Selecting Experimental Sites for External Validity*

Michael Gechter, Keisuke Hirano, Jean Lee, Mahreen Mahmud, Orville Mondal, Jonathan Morduch, Saravana Ravindran and Abu S. Shonchoy†

September 26, 2026

Abstract

The ability to generalize experimental findings usually depends on where the research took place and where the evidence is being applied. By choosing where to experiment, researchers influence the external validity of estimated treatment effects and, ultimately, the quality of decisions informed by them. We characterize the optimal selection of a small number of experimental sites to maximize the value of evidence for policymaking across a wider range of settings. We develop a Bayesian decision-theoretic framework that links sites through two sources of information: a smoothing prior based on proximal similarity and a structural prior derived from a micro-econometric model. Using data from South Asia, we show that experimentation alone does not necessarily improve policy choices: depending on how the estimates are extrapolated, some experimental evidence can be so unrepresentative that policy decisions generate large expected welfare losses. We quantify the welfare gains from Bayesian updating of evidence from optimally-selected sites, and we identify when replicating experiments in multiple sites can increase welfare gains within a given budget.

Keywords: external validity, experimental design, field experiments, mobile money, structural modeling, statistical decision theory

JEL codes: C9, C11, O15, O16, O18

*This paper previously circulated as "Site Selection for External Validity: Theory and an Application to Mobile Money in South Asia." This work was supported by the Bill and Melinda Gates Foundation. Hirano received support from the National Science Foundation through grant SES-2117260. Ravindran is funded by a startup grant at the Lee Kuan Yew School of Public Policy, National University of Singapore. We are grateful for research assistance from Caio Figueiredo, Nihal Mehta, Preksha Jain and Shashank Sreedharan. We thank Seth Garz and participants of workshops at the University of Chicago, Amsterdam, Duke, Florida State, Georgetown, Harvard, Innovations for Poverty Action, the Journées Louis-André Gérard-Varet, Michigan State, Singapore NUS, the Penn State - Cornell IO/Econometrics Conference, QUT Brisbane, Tennessee, Uber, the World Bank, UN-FAO, Yale, Y-RISE, and Harvard for helpful comments.

†Gechter: University of Arkansas (email: mgechter@uark.edu); Hirano: Pennsylvania State University (email: kuh237@psu.edu); Lee: The World Bank (email: jlee20@worldbank.org); Mahmud: University of Exeter (email: m.mahmud@exeter.ac.uk); Mondal: Bates White (email: orville.dm@gmail.com); Morduch: New York University (email: jonathan.morduch@nyu.edu); Ravindran: National University of Singapore (email: saravana@nus.edu.sg); Shonchoy: Florida International University (email: shonchoy@fiu.edu).

1 Introduction

Evidence-based policy requires reliable and generalizable estimates of treatment effects (Al-Ubaydli, List, and Suskind 2017b; Bates and Glennerster 2017; Allcott 2015; Vivalt 2020). A large literature describes research designs that optimize the internal validity of treatment effects (see, e.g., Angrist and Pischke, 2010; Abadie and Cattaneo, 2018), but external validity is usually considered only after treatment effects have been estimated (Campbell and Stanley, 1966; Cook and Shadish, 1994).

We build from the idea that researchers designing field experiments can influence the external validity of their results by choosing where to experiment. We start with a set of policy choices that will ultimately be made in different settings, and we then work backward to determine where experiments should take place to be most informative for those choices. Our framework brings external validity into research design by taking a Bayesian decision-theoretic approach to learning and by framing external validity through a policy lens. The choice of setting recognizes that selecting a given site in which to experiment might deliver estimates that are more relevant for some decision-makers but less relevant for others. The framework allows potential research settings to be ranked by how evidence generated in each site affects policy decisions, and thus expected welfare, in other sites.

We begin our analysis from the point of view of a social planner who must use the results from a single experiment, or a limited set of coordinated experiments, to determine if a program should be implemented in other places. Following Deaton and Cartwright (2018), our notion of external validity departs from the narrow question of whether a treatment effect from one site is expected to be similar elsewhere (see, e.g., Campbell and Stanley, 1966). Instead, we ask how informative a treatment effect estimated in a given site is for expected treatment effects in other sites, taking into account knowledge about the structure of economic relationships, the nature of economic contexts, and previous empirical evidence. Based on this notion of external validity, we consider the experimental design problem of choosing the sites in which to run experiments to make more effective decisions broadly.

Our quasi-Bayesian framework for this experimental design problem allows for various prior assumptions about the heterogeneity in site-level effects. We develop a novel method to select experimental sites that incorporates structural econometric modeling and estimation applied to a pilot data set in combination with other prior judgments about the likely similarity of treatment effects across sites. Formally incorporating this pilot data in our design problem leads to a multi-stage or adaptive framework, where past experimental data can influence, and improve, future site selection.

One version of this approach uses a "smoothing" prior that encodes the intuition that sites

with proximally similar characteristics (such as origin–destination distance and average household income)[1] should have similar treatment effects. If there are many site characteristics, however, the analysis can be sensitive to the inclusion and weighting of different characteristics.

A second approach is structural. Tailored to our application, we introduce a model for migrant remittance decisions that provides a microeconomic foundation for site-level effect heterogeneity and allows us to link the experimental data from Lee, Morduch, Ravindran, Shonchoy, and Zaman (2021), which we regard as a pilot experiment, to the experimental design problem. In particular, we specify and estimate a structural model using the experimental data. We then use the parameter estimates and their uncertainty to construct a joint prior for the ATEs at all the prospective experimental sites. This allows us to draw on the detailed individual-level data in the pilot experiment to assess the *a priori* similarity of different sites. While a number of researchers have used structural modeling methods to analyze data from randomized controlled trials in development economics (see, for example, the survey articles by Todd and Wolpin 2010, 2023), the potential for applying structural approaches to *experimental design* has been relatively unexplored.

Incorporating structural modeling can greatly increase the prior information available for the experimental design step, but the resulting design could be fragile to misspecification of the structural model. For this reason, we also develop a hybrid joint prior for the potential experimental sites that combines the two approaches above, mixing between the prior generated by the structural model and the smoothing prior that allows sites with similar characteristics to have similar predicted treatment effects.

In our application, the question is whether to promote the adoption of a new financial technology initially evaluated by Lee et al. (2021). The framework is used to choose a small set of migration corridors from among hundreds of candidates across India, Bangladesh, and Pakistan in which to experimentally evaluate the financial technology. Lee et al. (2021) reported on the original intervention, which was implemented in a single migration corridor in Bangladesh. The intervention made it easier for migrant workers to send and receive remittances, causing an increase in the volume of urban-to-rural remittances and a reduction of rural poverty. It is unclear, however, when and where those results might generalize.

The goal of the new set of experiments is to derive policy recommendations for adopting the intervention in *all* the candidate sites. In our context a "site" is a migration corridor consisting of an origin and destination district (a geographic area roughly equivalent to a US county). While the experiments provide direct estimates of the average treatment effects (ATEs) for the chosen sites, the goal is to extrapolate from those estimates to gain knowledge about the overall

[1] The notion of proximal similarity, due to Campbell (1986), refers to similarity on the dimensions likely to matter for the treatment response, rather than geographic proximity.

distribution of site-level ATEs across Bangladesh, India, and Pakistan. The ATEs of the intervention may vary across sites due to differences in the characteristics of migrants associated with the site, other site-level characteristics, and further unobserved factors. The problem captures one of the essential challenges of external validity: the number of potential corridors is large and their characteristics vary substantially in ways that could alter the effect of the intervention.[2]

In our analysis, we take advantage of the prior experimental study that was conducted in Bangladesh (Lee et al., 2021). Since many field experiments have a pilot phase, or are replications or extensions of previous experiments in different contexts, such prior information may often be available to the designer of the new experiment. Given the large number of possible sites and the limited scope for experimentation, being able to incorporate this prior experimental evidence adds critical predictive power. The framework allows us to link the prior evidence and the effects at the chosen experimental sites to the effects at the remaining sites.[3]

Evidence from the small set of new experiments informs a decision rule for each of the hundreds of other migration corridors. The rule maps each corridor's characteristics to an up-or-down recommendation for whether the technology adoption program should also be implemented there. We measure the effects of the intervention, and the overall implications of the site selection, through a constant elasticity of substitution social welfare function (c.f. Deaton 1997, Alderman, Behrman, and Tasneem 2019) for rural household members, where we proxy individual utility by per capita expenditure.

We find that the sites selected through the welfare-optimizing approach are relatively centrally located in the space of characteristics of all sites, where our model specifies the appropriate notion of distance in characteristic space. This is because the optimal sites for experimentation are those that are most informative about policy decisions in the other sites—so the optimal sites will tend to be similar to other sites along relevant dimensions. The optimal choices can thus be very different from sites where the intervention is likely to have the highest expected treatment effect or the sites where the efficacy of the intervention can be best demonstrated.

Relative to using only the information from the Lee et al. (2021) experiment, as embodied in the prior, there are large welfare gains to additional experimentation in other sites, but we show that the marginal gains diminish in the number of experimental sites. Under a fixed budget,

[2] Muralidharan and Niehaus (2017) discuss the scaleup from experimentation in a pilot experiment in one site to several sites, raising concerns about spillover effects and differences in small vs. large scale interventions, which we do not address directly, although our approach does not preclude including either concern in experimental design. Similarly, Al-Ubaydli, List, LoRe, and Suskind (2017a) discuss other factors in scaling, drawing on insights from the medical literature on nonadherence to treatment even when it has been proven effective.

[3] Alternatively, the smoothing and structural approaches can be seen as ways of specifying a prior in terms familiar to economists in the absence of pilot data.

the advantages of experimenting in a broader range of settings are balanced against losses in statistical power in each site. The welfare gains are concentrated among experimental sites where the suitability of the treatment is *a priori* fundamentally uncertain.

