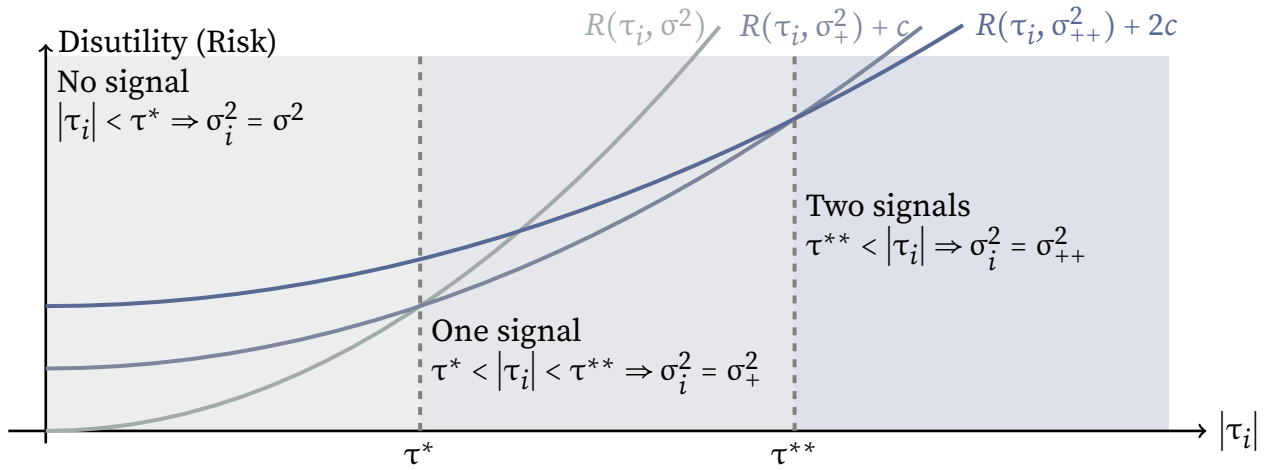
<sup>27</sup>Notice that since  $\alpha_i \equiv \frac{\sigma_X^2}{\sigma_S^2 + \sigma_X^2}$ ,  $\alpha_i$  and  $\sigma_X^2$  move together. That is, holding fixed  $\sigma_S^2$ , an increase in  $\sigma_X^2$  implies an increase in  $\alpha_i$  and vice-versa.

process of costly information acquisition. In Appendix B.2, I solve a special case of this model where the subjective risk is given by the expected quadratic loss  $R(\tau_i, \sigma^2) = \tau_i^2 \sigma^2$ . Parameterizing the loss function makes it possible to solve analytically for the learning rate  $\alpha_i$  and variance of the prior  $\sigma_i^2$  as a function of the causal effects of beliefs  $\tau_i$ .

People have inaccurate and imprecise beliefs precisely because they have small individual partial effects (small  $|\tau_i|$ ); when beliefs are an important determinant of the behaviors (large  $|\tau_i|$ ), people exert effort to form accurate and precise beliefs. In this environment, parameters with weights proportional to the strength of the shift in beliefs will be attenuated and underestimate the magnitude of the average effect.

Alternative models of the relationship between belief updating and the effects of beliefs on behaviors can be used to relate causal parameters estimated using standard specifications to the APE. For example, Fuster et al. (2022) allow variation in the learning rate to come from a more complicated model that adds dynamics of rational inattention to costly information acquisition.

FIGURE B.1. People with Large Effects of Beliefs  $\tau_i$  Form Precise Beliefs



*Notes:* This figure plots the loss as a function of  $|\tau_i|$  after seeing no signals, one signal, and two signals. The assumptions on  $R_i$  ensure that each pair of lines crosses exactly once. Since  $R(\tau_i, \sigma^2) = R(\tau_i, \sigma_+^2)$  when  $\tau_i = 0$ ,  $R(\tau_i, \sigma^2) < R(\tau_i, \sigma_+^2) + c$ . If  $\sigma_{++}^2 > 0$ , these curves are all strictly increasing in  $|\tau_i|$  by assumption. Additionally, since  $\sigma^2 > \sigma_+^2 > \sigma_{++}^2$ , then  $R(\tau_i, \sigma^2)$  is steeper than  $R(\tau_i, \sigma_+^2)$ , which is steeper than  $R(\tau_i, \sigma_{++}^2)$  by the assumption that  $\frac{\partial^2 R}{\partial \sigma^2 \partial |\tau_i|} > 0$ .

## Contents of Online Appendix

<b>C</b>	<b>Proofs and Derivations</b>	<b>A.1</b>
C.1	Derivations of Weights . . . . .	A.1
C.1.1	Weights in the Panel Specification . . . . .	A.1
C.1.2	Weights in the Active Control Specification . . . . .	A.2
C.1.3	Weights in the Passive Control Specification . . . . .	A.3
C.2	Main Identification Results . . . . .	A.6
C.2.1	Identification in Panel Experiments . . . . .	A.7
C.2.2	Identification in Active Experiments . . . . .	A.9
C.2.3	Identification in Passive Experiments . . . . .	A.10
C.3	Linear Controls in a Reweighted Regression . . . . .	A.13
<b>D</b>	<b>Estimation Details</b>	<b>A.15</b>
D.1	Linear Belief Updating Simplifies Estimation . . . . .	A.15
D.1.1	Local Regressions in Panel Experiments . . . . .	A.15
D.1.2	Local Regressions in Active and Passive Control Experiments . . . . .	A.16
D.2	Trimming . . . . .	A.16
D.2.1	Trimming in Panel Experiments . . . . .	A.17
D.2.2	Trimming in Active Control Experiments . . . . .	A.17
D.2.3	Trimming in Passive Control Experiments . . . . .	A.17
D.3	Bandwidth Selection . . . . .	A.17
<b>E</b>	<b>Application Details</b>	<b>A.26</b>
E.1	Systematic Selection of Empirical Applications . . . . .	A.26
E.2	Application Details: Wiswall and Zafar (2015) . . . . .	A.27
E.2.1	Setting . . . . .	A.27
E.2.2	Specification of Interest . . . . .	A.28
E.2.3	Implementing the LLS Estimator . . . . .	A.29
E.3	Application Details: Armona, Fuster, and Zafar (2019) . . . . .	A.29
E.3.1	Setting . . . . .	A.29
E.3.2	Specification of Interest . . . . .	A.30
E.3.3	Implementing the LLS Estimator . . . . .	A.30
E.4	Application Details: Settele (2022) . . . . .	A.31
E.4.1	Setting . . . . .	A.31
E.4.2	Specification of Interest . . . . .	A.31

E.4.3	Implementing the LLS Estimator . . . . .	A.32
E.5	Application Details: Roth, Settele, and Wohlfart (2022) . . . . .	A.33
E.5.1	Setting . . . . .	A.33
E.5.2	Specification of Interest . . . . .	A.33
E.5.3	Implementing the LLS Estimator . . . . .	A.34
E.6	Application Details: Kumar, Gorodnichenko, and Coibion (2023) . . . . .	A.35
E.6.1	Setting . . . . .	A.35
E.6.2	Specification of Interest . . . . .	A.35
E.6.3	Implementing the LLS Estimator . . . . .	A.36
E.7	Application Details: Cantoni, Yang, Yuchtman, and Zhang (2019) . . . . .	A.36
E.7.1	Setting . . . . .	A.37
E.7.2	Specification of Interest . . . . .	A.37
E.7.3	Implementing the LLS Estimator . . . . .	A.37
E.7.4	Discussion . . . . .	A.38
<b>F</b>	<b>Information Experiments and the TSLS Estimator</b>	<b>A.40</b>
F.1	The Reduced Form Effect of Information Provision . . . . .	A.40
F.1.1	From the Effect of Information to the Effect of Beliefs . . . . .	A.41
F.1.2	Constructing TSLS Estimates . . . . .	A.41
F.2	Unconditional Instrument Monotonicity and Bayesian Updating . . . . .	A.42
F.2.1	Monotonicity in Active Designs . . . . .	A.42
F.2.2	Monotonicity in Passive Designs . . . . .	A.42
F.3	Strategies for Ensuring Non-Negative Weights in Passive Designs . . . . .	A.43
F.3.1	Sample Splitting Approach . . . . .	A.43
F.3.2	Exposure-Weighted Instruments . . . . .	A.43
F.4	Implementation When Priors Are Unobserved . . . . .	A.44
<b>G</b>	<b>Nonlinearity, Convex Combinations, and Discrete Slopes</b>	<b>A.46</b>
G.1	Convex Combinations and Magnitudes . . . . .	A.46
G.2	Discrete Slopes Versus Derivatives . . . . .	A.46
G.2.1	Identifying Derivatives Without Linearity . . . . .	A.47
G.3	Nonlinearity Affects Interpretation, Not the Case for Equal Weights . . . . .	A.47

## C. Proofs and Derivations

This section contains proofs and derivations.

### C.1. Derivations of Weights

This section provides derivations for the weights reported in Section 2. All three derivations use the definition of individual slopes from equation (5):

$$\beta_i \equiv \frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \quad (5)$$

This definition implies that  $G_i(X_i(A)) - G_i(X_i(B)) = \beta_i(X_i(A) - X_i(B))$  for any outcome function  $G_i(\cdot)$ .

#### C.1.1. Weights in the Panel Specification

ASSUMPTION 1. *Panel Assumptions.*

a. *Panel Outcomes: The panel outcome equation (2) holds.*

$$Y_{it} = G_i(X_{it}) + \gamma_t \quad (2)$$

b. *Relevance: There is variation in beliefs over time  $\text{Var}[\Delta X_i] > 0$ .*

c. *Existence: The relevant moments exist and are finite.*

The parsimonious specification in the panel data model in (7) is given by:

$$\beta^{Panel} = \frac{\text{Cov}[\Delta Y_i, \Delta X_i]}{\text{Var}[\Delta X_i]} \quad (34)$$

Substitute the outcome equation (2):

$$= \frac{\text{Cov}[G_i(X_{i1}) - G_i(X_{i0}) + \gamma_1 - \gamma_0, \Delta X_i]}{\text{Var}[\Delta X_i]} \quad (35)$$

Apply the definition of individual slopes. Since  $\Delta X_i = X_{i1} - X_{i0}$ , we have  $G_i(X_{i1}) - G_i(X_{i0}) = \beta_i \Delta X_i$ :

$$= \frac{\text{Cov}[\beta_i \Delta X_i + \gamma_1 - \gamma_0, \Delta X_i]}{\text{Var}[\Delta X_i]} \quad (36)$$

From definitions of covariance and variance;  $\text{Cov}(a, b) = \mathbb{E}[a(b - \mathbb{E}(b))]$ :

$$= \frac{\mathbb{E}[\beta_i \Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i])]}{\mathbb{E}(\Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i]))} \quad (37)$$

To express this as a weighted average of individual effects, rearrange:

$$= \mathbb{E} \left[ \beta_i \cdot \frac{\Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i])}{\mathbb{E}(\Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i]))} \right] \quad (38)$$

This gives the weights  $\omega_i(\text{Panel}) \propto \Delta X_i (\Delta X_i - \mathbb{E}[\Delta X_i])$ , which are normalized to integrate to one.

### C.1.2. Weights in the Active Control Specification

ASSUMPTION 2. *Active Control Assumptions.*

a. *Nonparametric Outcomes: The outcome model in equation (1) holds.*

$$Y_i = G_i(X_i) \quad (1)$$

b. *Learning rate updating: The belief potential outcomes in equation (4) hold.*

$$X_i(z) = \alpha_i (S_i(z) - X_i^0) + X_i^0 \quad (4)$$

c. *Relevance: There is variation in potential beliefs  $\mathbb{E}[X_i(A) - X_i(B)] \neq 0$ .*

d. *Random Assignment: The treatment  $Z_i$  is randomly assigned.*

e. *Existence: The relevant moments exist and are finite.*

Starting with the TSLS coefficient in the active control design:

$$\beta^{\text{TSLS}} = \frac{\mathbb{E}[Y_i | Z_i = A] - \mathbb{E}[Y_i | Z_i = B]}{\mathbb{E}[X_i | Z_i = A] - \mathbb{E}[X_i | Z_i = B]} \quad (39)$$

From the outcome equation (1) and random assignment:

$$= \frac{\mathbb{E}[G_i(X_i(A))] - \mathbb{E}[G_i(X_i(B))]}{\mathbb{E}[X_i(A)] - \mathbb{E}[X_i(B)]} \quad (40)$$

Apply the definition of individual slopes:  $G_i(X_i(A)) - G_i(X_i(B)) = \beta_i(X_i(A) - X_i(B))$ :

$$= \frac{\mathbb{E}[\beta_i(X_i(A) - X_i(B))]}{\mathbb{E}[X_i(A) - X_i(B)]} \quad (41)$$

To express this as a weighted average of individual effects, rearrange:

$$= \mathbb{E} \left[ \beta_i \cdot \frac{X_i(A) - X_i(B)}{\mathbb{E}[X_i(A) - X_i(B)]} \right] \quad (42)$$

This gives us the weights  $\omega_i(\text{Active}) \propto X_i(A) - X_i(B)$ , which are normalized to integrate to one.

