Identifying Causal Effects in Information Provision Experiments

Dylan Balla-Elliott

June, 2025

Information treatments often shift beliefs more for people with weaker belief effects. Since standard TSLS and panel specifications in information provision experiments have weights proportional to belief updating in the first-stage, this dependence attenuates existing estimates. This is natural if people whose decisions depend on their beliefs gather information before the experiment. I propose a local least squares estimator that identifies unweighted average effects in several classes of experiments under progressively stronger versions of Bayesian updating. In five of six recent studies, average effects are larger than–in several cases more than double–estimates in standard specifications.

JEL CODES: C26, C9, D83, D9

How do people make important decisions? Information provision experiments have emerged as a powerful tool to answer this question, revealing how people choose what to study (Wiswall and Zafar, 2015) or where to live (Bottan and Perez-Truglia, 2022b) They are also a credible way to generate identifying variation more generally. For example, rather than experimentally manipulating actual returns to education, Jensen (2010) provides information to shift beliefs about returns to education and observe how this affects schooling choices. Similarly, information about wages, home prices, or recession risk can generate variation in beliefs about these fundamentals without needing to change the fundamentals themselves (Armona et al., 2019; Jäger et al., 2023; Roth and Wohlfart, 2020).

In these experiments, researchers vary the information ("signal") shown to participants. Then, they typically estimate the effect of beliefs on behavior using panel or two-stage least squares (TSLS) regressions. It is well known that these estimators target weighted averages of individual causal effects.[1] In information provision experiments, these weights are proportional to the first-stage effect of information on beliefs. This creates a potential problem: strong dependence between belief updating and belief effects makes existing estimates substantially misrepresent average effects.

This paper demonstrates that attenuation arising from this dependence is empirically widespread. I therefore propose a local least squares (LLS) estimator that consistently estimates the average partial effect (APE), even when there is strong dependence between belief updating and belief effects. This estimator applies to several classes of experiments–panel, active control, and passive control–under plausible assumptions about belief updating.[2] As an intermediate step, researchers recover estimates of the dependence between belief updating and belief effects, which can reveal underlying mechanisms regardless of the ultimate choice of estimator and target parameter.

I apply the LLS estimator to six recent information provision studies published in leading economics journals.[3] In five of these six applications, the LLS estimates of the

---

[1] See the influential literature on TSLS following Imbens and Angrist (1994). Results of this flavor have also been specialized to a range of empirical settings. Difference-in-difference is a leading example (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2020).

[2] The LLS estimator applies immediately in the panel experiment. In experiments with active control groups, the LLS estimator identifies the APE under a Bayesian updating assumption. In experiments with passive control groups, the LLS estimator identifies the APE when the variance of the prior is elicited in addition to the mean and Bayesian updating is slightly strengthened. 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".)

[3] 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 and Wohlfart, 2020) 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

APE are meaningfully larger than the conventional estimators. In two cases the estimates more than double. Directly estimating the dependence between belief updating and belief effects reveals that in many cases the belief effects are largest for the groups with the smallest belief updates. This pattern suggests that workhorse TSLS and panel regressions systematically understate the average effect of beliefs on behavior in many empirical settings, since these groups with small updates get very little weight.

A simple model of endogenous information acquisition can explain this dependence between belief updating and belief effects. Consider a person for whom a particular belief particularly matters for a decision-say, a homeowner whose refinancing decision depends strongly on house price expectations. They would have a strong incentive to acquire precise information. When provided with experimental information, they would update their beliefs only slightly because their priors are already precise. In contrast, individuals for whom the belief is less consequential–those with weaker causal effects–may have less precise priors and update more strongly.

In concurrent and related work, Vilfort and Zhang (Forthcoming) study the interpretation of TSLS specifications in information provision experiments. They consider a general non-parametric model and provide conditions under which TSLS can have non-negative weights. They propose that researchers use knowledge of the priors and signals in passive designs to construct specifications with non-negative weights.[4] The present paper complements this by providing an alternative to TSLS that targets the APE directly.

The remainder of this paper is organized as follows. Section 1 presents the conceptual framework and introduces the three classes of experiments. Section 2 presents workhorse specifications that are weighted averages of individual effects with weights proportional to the first stage variation in beliefs. Section 3 presents the LLS estimator and establishes conditions that identify the APE. Section 4 presents six empirical applications demonstrating that TSLS attenuation is widespread. Section 5 concludes.

1.   Conceptual Framework

This paper focuses on experiments that study how beliefs affect behavior, rather than only how new information affects beliefs. I analyze three leading experimental designs: panel

---

person experiments with both active and passive control groups.

[4]Since they seek minimal assumptions that ensure that TSLS specifications have non-negative weights, the APE is not generally identified in the specifications they consider, except in the special cases when first-stage heterogeneity is uncorrelated with treatment effects. The stronger Bayesian updating assumption that I maintain in this paper is valuable because it enables the alternative LLS estimation strategy that directly targets the APE.

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.[5]

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 $Z \to S \to X \to Y$ structure allows us to study how exogenous variation in information provision translates into belief changes and ultimately behavioral responses.

1.1. Outcomes

The outcome equation is a linear model with heterogeneous coefficients on beliefs:

$$Y_i = \tau_i X_i + U_i \tag{1}$$

This is the canonical random coefficients model where Y_i is the outcome or behavior of interest, X_i is the belief and U_i is the structural error term. The heterogeneous coefficient τ_i is the heterogeneous belief effect. We will assume that the beliefs X_i are endogenous ($\mathbb{E}[X_i U_i] \neq 0$); Y_i can be arbitrarily affected by unobservables U_i. This model is structural in the sense that it generates potential outcomes $Y_i(x) = \tau_i x + U_i$.

The linearity assumption is a parsimonious way to introduce heterogeneity across agents but is not substantively important and plays no role in the identification arguments.[6]

1.2. The Average Partial Effect

A natural parameter of interest is the average partial effect (APE) of X_i on Y_i, denoted as $\mathbb{E}[\tau_i]$. On average, a one unit increase in beliefs causally shifts the outcome by the APE. In general, some researchers may prefer a LATE-like weighted average, and others may simply attempt to provide a proof of concept that any causal effect exists.[7] In practice, the empirical applications demonstrate that TSLS and panel estimators are often attenuated relative to the APE. In several cases the APE is more than twice as large. For researchers

---

[5]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.

[6]Section 3.2 shows how to interpret the main results with general potential outcomes $Y_i(x)$ generating $Y_i \equiv Y_i(X_i)$. Proofs in Appendix A.3 obtain under this general form.

[7]In a recent review, Mogstad and Torgovitsky (2024) note that these convex combination parameters are informative only about the sign of the individual effects, and only when every individual effect has the same sign. The APE is informative about the magnitude of the effect of X_i on Y_i, in addition to the sign.

seeking to summarize the strength of belief effects in a single number, a LATE-like weighted average may provide a substantial understatement in many empirically relevant settings.

Regardless of one's preferred target parameter, the difference between the APE and standard specifications is informative about the dependence between belief updating and belief effects. Understanding this relationship reveals underlying mechanisms of belief formation and can guide both the interpretation of existing results and the design of "nudges" or other policy interventions.

1.3. Belief Updating

Potential beliefs are a linear function of the prior X_i^0 and an experimental signal s:

$$X_i(s) = \alpha_i \left(s - X_i^0 \right) + X_i^0 \tag{2}$$

The heterogeneous coefficient on the signal is given by the heterogeneous learning rate α_i. This weighted average expression is a workhorse in the applied literature and seems to reflect belief updating well, at least in the context of information provision experiments.[8] Appendix A.1 shows how this linear updating rule can be microfounded in a normal-normal Bayesian updating.

1.4. Experimental Designs

This paper considers three broad classes of information provision experiments:

> **Panel:** The panel design uses contrasts within-individual 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.

> **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" 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 recieve 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.

---

[8]See for example (Balla-Elliott et al., 2022; Cavallo et al., 2017; Cullen et al., 2023; Cullen and Perez-Truglia, 2022; Fuster et al., 2022; Giaccobasso et al., 2022).

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. The treatment indicator $T_i \equiv \mathbb{1}\{Z_i = A\}$ indicates assignment to arm A. Finally, $S_i(z)$ is the signal that is shown to individual i in treatment arm z.[9]

Treatment is assigned randomly in the sense that Z_i is independent of the potential outcomes: the structural residual U_i, the prior X_i^0, the potential signals $S_i(\cdot)$, the learning rate α_i, and the belief effect τ_i. 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 with individuals in a way that is not assumed to be independent of the structural unobservable U_i.[10] 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

$$X_i(z) \equiv X_i(S_i(z)) = \alpha_i \left(S_i(z) - X_i^0 \right) + X_i^0 \tag{3}$$

Notice that in passive designs $X_i(B) = X_i^0$ since we set $S_i(B) \equiv X_i^0$ when treatment arm B receives no information. It is worth emphasizing that this is merely a notational device to ensure that the potential signals $S_i(z)$ are always defined. We will use the structural equations (1) and (3) to study common empirical specifications.

1.5. Adapting Notation for Panel Experiments

In panel experiments, the key identifying variation is within individuals. In order to highlight this in the notation, let time t have two periods, denoting pre ($t = 0$) and post ($t = 1$) information provision. Let

$$Y_{it} = \tau_i X_{it} + \gamma_t + U_i \tag{4}$$

---

[9]In the panel design, the researcher may randomly assign Z_i in the same way, or may chose 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.

[10]For example, consider when $S_i(A)$ is a high estimate of home value and $S_i(B)$ is a low estimate of the home value as in Bottan and Perez-Truglia (2022a). The researcher will randomly assign an individual to see a high or low signal, but the potential signal values are not random and indeed often depend directly on observable features (Balla-Elliott et al., 2022; Roth et al., 2022). The realized signal is only randomly assigned conditional on the potential signal values.

This is the standard panel model as used in the literature (e.g. Armona et al., 2019; Wiswall and Zafar, 2015), modified to allow for heterogeneity in the treatment effect τ_i.

Pre and post treatment beliefs are also elicited in the between-person designs. To link the within- and between-person designs, the pre-treatment belief is the prior, and the post-treatment belief is the posterior ($X_{i0} \equiv X_i^0$; $X_{i1} \equiv X_i$). Unlike the other two cases, we make no assumptions about how the treatment shifts beliefs.

Within and Between Person Designs Use Different Kinds of Variation. The key structure in the panel design is the assumption that different changes in outcomes are due only to different changes in beliefs.[11] There are no assumptions on the content of the belief update or restrictions on how beliefs would have changed under alternative signals. In contrast, the key structure in the between-person designs with active or passive controls is the belief updating model in equation (3).

2. Standard Specifications

This section shows that standard estimators in information provision experiments yield weighted averages of individual effects, with weights proportional to belief updating. This weighting scheme can lead to systematic attenuation when belief updating is negatively correlated with treatment effects.

I focus on representative simple specifications, though of course empirical researchers employ a variety of specifications..[12] These estimands are weighted averages of individual average effects τ_i:

$$\beta^{design} \equiv \mathbb{E}\left[\tau_i \times \omega_i(design)\right] \tag{5}$$

The precise form of these weights varies, but in all three cases, existing specifications weight individual effects τ_i in proportion to the first-stage belief updating. In all specifications, these weights integrate to one. Appendix A.2 contains derivations for all expressions in this section and Appendix E provides a more general discussion of TSLS in information experiments.

---

[11]The time trend γ_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 γ_t can be interacted with observables W_i to allow for these time trends to vary across observables, including the prior belief. Models without a time trend have the testable implication that $\mathbb{E}\left[\Delta Y_i | \Delta X_i = 0\right] = 0$; i.e. outcomes do not change for people who do not change their beliefs.