The framework makes explicit how learning takes place across sites, and it permits comparisons of alternative ways of selecting sites and extrapolating evidence. We compare the optimal results generated by the Bayesian framework to two alternative ways of choosing where to experiment. The first is choosing sites that *a priori* are expected to deliver the largest treatment effects. This kind of positively-biased site selection is documented, for example, in Allcott (2015)'s study of 111 energy-conservation experiments in the United States. The second strategy is site selection through random choice, a proposal by Allcott (2015) to mitigate bias in inference due to correlation between ATEs and site selection.[4]

The data from South Asia show that, relative to basing policies on experiments in optimal sites, large welfare losses result from instead using evidence from randomly-selected sites or, conversely, from sites with the largest expected treatment effects. Indeed, naively extrapolating evidence from experiments in the sites predicted to have the highest treatment effect under our prior can, in absolute terms, yield a substantial *negative* welfare effect from experimentation. The welfare losses are mitigated through extrapolation that is sensitive to context, but expected welfare is maximized only with optimal extrapolation of evidence from optimally-selected sites.

1.1 Background and Related Literature

Our work contributes to a growing literature on external validity and generalizability in economics, statistics, and other fields. Some authors, such as Pritchett and Sandefur (2015), Vivalt (2020), and Meager (2019), have primarily viewed external validity as holding when treatment effects are relatively similar across different contexts, and focus on measuring the extent of departure from homogeneity. In line with this focus, most previous work on research settings and generalization aims to address external validity by either running coordinated experiments in similar sites (to establish the basis for homogeneity) or, in contrast, by running experiments in sites that are deliberately dissimilar (Cook and Shadish, 1994). For example, Banerjee, Duflo, Goldberg, Karlan, Osei, Pariente, Shapiro, Thuysbaert, and Udry (2015) jointly design and analyze six RCTs, each in a different country, replicating a program that was first developed and analyzed in rural Bangladesh (Hashemi and De Montesquiou 2011; Bandiera, Burgess, Das, Gulesci, Rasul, and Sulaiman 2017).[5]

[4] In line with Allcott (2015), biases in site selection are judged against a benchmark of randomized site selection; e.g., Wang and Yang (2025), Mentges, Blowes, Hodapp, Hillebrand, and Chase (2021), and Stuart, Bell, Ebnesajjad, Olsen, and Orr (2017).

[5] Breza, Chandrasekhar, and Viviano (2026) re-analyze these RCTs to describe the heterogeneity of treatment effects and the nature of uncertainty about the evidence. The original program was ultimately scaled to more than

Other work, such as Stuart, Cole, Bradshaw, and Leaf (2011), O'Muircheartaigh and Hedges (2014), and Tipton (2013, 2014), focuses on generalizing across contexts by adjusting for differences in the distribution of covariates, under the assumption that conditional average treatment effects are stable. Unlike these papers, we allow for unobserved factors that could lead to differences across sites and focus on the small-number-of-sites case.[6]

The problem of transporting causal estimates and inferences across contexts has also been studied by Hotz, Imbens, and Mortimer (2005), Gechter (2023), Andrews, Fudenberg, Lei, Liang, and Wu (2023), Kuang, Xiong, Cui, Athey, and Li (2018), Dehejia, Pop-Eleches, and Samii (2021), Gechter, Samii, Dehejia, and Pop-Eleches (2019), Adjaho and Christensen (2022), Kallus and Zhou (2021), and Menzel (2023), among others. Our setup is particularly closely related to the problem considered in Ishihara and Kitagawa (2021), Montiel Olea, Qiu, and Stoye (2023), Aradillas Fernandez, Blanchet, Montiel Olea, Qiu, Stoye, and Tan (2024), and Epanomeritakis and Viviano (2025). While most of these papers impose *a priori* bounds on the differences between site-level populations to facilitate extrapolation, our strategy is to craft a Bayesian prior that allows us to model varying degrees of similarity across sites.

Glennerster (2017) and Chassang and Kapon (2021), among others, discuss the importance of considering external validity when designing experiments. O'Muircheartaigh and Hedges (2014) and Tipton (2014) discuss how experiments should be designed so that the distribution of covariates in the experiment is sufficiently rich for covariate-based extrapolation to other contexts. Egami and Lee (Forthcoming) also treat site selection as a design problem with a small number of sites, choosing experimental sites so that the observed characteristics of each non-selected site are well approximated by a weighted average of the selected sites' characteristics. The corresponding weighted average of experimental findings then minimizes the worst-case mean squared error, among weighted-average estimators, for the average treatment effect across all sites. Our approach instead models how treatment effects covary across sites, including through unobserved factors, incorporates evidence from prior experiments, and targets the welfare consequences of site-specific policy decisions rather than an average effect. There has been relatively little work on experimental design that accounts for more general forms of heterogeneity across sites. Our approach is Bayesian and sequential, in that we use data from a

2.1 million households in Bangladesh and over 1.1 million households in 14 other countries (JPAL, 2022).

[6]We focus on settings with few experimental sites because even in very well-resourced trials the number of manageable sites is small relative to the number of potential sites. The total number of trials evaluating the aforementioned program targeting extreme poverty was only 10. Other examples of multi-site trials have similarly few sites. Angrist, Ainomugisha, Bathena, Bergman, Crossley, Cullen, Letsomo, Matsheng, Marlon Panti, Sabarwal, and Sullivan (2023) and Angrist, Bergman, and Matsheng (2022) report on six RCTs of remote instruction for primary school children during the school shut-downs caused by COVID-19 and Cusolito, Darova, and McKenzie (2023) evaluate an intervention in six Balkan countries that provides skills to help small businesses compete in the export sector.

pilot experiment to inform our choice of future experimental sites. Finan and Pouzo (2023) consider sequential experimentation in one site or context, using various sources of prior information including past data from other contexts to inform the experimental stopping rule and policy recommendation. Other recent work in sequential experimental design in economics includes Hahn, Hirano, and Karlan (2011), Tabord-Meehan (2018), Kasy and Sautmann (2021), Caria, Gordon, Kasy, Quinn, Shami, and Teytelboym (2021), and Athey, Bergstrom, Hadad, Jamison, Ozler, Parisotto, and Sama (2021). Unlike many of these other works, our key experimental design choice is in the selection of experimental sites.[7]

In the next section, we explain our proposed methodology at a general level that could be ported to other applications. We set up experimental site selection as a formal decision-theoretic problem that embodies our notion of external validity, and propose a Bayesian approach that allows us to tractably incorporate prior experimental evidence and structural econometric analyses. Then, in Section 3, we begin to focus on our specific application by developing a model for migrant labor supply and remittance choices. This model is taken to the prior experimental data in Section 4. Section 5 then implements the full experimental design analysis for Bangladesh, Pakistan, and India.

2 Experimental Site Selection for External Validity

2.1 Setup and Design Problem

We first introduce a framework for analyzing experimental site choice with external generalizability as its key goal, by formalizing a policy choice problem across all potential experimental sites. We have a set of potential experimental sites $s = 1, \ldots, S$. In our application, each site s represents a migration corridor consisting of a home–destination district pair. For each site s, we observe site characteristics $V_s \in \mathcal{V}$. The variable V_s could be high-dimensional, for example to capture the distributions of demographic and other variables at the site.

The true (but unknown) effect at site s is τ_s. Let $\tau = (\tau_1, \ldots, \tau_S)'$. If we select a site s and run a randomized controlled trial at that site, we will obtain an estimate $\hat{\tau}_s$ of the site-level effect, where

$$\hat{\tau}_s = \tau_s + \epsilon_s, \quad \epsilon_s \overset{\text{ind}}{\sim} N(0, \sigma^2_{\epsilon,s}).$$

We assume that the $\sigma^2_{\epsilon,s}$ are known (or user-specified and fixed in advance) for the remainder of

[7] Site selection is an important concern for clinical trials in medicine; see for example Potter, Donovan, Weiss, Gardin, Lindblad, Wakim, and Dodd (2011) and Hurtado-Chong, Joeris, Hess, and Blauth (2017). However, this literature has primarily focused on choosing sites to ensure sufficient enrollment of experimental subjects, and timely and high-quality implementation of experimental treatments. Related issues are discussed in the context of social science field experiments by Glennerster (2017).

the analysis. In our application, this assumption is plausible, as we fix the sample size at each site in advance, and we have prior data that is informative about individual-level treatment effects.[8]

The experimental design problem is to choose a *subset* of sites, $\mathscr{S} \subset \{1,\ldots,S\}$, on which to run experiments. We are given a feasible set of subsets $\mathscr{A}$, and must choose $\mathscr{S} \in \mathscr{A}$. In the case where we are choosing a single site, the choice set $\mathscr{A}$ would consist of all the singleton sets $\{1\},\{2\},\ldots,\{S\}$. More generally, the feasible set $\mathscr{A}$ can incorporate various constraints; for example it may limit the total number of experimental sites, or rule out certain combinations of sites. After choosing $\mathscr{S}$ we will observe $\hat{\tau}_s$ only for $s \in \mathscr{S}$. Let $\hat{\tau}_{\mathscr{S}} := \{\hat{\tau}_s : s \in \mathscr{S}\}$ denote the observed experimental estimates for the sites in $\mathscr{S}$.

Although we are constrained to experiment on a limited subset of the sites, we want to choose the experimental sites to maximize the information gained about the effects across all sites. We operationalize this goal by focusing on the entire vector of site-level effects $\tau = (\tau_1,\ldots,\tau_S)'$ and posing a hypothetical policy choice problem as a sequential decision problem. The social planner, after observing $\hat{\tau}_s$ for the subset of sites $s \in \mathscr{S}$, chooses treatments for every site $s \in \{1,\ldots,S\}$, with the goal of maximizing aggregate welfare. This sequential decision problem is outlined in Figure 1.