### C.1.3. Weights in the Passive Control Specification

ASSUMPTION 3. *Passive Control Assumptions.*

a. *Nonparametric Outcomes: The outcome model in equation (1) holds.*

$$Y_i = G_i(X_i) \quad (1)$$

b. *Learning rate updating: The belief potential outcomes in equation (4) hold.*

$$X_i(z) = \alpha_i(S_i(z) - X_i^0) + X_i^0 \quad (4)$$

- c. *Relevance: There is variation in potential beliefs  $\mathbb{E}[X_i(A) - X_i(B)] \neq 0$ .*
- d. *Random Assignment: The treatment  $Z_i$  is randomly assigned.*
- e. *Existence: The relevant moments exist and are finite.*
- f. *Passive control: Treatment arm B does not receive any signal:  $S_i(B) \equiv X_i^0$ .*

In the passive control design, the exposure-weighted instrument is defined as:

$$T_i^{ex} \equiv T_i(S_i(A) - X_i^0) \quad (14)$$

Since we are interested in coefficients on  $T_i^{ex}$  in regressions that control for  $S_i(A) - X_i^0$ , we can appeal to FWL and instead consider the coefficients on the residualized  $\tilde{T}_i^{ex}$ . To construct this residual, regress  $T_i^{ex}$  on  $(S_i(A) - X_i^0)$  and a constant:

$$\theta = \frac{\text{Cov}(T_i^{ex}, S_i(A) - X_i^0)}{\text{Var}(S_i(A) - X_i^0)} \quad (43)$$

$$= \frac{\mathbb{E}[T_i(S_i(A) - X_i^0)^2] - \mathbb{E}[T_i]\mathbb{E}[(S_i(A) - X_i^0)^2]}{\text{Var}(S_i(A) - X_i^0)} \quad (44)$$

Since  $T_i$  is binary and independent of  $(S_i(A) - X_i^0)$  by random assignment:

$$\theta = \frac{\mathbb{E}[T_i] \text{Var}(S_i(A) - X_i^0)}{\text{Var}(S_i(A) - X_i^0)} = \mathbb{E}[T_i] \quad (45)$$

The recentered instrument is then the residual from this regression:

$$\tilde{T}_i^{ex} = T_i^{ex} - \theta(S_i(A) - X_i^0) \quad (46)$$

$$= (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \quad (47)$$

Since  $\mathbb{E}[\tilde{T}_i^{ex}] = 0$ , the TSLS coefficient is:

$$\beta^{\text{Passive}} = \frac{\mathbb{E}[\tilde{T}_i^{ex} Y_i]}{\mathbb{E}[\tilde{T}_i^{ex} X_i]} \quad (48)$$

The denominator is:

$$\mathbb{E}[\tilde{T}_i^{ex} X_i] = \mathbb{E}[(T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \cdot X_i] \quad (49)$$

Plugging in the potential beliefs for  $X_i$ :

$$= \mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[(S_i(A) - X_i^0)(X_i(A) - X_i^0)] \quad (50)$$

Using the definition of  $X_i(A)$  from (4) to simplify further yields:

$$= \mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2] \quad (51)$$

For the numerator, apply the outcome equation and random assignment:

$$\mathbb{E}[\tilde{T}_i^{ex} Y_i] = \mathbb{E}[(T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \cdot G_i(X_i)] \quad (52)$$

$$= \mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[(S_i(A) - X_i^0)(G_i(X_i(A)) - G_i(X_i^0))] \quad (53)$$

Apply the definition of individual slopes. In passive designs  $X_i(B) = X_i^0$ , so  $G_i(X_i(A)) - G_i(X_i^0) = \beta_i(X_i(A) - X_i^0) = \beta_i\alpha_i(S_i(A) - X_i^0)$ :

$$= \mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[\beta_i\alpha_i(S_i(A) - X_i^0)^2] \quad (54)$$

Thus, the TSLS coefficient is:

$$\beta^{\text{Passive}} = \frac{\mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[\beta_i\alpha_i(S_i(A) - X_i^0)^2]}{\mathbb{E}[T_i](1 - \mathbb{E}[T_i])\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2]} \quad (55)$$

$$= \mathbb{E}\left[\beta_i \cdot \frac{\alpha_i(S_i(A) - X_i^0)^2}{\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2]}\right] \quad (56)$$

This gives us the weights  $\omega_i(\text{Passive}) \propto \alpha_i(S_i(A) - X_i^0)^2$ , which are normalized to

integrate to one.

## C.2. Main Identification Results

The key identification insight across all three designs is that by appropriately conditioning on observables, we can isolate variation in beliefs that is driven solely by exogenous treatment assignment. This creates local comparisons where beliefs effectively take only two values, making each regression equivalent to a simple difference in conditional means. This section proves that these local regressions recover average partial effects.

**PROPOSITION 1 (Binary Regression Property).** *Consider a linear regression of  $Y$  on  $X$  where  $X$  takes only two values,  $x_1$  and  $x_2$ . Then the regression coefficient  $\beta$  equals:*

$$\beta = \frac{\mathbb{E}[Y | X = x_2] - \mathbb{E}[Y | X = x_1]}{x_2 - x_1} \quad (57)$$

**PROOF.** The regression coefficient is defined as:

$$\beta = \frac{\text{Cov}(Y, X)}{\text{Var}(X)} \quad (58)$$

Let  $p = \Pr[X = x_2]$ . Then:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \quad (59)$$

$$= p(1 - p)(x_2 - x_1)^2 \quad (60)$$

For the covariance:

$$\text{Cov}(Y, X) = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])] \quad (61)$$

$$= p(1 - p)(x_2 - x_1)(\mathbb{E}[Y | X = x_2] - \mathbb{E}[Y | X = x_1]) \quad (62)$$

Therefore:

$$\beta = \frac{\text{Cov}(Y, X)}{\text{Var}(X)} = \frac{\mathbb{E}[Y | X = x_2] - \mathbb{E}[Y | X = x_1]}{x_2 - x_1} \quad (63)$$

□

### C.2.1. Identification in Panel Experiments

ASSUMPTION 1A. *Maintain the panel assumptions 1. Additionally*

- i. *Either  $\mathbb{P}[\Delta X_i = 0] > 0$  (control group exists), or  $\Delta X_i$  has positive density in a neighborhood of zero (as in Graham and Powell, 2012).*
- ii. *Nonlinear outcome: Relax the outcome equation to*

$$Y_{it}(x) = G_i(x) + \gamma_t \quad (2)$$

PROPOSITION 2 (Panel Identification). *Under Assumption 1A, for any  $x \neq 0$  in the support of  $\Delta X_i$ :*

$$\frac{\mathbb{E}[G_i(X_i^0 + x) - G_i(X_i^0) | \Delta X_i = x]}{x} = \frac{\mathbb{E}[\Delta Y_i | \Delta X_i = x] - \mathbb{E}[\Delta Y_i | \Delta X_i = 0]}{x} \quad (64)$$

*If  $G_i(X_{it}) = \tau_i X_{it} + U_i$  as in (2), the estimand simplifies further to  $\mathbb{E}[\tau_i | \Delta X_i = x]$ .*

PROOF. By Proposition 1, the regression of  $\Delta Y_i$  on  $\Delta X_i$  conditional on  $\Delta X_i \in \{0, x\}$  has coefficient:

$$\beta(x) = \frac{\mathbb{E}[\Delta Y_i | \Delta X_i = x] - \mathbb{E}[\Delta Y_i | \Delta X_i = 0]}{x} \quad (65)$$

For individuals with  $\Delta X_i = x$ , we have  $X_{i1} = X_{i0} + x$ . Thus:

$$\mathbb{E}[\Delta Y_i | \Delta X_i = x] = \mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) | \Delta X_i = x] + \Delta \gamma \quad (66)$$

For those with  $\Delta X_i = 0$ , we have  $X_{i1} = X_{i0}$ , giving:

$$\mathbb{E}[\Delta Y_i \mid \Delta X_i = 0] = \mathbb{E}[G_i(X_{i0}) - G_i(X_{i0}) + \Delta \gamma \mid \Delta X_i = 0] \quad (67)$$

$$= \Delta \gamma \quad (68)$$

Taking the difference:

$$\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] - \mathbb{E}[\Delta Y_i \mid \Delta X_i = 0] = \mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) \mid \Delta X_i = x] \quad (69)$$

Dividing by  $x$  completes the proof:

$$\frac{\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] - \mathbb{E}[\Delta Y_i \mid \Delta X_i = 0]}{x} = \frac{\mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) \mid \Delta X_i = x]}{x} \quad (70)$$

□

The necessity of a control group (1A) is not unique to the LLS estimator, but is instead a necessary condition for the data to be informative about the  $\tau_i$ . Formally:

**PROPOSITION 3 (Necessity).** *If Assumption 1A.i fails, the identified sets for  $\gamma_t$  and each  $\tau_i$  are the real line.*

**PROOF.** Suppose Assumption 1A.i fails, such that  $\Delta X_i$  is bounded away from zero. Then for any candidate intercept  $a$ , define:

$$B_i(a) \equiv \frac{\Delta Y_i - a}{\Delta X_i} \quad (71)$$

The pair  $(a, B_i(a))$  is observationally equivalent to  $(\gamma_1 - \gamma_0, \tau_i)$  since they generate the same joint distribution of  $(\Delta Y_i, \Delta X_i)$  and satisfy  $\mathbb{E}[\Delta Y_i - a - B_i(a)\Delta X_i \mid \Delta X_i] = 0$ . We can repeat the exercise by first choosing any  $i'$  and any  $B_{i'}$ . Choose  $a(B_{i'}) \equiv \frac{\Delta Y_{i'}}{B_{i'}\Delta X_{i'}}$  and then choose the remaining  $B_i$  as above.

Thus the identified sets for  $\gamma_1 - \gamma_0$  and  $\tau_i$  are the real line. Choose an arbitrary  $\gamma_1 - \gamma_0$  or an arbitrary  $\tau_{i'}$  for some  $i'$  and there are values for the remaining parameters that rationalize the data.  $\square$

The “control group” is crucial to identify  $\gamma_t$  in this flexible model. If there is no control group it is necessary to consider adding additional assumptions. One solution would be simply to assume that  $\gamma_t = 0$  such that causal effects can be directly identified from with-individual first-differences.

### C.2.2. Identification in Active Experiments

ASSUMPTION 2A. *The active control design maintains assumptions 2 from above, with the following modifications:*

- i. *Relevance:  $S_i(A) \neq S_i(B)$  and  $\alpha_i > 0$ .*
- ii. *Nonlinear outcome: Use the general form of potential outcomes*

$$Y_i(x) = G_i(x) \tag{1}$$

PROPOSITION 4 (Active Control Identification). *Under Assumption 2, for any value  $c$  of the control vector  $C_i \equiv [\alpha_i \ X_i^0 \ S_i(A) \ S_i(B)]$ :*

$$\frac{\mathbb{E}[G_i(x_A) - G_i(x_B) \mid C_i = c]}{x_A - x_B} = \frac{\text{Cov}[Y_i, X_i \mid C_i = c]}{\text{Var}[X_i \mid C_i = c]} \tag{72}$$

where  $x_A$  and  $x_B$  are the deterministic belief values for individuals with  $C_i = c$ . In the special case where  $G_i(X_{it}) = \tau_i X_{it} + U_i$  as in (1), the estimand simplifies further to  $\mathbb{E}[\tau_i \mid C_i = c]$ .