[12]Vilfort and Zhang (Forthcoming) provide additional examples of TSLS specifications in active and passive control experiments.

2.1. A Representative Panel Specification

Armona et al. (2019) use a regression in first-differences (equivalent to a panel regression with individual and time fixed effects). Let ΔX_i denote the difference between the post- and pre-treatment beliefs, $X_{i1} - X_{i0}$. The regression specification is simply

$$\beta^{Panel} \equiv \frac{\text{Cov}\left[\Delta Y_i, \Delta X_i\right]}{\text{Var}\left[\Delta X_i\right]} \tag{6}$$

which has implied weights

$$\omega_i(Panel) \propto \Delta X_i(\Delta X_i - \mathbb{E}[\Delta X_i]) \tag{7}$$

The regression of ΔY_i on ΔX_i and a constant generically has negative weights for ΔX_i between zero and the mean $\mathbb{E}[\Delta X_i]$.[13]

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:

$$\beta^{Active} \equiv \frac{\mathbb{E}\left[Y \mid Z = A\right] - \mathbb{E}\left[Y \mid Z = B\right]}{\mathbb{E}\left[X \mid Z = A\right] - \mathbb{E}\left[X \mid Z = B\right]} \tag{8}$$

$$\omega_i(Active) \propto X_i(A) - X_i(B) \tag{9}$$

which under Bayesian Learning simplifies further to

$$\omega_i(Active) \propto \alpha_i(S_i(A) - S_i(B)) \tag{10}$$

These weights are non-negative under Bayesian updating 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.

2.3. A Representative Passive Control Specification

Cullen and Perez-Truglia (2022) use an IV specification where the instrument is an indictor for assignment to the information treatment interacted with the initial gap in beliefs.[14]

$$T_i^{ex} \equiv T_i(S_i(A) - X_i^0) \tag{11}$$

---

[13]Suppose $\Delta X_i \geq 0$. Then a "sign flip", when the estimate is negative and every individual has a positive effect, occurs when people with small ΔX_i have very large τ_i. Then people with large ΔX_i (and very small positive effects) can have smaller ΔY_i than people with small ΔX_i (and very large positive effects), leading to a negative slope estimate.

[14]Vilfort and Zhang (Forthcoming) point out that similar specifications that also include the treatment indictor as an excluded instrument have negative weights.

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.[15]

$$\widetilde{T}_i^{ex} \equiv (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \tag{12}$$

The TSLS coefficient is then given by

$$\beta^{Passive} \equiv \frac{\text{Cov}\left[\widetilde{T}_i^{ex}, Y_i\right]}{\text{Cov}\left[\widetilde{T}_i^{ex}, X_i\right]} \tag{13}$$

$$\omega_i(Passive) \propto (X_i(A) - X_i(B))(S_i(A) - X_i^0) \tag{14}$$

which under Bayesian Learning simplifies further to

$$\omega_i(Passive) \propto \alpha_i(S_i(A) - X_i^0)^2 \tag{15}$$

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

2.4. 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 are "weakly causal".

Why might TSLS be attenuated?. Since standard specifications weight individual effects by the strength of belief updating, they are attenuated when belief updates are negatively correlated with belief effects. This could happen if people rationally form precise priors when these beliefs strongly affect decisions. 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. The applications that in Section 4 demonstrate that this pattern holds empirically in a range of contexts. Appendix D formalizes this mechanism in a simple model where individuals trade off the cost of acquiring information against the risk of making decisions with imprecise beliefs.

Alternative Approaches to Interpretation and Estimation. Vilfort and Zhang (Forthcoming) show that TSLS specifications in passive control experiments can still have non-negative

---

[15]To see this, notice that random assignment implies that $\mathbb{E}\left[T_i^{ex} \mid S_i(A) - X_i^0\right] = \mathbb{E}[T_i]\left(S_i(A) - X_i^0\right) = \mathbb{L}\left[T_i^{ex} \mid S_i(A) - X_i^0\right]$. By FWL $\widetilde{T}_i^{ex} \equiv T_i^{ex} - \mathbb{L}\left[T_i^{ex} \mid S_i(A) - X_i^0\right] = (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0)$.

weights under weaker assumptions than Bayesian updating. This section highlights that, even under a stronger Bayesian updating assumption, these TSLS specifications do not recover an unweighted average across people.[16] Section 3 shows that this Bayesian updating assumption can instead be used to construct an alternative estimator that does recover a simple average across people.

3. The Local Least Squares Estimator

This section presents a local least squares (LLS) estimator that consistently recovers the average partial effect (APE). Intuitively, the goal is to construct a vector of controls such that–conditional on these controls–there are only two possible beliefs and the only remaining variation in beliefs comes from treatment assignment. Appendix A.3 provides proofs for the results in this section.

3.1. Identification of the APE

The key intuition behind LLS is to construct "local" regressions that use only exogenous (i.e. experimental) variation in beliefs. Graham and Powell (2012) and Masten and Torgovitsky (2016) show how to construct these "local" regressions in panel and IV settings more generally. In this setting, Bayesian updating means that people who have the same learning rate, the same prior, and the same potential signals have the same potential beliefs; the only variation in their actual beliefs comes from the random assignment to the actual signal.[17]

   I present identification results for three experimental designs in increasing order of the strength of the identifying assumptions.

---

[16]Bayesian updating is a stronger assumption in the sense that it implies, but it not necessarily implied by, the kind of "updating toward the signal" behavior that Vilfort and Zhang (Forthcoming) directly assume. In this sense, the APE is not generally identified by their TSLS approach, even under the stronger Bayesian updating and linear outcome assumptions.

[17]Notice that the general practice of linearly controlling for the prior belief and the signals does not eliminate all of the endogeneity in the potential beliefs, since people with the same prior and signals can still have different beliefs if they have different learning rates. Intuitively, the bias in the existing specifications can be seen as (partially) coming from correlation between the heterogeneity in the learning rate and the treatment effects. Heterogeneity in the first stage thus generates "endogenous" (i.e. non-experimental) variation in posterior beliefs.

3.1.1. Local Regressions in Panel Experiments

Panel designs require no assumptions on belief updating. Under the panel outcome equation (4), for any belief change $x \neq 0$:

$$\mathbb{E}\left[\tau_i \mid \Delta X_i = x\right] \equiv \frac{\text{Cov}\left[\Delta Y_i, \Delta X_i \mid \Delta X_i \in \{0, x\}\right]}{\text{Var}\left[\Delta X_i \mid \Delta X_i \in \{0, x\}\right]} \tag{16}$$

The right hand side is a feasible local regression of ΔY_i on ΔX_i using only observations with $\Delta X_i = x$ or $\Delta X_i = 0$. This requires that some individuals have (close to) zero change in beliefs.[18] Iterating over x, we obtain the average partial effect $\mathbb{E}\left[\mathbb{E}\left[\tau_i \mid \Delta X_i\right]\right] = \mathbb{E}\left[\tau_i\right]$.

The panel approach works with any information treatment–including qualitative treatments, or bundles of multiple signals–because identification relies only on the assumption that different changes in outcomes are due only to different changes in beliefs.

3.1.2. Local Regressions in Active Control Experiments

Active designs rely on the Bayesian updating assumption (3) and identify learning rates directly from observed belief updates: $\alpha_i = \left(X_i - X_i^0\right) / \left(S_i - X_i^0\right)$. Conditioning on the learning rate α_i, the prior X_i^0, and signal values $\left[S_i(A)\ S_i(B)\right]$ ensures that the only variation in beliefs comes from random assignment. Under the linear outcome equation (1) and Bayesian updating (3):

$$\mathbb{E}\left[\tau_i \mid C_i = c\right] \equiv \frac{\text{Cov}\left[Y_i, X_i \mid C_i = c\right]}{\text{Var}\left[X_i \mid C_i = c\right]} \tag{17}$$

where $C_i \equiv \left[\alpha_i\ X_i^0\ S_i(A)\ S_i(B)\right]$ is the control vector. The regression is feasible when $\left(S_i - X_i^0\right) \neq 0$ and $\text{Var}\left[X_i \mid C_i = c\right] > 0$. This excludes cases with no learning ($\alpha_i = 0$) or identical signals ($S_i(A) = S_i(B)$). Iterating over c yields $\mathbb{E}\left[\mathbb{E}[\tau_i \mid C_i]\right] = \mathbb{E}\left[\tau_i\right]$.

3.1.3. Local Regressions in Passive Control Experiments

Passive designs also rely on the Bayesian updating assumption (3), 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:

---

[18]This is an easily verifiable condition in the data. For example, it is straightforwardly satisfied if $P\left[\Delta X_i = 0\right] > 0$; i.e. if there are people who don't change their beliefs. Technically, if δX_i is continuous is it possible to have $P\left[\Delta X_i = 0\right] = 0$, even while δX_i has positive mass in any neighborhood around zero, which will be sufficient to estimate $\mathbb{E}\Delta Y_i \mid \Delta X_i = 0$. See Graham and Powell (2012) for a detailed discussion of technical considerations associated with continuous ΔX_i without a point mass at zero.

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

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

In either case, under the linear outcome equation (1) and Bayesian updating (3):

$$\mathbb{E}\left[\tau_i \mid C_i = c\right] \equiv \frac{\operatorname{Cov}\left[Y_i, X_i \mid C_i = c\right]}{\operatorname{Var}\left[X_i \mid C_i = c\right]} \tag{18}$$

Appendix A.3.3 formally states the assumptions in both of these cases.

3.2. Linearity of the Outcome Equation is Not Necessary

The random coefficients model in (1) is a parsimonious way to model treatment effect heterogeneity but is not essential. With arbitrary potential outcomes

$$Y_{it}(x) = G_i(x) + \gamma_t \tag{19}$$

the local regressions recover average slopes of individual response functions $G_i(\cdot)$ between the individual potential beliefs $X_i(A)$ and $X_i(B)$.[19] Random assignment with respect to potential outcomes now means that $G_i(\cdot) \perp\!\!\!\perp Z_i$. The primary difference in interpretation is that LLS estimates are now particular to the observed belief distribution; they remain proper unweighted averages across people. Identification comes from conditioning on heterogeneity in potential beliefs and not assuming linear outcomes. The proofs in Appendix A.3 use this general outcome equation (19).

3.3. Implementation and Practical Considerations

Conditioning on high-dimensional control vectors is often impractical in experimental samples. The linear structure of Bayesian updating makes it 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.

---

[19]Note that in the cross section with a single t, γ_t can be absorbed into the $G_i(x)$ without loss of generality.

Roth et al. (2022)), it is also necessary to reweight by the inverse of the exposure. This weighted local regression recovers $\mathbb{E}\left[\tau_i \mid \alpha_i\right]$. Appendix A.4 shows that this modified local regression is sufficient and Appendix B provides general implementation guidance for each design.

3.4. Discussion

The LLS estimator provides a practical solution to the attenuation problem identified in Section 2. 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. This suggests that researchers interested in implementing an LLS estimator may find active designs more attractive since they reveal more information about belief updating.[20]

4. Empirical Applications

This section demonstrates that attenuation due to dependence between belief updating and belief effect is empirically relevant. I analyze six recent studies from leading economics journals, covering panel, active control, and passive control experiments. For each study, I compare standard specifications (panel or TSLS) to LLS estimates of the average partial effect (APE).[21] I omit additional demographic controls from all estimates for simplicity.[22] See Appendix B for estimation details.

    Table 1 contrasts LLS estimate with estimates recovered by the standard specification in each study. In five of the six studies, existing TSLS and panel estimators substantially understate the strength of the causal effects. Figure 1 plots an estimate of the CAPE curve for

---

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

[21]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.