Figure 1: Basic Site Selection Problem

1. Observe all the site characteristics $V = (V_1,\ldots,V_S)$.
2. Choose $\mathscr{S} \in \mathscr{A}$.
3. Observe $\hat{\tau}_{\mathscr{S}}$.
4. Choose a vector of binary treatments $T = (T_1,\ldots,T_S)$.
5. Evaluate T by a social welfare function $W(\tau, T)$.

The key design decision is in Step 2. From the standpoint of the experimental designer, Steps 3–5 are prospective, and serve to quantify the designer's objectives. Note that, in Step 4, we are choosing whether or not to implement the intervention separately across *all* sites, not just the experimental sites in $\mathscr{S}$. Thus a site-selection subset $\mathscr{S}$ is desirable if it will allow the social planner to make good policy decisions across observed and unobserved sites. This prospective

[8]Note that any clustering in the sampling design of the trial will be reflected in $\sigma_{\epsilon,s}$ but does not affect the estimand τ_s.

policy choice problem embodies our concern for the external validity or generalizability of the choice of a small number of experimental sites. As a consequence, we are interested in the full vector $\tau = (\tau_1, \ldots, \tau_S)'$ of site-level treatment effects, not some aggregate such as the average effect across all sites. We will discuss the specification of the social welfare function $W(\tau, T)$ in more detail below, but before doing so, we will extend this basic framework to allow us to incorporate prior information in the form of a pilot experimental study, and develop a solution concept for the extended problem.[9]

The basic decision problem outlined in Figure 1 is difficult in part because the decision-relevant parameter τ may be high-dimensional, yet the experimental process will provide direct information on only a limited number of its elements. This reflects the general challenge of external validity in empirical studies, but in cases where the total number of experimental sites is limited, it will be especially helpful to incorporate additional sources of information and modeling assumptions. We will do so by using prior experimental evidence in conjunction with structural modeling techniques that facilitate extrapolation of likely effects across all sites.

In our application to evaluating the impacts of mobile money training, we can take advantage of the previous experimental study conducted by Lee et al. (2021) on a different experimental site. We can view this as an additional site $s = 0$. For this site, we have observed not only its simple treatment effect estimate $\hat{\tau}_0$, but also a rich micro-level data set with individual characteristics, assigned treatments, and outcomes which permit more detailed analysis and extrapolation to other sites. We therefore augment the original decision problem with an initial step in which the previous randomized control trial, which we will refer to as the pilot RCT, is made available to the designer. This extended site selection problem is outlined in Figure 2.

2.2 Bayesian Solution

We propose to solve the extended experimental design problem by adopting a (quasi-) Bayesian "preposterior" analysis, in which a prior distribution for the unknown vector τ is used to evaluate the expected welfare gains from alternate possible choices of the site selection subset $\mathscr{S}$. Specifically, after Steps 0 and 1 in Figure 2, we will construct a prior distribution Π for the S-dimensional vector τ. The prior Π will be based on the preliminary experimental analysis and other modeling assumptions, as well as the vector of site characteristics $V = (V_1, \ldots, V_S)$. (For notational ease, we suppress the dependence of Π and other quantities on the characteristics V in the sequel.) We will discuss the specification of the prior in more detail in the next subsection,

[9] We assume the vector of site-specific treatment effects τ is invariant with respect to the vector of implementation decisions T. It is in this sense that we abstract from one dimension of "scaling up," where, say, the number of treated sites may modify treatment effects through general equilibrium effects. This consideration could in principle be added to our framework.

Figure 2: Extended Site Selection Problem

0. Pilot RCT: for site $s = 0$, observe individual characteristics, randomized treatments, and experimental outcomes.
1. Observe all the site characteristics $V = (V_1, \ldots, V_S)$.
2. Choose $\mathscr{S} \in \mathscr{A}$.
3. Observe $\hat{\tau}_{\mathscr{S}}$.
4. Choose a vector of binary treatments $T = (T_1, \ldots, T_S)$.
5. Evaluate T by a social welfare function $W(\tau, T)$.

but for now we allow for general forms of Π.

We can solve the problem using backward induction. Specifically, given a choice for $\mathscr{S} \in \mathscr{A}$ and observation of $\hat{\tau}_{\mathscr{S}}$, we update the prior $\Pi(\tau)$ for the vector of treatment effects to its posterior $\Pi(\tau \mid \hat{\tau}_{\mathscr{S}})$. The vector of treatments that maximize posterior expected welfare is given by

$$T^* = T^*(\Pi, \hat{\tau}_{\mathscr{S}}) = \arg\max_T \int W(\tau, T) d\Pi(\tau \mid \hat{\tau}_{\mathscr{S}}), \tag{1}$$

where the maximization is over all S-vectors T of binary treatments.

The optimal choice $\mathscr{S}^*$ of the site-selection subset can then be obtained by solving:

$$\mathscr{S}^* = \arg\max_{\mathscr{S} \in \mathscr{A}} \int \left[\int W(\tau, T^*(\Pi, \hat{\tau}_{\mathscr{S}})) dF(\hat{\tau}_{\mathscr{S}} \mid \mathscr{S}, \tau) \right] d\Pi(\tau). \tag{2}$$

To understand this equation, it may help to consider how to evaluate the right hand side of Equation (2) through simulation. We could generate draws for the site effects τ from the prior $\Pi(\tau)$, reflecting the outer integration. For each draw of τ, we can then simulate the observed site-level estimates $\hat{\tau}_{\mathscr{S}}$ from the sampling distribution $F(\hat{\tau}_{\mathscr{S}} \mid \mathscr{S}, \tau)$ that appears in the inner integral, find the optimal treatment choice T^* based on the observed estimates, and then evaluate the welfare of that choice. Repeating this process and averaging over the draws will approximate the *ex ante* or "preposterior"[10] expected welfare that the policymaker can anticipate for a particular choice of experimental sites $\mathscr{S}$, incorporating both the prior uncertainty about

[10] See Berger (1993), Chapter 7, for an introduction to preposterior Bayesian analysis.

the true τ and the sampling variability of the experimental estimates $\hat{\tau}_{\mathscr{S}}$.[11] [12] To make this computationally feasible, the posterior distribution $\Pi(\tau \mid \hat{\tau}_{\mathscr{S}})$, which appears in the definition of T^* in Equation (1), should be easy to calculate.

While the actual solution to Equation (2) will depend on the specific choice of the prior Π and the welfare function $W(\cdot)$, some intuition is possible even in this general form. Consider including a potential site s' in the set $\mathscr{S}$ of experimental sites. Observing $\hat{\tau}_{s'}$ will typically improve the decision $T_{s'}$ for whether or not to treat that site based on the precision of the estimate $1/\sigma^2_{\epsilon,s'}$. It can also improve the decision for other sites $s \neq s'$ if there is sufficient predicted similarity between τ_s and $\tau_{s'}$, represented as cross-site dependence in the joint prior Π, so that there is learning (posterior updating) about τ_s from $\hat{\tau}_{s'}$. However, such cross-site learning is only relevant for sites s that are "marginal" in the sense that it is not already clear from the prior whether or not the treatment is effective at that site. Taking these considerations together, we expect the solution in (2) to choose a subset of sites that, together, maximize learning about the welfare-relevant aspects of τ_s across all the *a priori* marginal sites.

2.3 Prior Specifications and Updating Rules

Next, we discuss the specification of the prior Π. To solve Equations (1) and (2) above, a key object is the posterior $\Pi(\tau \mid \hat{\tau}_{\mathscr{S}})$, which represents the updated beliefs about the entire vector $\tau = (\tau_1, \ldots, \tau_S)$ based on the observation of a limited set of sites $\mathscr{S}$. If the prior Π is independent across the elements $(\tau_1, \ldots, \tau_S)$, then it will not be updated for any components that are not selected in $\mathscr{S}$. And if we are restricted to choose a small number of sites so that the number of elements in $\mathscr{S}$ is much smaller than S, then the specification of the prior Π will be important for updating beliefs about unobserved sites. This motivates using the observed site characteristics V_s, combined with modeling assumptions about the relationship between V_s and τ_s, to generate a prior with dependence across elements of τ. We will consider a number of priors, but all of them can be expressed as an S-dimensional multivariate normal distribution

$$\Pi(\tau) = N_S\left(\mu_\tau, \Sigma_\tau\right),$$

where the choice of μ_τ and Σ_τ will determine how the posterior extrapolates from observed to unobserved sites.

[11] See Appendix K for a step-by-step walkthrough of the 8-pointed algorithm used in our application.

[12] In principle we can also allow for randomization across site-selection subsets, but given our solution concept and specifications for the priors and payoffs, the optimal choice will be generically nonrandomized.

2.3.1 Smoothing Prior

One class of priors we consider places smoothness restrictions on the relationship between V_s and τ_s, which in turn implies that sites with similar characteristics will be expected to have similar treatment effects. Specifically, let

$$\tau \mid V \sim N_S\left(\mu, \Sigma(V)\right) = \Pi_1,$$

where μ is some $S \times 1$ vector of prior means, and $\Sigma(V)$ is a covariance matrix with elements

$$\mathbb{C}(\tau_s, \tau_{s'}) = c_0 \cdot \exp\left(-c_1^{-1} \left\| h(V_s) - h(V_{s'}) \right\|^2\right), \tag{3}$$

where $h(V_s)$ is some low-dimensional function of the characteristics V_s. This will result in the posterior for τ_s being a weighted average of estimates $\hat{\tau}_{s'}$ based on the distance of $h(V_s)$ to $h(V_{s'})$, similarly to conventional kernel regression estimators. Here $c_0 > 0$ and $c_1 > 0$ are tuning parameters for the prior. c_1 controls the amount of local smoothing, and c_0 controls the overall scale of the variance matrix.[13]

2.3.2 Structural Prior

The smoothing prior above is a standard one in the literature on Gaussian smoothing methods (e.g., Rasmussen and Williams (2006)). However, this conventional approach may be sensitive to the choice of the function $h(V_s)$ and the tuning parameters, especially in settings like ours where the dimension of V_s is high. Moreover, we would like to take full advantage of the rich individual-level data in the pilot RCT, and insights from microeconomic modeling, to incorporate as much prior information as possible to aid in the choice of sites. To do so, we will construct an alternate prior based on a structural model of a migrant's choice of labor, consumption, and remittances. This model is fully detailed in Section 3.