PROOF. Since  $C_i$  includes  $\alpha_i$ ,  $X_i^0$ ,  $S_i(A)$ , and  $S_i(B)$ , the potential beliefs take the same value

for all individuals with  $C_i = c$ .

$$X_i(A) = X_i^0 + \alpha_i(S_i(A) - X_i^0) \quad (73)$$

$$X_i(B) = X_i^0 + \alpha_i(S_i(B) - X_i^0) \quad (74)$$

Thus, conditional on  $C_i = c$ , the observed belief  $X_i$  equals either  $X_i(A) = x_A$  or  $X_i(B) = x_B$  depending solely on the randomly assigned treatment  $Z_i$ . By Proposition 1, the regression of  $Y_i$  on  $X_i$  conditional on  $C_i = c$  has coefficient:

$$\beta(c) = \frac{\mathbb{E}[Y_i | X_i = x_A, C_i = c] - \mathbb{E}[Y_i | X_i = x_B, C_i = c]}{x_A - x_B} \quad (75)$$

Relevance guarantees that  $x_A \neq x_B$  and therefore  $X_i = x_A$  if and only if  $Z_i = A$ , and  $X_i = x_B$  if and only if  $Z_i = B$ . This yields

$$\beta(c) = \frac{\mathbb{E}[Y_i | Z_i = A, C_i = c] - \mathbb{E}[Y_i | Z_i = B, C_i = c]}{x_A - x_B} \quad (76)$$

Then, since  $Z_i$  is randomly assigned, we have:

$$\mathbb{E}[Y_i | Z_i = A, C_i = c] - \mathbb{E}[Y_i | Z_i = B, C_i = c] = \mathbb{E}[G_i(x_A) - G_i(x_B) | C_i = c] \quad (77)$$

Dividing by  $x_A - x_B$  completes the proof:

$$\frac{\text{Cov}[Y_i, X_i | C_i = c]}{\text{Var}[X_i | C_i = c]} = \frac{\mathbb{E}[G_i(x_A) - G_i(x_B) | C_i = c]}{x_A - x_B} \quad (78)$$

□

### C.2.3. Identification in Passive Experiments

Once the control vector  $C_i$  is available, the proof in the passive case is identical to the active case. By convention, we set  $S_i(B) = X_i^0$  in the passive case, so  $S_i(B)$  can be omitted from

the control vector  $C_i$ . The difference lies in constructing the first element of the control vector  $C_i$ . The identification challenge in the passive case is that the learning rate  $\alpha_i$  is unknown for the control group that does not receive information. There are two possible approaches in this case

**ASSUMPTION 6.** *Common signal variance and observed prior variance.*

- a. Let  $\alpha_i = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_S^2}$  with  $\sigma_S^2$  common across individuals.
- b. The researcher knows  $\sigma_{X_i}^2$ .

In normal-normal Bayesian updating,  $\alpha_i = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_S^2}$ , where  $\sigma_{X_i}^2$  is the variance of the prior belief  $X_i^0$  and  $\sigma_S^2$  is the variance of the signal  $S_i$ . The first assumption, that  $\sigma_S^2$  is common, means that people all think the signal is equally informative. The second assumption is about the design of the experiment and simply states that the variance of the prior distribution is elicited as in Kumar et al. (2023).

**ASSUMPTION 7.** *Belief updates can be predicted from observables (i.e. no unobservable heterogeneity in updating).*

- a. There is some function  $f$  with (estimable) parameters  $\theta$  such that  $X_i(A) = f(\theta, W_i)$

For example, if  $f$  is a linear function of  $W_i$  as in Balla-Elliott et al. (2022) and Cantoni et al. (2019), then  $X_i(A) = W_i' \theta$ . Since  $Z_i$  is randomly assigned,  $\theta$  is identified from a regression on the sample assigned to  $A$ .

If assumption 6 does not hold, researchers who would like to estimate the APE must make a strong assumption that there are sufficiently rich covariates to predict all of the heterogeneity in belief updating. This is in contrast with the active control designs, that use the observed updates as a “revealed preference” measure of peoples’ learning rates.

**ASSUMPTION 3A.** *The passive control design maintains assumptions 3 from above, with the following modifications:*

- i. Relevance:  $S_i(A) \neq X_i^0$  and  $\alpha_i > 0$ .

ii. *Nonlinear outcome: Use the general form of potential outcomes*

$$Y_i(x) = G_i(x) \quad (1)$$

iii. *Inferred Learning Rate: Either assumption 6 or 7 holds*

Assumption 3A for the passive case contains Assumption 2A for the active case, and adds 3A.iii since the learning rate is not directly identified for the control group.

**PROPOSITION 5 (Passive Control Identification).** *Under Assumption 3A, for any value  $c$  of the control vector  $C_i$  implied by either 6 or 7*

$$\frac{\mathbb{E}[G_i(x_A) - G_i(x_B) \mid C_i = c]}{x_A - x_B} = \frac{\text{Cov}[Y_i, X_i \mid C_i = c]}{\text{Var}[X_i \mid C_i = c]} \quad (79)$$

**PROOF.** Under Assumption 6,  $\alpha_i$  is a one-to-one function of  $\sigma_{X_i}^2$ . Thus conditioning on  $\sigma_{X_i}^2$  or its rank is equivalent to conditioning on  $\alpha_i$  and so conditional on  $C_i \equiv [\text{rank}(\sigma_{X_i}^2) \ X_i^0 \ S_i(A)]$ ,  $X_i(A)$  and  $X_i(B)$  are deterministic. The rest of the proof is identical to the active case.

Under Assumption 7,  $X_i(A)$  in the control group is known from  $f(\theta, W_i)$ . To maintain similar arguments as the other cases, notice then that this implies that  $\alpha_i$  is identified from  $\frac{f(\theta, W_i) - X_i^0}{S_i(A) - X_i^0}$  for the control group and directly from  $\frac{X_i - X_i^0}{S_i(A) - X_i^0}$  for the treated group. Then, conditional on  $C_i \equiv [\alpha_i \ X_i^0 \ S_i(A)]$ ,  $X_i(A)$  and  $X_i(B)$  are deterministic. The rest of the proof is identical to the active case.

□

In each case, integrating over the distribution of the conditioning variables recovers an average partial effect  $\mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)}\right]$ . In the linear case, we recover the average coefficient  $\mathbb{E}[\tau_i]$ .

### C.3. Linear Controls in a Reweighted Regression

This section shows that a reweighted linear regression that controls for  $\alpha_i$  nonparametrically but only controls linearly for  $X_i^0, S_i(A), S_i(B)$  also identifies the APE under the maintained assumptions.

**PROPOSITION 6 (Linear Controls with Reweighting).** *Consider the active control design with nonlinear potential outcomes  $Y_i = G_i(X_i)$ . Let  $W_i = [X_i^0 \ S_i(A) \ S_i(B)]'$ . Under Assumption 2A, conditional on  $\alpha_i$ , the weighted regression of  $Y_i$  on  $X_i$  and  $W_i$  with weights proportional to  $(S_i(A) - S_i(B))^{-2}$  yields a coefficient on  $X_i$  that identifies:*

$$\mathbb{E} \left[ \frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \middle| \alpha_i \right] \quad (80)$$

*In the special case where  $G_i(x) = \tau_i x + U_i$ , this estimand simplifies further to  $\mathbb{E}[\tau_i \mid \alpha_i]$ .*

*The analogous result holds for the passive design under Assumption 3A, with  $S_i(B) = X_i^0$  by convention. The reweighted regression then has weights proportional to  $(S_i(A) - X_i^0)^{-2}$ .*

**PROOF.** Consider the active design; the passive case follows analogously with  $S_i(B) = X_i^0$ . Appealing to FWL, consider the coefficient on  $\tilde{X}_i$ , the residual from the projection of  $X_i$  onto  $W_i = [X_i^0 \ S_i(A) \ S_i(B)]'$  conditional on  $\alpha_i$ . That is:

$$\tilde{X}_i = X_i - \mathbb{L}_{\alpha_i}[X_i \mid W_i] = X_i - \mathbb{E}[X_i \mid W_i, \alpha_i] \quad (81)$$

The second equality uses the fact that, under learning rate updating (4), the true conditional expectation is linear in  $W_i$  conditional on  $\alpha_i$ :

$$\mathbb{E}[X_i \mid W_i, \alpha_i] = (1 - \alpha_i)X_i^0 + \alpha_i S_i(B) + \mathbb{E}[T_i \mid \alpha_i] (S_i(A) - S_i(B)) \quad (82)$$

Thus the residual is with respect to the true conditional expectation and not only the linear projection. The notation  $\mathbb{L}_{\alpha_i}[X_i \mid W_i]$  is meant to highlight the fact that linear projection is

onto  $W_i$  after conditioning on  $\alpha_i$ . Writing  $X_i$  in a similar form shows that

$$X_i = (1 - \alpha_i)X_i^0 + \alpha_i S_i(B) + T_i \alpha_i (S_i(A) - S_i(B)) \quad (83)$$

$$\tilde{X}_i \equiv X_i - \mathbb{E}[X_i | W_i, \alpha_i] = \alpha_i (T_i - \mathbb{E}[T_i]) (S_i(A) - S_i(B)) \quad (84)$$

The weighted coefficient from regressing  $Y_i$  on  $\tilde{X}_i$  with weights  $(S_i(A) - S_i(B))^{-2}$  is thus:

$$\beta_\alpha = \frac{\mathbb{E}[Y_i \tilde{X}_i (S_i(A) - S_i(B))^{-2} | \alpha_i]}{\mathbb{E}[\tilde{X}_i^2 (S_i(A) - S_i(B))^{-2} | \alpha_i]} \quad (85)$$

$$= \frac{\mathbb{E}[Y_i \cdot \alpha_i (T_i - \mathbb{E}[T_i]) (S_i(A) - S_i(B))^{-1} | \alpha_i]}{\mathbb{E}[\alpha_i^2 (T_i - \mathbb{E}[T_i])^2 | \alpha_i]} \quad (86)$$

$$= \frac{\mathbb{E}[Y_i \cdot (T_i - \mathbb{E}[T_i]) (S_i(A) - S_i(B))^{-1} | \alpha_i]}{\alpha_i \mathbb{E}[(T_i - \mathbb{E}[T_i])^2 | \alpha_i]} \quad (87)$$

Now, we compute the numerator:

$$\mathbb{E}\left[Y_i \cdot \frac{(T_i - \mathbb{E}[T_i])}{(S_i(A) - S_i(B))} \mid \alpha_i\right] = \mathbb{E}\left[G_i(X_i) \cdot \frac{(T_i - \mathbb{E}[T_i])}{(S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (88)$$

$$= \mathbb{E}[T_i](1 - \mathbb{E}[T_i]) \cdot \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{S_i(A) - S_i(B)} \mid \alpha_i\right] \quad (89)$$

Note that the denominator simplifies to  $\alpha_i \mathbb{E}[(T_i - \mathbb{E}[T_i])^2 | \alpha_i] = \alpha_i \mathbb{E}[T_i](1 - \mathbb{E}[T_i])$  since  $T_i$  is Bernoulli. Substituting both into the expression for  $\beta_\alpha$ :

$$\beta_\alpha = \frac{\mathbb{E}[T_i](1 - \mathbb{E}[T_i])}{\alpha_i \mathbb{E}[T_i](1 - \mathbb{E}[T_i])} \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{(S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (90)$$

$$= \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{\alpha_i (S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (91)$$

Given that  $X_i(A) - X_i(B) = \alpha_i (S_i(A) - S_i(B))$ , the denominator simplifies further to:

$$\beta_\alpha = \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \mid \alpha_i\right] \quad (92)$$

This completes the proof. The derivation for the passive case is analogous, with  $S_i(B) = X_i^0$

by convention. The weights are then proportional to  $(S_i(A) - X_i^0)^{-2}$ . □

## **D. Estimation Details**

This section provides estimation details, including implementation protocols for each experimental design with specific guidance on the specification of the “local” regression, trimming, and bandwidth selection.