[22]Settele (2022) also includes probability weights in the original paper that I ignore for simplicity.

each study: $\mathbb{E}\left[\tau_i \mid \Delta X_i\right]$ in panel experiments (Panel A) and $\mathbb{E}\left[\tau_i \mid \text{rank}(\alpha_i)\right]$ in active and passive control experiments (Panels B and C). These curves directly reveal the dependence between belief updating and belief effects. Consistent with the information acquisition mechanism, attenuation occurs because people with the strongest causal effects of beliefs tend to have smaller belief updates.

4.1. Results from Panel Experiments

Wiswall and Zafar (2015) study how beliefs about field-specific earnings affect college students' major choices. The panel estimate of 0.32 (s.e. 0.086) is substantially smaller than the LLS estimate of 0.87 (s.e. 0.33), with the LLS estimate being 170% larger.

Armona et al. (2019) study how beliefs about home prices affect investment decisions. The panel estimate of 1.15 (s.e. 0.234) is smaller than the LLS estimate of 1.8 (s.e. 0.381), with the LLS estimate being over 50% larger.

4.2. Results from Active Control Experiments

Settele (2022) studies how beliefs about the gender wage gap affect support for gender equality policies. The TSLS estimate of 0.096 (s.e. 0.033) is substantially smaller than the LLS estimate of 0.16 (s.e. 0.042), with the LLS estimate being 66% larger.

Roth et al. (2022) examine the relationship between recession expectations and subjective personal unemployment risk. Their TSLS estimate of 0.755 color (s.e. 0.433) is somewhat smaller than the LLS estimate of 0.882 (s.e. 0.379), with the LLS estimate being 17% larger.

4.3. Results from Passive Control Experiments

Kumar et al. (2023) beliefs about GDP growth affect employment decisions. The TSLS estimate 0.466 (s.e. 0.19) is smaller than the LLS estimate 1.787 (s.e. 0.409), with the LLS estimate being 284% larger.

Cantoni et al. (2019) study how beliefs about others' protest participation affect one's own willingness to participate. The TSLS estimate (0.68, s.e. 0.253) and the LLS estimate (0.18, s.e. 0.133) are both quite noisy, making it difficult to draw strong conclusions about the direction or magnitude of any difference. The difference between the TSLS and LSS estimates is suggestive evidence that people with larger belief effects had *larger* belief updates. However, the CAPE curve in Panel C.ii of Figure 1 reveals only modest variation

across learning rate ranks, with quite wide confidence intervals.[23] It is perhaps unsurprising that the noisiest results are those that require the strongest assumptions: it may be the case that even with rich observables it is difficult to correctly model the heterogeneity in belief updating.

4.4. Discussion

The CAPE curves in Figure 1 reveal a consistent pattern across several experimental designs: in five of six applications, individuals who update their beliefs least have the strongest causal effects. This provides direct empirical support for models of endogenous information acquisition where people with decision-relevant beliefs invest in forming precise priors.

The difference between LLS and standard estimates is often substantial: 170% larger for Wiswall and Zafar (2015), 66% larger for Settele (2022), and 284% larger for Kumar et al. (2023). The magnitude of attenuation across a wide range of contexts–from educational choices to macroeconomic expectations–suggests this may be a pervasive feature of information provision experiments.

5. 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, average effect of beliefs on behavior are larger–in two cases more than double–estimates from standard specifications.

The local least squares estimator proposed in this paper offers a practical solution. It consistently recovers the average partial effect under plausible assumptions about belief updating. It can also be used to recover the entire CAPE curve, offering a richer picture of the relationship between belief effects and belief updating. Understanding this relationship reveals underlying mechanisms of belief formation and can guide both the interpretation of existing results and the design of future information provision experiments.
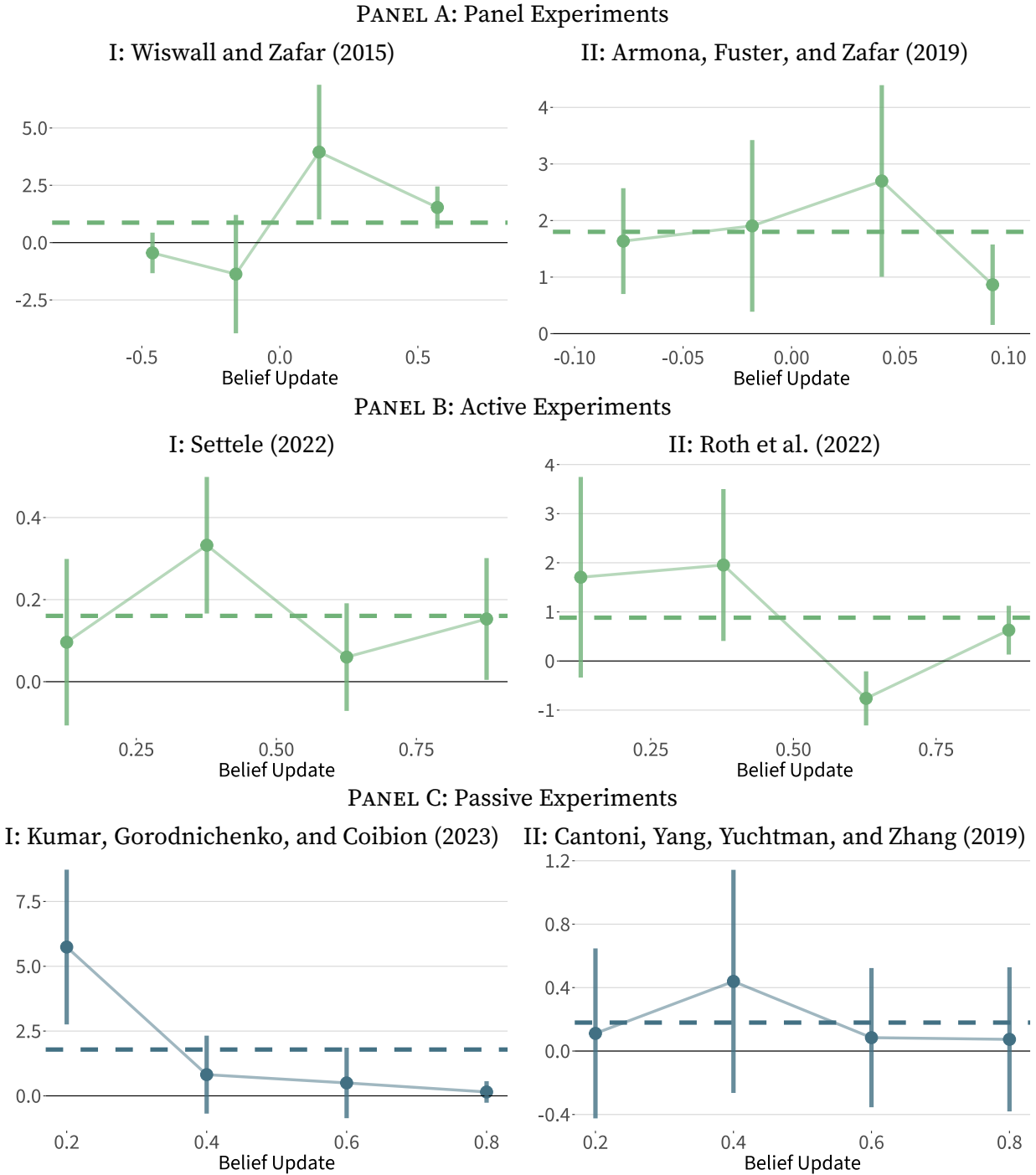
---

[23]Concerns of attenuation are only one reason among many to consider using the LSS estimator. The estimator consistently recovers the APE regardless of the sign of dependence between belief updating and treatment effects. The pattern of attenuation observed in five of six applications is an empirical finding, not a mechanical feature of the estimator.

TABLE 1. LLS and Standard Specifications in Six Studies

	PANEL A: Panel Experiments
	Wiswall and Zafar (2015)
LLS	0.87 (0.33)
Paper (FE)	0.32 (0.086)
Bandwidth	0.05
	Armona, Fuster, and Zafar (2019)
LLS	1.8 (0.381)
Paper (FE)	1.147 (0.234)
Bandwidth	0.025
	PANEL B: Active Experiments
	Settele (2022)
LLS	0.16 (0.042)
Paper (TSLS)	0.096 (0.033)
Bandwidth	0.01
	Roth, Settele, and Wohlfart (2022)
LLS	0.882 (0.366)
Paper (TSLS)	0.755 (0.435)
Bandwidth	0.075
	PANEL C: Passive Experiments
	Kumar, Gorodnichenko, and Coibion (2023)
LLS	1.787 (0.469)
Paper (TSLS)	0.466 (0.19)
Bandwidth	0.025
	Cantoni, Yang, Yuchtman, and Zhang (2019)
LLS	0.18 (0.133)
Paper (TSLS)	0.68 (0.253)
Bandwidth	0.1

Notes: This table presents estimates of the (unweighted) average partial effect of beliefs (APE) on outcomes to common first-difference (FD) or two-stage least squares (TSLS) weighted averages across all six replication studies. 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. Bootstrap standard errors are reported in parentheses. Appendix B discusses implementation details and reports results for alternative choices of bandwidth.

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

PANEL A: Panel Experiments

I: Wiswall and Zafar (2015)

II: Armona, Fuster, and Zafar (2019)

PANEL B: Active Experiments

I: Settele (2022)

II: Roth et al. (2022)

PANEL C: Passive Experiments

I: Kumar, Gorodnichenko, and Coibion (2023)

II: Cantoni, Yang, Yuchtman, and Zhang (2019)

Notes: This figure displays dependence between belief updating and belief effects in six studies across threee classes of experiments. Panel A reports estimates conditional on the individual belief update: $\mathbb{E}\left[\tau_i \mid \Delta X_i\right]$. Panels B and C report estimates conditional on the rank of the individual learning rate: $\mathbb{E}\left[\tau_i \mid \operatorname{rank}\left(\alpha_i\right)\right]$. Confidence intervals displayed are twice the bootstrap standard errors. Bootstrap standard errors are the standard deviation of the bootstrap distribution of the CAPE estimate in a particular bin. In five of the six studies, the TSLS or panel estimates are attenuated relative to the LLS estimate of the unweighted average partial effect (APE). Attenuation arises when effects are strongest for people with the smallest belief updates, who recieve less weight in the standard specifications.

References

Akesson, Jesper, Robert Hahn, Robert Metcalfe, and Itzhak Rasooly (2022). "Race and Redistribution in the United States: An Experimental Analysis", w30426. DOI: 10.3386/w30426 (p. A.1).

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.23).

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.35).

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. 1, 6, 7, 13, 15, 16, i, A.14, A.16, A.22, A.23).

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. 4, 5, 11, A.1, A.8).

Bhuller, Manudeep and Henrik Sigstad (2024). *2SLS with Multiple Treatments*. DOI: 10.48550/arXiv.2205.07836 (p. A.27).

Bottan, Nicolas L. and Ricardo Perez-Truglia (2022a). "Betting on the House: Subjective Expectations and Market Choices", p. 53. DOI: 10.3386/w27412 (p. 5).

— (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).

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. 1, 11, 13, 15, 16, ii, A.8, A.14, A.20, A.28).

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. 4, A.1).

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. 4).

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. 4, 7, A.1).

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. 4, A.29, A.34).

Giaccobasso, 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. 4, A.35).

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. 9, 10, A.5, A.12).

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).

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. 12).

Hansen, Bruce E. (2022). *Econometrics*. Princeton: Princeton University Press (p. A.13, A.14).