For now, to explain how the structural analysis fits into our design problem, suppose we have an economic model for the underlying outcomes, where the model parameters are given by a vector $\theta \in \Theta$. For every site s, the structural model predicts average treatment effect

$$\tau_s = g(\theta, V_s),$$

and we use $g(\theta, V)$ to denote the S-vector of predicted average treatment effects.

Based on the pilot experiment (in Step 0 of Figure 2), we have a posterior distribution for

[13] Appendix D gives an example of how c_1 governs how local the learning from experimental results is in the context of the applications.

θ, which we will denote $p(\theta)$. The distribution $p(\theta)$ combined with the mapping $g(\theta, V)$ generates a distribution for the vector τ. This could be implemented by drawing $\theta \sim p(\theta)$ and then forming $\tau = g(\theta, V)$.

To maintain tractability, we will assume that the induced prior for τ can be approximated well by a multivariate normal distribution with mean $\tilde{\mu}$ and variance $\tilde{\Sigma}$:

$$\tau \mid V \sim N_S\left(\tilde{\mu}, \tilde{\Sigma}\right) = \Pi_2.$$

2.3.3 Mixed Prior

The structural prior imposes the structural model exactly across all sites, which is likely to be unrealistic. We would like to allow for some discrepancy between the structural model $g(\theta, V)$ and the true effects τ_s. We do so by mixing the smoothing and structural priors. Specifically, for some value $b \in (0,1)$, we take the prior to be normal with mean $\tilde{\mu}$ (given by the structural model) and variance matrix a convex combination[14] of the smoothing and structural variances:

$$\Pi_3 = N_S\left(\tilde{\mu}, b\tilde{\Sigma} + (1-b)\Sigma(V)\right).$$

2.3.4 Posterior Updating

All of the choices for the prior we outlined above are multivariate normal distributions for the vector τ: for some mean vector μ_τ and variance matrix Σ_τ, the priors specify that $\tau \sim N_S(\mu_\tau, \Sigma_\tau)$. Recall that every $\hat{\tau}_s$ is also normally distributed and centered at τ_s. This joint multivariate normality implies that the posterior distribution of the entire vector τ given the observed estimates at the *selected* sites, $\hat{\tau}_{\mathscr{S}}$, will also be multivariate normal. In Appendix C, we provide explicit expressions for these posterior means and variances, and illustrate the posterior updating formulas in a simple example.

2.4 Specification of Welfare Function

To complete the specification of the experimental design problem, we need to specify the welfare function $W(\tau, T)$. The general solution method outlined above can be applied to any reasonable function W. Our analysis below will be based on the following welfare function:

$$W(\tau, T) = \sum_{s=1}^{S} T_s\left(\tau_s - \text{cost}_s\right),$$

[14] An alternative would be to start with the structural prior and add some fraction of the smoothing variance to represent deviations from the structural model, leading to a prior of the form $N_S(\tilde{\mu}, \tilde{\Sigma} + \tilde{b}\Sigma(V))$. More generally, we could consider various linear functions of the two variance components.

where cost_s is the cost associated with implementing the intervention at site s, in the same units as the treatment effect τ_s. We give more detail on how we define treatment effects and specify costs for our application below. Note that, given any predictive distribution for τ_s with mean $\bar{\tau}_s$, the optimal treatment is given by

$$T_s^* = \mathbf{1}(\bar{\tau}_s \geq \texttt{cost}_s). \tag{4}$$

An "oracle" treatment rule that uses knowledge of the true value of τ_s would simply assign treatment if the treatment effect τ_s is greater than or equal to cost_s.

3 A Model of Remittances

Section 2 has proposed a decision-theoretic formulation of the adaptive site-selection problem, and outlined its solution using a tractable class of prior distributions. A key element of our approach is to incorporate structural modeling to link the pilot experiment to the full set of potential future experimental sites. In this section, we develop an economic model of optimal labor supply, consumption, and remittance choices by a migrant whose cost of remitting income to their home household may be affected by the experimental intervention. We will then estimate this model in Section 4 and use it in our experimental design process.

3.1 The Migrant's Problem

Consider a worker who has migrated from his or her home (in a rural area) to a city to work. Some of the earnings from their work will be consumed by the migrant, and the rest sent back to their family for consumption at home. The migrant must choose how much they work, how much of their earnings to consume, and how much to remit back home to their family. Remittances may be sent using traditional means, such as via an agent or a friend who may be returning to the same village as where the migrant's family lives. Alternatively, the migrant can remit using a new method: "mobile money." The cost of sending remittances by mobile money does not vary with the distance between sender and recipient, and may be cheaper than traditional means of remittance. But the migrant may not be aware of the mobile money option, or know how to use it.

Suppose there is now a "treatment" that makes migrants aware of the potential benefits of mobile money and helps them learn how to use it. We wish to study the effects of such interventions on outcomes such as the amount remitted or consumption per capita of the migrant's family members back home.

Let C_m and C_h represent consumption by the migrant and per-capita consumption at home, respectively, and let L_m be the leisure of the migrant (in hours per day). We use w_m to denote the migrant's hourly wage, y_h for total income at home, and R for the remittance sent by the migrant. There are multiple prices: p_h is the price level at home relative to the city where the migrant works; and p_r gives the units of consumption that must be remitted by the migrant in order for their home family to receive one unit of consumption. Finally, let a_h denote the size (number of adults and children) of the migrant's family at home. In this section, we omit an "i" subscript indexing migrants. The i subscript will be introduced later when discussing the data and estimator. We will generally capitalize choice variables, and use lower case for non-choice variables.

The migrant's utility needs to possess a few key properties. We require the migrant to explicitly account for per capita expenditure by their family back home and we need the model to produce a remittance decision which leads the migrant to optimally remit nothing under appropriate conditions on the price to remit, while never remitting their entire income, both of which are features of the data. The utility specification below has these features and also leads to closed form expressions for all optimal quantities, which facilitates computation. The migrant chooses their consumption, leisure, and remittances to maximize utility, given by

$$\ln\left(C_m^{\alpha} L_m^{1-\alpha}\right) + \lambda a_h \ln\left(C_h^{\eta}\right),$$

subject to the following constraints:

$$\begin{aligned} C_m, L_m, R &\geq 0, \\ L_m &\leq 24, \\ C_m &= w_m(24 - L_m) - p_r R, \\ a_h \cdot C_h &= y_h + \frac{R}{p_h}. \end{aligned}$$

Here λ, α, and η are preference parameters. The migrant's utility depends on both their own consumption C_m and leisure L_m, as well as home consumption C_h. The parameter λ, which controls the relative weight placed on home utility vs. migrant's utility, will be specified in more detail below.

There are two key budget constraints in this model. The migrant receives income $w_m(24 - L_m)$ based on wages and hours worked (recall that L_m is leisure of the migrant). This income can be allocated to the migrant's consumption, or sent home as remittance R with price p_r. The price to remit a single unit of consumption, p_r, is assumed to be at least one, i.e. $p_r \geq 1$. This accounts for any transaction fees a migrant may have to incur when remitting. The home

consumption C_h depends on home income, plus any received remittance which has been normalized by the price index p_h, where typically $p_h \leq 1$. In Appendix E.1 we provide the optimal remittance amount R^*, as well as optimal migrant consumption and leisure (C_m^*, L_m^*) given p_r. Below, we develop a stochastic specification for p_r which will lead to a distribution for remittances.

3.2 Stochastic Specifications and Remittance Mode Choice

Here we take the basic model and build an empirical specification, which specifies the distributions of preferences and some other components of the model, and extends the model to include the choice of remittance mode.

The preference parameter λ, which determines the relative weighting of utility derived from migrant and home consumption, is specified as:

$$\lambda = \exp(\phi_0 + \phi_1 \text{male}),$$

where "male" is an indicator for the migrant's gender, discussed as a potential source of heterogeneity in treatment effects on remittance sending in Lee, Morduch, Ravindran, and Shonchoy (2022).

Next we specify the remittance price p_r. We suppose that the migrant can send remittances either through traditional means, or using the mobile money service, and that there is a treatment (the training intervention) $T \in \{0,1\}$, which can affect the non-monetary cost of using mobile money. Let $p_{r,\text{trad}}$ be the price of traditional remittance, and $p_{r,\text{mm}}(t)$ for $t \in \{0,1\}$ be the price of mobile money remittance after intervention t. For the price of traditional remittance, we specify that

$$p_{r,\text{trad}} = 1 + d\epsilon,$$

where d is the distance between the migrant and home locations, and $\epsilon \geq 0$ is a individual-specific stochastic shock.[15] Thus, the cost of remitting one unit is equal to one (the remittance itself) plus a positive amount $d\epsilon$ that increases with distance, capturing the increased risk of loss or direct cost of sending money a long way using traditional means. The shock ϵ reflects individual heterogeneity in the cost of remitting by traditional means with a low shock arising, for example, if the migrant can send money with a trusted friend who happens to be visiting home. We assume that ϵ follows an exponential distribution with mean $\bar{\epsilon}$. This ensures that ϵ is nonnegative in accordance with its interpretation as the increase in remittance cost with an

[15] When estimating the model as described in Section 4, we set $d = 1$ but account for other distances in the design stage.

additional unit of distance.