### **D.1. Linear Belief Updating Simplifies Estimation**

The panel design does not rely on the learning rate model and its local regression follows immediately from the identification discussion in Section 3.2.1. In the between-person designs (with active or passive controls), sample sizes are small enough that it is quite demanding to non-parametrically control for the learning rate, the prior, and potential signals. Thus it is convenient to take full advantage of the linearity in the belief updating process to condition only on the learning rate nonparametrically and control for the prior and potential signals linearly. In passive designs, or designs with person-specific high and low signals (i.e. Roth et al. (2022)), it is also necessary to reweight by the inverse of the exposure.

The specific specifications used for estimation are as follows:

#### **D.1.1. Local Regressions in Panel Experiments**

Conditional on the rank of the observed change in beliefs  $\Delta X_i$ , regress the change in the outcome  $\Delta Y_i$  on the change in beliefs  $\Delta X_i$  and a constant. This is exactly the local regression in Section 3.2.1.

### D.1.2. Local Regressions in Active and Passive Control Experiments

In active designs, the learning rate  $\alpha_i = (X_i - X_i^0) / (S_i - X_i^0)$  is directly observed for all individuals, since beliefs are elicited in both treatment arms. In passive designs, however, the learning rate is unobserved for the control group that receives no information. In this case, the learning rate must be imputed using one of the approaches described in Section 3.2.3: either from observed prior variance under common signal precision, or from predicted beliefs using rich observables. Online Appendix C.2.3 formally states the assumptions required in each case.

Given an observed or imputed learning rate, the regression procedure is the same in active and passive control experiments. Conditional on the rank of the observed learning rate  $\alpha_i$ , regress the outcome  $Y_i$  on the posterior belief  $X_i$ , the prior  $X_i^0$ , the signals  $S_i(A)$  and  $S_i(B)$ , and a constant. In the active case, if there is variation in the individual signals  $S_i(A), S_i(B)$ , weight the regression by  $(S_i(A) - S_i(B))^{-2}$ . In the passive case, weight the regression by  $(S_i(A) - X_i^0)^{-2}$ .

When signals are common across individuals (as in Settele (2022)), the signal controls are collinear with the constant and can be omitted. The controls for  $S_i(A)$  and  $S_i(B)$  are necessary only in designs with individual-specific signals (as in Roth et al. (2022)).

## D.2. Trimming

The estimator will perform poorly as the change in beliefs approaches zero. Trimming “away from zero” as in Graham and Powell (2012) thus can greatly improve the performance of the estimator in finite samples.<sup>28</sup>

---

<sup>28</sup>As in Graham and Powell (2012), we can impose some mild regularity conditions (i.e. smoothness and continuity) on the function  $\tau(c) = E[\tau_i | C_i = c]$  such that trimming does not affect the consistency of the estimators when the trimming thresholds are asymptotically zero.

### D.2.1. Trimming in Panel Experiments

Choose a threshold  $h^*$  and exclude treated observations with small nonzero changes in beliefs  $0 < |\Delta X_i| < h^*$ . Observations with  $\Delta X_i = 0$  (the control group in the panel design) are never trimmed. This is a special case of Graham and Powell (2012).

### D.2.2. Trimming in Active Control Experiments

Choose a threshold learning rate  $\alpha^*$  and exclude observations with a learning rate  $\alpha < \alpha^*$ . If there is variation in the individual signals  $S_i(A), S_i(B)$ , it is also important to choose a threshold  $s^*$  and exclude observations with  $(S_i(A) - S_i(B))^2 < s^*$  to ensure that the weights do not diverge (notice that when  $S_i(A) = S_i(B)$  the instrument is not relevant and  $(S_i(A) - S_i(B))^{-2}$  is not finite).

### D.2.3. Trimming in Passive Control Experiments

Choose a threshold learning rate  $\alpha^*$  and exclude observations with a learning rate  $\alpha < \alpha^*$ . Also, choose a threshold  $s^*$  and exclude observations with  $(S_i(A) - X_i^0)^2 < s^*$  to ensure that the weights do not diverge (notice that when  $S_i(A) = X_i^0$  the instrument is not relevant and  $(S_i(A) - X_i^0)^{-2}$  is not finite).

## D.3. Bandwidth Selection

Table D.1 presents Local Least Squares (LLS) estimates across all six applications alongside the original paper estimates for comparison. For each application, I report LLS estimates using four different bandwidth choices to illustrate the bias-variance tradeoff inherent in nonparametric estimation methods.

In the all applications, the conditioning variable (the learning rate or belief update) is transformed to ranks and normalized to the unit interval. Since the Epanechnikov kernel only has positive weight on the interval  $(-1, 1)$ , this makes the bandwidth directly

interpretable as the share of observations that receive positive weight in each local regression. To be explicit, for a bandwidth  $h$ , use  $K\left(\frac{R(\Delta X_i) - R(x)}{h/2}\right)$ , where  $R(\cdot)$  denotes the rank transformation and  $K$  is the Epanechnikov kernel. For example, a bandwidth of 0.05 roughly means that 5% of the data is used in each local regression; this is a parsimonious way to implement an adaptive bandwidth that gets larger in areas where there are fewer observations.

For the main analysis in the paper, the bandwidths range from 0.01 to 0.1. These bandwidths are small enough to minimize contamination from inappropriate comparisons across different treatment intensities, yet large enough to yield reasonably precise estimates. In most studies, the estimates are relatively stable across several bandwidths. More reassuringly, the CAPE curves are also qualitatively similar across bandwidths. For example, Figure D.3 shows that the CAPE estimates for Settele (2022) have a consistent peak in the second quartile and estimates in Figure D.5 (Kumar et al., 2023) consistently slope downwards.

Estimation in active and passive designs proceeds in multiple steps: first, estimate the learning rate  $\alpha_i$  (or its rank); second, estimate the “local” regressions over the grid of learning rates; third, aggregate the local estimates by bins of the learning rate to estimate the CAPE (as in Figure 1) or over the entire grid to estimate the APE (as in Table 1). Estimation in the panel case also proceeds in multiple steps, but skips estimation of the learning rate and begins directly by estimating local regressions conditional on the change in beliefs. It is important that the bootstrap resampling takes place before the first step so that the resulting standard errors reflect the uncertainty associated with the entire procedure. All standard errors in this paper are estimated using 1000 iterations of the Bayesian bootstrap with 1% of outliers dropped for stability (Hansen, 2022).

TABLE D.1. LLS and Fixed Effects Estimates

PANEL A: Panel Experiments Wiswall and Zafar (2015)				
Coefficient	0.695	0.721	0.808	0.379
Standard Error	(0.284)	(0.29)	(0.319)	(0.276)
Bandwidth	0.025	0.05	0.075	0.1
Armona, Fuster, and Zafar (2019)				
Coefficient	1.716	1.8	1.64	1.69
Standard Error	(0.377)	(0.387)	(0.384)	(0.367)
Bandwidth	0.01	0.025	0.05	0.1
PANEL B: Active Experiments Settele (2022)				
Coefficient	0.178	0.16	0.132	0.117
Standard Error	(0.061)	(0.042)	(0.037)	(0.035)
Bandwidth	0.005	0.01	0.025	0.05
Roth, Settele, and Wohlfart (2022)				
Coefficient	1.138	0.882	0.591	0.353
Standard Error	(0.373)	(0.365)	(0.352)	(0.322)
Bandwidth	0.05	0.075	0.1	0.15
PANEL C: Passive Experiments Kumar, Gorodnichenko, and Coibion (2023)				
Coefficient	1.368	1.787	2.036	2.214
Standard Error	(0.457)	(0.465)	(0.538)	(0.589)
Bandwidth	0.01	0.025	0.05	0.1
Cantoni, Yang, Yuchtman, and Zhang (2019)				
Coefficient	0.182	0.18	0.18	0.179
Standard Error	(0.236)	(0.164)	(0.133)	(0.12)
Bandwidth	0.025	0.05	0.1	0.2

*Notes:* This table presents estimates of the effect of beliefs on outcomes from all six replication studies. LLS estimates are presented at four different bandwidth choices. In all applications, the conditioning variable is transformed to ranks; these bandwidths thus have intuitive interpretation as the share of the data used in each local regression. Standard errors are reported in parentheses. They are the standard deviation of the bootstrap distribution with 1000 draws and 1% of outliers dropped for stability (Hansen, 2022).

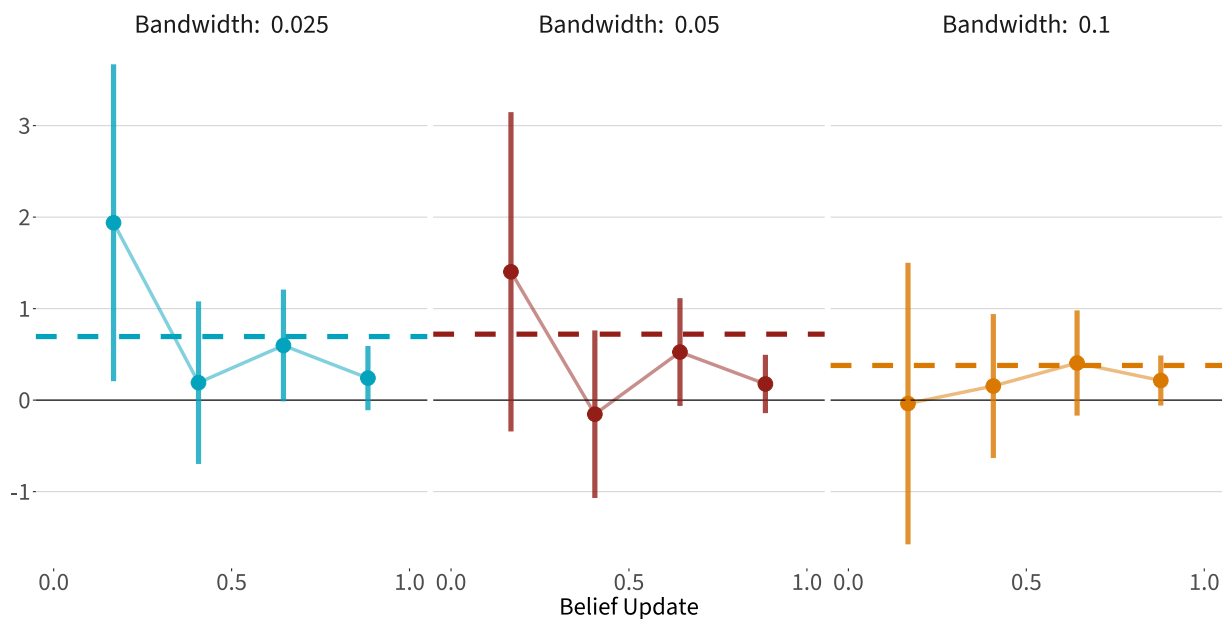


FIGURE D.1. Conditional Average Partial Effects in Wiswall and Zafar (2015), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $E[\tau_i | \Delta X_i = x]$  against the size of the belief update  $x$ . Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