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. A.1).

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 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).

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer (2023). "Worker Beliefs About Outside Options". *Working Paper* (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. A.2).

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. 1, 13–16, ii, A.8, A.13, A.14, A.19, A.26, A.27).

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. 12).

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. 9).

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. 3).

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. A.1).

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. 5, 12, 13, 15, 16, ii, A.2, A.11, A.14, A.18, A.25).

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. 1, A.2).

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. 1, 7, 12–16, i, A.13, A.14, A.17, A.23, A.24, A.26).

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 (Forthcoming). "Interpreting TSLS Estimators in Information Provision Experiments". *American Economic Review: Insights*. DOI: `10.48550/arXiv.2309.04793` (p. 2, 6–9, A.24, A.26, A.27, A.38).

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, 6, 13–16, i, A.14, A.15, A.21, A.22).

Contents of Online Appendix

A. Proofs and Derivations

A.1. Belief Potential Outcomes are Motivated by Bayesian Learning

The literature often motivates the weighted-average expression in (3) in a Bayesian learning model with normally distributed beliefs (Balla-Elliott et al., 2022; Cullen and Perez-Truglia, 2022). This section shows how these potential beliefs are generated by a Bayesian learning model and relate the key coefficient α_i to model primitives.

Consider a group of individuals with uncertain prior beliefs. The subjective probability that the variable X_i takes the value x is given by the density of the normal distribution $\mathcal{N}\left(X_i^0, \sigma_{iX}^2\right)$. We thus interpret X_i^0 as the mean of the prior distribution. As shorthand, we will call X_i^0 the prior belief of an individual i.

People then observe a signal S_i, which we model as a draw from a distribution $\mathcal{N}(S_i^*, \sigma_{iS}^2)$. The variances of these distributions reflect the subjective (inverse) precision of the prior and the signal. These variances are important only in their relative size. People for whom $\sigma_{iS}^2/\sigma_{iX}^2$ is large think their prior is more precise than the signal, whereas those for whom $\sigma_{iS}^2/\sigma_{iX}^2$ is small think that the signal is more precise than their prior.

The posterior is then a distribution

$$\mathcal{N}\left((1 - \alpha_i)\, X_i^0 + \alpha_i S_i, \frac{\sigma_{iS}^2 \sigma_{iX}^2}{\sigma_{iS}^2 + \sigma_{iX}^2}\right) \tag{20}$$

$$\text{where } \alpha_i \equiv \frac{\sigma_{iX}^2}{\sigma_{iS}^2 + \sigma_{iX}^2} \tag{21}$$

As with the prior, we will call the mean of this distribution the posterior X. Note that the mean of the posterior distribution is a weighted average of the prior X_i^0 and the signal S_i, where the weights are given by their relative precision.[24] We can also note that

$$\frac{\sigma_{iS}^2 \sigma_{iX}^2}{\sigma_{iS}^2 + \sigma_{iX}^2} < \sigma_{iX}^2$$

intuitively, the posterior distribution is more precise than the prior distribution.[25] We can then relate the prior X_i^0, the signal S_i and the posterior X_i through the equation

$$X_i = (1 - \alpha_i)X_i^0 + \alpha_i S_i \tag{22}$$

which generates the potential outcomes for beliefs in (3). There is some direct empirical

---

[24]A full discussion of this derivation can be found in introductory textbook treatments of Bayesian statistics like Robert (2007) or Hoff (2009).

[25]There is experimental evidence that people randomized to the group receiving a signal report greater confidence in their posterior beliefs (Akesson et al., 2022; Cavallo et al., 2017).

support for this Bayesian foundation of the linear updating model. For example, Roth et al. (2022) find that all the belief updating in their study is driven by people who report being "very unsure", "unsure" or "somewhat unsure" and that people who are "sure" or "very sure" do not update their beliefs. Similarly Roth and Wohlfart (2020) find that people who are less confident in their prior beliefs update roughly twice as much as people who are more confident. Kerwin and Pandey (2023) also find in a more general model that people with less precise priors update more in response to an information treatment.

A.2. Derivations of Weights

This section provides derivations for the weights reported in Section 2.

A.2.1. Weights in the Panel Specification

ASSUMPTION 1. *Panel Assumptions.*
  a. *Panel Outcomes : The panel outcome equation (4) holds.*

$$Y_{it} = \tau_i X_{it} + \gamma_t + U_i \tag{4}$$

  b. *Relevance: There is variation in beliefs over time* $\mathrm{Var}\left[\Delta X_i\right] > 0$.
  c. *Existence: The relevant moments exist and are finite.*

The parsimonious specification in the panel data model in (6) is given by:

$$\beta^{Panel} = \frac{\mathrm{Cov}\left[\Delta Y_i, \Delta X_i\right]}{\mathrm{Var}\left[\Delta X_i\right]} \tag{23}$$

Substitute the outcome equation (4):

$$= \frac{\mathrm{Cov}\left[\tau_i \Delta X_i + \gamma_1 - \gamma_0, \Delta X_i\right]}{\mathrm{Var}\left[\Delta X_i\right]} \tag{24}$$

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

$$= \frac{\mathbb{E}\left[\tau_i \Delta X_i \left(\Delta X_i - \mathbb{E}\left[\Delta X_i\right]\right)\right]}{\mathbb{E}\left(\Delta X_i \left(\Delta X_i - \mathbb{E}\left[\Delta X_i\right]\right)\right)} \tag{25}$$

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

$$= \mathbb{E}\left[\tau_i \cdot \frac{\Delta X_i \left(\Delta X_i - \mathbb{E}\left[\Delta X_i\right]\right)}{\mathbb{E}\left(\Delta X_i \left(\Delta X_i - \mathbb{E}\left[\Delta X_i\right]\right)\right)}\right] \tag{26}$$

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

A.2.2. Weights in the Active Control Specification

ASSUMPTION 2. *Active Control Assumptions.*

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

$$Y_i = \tau_i X_i + U_i \tag{1}$$

b. *Bayesian updating: The belief potential outcomes in equation (3) hold.*

$$X_i(z) = \alpha_i \left(S_i(z) - X_i^0 \right) + X_i^0 \tag{3}$$

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:

$$\beta^{\text{TSLS}} = \frac{\mathbb{E}[Y_i \mid Z_i = A] - \mathbb{E}[Y_i \mid Z_i = B]}{\mathbb{E}[X_i \mid Z_i = A] - \mathbb{E}[X_i \mid Z_i = B]} \tag{27}$$

From the outcome equation (1) and random assignment:

$$= \frac{\mathbb{E}[\tau_i X_i(A) + U_i] - \mathbb{E}[\tau_i X_i(B) + U_i]}{\mathbb{E}[X_i(A)] - \mathbb{E}[X_i(B)]} \tag{28}$$

$$= \frac{\mathbb{E}[\tau_i(X_i(A) - X_i(B))]}{\mathbb{E}[X_i(A) - X_i(B)]} \tag{29}$$

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

$$= \mathbb{E}\left[\tau_i \cdot \frac{X_i(A) - X_i(B)}{\mathbb{E}[X_i(A) - X_i(B)]} \right] \tag{30}$$

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

A.2.3. Weights in the Passive Control Specification

ASSUMPTION 3. *Passive Control Assumptions.*

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

$$Y_i = \tau_i X_i + U_i \tag{1}$$

b. *Bayesian updating: The belief potential outcomes in equation (3) hold.*

$$X_i(z) = \alpha_i \left(S_i(z) - X_i^0 \right) + X_i^0 \tag{3}$$

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:

$$T_i^{ex} \equiv T_i(S_i(A) - X_i^0) \tag{13}$$

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 $\widetilde{T}_i^{ex}$. To construct this residual, regress T_i^{ex} on $(S_i(A) - X_i^0)$ and a constant:

$$\theta = \frac{\text{Cov}(T_i^{ex}, S_i(A) - X_i^0)}{\text{Var}(S_i(A) - X_i^0)} \tag{31}$$

$$= \frac{\mathbb{E}[T_i(S_i(A) - X_i^0)^2] - \mathbb{E}[T_i]\mathbb{E}[(S_i(A) - X_i^0)^2]}{\text{Var}(S_i(A) - X_i^0)} \tag{32}$$

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

$$\theta = \frac{\mathbb{E}[T_i]\text{Var}(S_i(A) - X_i^0)}{\text{Var}(S_i(A) - X_i^0)} = \mathbb{E}[T_i] \tag{33}$$

The recentered instrument is then the residual from this regression:

$$\widetilde{T}_i^{ex} = T_i^{ex} - \theta(S_i(A) - X_i^0) \tag{34}$$

$$= (T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \tag{35}$$

Since $\mathbb{E}\left[\widetilde{T}_i^{ex}\right] = 0$, and $\mathbb{E}\left[\widetilde{T}_i^{ex}U_i\right] = 0$ from random assignment, the TSLS coefficient is simply:

$$\beta^{\text{Passive}} = \frac{\mathbb{E}[\widetilde{T}_i^{ex}\tau_i X_i]}{\mathbb{E}[\widetilde{T}_i^{ex}X_i]} \tag{36}$$

The denominator is

$$\mathbb{E}[\widetilde{T}_i^{ex}X_i] = \mathbb{E}[(T_i - \mathbb{E}[T_i])(S_i(A) - X_i^0) \cdot X_i] \tag{37}$$

Plugging in the potential beliefs for X_i and using $\mathbb{E}[T_i] = p$:

$$= p(1 - p)\mathbb{E}[(S_i(A) - X_i^0)(X_i(A) - X_i^0)] \tag{38}$$

Using the definition of $X_i(A)$ from (3) to simplify further yields:

$$= p(1 - p)\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2] \tag{39}$$

Similarly, for the numerator:

$$\mathbb{E}[\widetilde{T}_i^{ex}\tau_i X_i] = p(1 - p)\mathbb{E}[\tau_i\alpha_i(S_i(A) - X_i^0)^2] \tag{40}$$

Thus, the TSLS coefficient is:

$$\beta^{\text{Passive}} = \frac{p(1 - p)\mathbb{E}[\tau_i\alpha_i(S_i(A) - X_i^0)^2]}{p(1 - p)\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2]} \tag{41}$$

$$= \mathbb{E}\left[\tau_i \cdot \frac{\alpha_i(S_i(A) - X_i^0)^2}{\mathbb{E}[\alpha_i(S_i(A) - X_i^0)^2]}\right] \tag{42}$$

$$\tag{43}$$

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.

A.3. 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 β equals:*

$$\beta = \frac{\mathbb{E}[Y \mid X = x_2] - \mathbb{E}[Y \mid X = x_1]}{x_2 - x_1} \tag{44}$$

PROOF. The regression coefficient is defined as:

$$\beta = \frac{\text{Cov}(Y, X)}{\text{Var}(X)} \tag{45}$$

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

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \tag{46}$$

$$= p(1 - p)(x_2 - x_1)^2 \tag{47}$$

For the covariance:

$$\text{Cov}(Y, X) = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])] \tag{48}$$

$$= p(1 - p)(x_2 - x_1)(\mathbb{E}[Y \mid X = x_2] - \mathbb{E}[Y \mid X = x_1]) \tag{49}$$

Therefore:

$$\beta = \frac{\text{Cov}(Y, X)}{\text{Var}(X)} = \frac{\mathbb{E}[Y \mid X = x_2] - \mathbb{E}[Y \mid X = x_1]}{x_2 - x_1} \tag{50}$$

$\square$

A.3.1. 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 ΔX_i has positive density in a neighborhood of zero (as in Graham and Powell, 2012).*
ii. *Nonlinear outcome: Relax the outcome equation to*

$$Y_{it}(x) = G_i(x) + \gamma_t \tag{19}$$

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

$$\frac{\mathbb{E}\big[G_i(X_i^0 + x) - G_i(X_i^0) \mid \Delta X_i = x\big]}{x} = \frac{\mathbb{E}\big[\Delta Y_i \mid \Delta X_i = x\big] - \mathbb{E}\big[\Delta Y_i \mid \Delta X_i = 0\big]}{x} \tag{51}$$

If $G_i(X_{it}) = \tau_i X_{it} + U_i$ as in (4), the estimand simplifies further to $\mathbb{E}[\tau_i \mid \Delta X_i = x]$.