For the price of mobile money remittance, we specify that

$$p_{r,\text{mm}}(t) = 1 + \gamma + \exp(\delta o + \psi t)\xi.$$

Here, $\gamma \geq 0$ is the known monetary per-currency-unit-sent cost of remitting using mobile money. We describe the source of γ in each country in Appendix J.1. In addition to the formal price of mobile money, we assume there is a positive hassle cost represented by the term $\exp(\delta o + \psi t)\xi$. The price of remittance by mobile money depends on operator density o, which captures the ease with which mobile money transactions may be implemented. Use of mobile money to send remittances requires a physical agent at origin to withdraw funds. The greater the availability of such agents or operators, the easier it is to remit using a mobile money service. The parameter δ captures the effect of greater operator density on the price to remit using mobile money. The term ψ measures the change in the effective cost of mobile money remittance if the migrant receives the training intervention. Finally, we specify an individual-specific stochastic term ξ that generates variation in the hassle cost of mobile money remittance. By analogy with the specification for ϵ, we assume that ξ is exponentially distributed with mean $\bar{\xi}$, and that ξ is independent of ϵ.[16] The stochastic terms ϵ and ξ generate variation across individuals in their choice of remittance mode, as well as amount optimally remitted.

The migrant will choose whichever remittance mode is cheapest, so the effective price of remittance will be

$$\min\{p_{r,\text{trad}}, p_{r,\text{mm}}(t)\}.$$

Letting M denote the remittance mode choice, with $M = 1$ for mobile money and $M = 0$ for traditional remittance, we have $M^* = \mathbf{1}\{p_{r,\text{trad}} \geq p_{r,\text{mm}}(t)\}$.

After augmenting the basic model with the remittance mode choice, and specifying the stochastic determinants of preferences and remittance prices, the complete structural model implies a joint distribution for remittance amount R and remittance mode choice M given household/migrant characteristics (and given model parameters).[17] This joint distribution will be the basis for the estimation routine described in Section 4 when we use data from a randomized trial to fit the model described here.

[16] When we estimate the model parameters, $\bar{\xi}$ will not be separately identifiable from δ and ψ. In the experimental data, there is no variation in the operator density o, so we drop the term (δo) for estimation purposes, and then calibrate the value of δ as described in more detail below.

[17] We provide the specific expressions in Appendices E.2 and E.3.

4 Fitting the Structural Model to the Pilot Experiment

Lee et al. (2021) conducted an experiment in Bangladesh in 2014–2016 to evaluate an intervention that trained migrant workers in Dhaka and their family members at home to use a digital money transfer service. The study tracked 813 migrant-home family pairs whose home families were located in the district of Gaibandha and measured multiple outcomes of interest for both the migrants and their family members. The families chosen for the study were all "ultra poor" as defined by local administrative standards, and regularly received remittances from migrant family members in Dhaka. The experiment randomly assigned half the sample to a treatment arm that was given detailed training on how to use the mobile money application bKash to remit money. Members of the control arm were not given any training.

We use data from this experiment to fit the model from Section 3, as the first step in defining a Bayesian prior for the experimental design process described in Section 2. Since we do not model migrant unemployment, we drop the 34 households with migrants who report no earnings. We also drop the 116 households reporting no income at home because we believe these are in fact misreports. A large majority of these households have high consumption levels that greatly exceed the amount of remittances they receive, but do not report any debt. The loss in sample size is reflected in the uncertainty about the structural parameters as well as the implied site-level treatment effects. Appendix Table A1 displays summary statistics for the sample used for model estimation.

To take the model developed in Section 3 to the prior experimental data, we assume that the households $i = 1, \ldots, n$ in the experimental data are a random sample from the relevant population at that site. For each household we observe characteristics

$$X_i = (p_{hi}, a_{hi}, w_{mi}, y_{hi}, \text{male}_{mi}, o_i, d_i, T_i),$$

and we observe the household's amount remitted, and mode choice conditional on positive remittance:

$$Y_i = (R_i, M_i \cdot \mathbf{1}(R_i > 0)).$$

The full set of model parameters is $\{\psi, \phi_0, \phi_1, \bar{\epsilon}, \bar{\xi}, \alpha, \gamma, \eta, \delta\}$. While most of these parameters will be estimated by making explicit use of the model, as described below, γ, η, and δ will be normalized or estimated separately as described in Appendix G. We collect the remaining parameters into a vector θ:

$$\theta = \left(\psi, \phi_0, \phi_1, \bar{\epsilon}, \bar{\xi}, \alpha\right)',$$

and we will suppress the dependence of model-implied quantities on the other parameters for notational simplicity in what follows.

For the experimental site, there is a distribution of the covariates F_X, i.e. $X_i \overset{\text{iid}}{\sim} F_X$. The distribution F_X will in turn induce a distribution of outcomes Y_i through the structural model $F_{Y|X}(Y_i|X_i,\theta)$.[18] At the other sites, the distribution of covariates will differ from F_X, and this will lead to variation in the predicted site-level average treatment effects generated by the structural model.

4.1 Minimum Distance Estimator

We estimate the parameter vector θ by matching a vector of model-implied quantities $q(\theta)$ to corresponding sample conditional moments $\hat{\pi}$. First, we consider the conditional distributions of remittances R and remittance mode M. We fix a set of remittance values

$$\mathscr{R} = \{r_1, r_2, \ldots, r_{d_r}\} = \{0, 50, 100, 125, 150, 175, 200\}.$$

For each combination of (male, T) $= (m,t) \in \{0,1\}^2$ and $r \in \mathscr{R}$ we consider the following probabilities:

$$\mathbb{P}(R \leq r \mid \text{male} = m, T = t), \quad \mathbb{P}(M = 1, R > 0 \mid \text{male} = m, T = t).$$

These probabilities can be estimated directly by their sample analogs and compared to the values derived from the structural model evaluated at a parameter vector θ. The Lee et al. (2021) experiment asked migrants for the mode and amount of remittances sent over multiple months between the baseline and endline surveys. We first aggregate these up to the household level. The variable R_i is therefore the average monthly remittances sent by a migrant over the course of the survey, and the mobile money usage variable is aggregated to a probability of remitting via mobile money (bKash) conditional on remitting a positive amount.

We also include a component of $\hat{\pi}$ and $q(\theta)$ to estimate the parameter α. Solving the migrant's decision problem yields the following optimality condition:

$$\frac{\alpha}{1-\alpha} = \frac{C_m}{L_m \cdot w_m}.$$

We observe expenditure by a migrant E_m and consider it to be a noisy observation for consumption C_m as in:

$$E_m = C_m + \tilde{\epsilon}, \text{ where } \mathbb{E}[\tilde{\epsilon}] = 0.$$

Substituting E_m in the optimality condition and rearranging terms results in the following ex-

[18] We will also sometimes suppress the dependence of quantities on F_X in the expressions below for notational convenience.

pression for the parameter α:

$$\alpha = \frac{\mathbb{E}[E_m]}{\mathbb{E}[E_m] + \mathbb{E}[L_m \cdot w_m]} \tag{5}$$

We set one component of $q(\theta)$ as $q_j(\theta) = \alpha$, and the corresponding π_j equal to the sample analog of the expression above.

We estimate the parameter vector θ by classical minimum distance, following Newey and McFadden (1994).[19] The estimator solves

$$\min_{\theta} \left(\hat{\pi} - q(\theta)\right)' W_n \left(\hat{\pi} - q(\theta)\right),$$

where W_n is a weighting matrix with $W_n \xrightarrow{p} W$ for some W. In our application we set W_n equal to the identity matrix. Under standard regularity conditions, the estimator $\hat{\theta}$ will converge in probability to a pseudo-true value θ_0.[20] Its asymptotic distribution is given by

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta),$$

where the asymptotic variance matrix Σ_θ can be consistently estimated. Appendix F provides further details on the estimation method.

Table 1 displays estimated parameters, with standard errors in parentheses. Appendix H details the model fit by comparing treatment effects on the amount of daily remittances sent and the probability of remitting via mobile money, as well comparing the observed empirical distribution of remittances in the experimental data with the model generated distributions.

Table 1: Model Parameter Estimates

ψ	ϕ_0	ϕ_1	$\bar{\epsilon}$	$\bar{\xi}$	α
-0.5116	-2.2641	0.1012	0.2522	0.1341	0.3176
(0.2369)	(0.0772)	(0.042)	(0.2478)	(0.1618)	(0.0048)

Notes: standard errors in parentheses.

[19] In our application we have fully specified the conditional likelihood function for outcomes given covariates and parameters, so it would be possible to instead estimate θ by conditional maximum likelihood or calculate its full posterior distribution through Bayesian estimation. Under model misspecification, the probability limits of such likelihood-based estimates could differ from those of the minimum distance estimator. We take a minimum distance approach to provide a template for future applications adopting our approach where likelihood-based estimation is not possible.

[20] If the structural model does not hold exactly, as we generally take to be the case in our analysis, the pseudo-true parameter value θ_0 depends on the choice of the limiting weighting matrix W. We suppress this dependence in the notation.