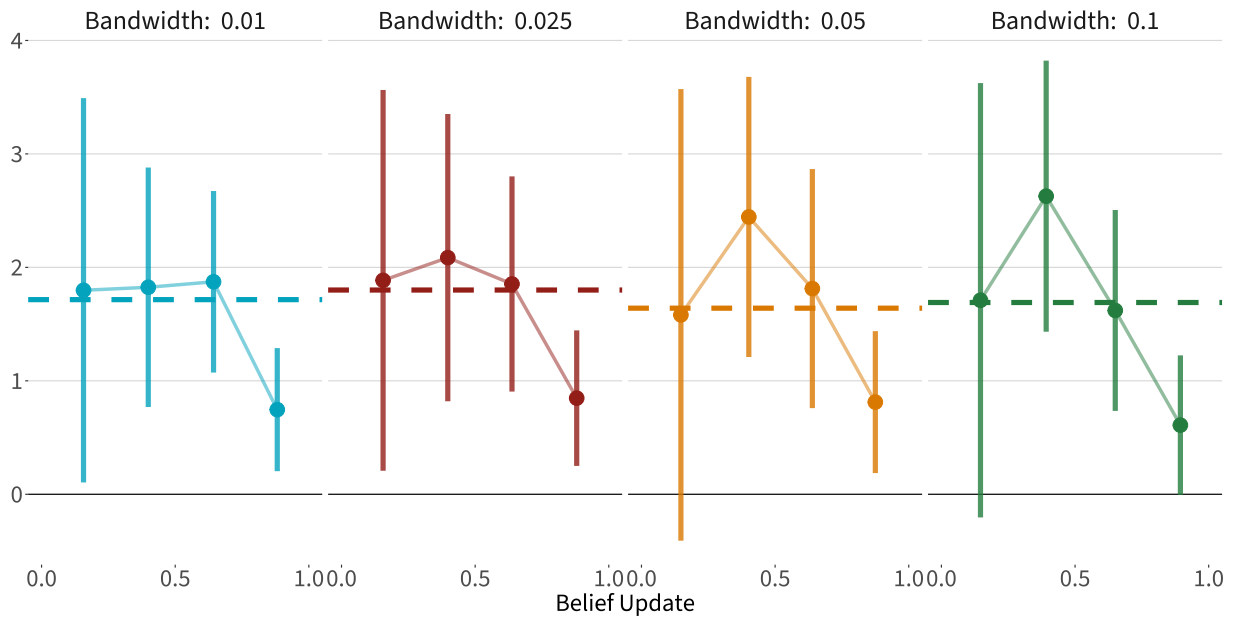


FIGURE D.2. Conditional Average Partial Effects in Armona et al. (2019), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $E[\tau_i | \Delta X_i = x]$  against the size of the belief update  $x$ . Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

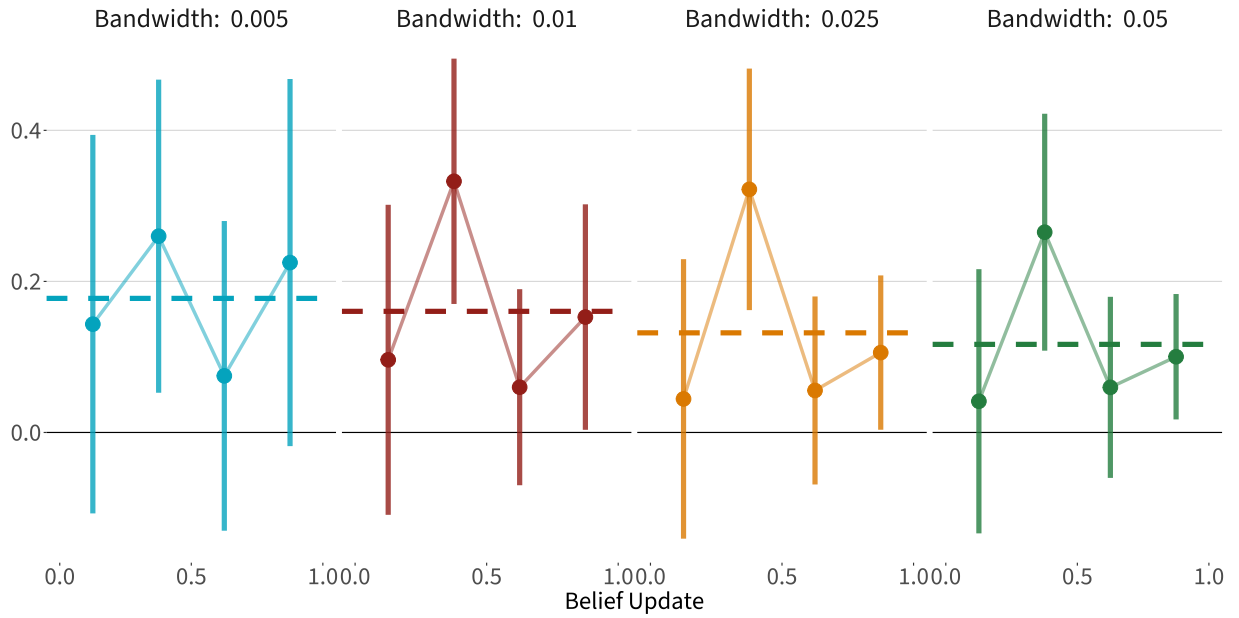


FIGURE D.3. Conditional Average Partial Effects in Settele (2022), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $\mathbb{E}[\tau_i | \text{rank}(\alpha_i)]$  against the rank of the individual learning rate. Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

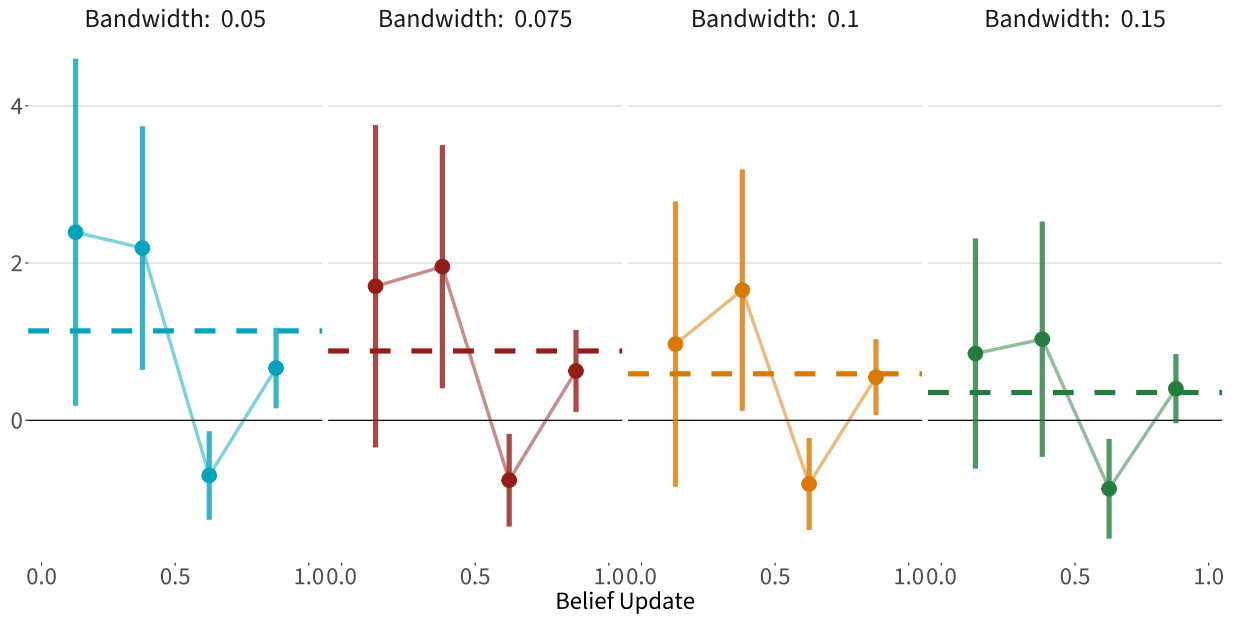


FIGURE D.4. Conditional Average Partial Effects in Roth et al. (2022), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $\mathbb{E}[\tau_i | \text{rank}(\alpha_i)]$  against the rank of the individual learning rate. Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

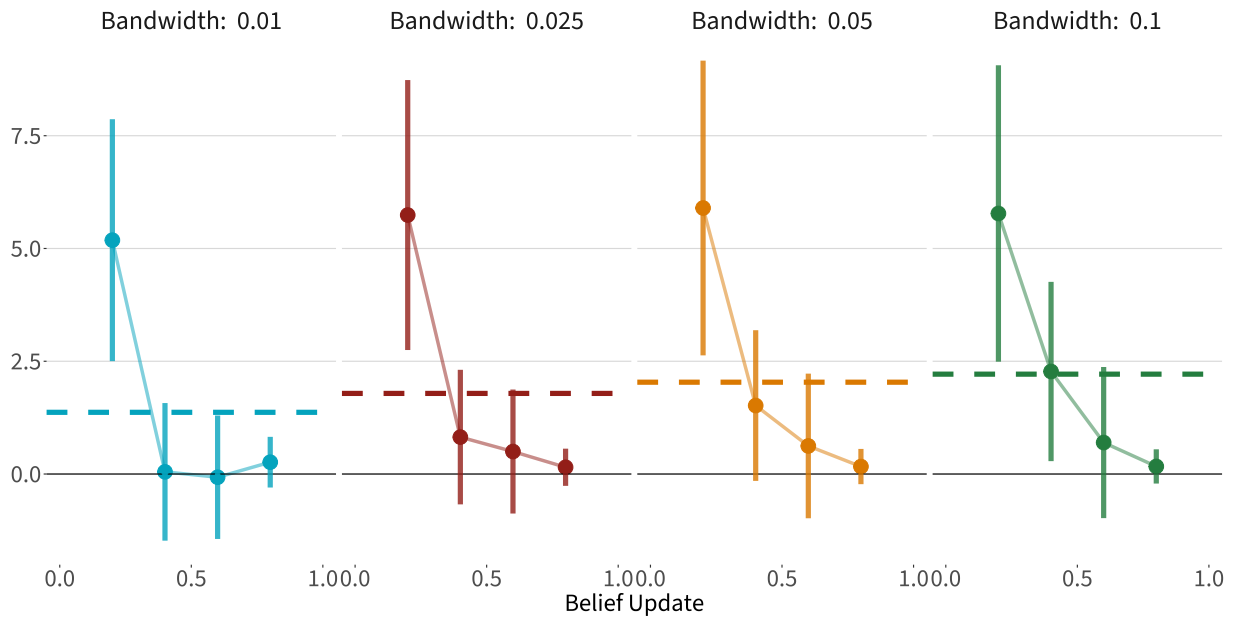


FIGURE D.5. Conditional Average Partial Effects in Kumar et al. (2023), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $\mathbb{E}[\tau_i | \text{rank}(\alpha_i)]$  against the rank of the individual learning rate. Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

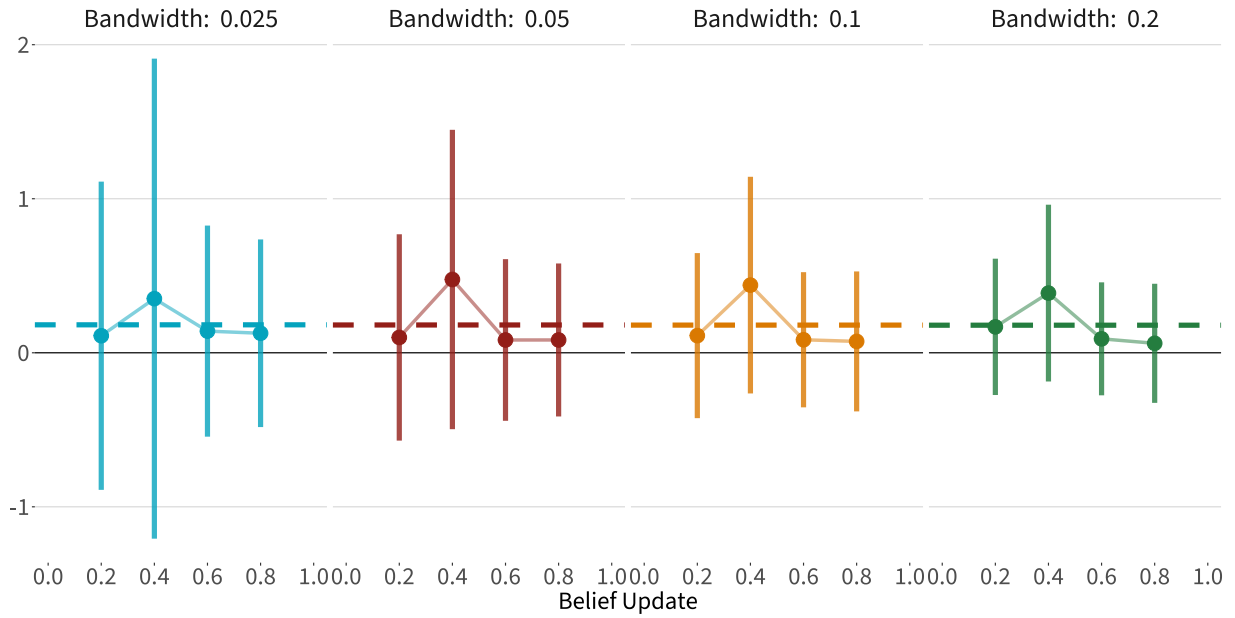


FIGURE D.6. Conditional Average Partial Effects in Cantoni et al. (2019), Several Bandwidths

*Notes:* This figure plots estimates of the conditional average partial effect  $\mathbb{E}[\tau_i | \text{rank}(\alpha_i)]$  against the rank of the individual learning rate. Each panel shows results for a different bandwidth choice. The dashed horizontal line in each panel shows the average partial effect (APE) estimated using that bandwidth. Confidence intervals displayed are twice the bootstrap standard errors. See Table D.1 for the point estimate and standard error of the APE.