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

$$\beta(x) = \frac{\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] - \mathbb{E}[\Delta Y_i \mid \Delta X_i = 0]}{x} \tag{52}$$

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

$$\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] = \mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) \mid \Delta X_i = x] + \Delta\gamma \tag{53}$$

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

$$\mathbb{E}[\Delta Y_i \mid \Delta X_i = 0] = \mathbb{E}[G_i(X_{i0}) - G_i(X_{i0}) + \Delta\gamma \mid \Delta X_i = 0] \tag{54}$$

$$= \Delta\gamma \tag{55}$$

Taking the difference:

$$\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] - \mathbb{E}[\Delta Y_i \mid \Delta X_i = 0] = \mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) \mid \Delta X_i = x] \tag{56}$$

Dividing by x completes the proof:

$$\frac{\mathbb{E}[\Delta Y_i \mid \Delta X_i = x] - \mathbb{E}[\Delta Y_i \mid \Delta X_i = 0]}{x} = \frac{\mathbb{E}[G_i(X_{i0} + x) - G_i(X_{i0}) \mid \Delta X_i = x]}{x} \tag{57}$$

$\square$

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 τ_i. Formally:

PROPOSITION 3 (Necessity). *If Assumption 1A.i fails, the identified sets for γ_t and each τ_i are the real line.*

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

$$B_i(a) \equiv \frac{\Delta Y_i - a}{\Delta X_i} \tag{58}$$

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 $B_{i'}$. Chose $a(B_{i'}) \equiv \frac{\Delta Y_i}{B_{i'}\Delta X_i}$ and then chose the remaining B_i as above.

Thus the identified sets for $\gamma_1 - \gamma_0$ and τ_i are the real line. Chose 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. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

The "control group" is crucial to identify γ_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.

A.3.2. Identification in Active Experiments

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

  i. *Relevance: $S_i(A) \neq S_i(B)$ and $\alpha_i > 0$.*

 ii. *Nonlinear outcome: Relax the outcome equation to the completely flexible*

$$Y_i(x) = G_i(x)$$

This is the cross-section analogue of $Y_{it}(x) = G_i(x) + \gamma_t$ (19) with $\gamma_t = 0$. This is without loss of generality since we can absorb the constant shift into G_i without changing differences $G_i(x) - G(x')$. With one time period, we can also eliminate the t subscript.

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

$$\frac{\mathbb{E}\left[G_i(x_A) - G_i(x_B) \mid C_i = c \right]}{x_A - x_B} = \frac{\mathrm{Cov}\left[Y_i, X_i \mid C_i = c \right]}{\mathrm{Var}\left[X_i \mid C_i = c \right]} \tag{59}$$

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 α_i, X_i^0, $S_i(A)$, and $S_i(B)$, the potential beliefs take the same value for all individuals with $C_i = c$.

$$X_i(A) = X_i^0 + \alpha_i(S_i(A) - X_i^0) \tag{60}$$
$$X_i(B) = X_i^0 + \alpha_i(S_i(B) - X_i^0) \tag{61}$$

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:

$$\beta(c) = \frac{\mathbb{E}[Y_i \mid X_i = x_A, C_i = c] - \mathbb{E}[Y_i \mid X_i = x_B, C_i = c]}{x_A - x_B} \tag{62}$$

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

$$\beta(c) = \frac{\mathbb{E}[Y_i \mid Z_i = A, C_i = c] - \mathbb{E}[Y_i \mid Z_i = B, C_i = c]}{x_A - x_B} \tag{63}$$

Then, since Z_i is randomly assigned, we have:

$$\mathbb{E}[Y_i \mid Z_i = A, C_i = c] - \mathbb{E}[Y_i \mid Z_i = B, C_i = c] = \mathbb{E}[G_i(x_A) - G_i(x_B) \mid C_i = c] \tag{64}$$

Dividing by $x_A - x_B$ completes the proof:

$$\frac{\text{Cov}[Y_i, X_i \mid C_i = c]}{\text{Var}[X_i \mid C_i = c]} = \frac{\mathbb{E}[G_i(x_A) - G_i(x_B) \mid C_i = c]}{x_A - x_B} \tag{65}$$

$\square$

A.3.3. 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 α_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* $\alpha_i = \frac{\sigma_{Xi}^2}{\sigma_{Xi}^2 + \sigma_S^2}$ *with* σ_S^2 *common across individuals.*
b. *The researcher knows* σ_{Xi}^2.

In normal-normal Bayesian updating, $\alpha_i = \frac{\sigma_{Xi}^2}{\sigma_{Xi}^2 + \sigma_S^2}$, where σ_{Xi}^2 is the variance of the prior belief X_i^0 and σ_S^2 is the variance of the signal S_i. The first assumption, that σ_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* θ *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,* θ *is identifed 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 modificiations:*

  i. *Relevance: $S_i(A) \neq X_i^0$ and $\alpha_i > 0$.*
 ii. *Nonlinear outcome: Relax the outcome equation to the completely flexible*

$$Y_i(x) = G_i(x)$$

This is the cross-section analogue of $Y_{it}(x) = G_i(x) + \gamma_t$ (19) with $\gamma_t = 0$. This is without loss of generality since we can absorb the constant shift into G_i without changing differences $G_i(x) - G(x')$. With one time period, we can also eliminate the t subscript.
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*

$$\frac{\mathbb{E}\left[G_i(x_A) - G_i(x_B) \mid C_i = c\right]}{x_A - x_B} \equiv \frac{\text{Cov}\left[Y_i, X_i \mid C_i = c\right]}{\text{Var}\left[X_i \mid C_i = c\right]} \tag{66}$$

PROOF. Under Assumption 6, α_i is a one-to-one function of σ_{Xi}^2. Thus conditioning on σ_{Xi}^2 or its rank is equivalent to conditioning on α_i and so conditional on $C_i \equiv \left[\text{rank}\left(\sigma_{Xi}^2\right) \; X_i^0 \; S_i(A)\right]$, $X_i(A)$ 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 α_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 $C_i \equiv \left[\alpha_i \; X_i^0 \; S_i(A)\right]$, $X_i(A)$ 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 $\mathbb{E}[\tau_i]$.

A.4. Linear Controls in a Reweighted Regression

This section shows that a reweighted linear regression that controls for α_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 α_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:*

$$\mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \,\middle|\, \alpha_i \right] \tag{67}$$

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; as 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 conditional on α_i. That is:

$$\tilde{X}_i = X_i - \mathbb{L}_{\alpha_i}[X_i \mid W_i] = X_i - \mathbb{E}[X_i \mid W_i, \alpha_i] \tag{68}$$

The second equality uses the fact that, under Bayesian updating (3), the true conditional expectation is linear in W_i conditional on α_i:

$$\mathbb{E}[X_i \mid W_i, \alpha_i] = (1 - \alpha_i)X_i^0 + \alpha_i S_i(B) + \mathbb{E}[T_i]\,\alpha_i\,(S_i(A) - S_i(B)) \tag{69}$$

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 α_i.

Writing X_i in a similar form shows that

$$X_i = (1 - \alpha_i)X_i^0 + \alpha_i S_i(B) + T_i \alpha_i\,(S_i(A) - S_i(B)) \tag{70}$$

$$\tilde{X}_i \equiv X_i - \mathbb{E}[X_i \mid W_i, \alpha_i] = \alpha_i\,(T_i - \mathbb{E}[T_i])\,(S_i(A) - S_i(B)) \tag{71}$$

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

$$\beta_\alpha = \frac{\mathbb{E}[Y_i \tilde{X}_i (S_i(A) - S_i(B))^{-2} \mid \alpha_i]}{\mathbb{E}[\tilde{X}_i^2 (S_i(A) - S_i(B))^{-2} \mid \alpha_i]} \tag{72}$$

$$= \frac{\mathbb{E}[Y_i \cdot \alpha_i (T_i - p)(S_i(A) - S_i(B))^{-1} \mid \alpha_i]}{\mathbb{E}[\alpha_i^2 (T_i - p)^2 \mid \alpha_i]} \tag{73}$$

$$= \frac{\mathbb{E}[Y_i \cdot (T_i - p)(S_i(A) - S_i(B))^{-1} \mid \alpha_i]}{\alpha_i p(1 - p)} \tag{74}$$

Now, we compute the numerator:

$$\mathbb{E}\left[Y_i \cdot \frac{(Z_i - p_z)}{(S_i(A) - S_i(B))} \mid \alpha_i\right] = \mathbb{E}\left[G_i(X_i) \cdot \frac{(T_i - p)}{(S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (75)$$

$$= p(1-p) \cdot \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{S_i(A) - S_i(B)} \mid \alpha_i\right] \quad (76)$$

Substituting this back into the expression for β_α:

$$\beta_\alpha = \frac{p(1-p)}{\alpha_i p(1-p)} \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{(S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (77)$$

$$= \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{\alpha_i(S_i(A) - S_i(B))} \mid \alpha_i\right] \quad (78)$$

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

$$\beta_\alpha = \mathbb{E}\left[\frac{G_i(X_i(A)) - G_i(X_i(B))}{X_i(A) - X_i(B)} \mid \alpha_i\right] \quad (79)$$

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 $(S_i(A) - X_i^0)^{-2}$. $\qquad\square$

B.   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.

B.1.   Linear Belief Updating Simplifies Estimation

In the replications in this paper and in many empirical settings, the sample size is small enough that it is quite demanding to non-parametrically control for the learning rate, the prior, and potential signals. Taking full advantage of the linearity in the belief updating process ((3)), it is sufficient to condition only on the learning rate and control for the prior 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:

B.1.1.   Local Regressions in Panel Experiments

Conditional on the rank of the observed change in beliefs $X_i - X_i^0$, regress the change in the outcome ΔY_i on the change in beliefs ΔX_i and a constant. This is exactly the local regression in Section 3.1.1.

B.1.2. Local Regressions in Active and Passive Control Experiments

Conditional on the rank of the observed learning rate α_i, regress the outcome Y_i on the posterior belief X_i, the prior X_i^0 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 $\left(S_i(A) - X_i^0\right)^{-2}$.

B.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.[26]

B.2.1. Trimming in Panel Experiments

Chose a threshold h^* and exclude observations with changes in beliefs $|\Delta X_i| < h^*$. This is a special case of Graham and Powell (2012).

B.2.2. Trimming in Active Control Experiments

Choose a threshold learning rate α^* 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 chose 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).

B.2.3. Trimming in Passive Control Experiments

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

---

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

B.3. Bandwidth Selection

Table B.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 inverval. 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 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 B.3 shows that the CAPE estimates for Settele (2022) have a consistent peak in the second quartile and estimates in Figure B.5 (Kumar et al., 2023) consistently slope downwards.