The negative estimated value for ψ shows that the treatment effectively reduces the hassle cost of remitting via mobile money. The negative estimate for ϕ_0 implies that migrants place a low weight on per-capita consumption across home household members relative to their own index of consumption and leisure. Specifically, the value $\hat{\phi}_0 = -2.26$ implies that female migrants put a weight of 0.905 on their own consumption-leisure index and the rest on home consumption. Our estimate of ϕ_1 has male migrants putting a weight of 0.897 on their consumption-leisure index. Based on Lee et al. (2022), we conjecture that this could be a result of migrant women having less bargaining power relative to their male co-migrants so that they are less able to remit to family members at home. Our estimate of $\hat{\alpha} \approx 1/3$ is in line with typical values for the consumption weight in consumption-leisure indices, e.g. Kydland and Prescott (1982).

4.2 Predicted Site-Level Effects Based on the Structural Model

The structural model can be used to predict the average effect of the intervention for every prospective site. In conjunction with a posterior distribution for the structural model parameters θ obtained from the pilot RCT data, we can thereby generate a joint prior for the vector of site-level effects $\tau = (\tau_1, \ldots, \tau_S)$. In our application, we face some additional computational and data limitations. To explain our approach, we first outline an idealized version of our simulation algorithm that abstracts from some of these complications, and then discuss the practical implementation that addresses the remaining issues.

Let $p(\theta)$ denote the posterior distribution of the structural parameters, based on fitting the structural model to the pilot bKash experimental data as discussed earlier in this section. To generate draws for the average effect τ_s at a site s, we need the values of the site-level characteristics (o, d, p_h). We also require the joint distribution of the characteristics $(w_{mi}, a_{hi}, y_{hi}, \text{male}_i)$ of households in corridor s. To simplify the notation, let $F_{X,s}$ denote the joint distribution of the site- and individual-level covariates at site s.[21]

We can generate draws for τ_s as follows. First, take a draw for the structural parameter: $\theta \sim p(\theta)$. Then, generate many draws $b = 1, \ldots, B$ for the site- and individual-level covariates: $X_{s,b} \sim F_{X,s}$, and also generate B draws for the model shocks ϵ and ξ from their exponential distributions, with the mean of ϵ depending on the draw for θ. ξ has mean one; the factor $e^{\delta o + \psi t}$ that multiplies it in the mobile money price carries the scale, with δ calibrated as described in Appendix G. Then, for each simulated individual b, we can generate their counterfactual optimal remittances and mode choice under both treatment and control, by setting $t = 0$ and $t = 1$ and using the formulas in Appendix E.

[21] This joint distribution will be degenerate for the site-level characteristics, putting probability one on their known values.

This results in a large number of draws $R_{s,b}(t), M_{s,b}(t), b = 1, \ldots, B$, for the outcome variables under both treatment and control at site s. We can then calculate the (approximate) value for τ_s. For example, if we are interested in the average effect of the treatment on remittances, we can set

$$\tau_s = \frac{1}{B} \sum_{b=1}^{B} \left[R_{s,b}(1) - R_{s,b}(0) \right].$$

(The actual definition of the treatment effect we work with in Section 5.1.2 is slightly more complicated, but the same idea can be applied to any notion of a treatment effect for which the structural model generates counterfactuals.) This generates one draw for τ_s, and repeating the entire process many times will generate a large number of draws for τ_s from the structural prior.

The full algorithm for simulating the structural prior extends this intuition in a number of ways. First, our minimum-distance estimation strategy described in Sections 4.1-4.1 does not provide the full Bayesian posterior distribution for θ. We instead adopt a limited-information perspective and interpret the minimum distance estimation strategy as providing an approximate or quasi-posterior for θ given by

$$\theta \sim N\left(\hat{\theta}, \widehat{\Sigma}_\theta / n, \right),$$

where $\hat{\theta}$ is the minimum distance estimate and $\widehat{\Sigma}_\theta$ is the estimated asymptotic variance-covariance matrix of its sampling distribution. In order to avoid the possibility of obtaining negative draws for $\bar{\epsilon}$ and $\bar{\xi}$, we transform θ by taking $\bar{\epsilon} \mapsto \ln(\bar{\epsilon})$ and $\bar{\xi} \mapsto \ln(\bar{\xi})$, leaving the remaining components of θ unchanged, using the Delta method to obtain the implied distribution.

Second, we need to sample jointly for *all* the site-level effects $\tau_1, \ldots, \tau_S$. For each draw of θ, we employ the same method to generate the implied value of τ_s across all sites $s = 1, \ldots, S$. When doing this, we can re-use the model shock draws ϵ_b, ξ_b across the sites to reduce simulation chatter.

Third, we face a significant data limitation that prevents us from directly estimating $F_{X,s}$, the joint distribution of individual (and site-level) characteristics by site. Ideally, we would have a data set that samples w_{mi}, a_{hi}, y_{hi}, and male$_i$ for migrants in each prospective site s. Unfortunately, there are no datasets representative at the district level which provide comprehensive data on households in rural areas linked to migrant members. We instead draw on representative data sets—which we will hereafter refer to as administrative data—to construct site-level marginal summary statistics, largely for either home households or migrants. In these data, we only regard the marginal distribution of gender among migrants, the marginal sample means and variances of migrant wage, home household size, and home household income conditional on migrant gender as being of sufficiently high quality for our purposes. We therefore need to make additional assumptions about the dependence structure of the individual-level variables

in order to specify their joint distribution.

To specify the joint distribution of individual characteristics by site, we assume that for each site, the vector $(\ln(w_{mi}), \ln(a_{hi}), \ln(y_{hi}))$ is jointly normally distributed conditional on migrant gender. We can estimate the means and the marginal variances of this distribution from the administrative data. To specify the conditional covariance terms, we estimate the correlation matrix for the same (logged) covariates from the pilot experimental data in Lee et al. (2021), separately by gender, and assume that the same correlation matrix holds in every prospective site s. This fully specifies $F_{X,s}$ through the probability of the migrant's gender, and the joint distribution of (w_{mi}, a_{hi}, y_{hi}) conditional on migrant's gender.

Table 2 demonstrates the results of our procedure for working with the administrative data. It shows the model implied treatment effect on daily remittances sent by migrants when using the full data from Lee et al. (2021) and compares it to the model-predicted value when restricted to only using the same summary statistics we calculate in the administrative data. The results are encouraging, with the model-implied treatment effect based on the summary statistics differing only by 8% from the one derived using the full microdata.

Table 2: Average Treatment Effects in Model and Data

Outcome	Model	Data
Remittance per Day (Lee et al. (2021) Microdata)	5.94	6.41
Remittance per Day (Summary Statistics Only)	6.44	6.41

Notes: amounts are in Bangladesh taka (1 2015 USD = 78 BDT). The Model column gives the model-implied average treatment effect when inputting the full experimental microdata (row 1) and only summary statistics as described in Section 4.2 (row 2). The Data column gives the unrestricted average treatment computed in the data.

5 Application: Designing Experiments in South Asia

The framework allows us to determine how many experiments to run and where. The choices take into account how the sites vary, prior information on economic relationships, and experimental budget constraints. The framework also allows for comparisons of counterfactual choices. How would policy decisions change—and with what consequences for expected welfare—if alternative sites were chosen? In parallel we can ask how expected welfare would change if policymakers do not follow the notions of optimal learning embedded in the Bayesian decision theory framework. At an extreme, what if policymakers base decisions on treatment effects estimated in experiments in other sites without adjusting for how the experimental sites compare to their own sites?

We apply these ideas to decisions about new randomized trials in Bangladesh, India, and Pakistan. Section 5.1 takes the approach described in Sections 2, 3, and 4 and supplements it with implementation choices needed to apply the framework in practice. Section 5.2 describes our results on site selection, starting with the most common choice experimenters face: where to experiment if you can only run one experiment.

We compare optimally-selected sites (in terms of expected welfare) to sites chosen in other ways: choosing sites with the largest expected treatment effects and, contrastingly, choosing them randomly. The comparisons show consistent welfare gains to optimal site selection. The framework allows us to additionally compare different modes of extrapolating a finding from one site to another, quantifying the benefits of Bayesian learning. Section 5.3 turns to choosing the optimal number of experimental sites.

5.1 Implementing the Site Selection Framework

In Section 5.1.1, we describe migration corridors and the initial steps we take to narrow the set of migration corridors for the analyses. Section 5.1.2 describes the social welfare function and the corresponding treatment effects of welfare interest. Using the social welfare function, we translate costs into welfare terms and describe the site-level data and prior distributions we use. Section 5.1.3 describes how we implement the priors and how they differ in updating predictions for non-experimental sites.

5.1.1 Defining and Narrowing the Set of Migration Corridors

The experimental intervention in the new sites will again be aimed at reducing the costs a migrant may face when remitting to their family in distant home districts. Migrants move to work in a "destination" district, having started in an "origin" district, where their parents and other family members usually still live. A migration corridor (what we have been calling a site) is defined as a paired destination district and origin district. We give site names in destination-origin format throughout the paper; e.g., Dhaka–Noakhali in Bangladesh. Districts are referred to by the name "district" in India and Pakistan, while Bangladesh refers to the same unit as a *zila*. For convenience we will refer to all such spatial units as districts in what follows.

While data is in principle available for all combinations of origin and destination districts in each country (Appendix J.1 discusses the construction of our dataset of site characteristics), we focus on a somewhat smaller subset. We have two main objectives in doing so. The first is to build in robustness to prior misspecification through geographic diversity: we require that the two corridors have origins in different states (India), provinces (Pakistan), and divisions (Bangladesh). The second concerns the practicality of survey and program administration: ori-

gin locations must be sufficiently dense in households with migrants to make it feasible to sample in that location, and we can only survey in areas where our potential enumeration partners work. We view potential program administrators as being subject to the same constraints as our enumeration partners. Program administrators, too, must have a sufficiently dense pool of migrants to be able to fill training sessions. Appendix J.2 provides details on the migrant density restrictions we use, and Appendix Table A2 lists summary statistics on the site characteristics entering the structural model for sites matching our criteria across the three countries.