## E. Application Details

This section provides additional information about the key specifications under consideration in each of the six applications.

### E.1. Systematic Selection of Empirical Applications

I identified papers for empirical reanalysis through a systematic search of top economics journals. On April 15, 2024, I searched the Web of Science database for papers published in the top five economics journals (American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, Review of Economic Studies), the Review of Economics and Statistics, the four American Economic Journals, and AER: Insights. The search identified papers containing “beliefs,” “information,” or “perception” together with “experiment” or “treatment” in their title or abstract. This yielded 344 papers and 22 duplicates.<sup>29</sup>

I applied a hierarchical set of exclusion criteria to identify papers suitable for reanalysis. The initial screen excluded papers that were not information provision experiments (228 papers). This left 116 experiments, which I sorted by citation count and classified according to the first exclusion criterion each failed. I required papers to study how beliefs affect outcomes, not just how information affects beliefs (the first stage) or how information affects outcomes (reduced form). The experimental design had to follow a prior-treatment-posterior-outcome structure and be one of three types compatible with our estimator: panel experiments, active control experiments, or passive control experiments. For passive control designs, I additionally required that the study elicit the variance or uncertainty of

---

<sup>29</sup>The Web of Science search query was (SO=(JOURNAL OF POLITICAL ECONOMY) OR SO=(AMERICAN ECONOMIC REVIEW) OR SO=(QUARTERLY JOURNAL OF ECONOMICS) OR SO=(REVIEW OF ECONOMIC STUDIES) OR SO=(ECONOMETRICA) OR SO=(REVIEW OF ECONOMICS "AND" STATISTICS) OR SO=(AMERICAN ECONOMIC JOURNAL\* OR AMERICAN ECONOMIC REVIEW INSIGHTS)) AND (TI=(Belief OR Information OR perception) OR AB=(Belief OR Information OR perception)) AND (TI=(Experiment OR Treatment) OR AB=(Experiment OR Treatment))

participants' prior beliefs, which is necessary to model heterogeneity in belief updating. Finally, I required publicly available replication data.

I sought two examples of each experimental design type. After identifying six papers meeting all criteria (two panel, two active control, two passive control), I stopped screening. The remaining 61 papers had fewer citations than the least-cited included paper.

Table E.1 shows the classification of all 344 papers. Among the 55 experiments I fully screened, the most common reasons for exclusion were incompatible experimental designs (25 papers), papers studying reduced-form effects of information without measuring beliefs (10 papers), and papers studying only belief updating without measuring outcome effects (8 papers). These design-incompatible experiments are experiments related to beliefs or information but using designs other than the ones studied in the paper. For example, Pallais (2014) randomly assigns workers to receive either detailed or coarse public evaluations and shows that inexperienced workers value this public information about their ability. One paper was excluded solely for lack of public replication data.

## **E.2. Application Details: Wiswall and Zafar (2015)**

Wiswall and Zafar (2015) study how beliefs about future earnings affect how college students choose majors. Their panel experimental design measures beliefs and outcomes before and after an information intervention.

### **E.2.1. Setting**

In their experiment, undergraduate students were surveyed about their beliefs regarding future earnings, as well as population averages. They were also surveyed about their probability of graduating with a particular college major. After eliciting these prior beliefs, students received information about the true population distributions of these attributes. Finally, they reported revised beliefs about future earnings and college major choices.

TABLE E.1. Systematic Paper Selection Results

Classification	Count	Description
<i>Initial search results</i>		
Total papers identified	344	
<i>Stage 1: Information provision experiment</i>		
Not experiment	228	Not an information provision experiment
Experiments to screen	116	
<i>Iterative Step: Screen Until Two Examples Found In Each Design</i>		
Eliminated by citation cutoff	61	Fewer citations than least-cited included paper
Experiments fully screened	55	
<i>Stage 2: Design compatibility and data availability</i>		
No belief measurement	10	Outcome effects only, no beliefs updating
No outcome measurement	8	Belief updating only, no outcome effects
Misc. design incompatible	25	Not prior-treatment-posterior-outcome
Passive, no variance	5	Passive control without prior uncertainty
No replication data	1	Replication package not publicly available
<b>Included in analysis</b>	<b>6</b>	<b>Met all criteria (2 panel, 2 active, 2 passive)</b>

*Notes:* This table shows the results of our systematic search of Web of Science conducted on April 15, 2024. The search covered papers published in the top five economics journals (American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, Review of Economic Studies), Review of Economics and Statistics, all American Economic Journals, and AER: Insights. Search terms were “beliefs,” “information,” or “perception” combined with “experiment” or “treatment” appearing in title or abstract. Papers were sorted by citation count within each category. After identifying two examples of each experimental design type (panel, active control, passive control), I stopped screening; remaining experiments had fewer citations than the least-cited included paper.

### E.2.2. Specification of Interest

The paper’s main econometric specification is a first-difference regression of the change in stated probability of choosing a major on the change in beliefs about earnings. The authors normalize major choice and earnings relative to humanities/arts, thus the key first-differenced variables are

$$\Delta Y_i = \ln(\pi_{k,i,\text{post}}/\pi_{\bar{k},i,\text{post}}) - \ln(\pi_{k,i,\text{pre}}/\pi_{\bar{k},i,\text{pre}}) \quad (93)$$

$$\Delta X_i = \ln(\omega_{k,i,\text{post}}/\omega_{\bar{k},i,\text{post}}) - \ln(\omega_{k,i,\text{pre}}/\omega_{\bar{k},i,\text{pre}}) \quad (94)$$

where  $\pi_{k,i}$  is the probability of majoring in field  $k$  and  $\omega_{k,i}$  is the expected earnings in field  $k$  for individual  $i$ , with  $\bar{k}$  representing humanities/arts. See page 814, equation 9 of Wiswall and Zafar (2015) for details.

This specification follows column 3 of Table 6.B of Wiswall and Zafar (2015). This specification restricts to the sample of freshmen and sophomores (who are more able to adjust their major) and trims out outliers who update beliefs by more than \$50,000. This is the specification with the largest point estimate (and t-statistic) in Table 6.

### **E.2.3. Implementing the LLS Estimator**

I also trim the sample to exclude very small updates (less than 0.05 in absolute value) that aren't exactly zero; this avoids regressions with very small variation in the regressors.<sup>30</sup> I also follow Wiswall and Zafar (2015) and include fixed effects for college major in the local regressions.

## **E.3. Application Details: Armona, Fuster, and Zafar (2019)**

Armona et al. (2019) study how past home price growth affects beliefs about home prices and how these expectations affect investment decisions. Their panel experimental design measures beliefs and outcomes before and after an information intervention.

### **E.3.1. Setting**

In their experiment, participants in an online survey were first asked about their beliefs regarding past and future home price changes in their zip code. After eliciting these prior beliefs, the researchers provided a random subset of respondents with factual information

---

<sup>30</sup>While point estimates are qualitatively similar without trimming away from zero, this trimming is important for the precision of estimates.

about past local home price changes. They then re-elicited expectations about future price changes from all participants, creating an experimental panel. The outcome is constructed from a portfolio allocation task; participants were also asked to assign money to a savings account or a housing fund, both before and after the information treatment.

### **E.3.2. Specification of Interest**

The paper's main econometric specification is a first-difference regression of the change in investment decisions (from the portfolio allocation task) on the change in beliefs about future home price growth.

Define  $\Delta Y_i$  as the change in the percentage allocation to the housing asset and  $\Delta X_i$  as the change in one-year-ahead home price expectations. For each individual  $i$ , we observe these changes directly as first differences:

$$\Delta Y_i = Y_{i1} - Y_{i0} \tag{95}$$

$$\Delta X_i = X_{i1} - X_{i0} \tag{96}$$

This specification follows columns 5-7 of Table 10 of Armona et al. (2019), with covariates omitted to focus on the key variable of interest.

### **E.3.3. Implementing the LLS Estimator**

The sample selection criteria are as follows. As in column (7) of Table 10 of Armona et al. (2019), the coefficient of interest is the coefficient on  $\Delta X_i$  among the treated group; the control group is omitted from the regression. I also trim the sample to exclude very small updates (less than 0.025 in absolute value) that aren't exactly zero to avoid regressions with very small variation in the regressors.

#### **E.4. Application Details: Settele (2022)**

Settele (2022) studies how beliefs about the gender wage gap affect support for policies aimed at reducing gender inequality. The active control experimental design provides all participants with information about the gender wage gap, but varies the information across treatment groups.

##### **E.4.1. Setting**

In the experiment, participants were first asked to report their beliefs about the gender wage gap. Then, participants were randomly assigned to see either a “high gap” truthful estimate (women earn 74% of men’s wages) or a “low gap” truthful estimate (women earn 94% of men’s wages). They were then asked to report their beliefs about the gender wage gap again after seeing the signal and were asked about their support for various gender-equality policies.

##### **E.4.2. Specification of Interest**

The paper’s main econometric specification uses a two-stage least squares (TSLS) regression, where assignment to the “high gap” treatment serves as an instrument for posterior beliefs about the gender wage gap. This specification follows column 7 of Table 5.C of Settele (2022). Posterior beliefs and the outcome are z-scored. The outcome in column 7 is a summary index constructed from demand for six gender-equality policies. The construction of the index is described in Online Appendix D.7 of Settele (2022) as follows:

*To adjust for multiple inference, I follow Anderson (2008) in applying a combined approach: First, I group the main outcome variables of interest into families and test for an overall treatment effect in a highly conservative way. Second, I test for a treatment effect on disaggregated outcomes within each family, allowing for more power in exchange for a small number of Type I errors. In the remainder of this section I describe the implemen-*

*tation of this combined approach and the intuition behind it (page 34, Online Appendix Settele, 2022).*

#### **E.4.3. Implementing the LLS Estimator**

The point estimate in the original paper is negative and seeks to measure the effect of “women’s relative earnings” on support for gender-equality policies. To make the discussion parsimonious across applications, we flip the sign of the belief variable so that point estimates are positive (unlike the original paper). The effect of interest can then be interpreted as the effect of “women’s earnings gap” on support for gender-equality policies.

The sample selection criteria are as follows. We can only estimate the learning rate for individuals with  $\text{prior} \neq \text{signal}$ , so we exclude people with  $\text{prior} = \text{signal}$ . Additionally, the local regression is not identified for individuals with  $\alpha = 0$ , so we exclude them as well.<sup>31</sup> Finally, also exclude individuals with negative learning rates (those whose posterior is farther from the signal than their prior), as their updating doesn’t follow reasonable updating patterns and thus the Bayesian learning structure does not hold on this sample.<sup>32</sup>

As discussed in Online Appendix D.1, it is sufficient to control non-parametrically for the learning rate  $\alpha_i$  and to control linearly for the remaining elements of the control vector  $[S_i(A), S_i(B), X_i^0]$ . Since the signals are common and  $S_i(A) = 74, S_i(B) = 94$  for all  $i$ , this simplifies further. The only remaining control variable is the prior  $X_i^0$  and there is no need to reweight. Following Settele (2022), I include fixed effects for the elicitation subgroup, since this is the level of randomization. Other controls and sampling weights are omitted. The local regression is thus a regression of  $Y_i$  on  $X_i, X_i^0$  and elicitation subgroup fixed

---

<sup>31</sup>Directly dividing the belief update by the difference between the signal and the prior leads to very noisy estimates of the learning rate, which causes the LLS estimator to behave poorly in the bootstrap. Thus, for each individual in the sample, I take a kernel-weighted average of the belief update and the exposure to the signal and use that ratio to construct the learning rate. Intuitively, instead of constructing the learning rate from the raw prior and posterior, I construct it from smoothed versions of the prior and posterior.