Estimation in active and passive designs proceeds in multiple steps: first, estimate the learning rate α_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 B.1. LLS and Fixed Effects Estimates

	PANEL A: Panel Experiments Wiswall and Zafar (2015)			
LLS	0.831 (0.318)	0.87 (0.33)	0.808 (0.319)	0.58 (0.313)
Paper (FE)	0.32 (0.086)	0.32 (0.086)	0.32 (0.086)	0.32 (0.086)
Bandwidth	0.025	0.05	0.075	0.1

	Armona, Fuster, and Zafar (2019)			
LLS	1.716 (0.368)	1.8 (0.381)	1.64 (0.383)	1.69 (0.359)
Paper (FE)	1.147 (0.234)	1.147 (0.234)	1.147 (0.234)	1.147 (0.234)
Bandwidth	0.01	0.025	0.05	0.1

	PANEL B: Active Experiments Settele (2022)			
LLS	0.178 (0.063)	0.16 (0.042)	0.132 (0.036)	0.117 (0.034)
Paper (TSLS)	0.096 (0.033)	0.096 (0.033)	0.096 (0.033)	0.096 (0.033)
Bandwidth	0.005	0.01	0.025	0.05

	Roth, Settele, and Wohlfart (2022)			
LLS	1.138 (0.383)	0.882 (0.366)	0.591 (0.357)	0.353 (0.33)
Paper (TSLS)	0.755 (0.435)	0.755 (0.435)	0.755 (0.435)	0.755 (0.435)
Bandwidth	0.05	0.075	0.1	0.15

	PANEL C: Passive Experiments Kumar, Gorodnichenko, and Coibion (2023)			
LLS	1.368 (0.462)	1.787 (0.469)	2.036 (0.537)	2.214 (0.588)
Paper (TSLS)	0.466 (0.19)	0.466 (0.19)	0.466 (0.19)	0.466 (0.19)
Bandwidth	0.01	0.025	0.05	0.1

	Cantoni, Yang, Yuchtman, and Zhang (2019)			
LLS	0.182 (0.236)	0.18 (0.164)	0.18 (0.133)	0.179 (0.12)
Paper (TSLS)	0.68 (0.253)	0.68 (0.253)	0.68 (0.253)	0.68 (0.253)
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 for different bandwidth choices 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).
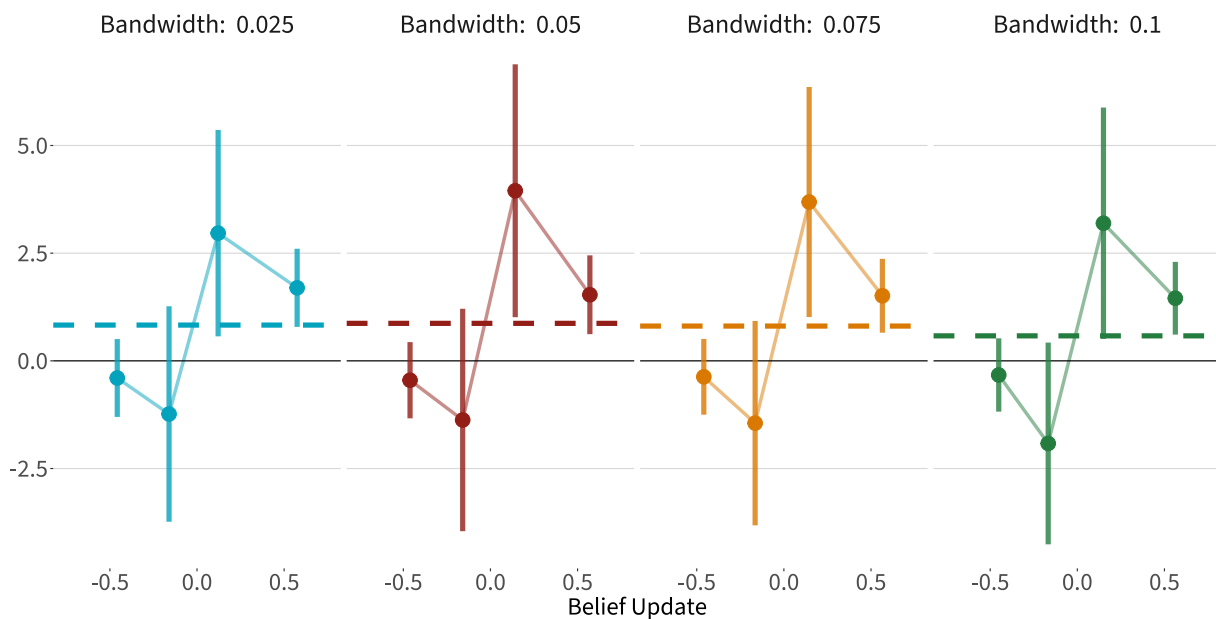
FIGURE B.1. Conditional Average Partial Effects in Wiswall and Zafar (2015), Several Bandwidths

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 B.1 for the point estimate and standard error of the APE.

FIGURE B.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 B.1 for the point estimate and standard error of the APE.
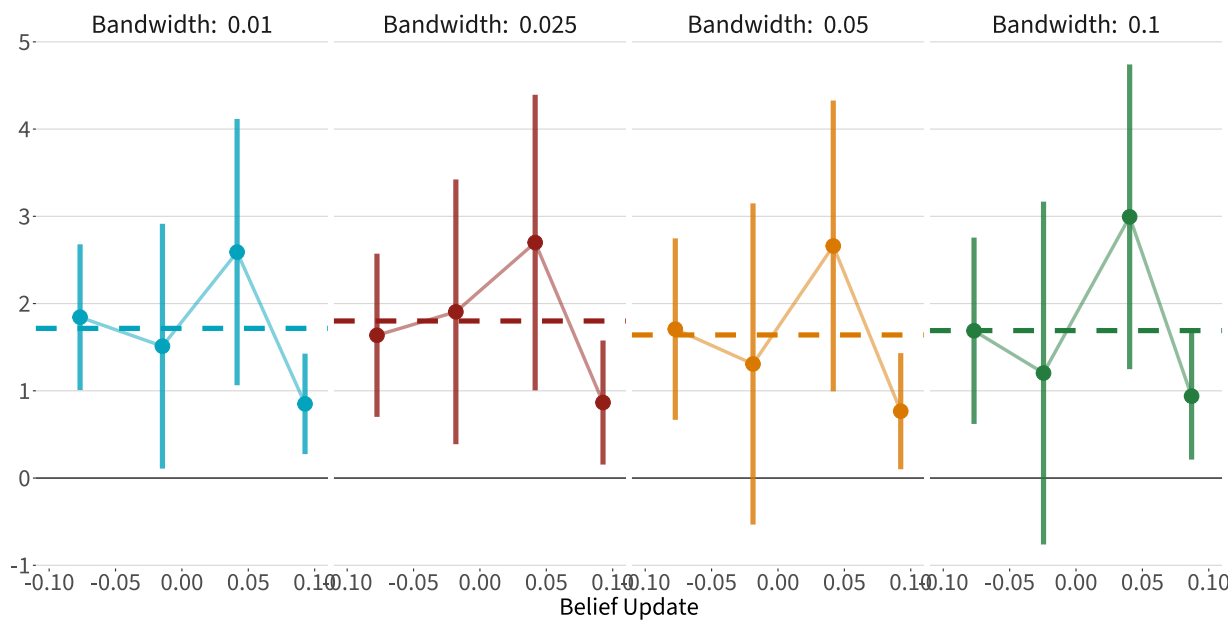
FIGURE B.3. Conditional Average Partial Effects in Settele (2022), Several Bandwidths

Notes: This figure plots estimates of the conditional average partial effect $\mathbb{E}\left[\tau_i \mid \text{rank}\left(\alpha_i\right)\right]$ 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 B.1 for the point estimate and standard error of the APE.
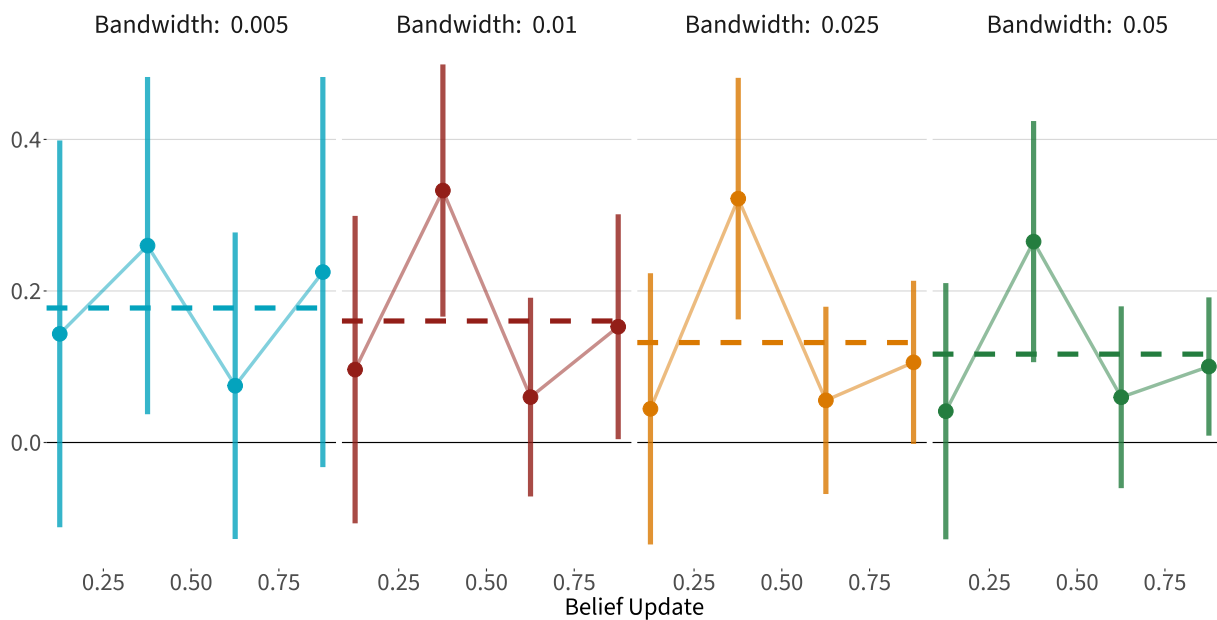
FIGURE B.4. Conditional Average Partial Effects in Roth et al. (2022), Several Bandwidths