5.1.2 Specifying Social Welfare

The household-level outcome of interest in this application is the additive log social welfare associated with each of its home members' utility,[22]

$$\sum_{j \in HH_i} \ln u_j.$$

As in Deaton (1997), we assume the utility attained by an individual j in home household i is household per-capita consumption which, after accounting for remittances, is

$$u_j = \frac{1}{a_{h,i}} \left(y_{h,i} + \frac{R_i}{p_h} \right).$$

Since all members of the household have the same per-capita consumption, the outcome of interest may be written as:

$$Y_i(t) = a_{h,i} \ln \left[\frac{1}{a_{h,i}} \left(y_{h,i} + \frac{R_i^t}{p_h} \right) \right],$$

where the dependence on the treatment status $t \in \{0,1\}$ is made explicit in the notation R_i^t.

For site s, the average treatment effect in welfare terms equals

$$\tau_s = \mathbb{E}_\theta \left[Y_i(1) - Y_i(0) \right],$$

where the θ subscript indicates the dependence of the distribution of remittances on the model parameters, as described in Section 3.2. The expectations operator above also aggregates over the joint distribution of $(a_{h,i}, y_{h,i}, w_{m,i})$ at site s, as detailed in Section 4.2. τ_s approximates the percentage change in consumption across the average migrant's family members in site s due

[22] This is a special case of additive CES social welfare (c.f. Deaton 1997, Alderman et al. 2019). We focus on log welfare for transparency and compatibility with the log per-capita consumption outcomes reported in Lee et al. (2021), but could consider alternative parameterizations. This is isomorphic to specifying j's utility from consumption as log with utilitarian social welfare. We abstract from the migrant's utility gain from a decrease in the effective price of remitting via mobile money decreasing due to the training, but this could easily be incorporated.

to the training program.

One last point to handle is that household size $a_{h,i}$ can be affected by the treatment. In Lee et al. (2021) household size declines slightly with treatment as some additional household members migrate. We consider these members to benefit from the increase in per-capita consumption at home, so we measure the treatment effect in a site s as

$$\mathbb{E}_\theta\left[a_{h,0i}^{\text{baseline}} \cdot \ln\left(\frac{\text{home expenditure}_{1i}^{\text{endline}}}{a_{h,1i}^{\text{endline}}}\right)\right] - \mathbb{E}_\theta\left[a_{h,0i}^{\text{baseline}} \cdot \ln\left(\frac{\text{home expenditure}_{0i}^{\text{endline}}}{a_{h,0i}^{\text{endline}}}\right)\right],$$

where the 0 or 1 preceding the household index i denotes treatment.

A site will be recommended for treatment in the policymaking stage if the perceived welfare benefit of treatment there (τ_s) exceeds the cost. We translate the per-migrant cost of the program, given in Table 3 of Lee et al. (2021) as 885.84 taka, into welfare terms by supposing the cost is homogeneous across sites within a country and the program is financed by a uniform consumption tax in the country. We provide the details of our calculation in Appendix I, which yields welfare-unit treatment costs of 0.183, 0.145, and 0.153 for Bangladesh, Pakistan, and India, respectively.

5.1.3 Describing the Priors

Mixed prior Our preferred prior to provide the extrapolation benefits of the structural model and the robustness of the smoothing prior is the mixed prior described in Section 2.3, with equal weight placed on the structural prior and a smoothing prior. The mean parameter for the prior is the vector of average treatment effects generated by the structural model. The covariance matrix for the smoothing prior is constructed as in Section 2.3.1 using two site level characteristics: the distance between destination and origin; and home household income. These variables are important both in our fully-specified model and according to general economic reasoning. The benefits of constant-price mobile money relative to traditional methods like migrant travel are increasing in the distance between the destination and origin points of a corridor, and the marginal value to the migrant of income sent home is higher when the home household is poorer. We standardize both variables to have zero mean and unit standard deviation.

Migrant wage at destination would also seem important but does not vary much in Bangladesh since there are so few destinations among high-migrant-density corridors. Furthermore, including more variables in the smoothing prior increases the sensitivity of site selections to its tuning parameters, specifically the measure of distance between site characteristics.[23] This is

[23] In Pakistan and India there are more destinations and wage at destination varies more across corridors (see

a key difference with the structural prior, whose parametric structure specifies the relevant notion of distance explicitly.

We set $h(\cdot)$ in the definition of the smoothing prior covariance matrix in Section 2.3.1 to the identity function and use Euclidean distance as the norm. We set c_1 to 1. The characteristics are standardized so distance in characteristic space is measured in cross-site standard deviations and c_1 in squared standard deviations; $c_1 = 1$ implies that two sites whose characteristics differ by one cross-site standard deviation, for example in destination–origin distance alone, have correlation $\exp(-1) \approx 0.37$ under the smoothing prior. The middle panel of Appendix Figure A7 shows the updating weights this value implies.[24] We choose c_0 such that the scale of covariance values implied by the smoothing prior is similar to those of the structural prior. In particular, c_0 is chosen to approximately minimize the maximum absolute difference in variance values as defined by the structural and smoothing priors.[25]

While most empirical results shown below are based on this mixed prior, we also consider other priors with different weights on the structural model.

Structural prior The pure structural prior we consider is a mixed prior with zero weight on the smoothing prior.

Smoothing prior We also consider a pure smoothing prior. We intend it to represent an off-the-shelf application of the ideas behind the smoothing prior, showing an alternative that does not involve undertaking the construction of a structural model. Instead of setting the prior mean to the structural-model-predicted vector of site treatment effects the way we do in all mixed priors, we set the mean parameter in the pure smoothing prior to be constant across all sites and equal to the estimated treatment effect (in welfare terms) in the pilot bKash experiment of Lee et al. (2021). As in the smoothing component of the mixed prior, we construct the covariance matrix for the pure smoothing prior using destination–origin distance and home income, but we choose the parameter c_0 governing the overall prior scale so that the site-level variances match the sampling variance for the estimated treated effect in the pilot experiment. This gives us $c_0 \approx 0.13$.

Table A2). We keep the number of characteristics at two for consistency with Bangladesh and because of the aforementioned concerns about dimensionality in the smoothing prior.

[24] Appendix D.2 discusses how c_1 can alternatively be identified from expert forecasts.

[25] Following this approach, we set $c_0 = 0.1$ for Pakistan, while for India we use $c_0 = 0.073$. In Bangladesh, data limitations require us to use predicted values from a wage regression to approximate migrant wage at destination, as described in Appendix J.1. This makes mean migrant wages much more homogeneous in Bangladesh than in the other two countries (see Table A2), which results in a structural prior with understated uncertainty. Therefore we set c_0 in Bangladesh to be the same as in the pure smoothing prior.

Distribution of $\hat{\tau}$ given τ Recall that the distribution of $\hat{\tau}$ conditional on the true vector of site-level treatment effects τ is

$$\hat{\tau} \mid \tau \sim N(\tau, \Sigma_\epsilon).$$

In our implementation, we set $\Sigma_\epsilon = \text{diag}\{\sigma_\epsilon^2, \ldots, \sigma_\epsilon^2\}$ with the sampling error σ_ϵ^2 defined as

$$\sigma_\epsilon^2 = \frac{\hat{\sigma}^2_{\text{BKash}}}{n_{\text{exp}}},$$

where $\hat{\sigma}^2_{\text{BKash}}$ is the estimated variance of the welfare treatment effect from the Lee et al. (2021) data and n_{exp} is the number of households sampled in each corridor for the new experiment. The value of n_{exp} depends on the number of corridors selected for experimentation, as described in Appendix M.2.

Comparing the priors As discussed above, the parametric model underlying the structural prior allows experimental results in a given site to be informative for treatment effects in sites with quite different characteristics. We can visualize this by examining how the mean predicted treatment effect in other sites changes as we change the value of the treatment effect in a reference site. This amounts to plotting $\mathbb{E}[\tau_s|\tau_{s'}]$ (where expectations are taken under the prior) as a function of $\tau_{s'}$. Figure 3 displays such curves for all sites s under the structural prior, with the treatment effect for the reference site Dhaka–Noakhali on the horizontal axis. Dhaka–Noakhali is a migration corridor between Noakhali on the Bay of Bengal and Dhaka, the capital of Bangladesh and the main location of garment factories.[26] Most of the lines have substantial positive slopes, meaning that a large treatment effect in Dhaka–Noakhali is expected to be associated with a large and positive predicted treatment effect in most other sites. Figure 4 plots the same relationships under the smoothing prior. Here the predicted treatment effect in Dhaka–Noakhali is strongly positively correlated with the predicted effects in only a few other sites, specifically only those sites with similar average home household income and destination–origin distance. The horizontal axes of the two plots, which are set by the range of 1,000 draws from each prior, are also of note. It is evident that the structural prior puts very little probability on Dhaka–Noakhali having a negative welfare ATE, while the smoothing prior remains agnostic. We provide analogous plots for Pakistan, yielding the same qualitative conclusions, in Appendix B.1.

[26] Patterns are similar regardless of the choice of reference site.