<sup>32</sup>Vilfort and Zhang (2025) show that updating “towards the signal” is predicted by a much broader class of models than the Bayesian model. One reasonable interpretation is that these individuals are simply failing an “attention check”.

effects conditional on (the rank of)  $\alpha_i$ .

### **E.5. Application Details: Roth, Settele, and Wohlfart (2022)**

Roth et al. (2022) study how perceived exposure to macroeconomic risk affects households' demand for macroeconomic information. Their active control experimental design exploits sampling variation between two official census surveys to create exogenous variation in beliefs about exposure to unemployment risk.

#### **E.5.1. Setting**

In this experiment, participants first reported their prior beliefs about how the Great Recession affected unemployment rates among similar people. Then, participants were randomly assigned to receive truthful information about actual unemployment rate changes during the Great Recession based on data from either the American Community Survey (ACS) or the Current Population Survey (CPS). Sampling variation and procedural differences between these two surveys generate variation in the signals.

After receiving this information treatment, participants reported their posterior beliefs about their personal probability of becoming unemployed during the next recession. Finally, respondents chose between receiving expert forecasts about four different macroeconomic variables: recession likelihood, inflation, government bond returns, or government spending, or receiving no forecast at all.

#### **E.5.2. Specification of Interest**

The paper's main econometric specification uses a two-stage least squares (TSLS) regression where the difference in unemployment increase information between ACS and CPS data serves as an instrument for posterior beliefs about personal unemployment risk during the next recession. I replicate the main specification where the outcome variable is the probability of choosing to receive a recession forecast (multiplied by 100 so that the

final estimates are in percentage point units). Since there is individual level variation in the potential signals, this estimand does not simplify to the expression given in 11. Instead, this estimand targets a weighted average of  $\tau_i$  with weights  $\omega_i \propto \alpha_i(S_i(A) - S_i(B))^2$ .

More formally, the instrument is

$$T_i^\Delta \equiv \begin{cases} S_i(A) - S_i(B) & \text{if } Z_i = A \\ S_i(B) - S_i(A) & \text{if } Z_i = B \end{cases} \quad (97)$$

and the TSLS estimand is

$$\frac{\text{Cov}[T_i^\Delta, Y_i]}{\text{Cov}[T_i^\Delta, X_i]} = \mathbb{E} \left[ \tau_i \cdot \frac{\alpha_i(S_i(A) - S_i(B))^2}{\mathbb{E}[\alpha_i(S_i(A) - S_i(B))^2]} \right] \quad (98)$$

### E.5.3. Implementing the LLS Estimator

As in Settele (2022), we implement the LLS estimator using the two-step approach. The signals vary across participants based on their demographic characteristics, so we weight the local regressions by the inverse of the squared exposure  $(S_i(A) - S_i(B))^{-2}$  to account for this variation in instrument strength.

The estimation of the learning rate and the sample restrictions are identical to Settele (2022), as discussed in E.4.3. I use a smoothed estimate of the learning rate and exclude individuals with  $\alpha \leq 0$ . Additionally, since there are individual specific signals, I trim individuals with very small variation in the potential signals and require that  $(S_i(A) - S_i(B))^2 > 0.25$ . This ensures that the weights proportional to  $(S_i(A) - S_i(B))^{-2}$  are well behaved.

The local regression is thus a regression of  $Y_i$  on  $X_i, X_i^0, S_i(A), S_i(B)$  conditional on (the rank of)  $\alpha_i$ , with weights proportional to  $(S_i(A) - S_i(B))^{-2}$ . The linear controls for  $X_i^0, S_i(A), S_i(B)$ , are sufficient to ensure that the residual variation is mean independent of the error term  $U_i$ . The weights ensure that each covariate group receives equal weight in the local regression so that the estimand retains its interpretation as an unweighted

average.

## **E.6. Application Details: Kumar, Gorodnichenko, and Coibion (2023)**

Kumar et al. (2023) study how firms' macroeconomic forecasts affect their economic decisions. The passive experiment provided a random subset of participants with a macroeconomic forecast.

### **E.6.1. Setting**

In this experiment, participating firms were first asked to report their prior beliefs about GDP growth. Then, participants were randomly assigned to one of three treatment groups receiving different types of information about macroeconomic forecasts, or to a control group receiving no information. Finally, they reported revised beliefs about GDP growth as well as actual firm decisions six months later.

Like Vilfort and Zhang (2025), I exclude the treatment groups that were designed to shift the second moment of beliefs and use only the first treatment group that provided information about the level of GDP growth.<sup>33</sup> The analysis in this paper uses only comparisons between a single treatment arm and the control.

### **E.6.2. Specification of Interest**

The main econometric specification I replicate is a simplified version of the system of equations given in equations 3 and 4'. Instead of using all treatment arms to instrument for both the posterior mean and posterior uncertainty, I use only the first treatment arm to instrument for the posterior mean. I interact the treatment indicator with the sign of

---

<sup>33</sup>As Vilfort and Zhang (2025) also discuss, belief experiments with multiple information treatments that induce variation in both the level and the uncertainty of beliefs are delicate to interpret when effects of both the mean and the effect of the uncertainty are heterogeneous. In general, TSLS specifications with multiple endogenous variables can be difficult to interpret (Bhuller and Sigstad, 2024).

the difference between the signal and the prior.<sup>34</sup> This specification is similar in spirit to the estimates in Table 3 of Kumar et al. (2023).

### **E.6.3. Implementing the LLS Estimator**

Kumar et al. (2023) elicit not only the mean of the prior belief, but also the variance. The implementation of the LLS estimator in this application thus follows Case 1 discussed in Section 3.1. Under the assumption that individuals agree on the variance of the signal, the rank of the learning rate is simply the rank of the prior variance; conditioning on the rank of the prior variance is sufficient to condition on the learning rate.

I trim individuals with very small variation in the exposure to the signal and require that  $(S_i - X_i^0)^2 > 0.01$ . This ensures that the weights proportional to  $(S_i - X_i^0)^{-2}$  are well behaved.

The local regression is thus a regression of  $Y_i$  on  $X_i, X_i^0$  conditional on (the rank of)  $\sigma_{X_i}^2$ , with weights proportional to  $(S_i - X_i^0)^{-2}$ . The linear control for  $X_i^0$ , is sufficient to ensure that the residual variation is mean independent of the error term  $U_i$ . The weights ensure that the covariate groups receive equal weight in the inner regression so that our estimand retains its interpretation as an unweighted average. To make the CAPE curves presented in Figure 1 Panel C.i and Figure D.5 more comparable to those in other designs, I estimate  $\mathbb{E}(\text{rank}(\alpha) \mid \text{rank}(\sigma_{X_i}^2))$  on the treated group and use this for the x-axis of the binned estimates.

### **E.7. Application Details: Cantoni, Yang, Yuchtman, and Zhang (2019)**

Cantoni et al. (2019) study how beliefs about others' participation in protests affect an individual's own protest decisions. The passive experiment provided a random subset of

---

<sup>34</sup>Vilfort and Zhang (2025) also replicate these results and use only the first treatment arm. They show that results are similar in specifications that interact treatment with the actual difference between the signal and the prior and those that only interact it with the sign of the difference. Results can be different, however, in specifications that also include the un-interacted treatment indicator, since specifications can have negative weights.

participants with truthful information about the planned participation of their classmates.

### **E.7.1. Setting**

In this experiment, participating students were asked to report prior beliefs about their classmates' participation in an upcoming political protest. Then, one day before the protest, a random subset of participants were provided with truthful information about the planned participation of their classmates. Finally, after the protest, they collected data on participants' actual protest behavior.

### **E.7.2. Specification of Interest**

The paper's main econometric specification uses a two-stage least squares (TSLS) regression where treatment indicator, interacted with the sign of the difference between the prior and the signal, is an instrument for posterior beliefs. This specification targets a weighted average of  $\tau_i$  with weights  $\omega_i \propto \alpha_i |S_i - X_i^0|$ .

The TSLS estimand is

$$\frac{\text{Cov} \left[ \text{sign}(S_i - X_i^0) T_i, Y_i \right]}{\text{Cov} \left[ \text{sign}(S_i - X_i^0) T_i, X_i \right]} \quad (99)$$

### **E.7.3. Implementing the LLS Estimator**

Cantoni et al. (2019) collect a rich set of observables in their survey, which they use to predict prior beliefs in a supplemental analysis (Online Appendix Table A.5). The implementation of the LLS estimator in this application thus follows Case 2 discussed in Section 3.1. Under the assumption that the counterfactual belief update in the passive control group can be predicted from rich observables, these estimates can be used to predict the (latent) learning rate in the control group. Then, the estimated learning rate can be used in the place of the observed learning rate in an active design.

I use the replication package provided by the authors to directly replicate the prediction exercise in Appendix Table A.5, directly predicting the learning rate instead of the prior belief. Then, I impose the same restrictions as in the active cases. In particular, I restrict to learning rates strictly greater than zero. Like in E.6.3, I trim individuals with very small variation in the exposure to the signal and require that  $(S_i - X_i^0)^2 > 0.01$ .

The local regression is thus a regression of  $Y_i$  on  $X_i, X_i^0$  conditional on (the rank of)  $\tilde{\alpha}_i$ , with weights proportional to  $(S_i - X_i^0)^{-2}$ . Recall that I use the notation  $\tilde{\alpha}_i$  to emphasize that the learning rate is predicted in the control group. The linear control for  $X_i^0$  is sufficient to ensure that the residual variation is mean independent of the error term  $U_i$ . The weights ensure that the covariate groups receive equal weight in the inner regression so that our estimand retains its interpretation as an unweighted average. To estimate standard errors, we use the empirical bootstrap with 1000 iterations.

#### **E.7.4. Discussion**

The TSLS estimate and the LLS estimate are both quite noisy, making it difficult to draw strong conclusions about the direction or magnitude of any difference. However, if one takes the point estimates literally, it would suggest a different model of the dependence between belief updating and belief effects. Suppose that this is a setting where it is difficult for anyone to form precise beliefs so that uncertainty is widespread. Then, the relevant heterogeneity in updating may come from inattention: people who use the information in their decisions spend time carefully interpreting the signal and incorporating it into their beliefs. In contrast, people whose decisions don't depend on these beliefs may mostly ignore the signal and update their beliefs only slightly. A model where agents choose both how much information to acquire at baseline and how much to pay attention to new information as in Fuster et al. (2022) may be the appropriate theoretical generalization to unify the results across all six studies. An interesting task for future research would be to use the empirical tools provided in this paper to discipline models where the correlation

between belief updating and the belief effects is ex ante ambiguous.

## F. Information Experiments and the TSLS Estimator

This appendix provides discussion of the interpretation of TSLS estimators in information provision experiments. The challenges with obtaining unconditional monotonicity motivate the representative specifications discussed in Section 2, which have non-negative weights under a weaker conditional monotonicity assumption. While the weighted average interpretation of TSLS estimands is well-established (Angrist and Imbens, 1995), this section examines the specific implications for information experiments and relates them to a workhorse learning rate updating assumptions. Section F.4 provides a novel strategy to ensure non-negative weights when priors are not elicited.

### F.1. The Reduced Form Effect of Information Provision

In active designs, the reduced form effect of treatment is the effect of being assigned to see the signal in arm  $A$  rather than the signal in arm  $B$ . In passive designs, this is the effect of being assigned to see new information. Consider the simple OLS regression of the outcome  $Y_i$  on the treatment indicator  $T_i \equiv \mathbb{1}\{Z_i = A\}$ .

$$\beta^{RF} \equiv \frac{\text{Cov}[T_i, Y_i]}{\text{Var}[T_i]} \quad (100)$$

$$= \mathbb{E}[\tau_i (X_i(A) - X_i(B))] \quad (101)$$

The reduced form effect of assignment to arm  $A$  on the outcome is the expectation of the individual effect of beliefs on behaviors  $\tau_i$  scaled by the individual effect of the information treatment on beliefs  $X_i(A) - X_i(B)$ . If all  $\tau_i$  have the same sign, the reduced form effect of treatment assignment on the outcome will be informative of the sign of the effect of beliefs on behaviors only if the  $X_i(A) - X_i(B)$  are all positive or all negative. If the first stage effect on beliefs is positive for some people and negative for others, then the average effect of the information treatment on behaviors can be close to zero, even if the effect

of beliefs on behaviors is large and the individual first stage effects of the information treatment on beliefs are large.