Notes: This figure plots estimates of the conditional average partial effect $\mathbb{E}\left[\tau_i \mid \mathrm{rank}\left(\alpha_i\right)\right]$ 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 B.1 for the point estimate and standard error of the APE.
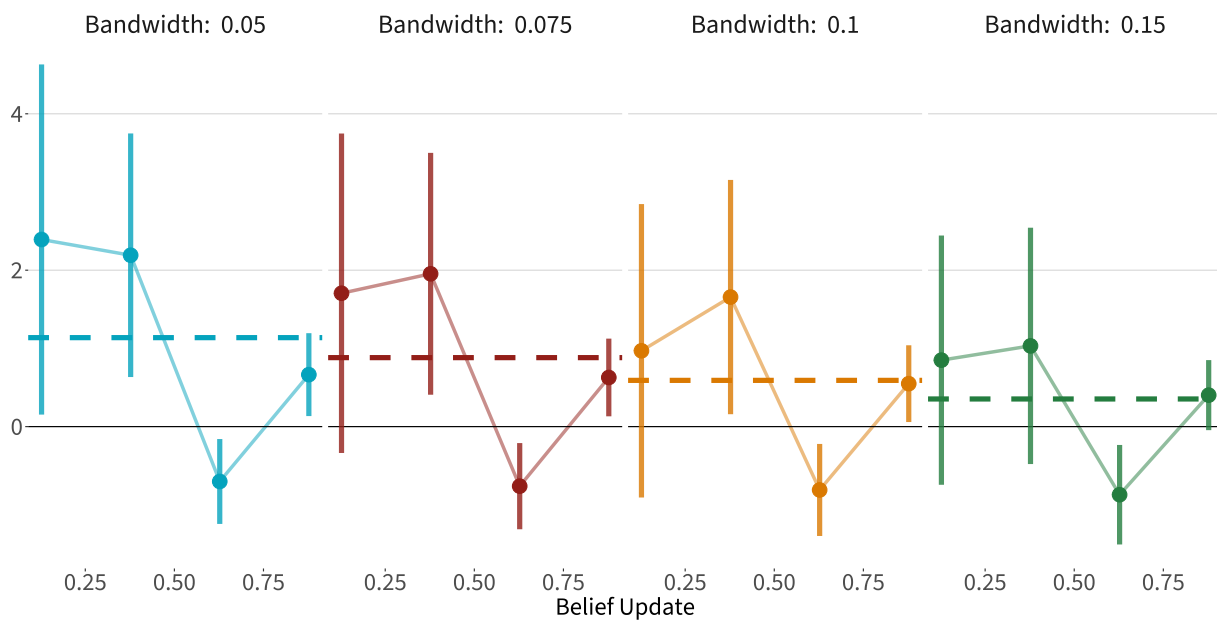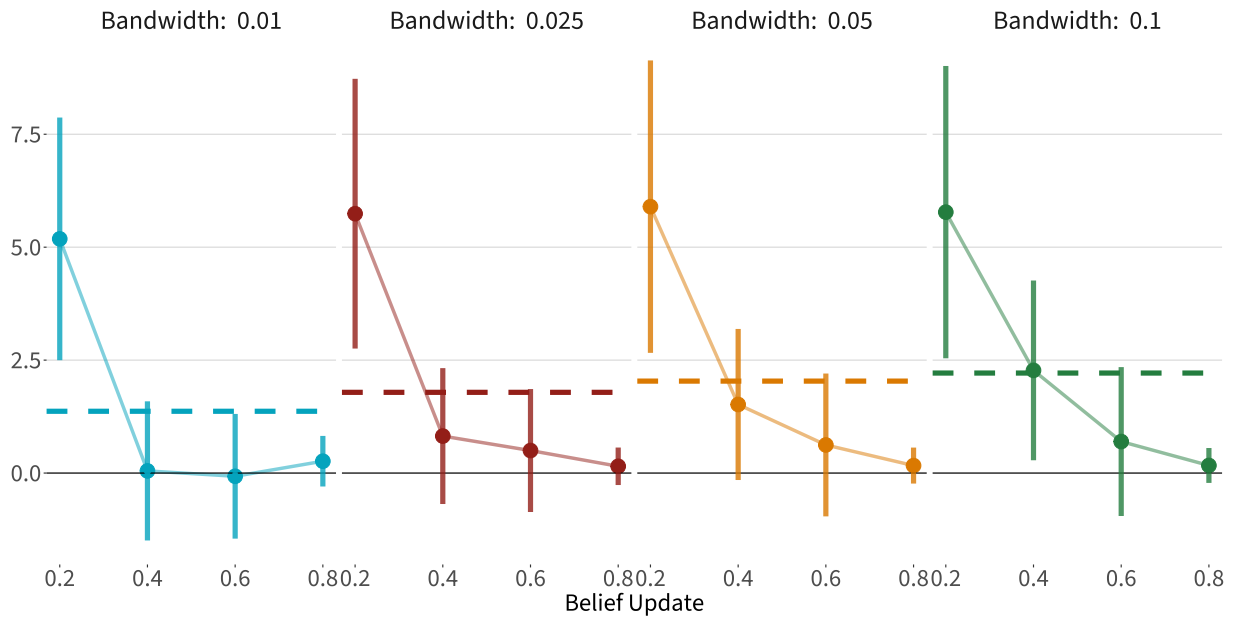
FIGURE B.5. Conditional Average Partial Effects in Kumar et al. (2023), Several Bandwidths

Notes: This figure plots estimates of the conditional average partial effect $\mathbb{E}\left[\tau_i \mid \text{rank}(\alpha_i)\right]$ 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 B.1 for the point estimate and standard error of the APE.

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

Notes: This figure plots estimates of the conditional average partial effect $\mathbb{E}[\tau_i \mid \text{rank}(\alpha_i)]$ 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 B.1 for the point estimate and standard error of the APE.
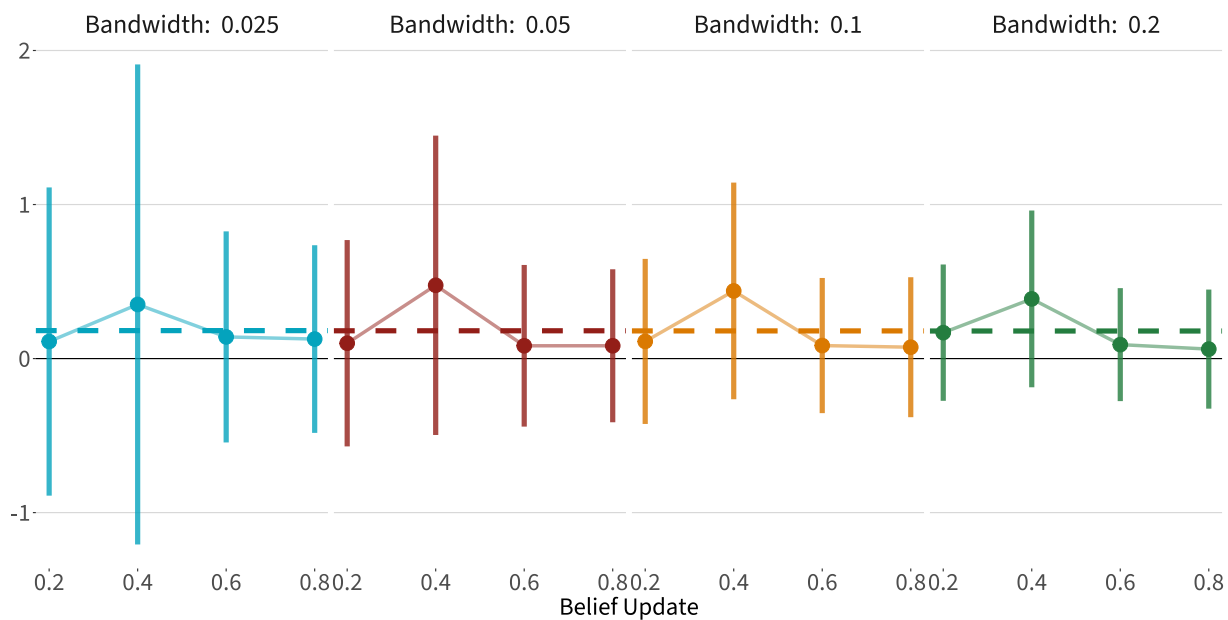
C.   Application Details

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

C.1.   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.

C.1.1.   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.

C.1.2.   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

$$\Delta Y_i = \ln(\pi_{k,i,\text{post}}/\pi_{\bar{k},i,\text{post}}) - \ln(\pi_{k,i,\text{pre}}/\pi_{\bar{k},i,\text{pre}}) \tag{80}$$

$$\Delta X_i = \ln(\omega_{k,i,\text{post}}/\omega_{\bar{k},i,\text{post}}) - \ln(\omega_{k,i,\text{pre}}/\omega_{\bar{k},i,\text{pre}}) \tag{81}$$

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.

C.1.3. 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.[27] I also follow Wiswall and Zafar (2015) and include fixed effects for college major in the local regressions.

C.2. 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.

C.2.1. 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 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.

C.2.2. 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 Y_i as the change in the percentage allocation to the housing asset and X_i as the change in one-year-ahead home price expectations. For each individual i, we observe these changes directly as first differences:

$$\Delta Y_i = Y_{i,\text{post}} - Y_{i,\text{pre}} \tag{82}$$

$$\Delta X_i = X_{i,\text{post}} - X_{i,\text{pre}} \tag{83}$$

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.

---

[27]While point estimates are qualitatively similar without trimming away from zero, this trimming is important for the precision of estimates.

C.2.3. 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 ΔX_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.

C.3. 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.

C.3.1. 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.

C.3.2. 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 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

C.3.3.   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 prior $\neq$ signal, so we exclude people with prior = signal. Additionally, the local regression is not identified for individuals with $\alpha = 0$, so we exclude them as well.[28] 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.[29]

As discussed in Appendix B.1, it is sufficient to control non-parametrically for the learning rate α_i and to control linearly for the remaining elements of the control vector $\left[S_i(A), S_i(B), X_i^0 \right]$. 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 effects conditional on (the rank of) α_i.

---

[28]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.

[29]Vilfort and Zhang (Forthcoming) 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".

C.4. 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.

C.4.1. 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.

C.4.2. 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 10. Instead, this estimand targets a weighted average of τ_i with weights $\omega_i \propto \alpha_i (S_i(A) - S_i(B))^2$.

More formally, the instrument is

$$T_i^\Delta \equiv \begin{cases} S_i(A) - S_i(B) & \text{if } Z_i = A \\ S_i(B) - S_i(A) & \text{if } Z_i = B \end{cases} \tag{84}$$

and the TSLS estimand is

$$\frac{\text{Cov}\left[T_i^\Delta, Y_i\right]}{\text{Cov}\left[T_i^\Delta, X_i\right]} = \mathbb{E}\left[\tau_i \cdot \frac{\alpha_i(S_i(A) - S_i(B))^2}{\mathbb{E}\alpha_i(S_i(A) - S_i(B))^2}\right] \tag{85}$$

C.4.3. 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 C.3.3. 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) α_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.

C.5. 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.

C.5.1. Setting

In this experiment, participating firms were first asked to report their prior beliefs about GDP growth. Then, participants were then 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 (Forthcoming), 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.[30] The analysis in this paper uses only comparisons between a single treatment arm and the control.

C.5.2. 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 the difference between the signal and the prior.[31] This specification is similar in spirit to the estimates in Table 3 of Kumar et al. (2023).

C.5.3. 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 $\left(S_i - X_i^0\right)^2 > 0.01$. This ensures that the weights proportional to $\left(S_i - X_i^0\right)^{-2}$ are well behaved.

The local regression is thus a regression of Y_i on X_i, X_i^0 conditional on (the rank of) σ_{Xi}^2, with weights proportional to $\left(S_i - X_i^0\right)^2$. The linear controls 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 recieve 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 B.5 more comparable to those in other designs, I estimate $\mathbb{E}\mathrm{rank}(\alpha) \mid \mathrm{rank}(\sigma_{Xi}^2)$ on the treated group and use this for the x-axis of the

---

[30]As Vilfort and Zhang (Forthcoming) 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).

[31]Vilfort and Zhang (Forthcoming) 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.

binned estimates.

C.6. Application Details: Cantoni, Yang, Yuchtman, and Zhang (2019)

Cantoni et al. (2019) study how beliefs about others' participation in protests affect an individuals' own protest decisions. The passive experiment provided a random subset of participants with truthful information about the planned participation of their classmates.

C.6.1. 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.

C.6.2. 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 τ_i with weights $\omega_i \propto \alpha_i |S_i - X_i^0|$.

The TSLS estimand is

$$\frac{\text{Cov}\left[\text{sign}\left(S_i - X_i^0\right) T_i, Y_i\right]}{\text{Cov}\left[\text{sign}\left(S_i - X_i^0\right) T_i, X_i\right]} \tag{86}$$

C.6.3. 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 C.5.3, I trim individuals with very small variation in the exposure to the signal and require that $\left(S_i - X_i^0\right)^2 > 0.01$.

The local regression is thus a regression of Y_i on X_i, X_i^0 conditional on (the rank of) $\widetilde{\alpha}_i$, with weights proportional to $\left(S_i - X_i^0\right)^2$. Recall that I use the notation $\widetilde{\alpha}_i$ 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 recieve 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.

C.6.4. 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 constrast, 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.

D. 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.

D.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 μ_i and variance σ_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 τ_i faces when their belief variance is σ_i^2. Note that R depends on the distribution F_i only through its variance σ_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 $\frac{\partial R}{\partial |\tau|^2} = 0$ 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 σ_+^2 is less than the variance of the initial beliefs σ^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

FIGURE D.1. People with Large Effects of Beliefs τ_i Form Precise Beliefs
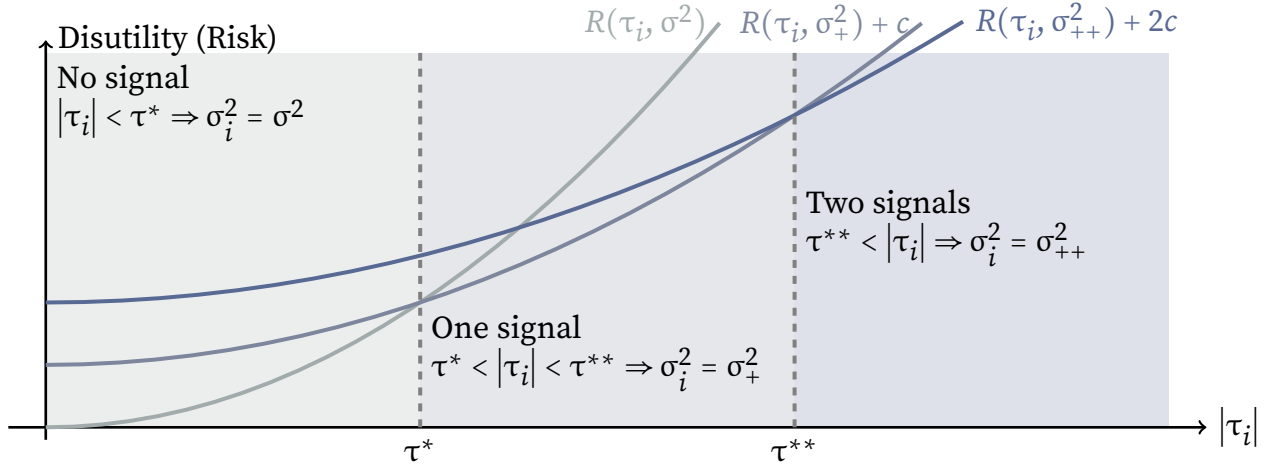
Notes: This figure plots the loss as a function of $|\tau_i|$ after seeing no signals, one signal, and two signals. The assumptions on R_i ensure that each pair of lines crosses exactly once. Since $R(\tau_i, \sigma^2) = R(\tau_i, \sigma_+^2)$ when $\tau_i = 0$, $R(\tau_i, \sigma^2) < R(\tau_i, \sigma_+^2) + c$. 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 $\frac{\partial^2 R}{\partial \sigma^2 \partial |\tau_i|} > 0$.

with variance σ^2, her loss can be given recursively by

$$V(\tau_i, \sigma^2) = \min\left\{R(\tau_i, \sigma^2), V(\tau_i, \sigma_+^2) + c\right\} \tag{87}$$

Given the assumptions we have made on R, for any beliefs with $\sigma^2 > 0$, there is some threshold value τ^* 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 D.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.

D.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 π_i. The initial variance of their beliefs is $\sigma^2_{X_0}$ so that their beliefs are represented by the normal $\mathcal{N}(\pi_i, \sigma^2_{X_0})$. For simplicity, $\sigma^2_{X_0}$ is common. We will consider signals S drawn from a normal distribution $\mathcal{N}(\mu_S, \sigma^2_S)$. 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 $L_i(y, x) = D\big(y, Y_i(x)\big)$, 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 τ_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. $\int L_i(y, x)dF(x)$).

When beliefs are given by the normal $\mathcal{N}(\overline{X}, \sigma^2_{X_0})$, the choice of Y that minimizes expected loss is simply $Y^* \equiv Y_i(\overline{X}) = \tau_i \overline{X} + U_i$. We can use this to further simplify the expression for expected loss and write

$$\int L_i(Y^*, x)dF(x) = \int D(Y_i(\overline{X}), Y_i(x))dF(x) \tag{88}$$

$$= \int \big((\tau_i \overline{X} + U_i) - (\tau_i x + U_i)\big)^2 dF(x) = \tau_i^2 \sigma_X^2 \tag{89}$$

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 D.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 we have been working with throughout. When

a person's beliefs are given by $\mathcal{N}(\overline{X}, \sigma_x^2)$, her loss is given recursively by

$$V_i(\overline{X}, \sigma_x^2) = \min\left\{\mathbb{E}_X[L_i(Y_i(\overline{X}), x)], \mathbb{E}_S[V_i(X'(s)), \sigma_{X'}^2] + C\right\} \tag{90}$$

Where $\sigma_{X'}^2 = \frac{\sigma_X^2 \sigma_S^2}{\sigma_X^2 + \sigma_S^2}$ and the expectation $\mathbb{E}[S]$ is the expectation with respect to the signal. Notice that in this model, 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.

Solving this recursive problem gives the equilibrium condition

$$\tau_i^2 \sigma_X^2 = \tau_i^2 \sigma_{X'}^2 + C \tag{91}$$

In equilibrium, agents will be indifferent between paying the fixed cost to obtain new information and living with the uncertainty they have.[32] Replacing $\sigma_{X'}^2$ with its definition, and recalling that $1 - \frac{\sigma_S^2}{\sigma_S^2 + \sigma_X^2} = \alpha_i$ we obtain the following equality

$$\alpha_i \tau_i^2 \sigma_X^2 = C \tag{92}$$

Agents for whom the outcome is very sensitive to the beliefs (τ_i^2 is very large) will update their information until $\sigma_X^2 \alpha_i$ is small.[33] On the other hand, agents for whom the outcome is not sensitive to beliefs (τ_i^2 is small) will stop after seeing fewer signals, so that $\sigma_X^2 \alpha_i$ is relatively large.

We can see in this toy model 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 is directly depends on the causal effects of interest.

D.3. Using Models of Belief Updating to Interpret Empirical Estimates

The class of parameters that are targeted by existing standard specifications depend not only on the causal effects of beliefs on outcomes τ_i, but also on heterogeneity in the way

---

[32]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.

[33]Notice that since $\alpha_i \equiv \frac{\sigma_X^2}{\sigma_S^2 + \sigma_X^2}$, α_i and σ_X^2 move together. That is, holding fixed σ_S^2, an increase in σ_X^2 implies an increase in α_i and vice-versa.

that beliefs are updated in response to new information.

In the model proposed in this section, beliefs are formed endogenously through a process of costly information acquisition. In Appendix D.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 α_i and variance of the prior σ_i^2 as a function of the causal effects of beliefs τ_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. Any model that makes predictions about the covariance between the learning rate α_i and the causal effect of beliefs on behavior τ_i can be used to make predictions about the difference between estimates from standard specifications and the APE.

E. 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 Bayesian updating assumptions. Section E.4 provides a novel strategy to ensure non-negative weights when priors are not elicited.

E.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 $T_i \equiv \mathbb{1}\{Z_i = A\}$.

$$\beta^{RF} \equiv \frac{\mathrm{Cov}\left[T_i, Y_i\right]}{\mathrm{Var}\left[T_i\right]} \tag{93}$$

$$= \mathbb{E}\left[\tau_i \left(X_i(A) - X_i(B)\right)\right] \tag{94}$$

The reduced form effect of assignment to arm A on the outcome is the expectation of the individual effect of beliefs on behaviors τ_i scaled by the individual effect of the information treatment on beliefs $X_i(A) - X_i(B)$. If all τ_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 beliefs 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.

E.1.1. From the Effect of Information to the Effect of Beliefs

As Giaccobasso 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.

E.1.2. Constructing TSLS Estimates

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

$$\beta^{TSLS} \equiv \frac{\beta^{RF}}{\beta^{FS}} = \frac{\text{Cov}\left[T_i, Y_i\right]}{\text{Cov}\left[T_i, X_i\right]} \tag{95}$$

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

$$\beta^{TSLS} = \frac{\mathbb{E}\left[Y_i \mid T_i = 1\right] - \mathbb{E}\left[Y_i \mid T_i = 0\right]}{\mathbb{E}\left[X_i \mid T_i = 1\right] - \mathbb{E}\left[X_i \mid T_i = 0\right]} \tag{96}$$

Substituting the linear outcome equation (1) yields

$$\beta^{TSLS} = \frac{\mathbb{E}\left[\tau_i\left(X_i(A) - X_i(B)\right)\right]}{\mathbb{E}\left[X_i(A) - X_i(B)\right]} \tag{97}$$

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 τ_i and the first stage variation $X_i(A) - X_i(B)$:

$$\frac{\mathbb{E}\left[\tau_i\left(X_i(A) - X_i(B)\right)\right]}{\mathbb{E}\left[\left(X_i(A) - X_i(B)\right)\right]} = \mathbb{E}\left[\tau_i\right] + \frac{\text{Cov}\left[\tau_i, \left(X_i(A) - X_i(B)\right)\right]}{\mathbb{E}\left[\left(X_i(A) - X_i(B)\right)\right]} \tag{98}$$

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

E.2. Unconditional Instrument Monotonicity and Bayesian Updating

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

E.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.

E.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 (95) is not widely used in practice; instead researchers use one of two main strategies to ensure positive weights.

E.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.

E.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 $S_i(A) - X_i^0 > 0$:

$$\beta_+^{\text{split}} = \frac{\text{Cov}\left[T_i, Y_i \mid S_i(A) - X_i^0 > 0\right]}{\text{Cov}\left[T_i, X_i \mid S_i(A) - X_i^0 > 0\right]} \tag{99}$$

$$= \mathbb{E}\left[\tau_i \cdot \frac{\alpha_i |S_i(A) - X_i^0|}{\mathbb{E}[\alpha_i |S_i(A) - X_i^0| \mid S_i(A) - X_i^0 > 0]} \mid S_i(A) - X_i^0 > 0\right] \tag{100}$$

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

E.3.2. Exposure-Weighted Instruments

An example of the exposure-weighted instrument is presented in Section 2.3.

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

The recentering is implicit since in practice researchers use the interaction as an instrument and control for the uninteracted exposure. These weights proportional to

$\alpha_i(S_i(A) - X_i^0)^2$ are non-negative under Bayesian Learning 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 (Forthcoming) show that implementation of these specifications requires care, as including both the exposure-weighted instrument and the treatment indicator can result in misspecification.

E.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 β_+^{split} and β_-^{split} in Appendix E.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 3 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 E.3.1 are not feasible since they use the prior to split the sample, the following regressions are feasible and equivalent.

$$\beta_+^{\text{split}} = \widetilde{\beta}_+^{\text{split}} \equiv \frac{\text{Cov}\left[T_i, Y_i \mid S_i(A) - X_i > 0\right]}{\text{Cov}\left[T_i, X_i \mid S_i(A) - X_i > 0\right]} \tag{101}$$

$$\beta_-^{\text{split}} = \widetilde{\beta}_-^{\text{split}} \equiv \frac{\text{Cov}\left[T_i, Y_i \mid S_i(A) - X_i < 0\right]}{\text{Cov}\left[T_i, X_i \mid S_i(A) - X_i < 0\right]} \tag{102}$$