### **F.1.1. From the Effect of Information to the Effect of Beliefs**

As Giacobasso et al. (2022) note, reduced form estimates can be difficult to interpret since they combine the causal effects of beliefs on behaviors with the first stage effects of the information provision on beliefs. The reduced form can therefore be small if beliefs have only a weak effect on behavior, or if the information provision has only a weak effect on beliefs.

The reduced form is most directly policy-relevant when the counterfactual of interest concerns information provision per se rather than belief changes more generally. However, when the relationship of interest is the effect of beliefs on behavior, researchers typically normalize the reduced form effect by the first stage effect and report TSLS estimates.

### **F.1.2. Constructing TSLS Estimates**

To motivate the specifications in Section 2, we consider the simplest TSLS estimand that directly uses treatment assignment  $T_i$  to instrument for beliefs.

$$\beta^{TSLS} \equiv \frac{\beta^{RF}}{\beta^{FS}} = \frac{\text{Cov}[T_i, Y_i]}{\text{Cov}[T_i, X_i]} \quad (102)$$

where  $\beta^{FS} \equiv \text{Cov}[T_i, X_i] / \text{Var}[T_i]$ . For the binary treatment indicator, this becomes

$$\beta^{TSLS} = \frac{\mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0]}{\mathbb{E}[X_i | T_i = 1] - \mathbb{E}[X_i | T_i = 0]} \quad (103)$$

Substituting the linear outcome equation (1) yields

$$\beta^{TSLS} = \frac{\mathbb{E}[\tau_i(X_i(A) - X_i(B))]}{\mathbb{E}[X_i(A) - X_i(B)]} \quad (104)$$

In the presence of heterogeneous effects, TSLS does not generally recover the average of the individual treatment effects. The TSLS coefficient depends on the covariance between individual belief effects  $\tau_i$  and the first stage variation  $X_i(A) - X_i(B)$ :

$$\frac{\mathbb{E}[\tau_i (X_i(A) - X_i(B))]}{\mathbb{E}[(X_i(A) - X_i(B))]} = \mathbb{E}[\tau_i] + \frac{\text{Cov}[\tau_i, (X_i(A) - X_i(B))]}{\mathbb{E}[(X_i(A) - X_i(B))]} \quad (105)$$

The covariance term is the “bias” relative to the APE  $\mathbb{E}[\tau_i]$  and motivates the LLS estimator developed in Section 3.

## **F.2. Unconditional Instrument Monotonicity and Bayesian Updating**

The weights derived in Section F.1.2 are non-negative when unconditional monotonicity holds. This section examines when Bayesian updating ensures monotonicity across different experimental designs.

### **F.2.1. Monotonicity in Active Designs**

In active designs, monotonicity follows directly from Bayesian updating when signals are ordered such that  $S_i(A) \geq S_i(B)$ . Since  $X_i(A) - X_i(B) = \alpha_i(S_i(A) - S_i(B))$  and  $\alpha_i \in (0, 1)$  under Bayesian updating, the sign of the first stage is determined by  $\text{sign}(S_i(A) - S_i(B))$ . The immediacy of monotonicity in active designs should be considered one advantage of this design relative to passive designs.

### **F.2.2. Monotonicity in Passive Designs**

In passive designs, unconditional monotonicity requires that  $S_i(A) - X_i^0$  has the same sign for all participants—either the signal is above everyone’s prior or below everyone’s prior. This is often empirically implausible; in all six empirical examples considered in this paper, we observe participants with priors both above and below the signal. This is why the simple specification (102) is not widely used in practice; instead researchers use

one of two main strategies to ensure positive weights.

### F.3. Strategies for Ensuring Non-Negative Weights in Passive Designs

When unconditional monotonicity fails, researchers can construct specifications with non-negative weights by incorporating information about priors and signals.

#### F.3.1. Sample Splitting Approach

Researchers can split the sample based on whether the signal is above or below each participant's prior, then estimate separate TSLS regressions within each subsample. For participants with  $S_i(A) - X_i^0 > 0$ :

$$\beta_+^{\text{split}} = \frac{\text{Cov}[T_i, Y_i \mid S_i(A) - X_i^0 > 0]}{\text{Cov}[T_i, X_i \mid S_i(A) - X_i^0 > 0]} \quad (106)$$

$$= \mathbb{E} \left[ \tau_i \cdot \frac{\alpha_i |S_i(A) - X_i^0|}{\mathbb{E}[\alpha_i |S_i(A) - X_i^0| \mid S_i(A) - X_i^0 > 0]} \mid S_i(A) - X_i^0 > 0 \right] \quad (107)$$

A symmetric expression applies for  $S_i(A) - X_i^0 < 0$ . Both specifications yield non-negative weights under Bayesian updating since  $\alpha_i > 0$ .

#### F.3.2. Exposure-Weighted Instruments

An example of the exposure-weighted instrument is presented in Section 2.2.3.

$$\tilde{T}_i^{\text{ex}} \equiv (T_i - \mathbb{E}[T_i])(S_i(A) - S_i(B))$$

The recentering is implicit since in practice researchers use the interaction as an instrument and control for the un-interacted exposure. Recall the notational device that in the passive design  $S_i(B) = X_i^0$ . These weights proportional to  $\alpha_i(S_i(A) - X_i^0)^2$  are non-negative under Bayesian updating and in a general class of updating models when the

monotonicity assumption holds:  $\text{sign}(X_i(A) - X_i(B)) = \text{sign}(S_i(A) - S_i(B))$ .

Vilfort and Zhang (2025) show that implementation of these specifications requires care, as including both the exposure-weighted instrument and the treatment indicator can result in misspecification.

#### F.4. Implementation When Priors Are Unobserved

Some experiments do not elicit prior beliefs directly. Under Bayesian updating, the direction of the belief update can be inferred from the posterior belief and the signal. If the posterior lies between the prior and signal, then  $\text{sign}(S_i(A) - X_i) = \text{sign}(S_i(A) - X_i^0)$ , allowing sample splitting even when priors are unobserved. This assumption identifies the same causal parameters that are targeted by  $\beta_+^{\text{split}}$  and  $\beta_-^{\text{split}}$  in Online Appendix F.3.1.

Since the control group that is not shown a signal, we directly observe their prior: recall that  $X_i(B) = X_i^0$  in passive designs. Since the signal is known, we can directly condition on the sign of  $(S_i(A) - X_i^0)$ . The prior for the treated group is unknown and we observe only  $X_i(A)$ . But since we can rewrite the potential outcome equation in 4 as

$$S_i(A) - X_i(A) = (1 - \alpha_i)(S_i(A) - X_i^0)$$

and since  $\alpha \in (0, 1)$  then

$$S_i(A) - X_i(A) > 0 \iff (S_i(A) - X_i^0) > 0$$

We used the Bayesian updating structure, but note this could be relaxed to include any model of updating such that the posterior lies between the prior and the signal.

Thus, although the regressions in Section F.3.1 are not feasible since they use the prior to split the sample, the following regressions are feasible and equivalent.

$$\beta_+^{\text{split}} = \tilde{\beta}_+^{\text{split}} \equiv \frac{\text{Cov}[T_i, Y_i \mid S_i(A) - X_i > 0]}{\text{Cov}[T_i, X_i \mid S_i(A) - X_i > 0]} \quad (108)$$

$$\beta_-^{\text{split}} = \tilde{\beta}_-^{\text{split}} \equiv \frac{\text{Cov}[T_i, Y_i \mid S_i(A) - X_i < 0]}{\text{Cov}[T_i, X_i \mid S_i(A) - X_i < 0]} \quad (109)$$

## G. Nonlinearity, Convex Combinations, and Discrete Slopes

This appendix provides additional detail on the interpretation of the LLS estimand under nonlinear outcome functions. The main text (Section 4) establishes that LLS delivers a representative average regardless of functional form, while linearity adds a structural interpretation permitting extrapolation. This appendix elaborates on three related points.

### G.1. Convex Combinations and Magnitudes

When simple averages are not identified, researchers sometimes target parameters from a broader class of convex combinations (weighted averages with non-negative weights that sum to one). Under a weaker “signal monotonicity” assumption, the unweighted average is not generally identified by LLS, but TSLS can still identify a convex combination of individual effects (Vilfort and Zhang, 2025).

Mogstad and Torgovitsky (2024) note that convex combinations are informative only about the sign of effects, and only when every individual effect has the same sign. Parameters that can be an arbitrary convex combination are generally uninformative about magnitudes. An equally weighted average, by contrast, is informative about both sign and magnitude. This is a key advantage of targeting the LLS estimand under learning-rate updating: it delivers not just a convex combination, but a specific, interpretable average.

### G.2. Discrete Slopes Versus Derivatives

With only two potential beliefs per person, we observe only the average slope  $\beta_i$  over each person’s belief interval  $[X_i(B), X_i(A)]$ , not the full shape of  $G_i(\cdot)$ .<sup>35</sup>

The difference between the average slope  $\mathbb{E}[\beta_i]$  and the non-parametric ATE  $\mathbb{E}[G'_i(X_i)]$  is the difference between slopes over discrete changes and derivatives. The ATE averages derivatives;  $\mathbb{E}[\beta_i]$  averages slopes over discrete changes. In this sense,  $\mathbb{E}[\beta_i]$  can be

---

<sup>35</sup>In general, binary instruments do not identify the nonlinearity of individual response functions without further assumptions (Brinch et al., 2017).

interpreted as a discrete approximation to the ATE.<sup>36</sup>

As the potential beliefs get closer together, the discrete slope converges to the derivative:  $\beta_i \rightarrow G'_i(X_i)$  when  $X_i(A) - X_i(B) \rightarrow 0$ . This would require picking signals very close to each other (in the active case) or very close to the prior (in the passive case). This is likely unattractive in practice, since using a pair of signals that are close to each other would lead to a weak first stage.

### **G.2.1. Identifying Derivatives Without Linearity**

Richer experimental variation could identify derivative-based parameters without assuming linearity. Experiments with  $K$  signal values can identify degree- $(K - 1)$  polynomial approximations to the response function (Masten and Torgovitsky, 2016). Experiments with continuously distributed signals can identify the average structural function  $\mathbb{E}[G_i(x)]$  using nonparametric control function methods (Heckman and Vytlacil, 2007; Imbens and Newey, 2009).

These approaches require substantially richer data than the two-arm experiments typical in current practice. The practical implication is that researchers using standard experimental designs face a choice: either maintain linearity and interpret  $\mathbb{E}[\beta_i]$  as the ATE/APE, or relax linearity and interpret  $\mathbb{E}[\beta_i]$  as a representative average of slopes over the discrete belief changes induced by the experiment.

### **G.3. Nonlinearity Affects Interpretation, Not the Case for Equal Weights**

The key point from Section 4 bears repeating: nonlinearity affects the interpretation of individual slopes  $\beta_i$ , but it does not affect the difference between LLS and TSLS. Both estimators aggregate the same individual slopes; they differ only in how they weight those

---

<sup>36</sup>Researchers may also be interested in an alternative parameter  $\mathbb{E}[G'_i(X_i^0)]$ , where derivatives are evaluated at the prior rather than the posterior. Under linearity,  $G'_i(x) = \tau_i$  for all  $x$ , so  $\mathbb{E}[\beta_i]$ , the ATE, and this alternative all coincide. Under nonlinearity, the gap between the discrete-slope parameter  $\mathbb{E}[\beta_i]$  and derivative-based parameters depends on both the curvature of  $G_i(\cdot)$  and the width of the belief interval  $X_i(A) - X_i(B)$ .

slopes. LLS uses equal weights; TSLS uses weights proportional to belief updating.

Whether or not outcomes are linear in beliefs, researchers who want a representative summary of heterogeneous effects should prefer equal weights. Linearity is an additional assumption that strengthens the interpretation of the equally-weighted average; it is not a precondition for preferring equal weights over unequal weights.