Figure 3: Expected Welfare TE (Structural Prior)

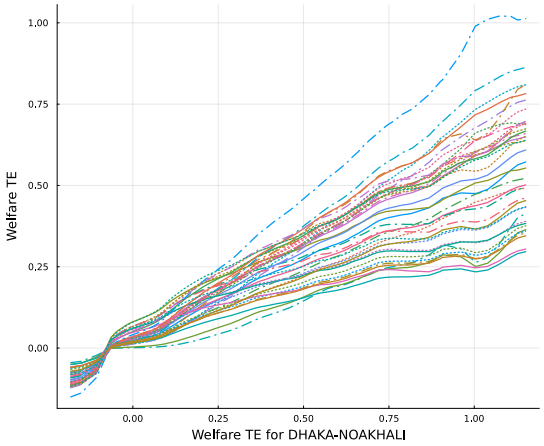
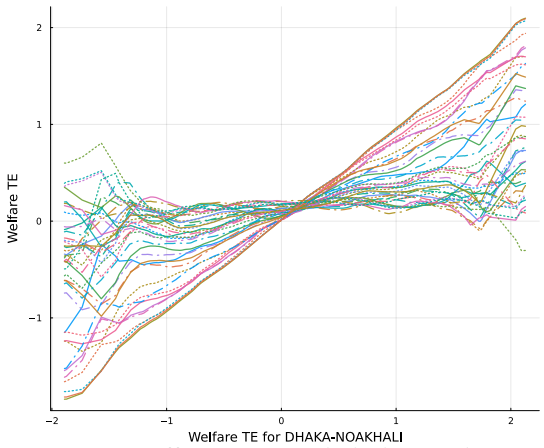


Figure 4: Expected Welfare TE (Smoothing Prior)



Notes: Each line in the figures shows the expected welfare treatment effect in a Bangladesh corridor conditional on knowing the welfare treatment effect in the Dhaka-Noakhali corridor is equal to the value on the X-axis.

5.2 Selecting Experimental Sites

In Section 5.2.1, we consider the choice of a single site that optimizes external validity given the policy context. In Section 5.2.2, we describe the nature of the migration corridors selected when it is possible to run two experiments in each country, which is the case we can actually take to the field in South Asia.

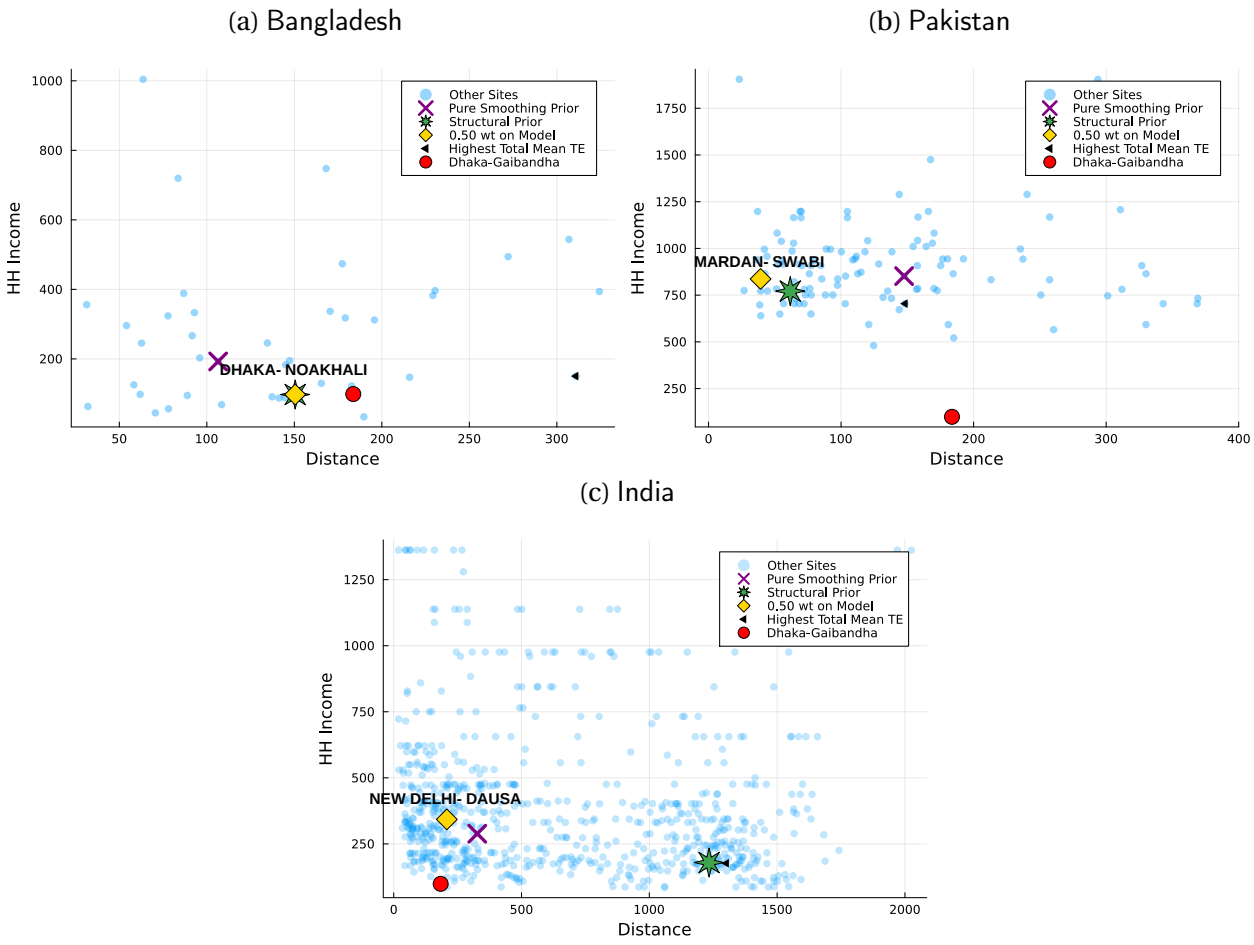
5.2.1 Choosing a Single Experimental Site

We first consider the problem of choosing a single experimental site in each of the three countries, drawing on the pilot experiment to construct the prior distributions. The site for the original experiment in Bangladesh (the Dhaka–Gaibandha migration corridor) was chosen in part because it was a promising location for the intervention. Lee et al. (2021) find relatively large treatment effects, which likely reflects particular qualities of the Dhaka–Gaibandha corridor. First, the distance between Gaibandha and Dhaka makes it costly to travel between the two sites, and the ability to send money digitally, with a price that does not depend on distance, is thus particularly appealing. Second, the high poverty rate in Gaibandha meant that getting extra remittances could be expected to make a larger marginal impact than in better-off districts. Third, the infrastructure for mobile money was already in place, led by bKash, the largest provider. And, fourth, barriers existed to the adoption of bKash that could be reduced through a basic training intervention.

The question after the experiment was whether the results in Lee et al. (2021) would generalize to other contexts. Panel (a) of Figure 5 shows other migration corridors in Bangladesh. The

red circle which locates the Dhaka–Gaibandha corridor is on the periphery of the main cluster of sites in the figure. The horizontal axis gives the distance between the two ends of each corridor, and Dhaka–Gaibandha is among the longer corridors. The vertical axis is average income among households in the home district who report having sent an internal migrant, and Gaibandha is on the lower end compared to others.

Figure 5: Optimal Experimental Sites Under Alternative Decision Rules, One Site Per Country



Notes: Light dots represent candidate sites. HH Income refers to the average income in 2015 Bangladesh taka (1 2015 USD = 78 BDT), excluding remittances, of a household at migration corridor origin reporting having sent an internal migrant. Distance is km between corridor origin and destination as the crow flies. See Appendix J.1 for details. Site names are in destination-origin format and refer to the sites selected under our preferred mixed prior with 50% weight on the structural model.

The black triangle in Panel (a) indicates the corridor where the highest total mean treatment effect from the intervention is expected, according to the mixed prior. It is even more peripheral than the original site in Gaibandha. The site marked by the black triangle is a relatively poor district like Gaibandha, but it is twice as far from Dhaka. Choosing a site for additional experimentation which is similar to Dhaka–Gaibandha—or similar to the site where the treatment

effect is expected to be largest—would lead to choices that are unlike most other migration corridors in Bangladesh.

The single site in Bangladesh chosen using the pure smoothing prior depends on the distribution across sites of the average income of migrants' home households and the length of the corridor, the two axes of the figure. The purple X in the figure shows the corridor chosen with the pure smoothing prior and it is, by design, in the middle of the other corridors. The chosen site is far more typical than the Dhaka–Gaibandha corridor; the distance is shorter, and on average the population is better off economically.[27]

The smoothing prior narrows attention to the two dimensions of the figure, but other dimensions can matter as well (as described in Section 2.3). The microeconomic model in Section 3, reflected in the pure structural prior, incorporates the migrants' choices of labor, consumption, and remittances to predict the vector of average treatment effects across sites. The richer information yields a choice of site that is closer to the original choice of the Dhaka–Gaibandha corridor. The optimal choice, marked by the green 8-pointed star, is the Dhaka–Noakhali corridor, which is more distant from Dhaka and poorer than the site chosen with the pure smoothing prior. The mixed prior, which puts equal weight on both the smoothing and the structural prior (and which is the preferred specification), also selects the Dhaka–Noakhali corridor (shown by the yellow diamond).

Panels (b) and (c) of Figure 5 show optimal single experimental site selections in Pakistan and India. In both contexts, choosing migration corridors for experimentation with distances and home incomes as close as possible to the Dhaka–Ghaibandha corridor, as might be dictated by the logic of pure replication, would again lead to peripheral choices. So would choosing the corridor where the total mean treatment effects are expected to be highest. Instead, our approach again optimally selects sites that are similar to other sites; when just basing choices on the pure smoothing prior, the sites are centrally located in terms of home household income and corridor distance. The sites selected when incorporating the structural prior again diverge from those chosen just with the smoothing prior, reflecting the additional dimensions of comparison. But, with either prior, the figures show that choosing a single site favors selection of locations that are generally central among candidate sites, not necessarily geographically but in terms of relevant characteristics.

5.2.2 Choosing Two Sites

Panel (c) of Figure 5 points to a limitation of conducting only a single experiment. Sites in India are dispersed widely, with two main clusters of points. By the logic underlying the smoothing

[27] We lay out in detail the simulation algorithm underlying these choices and those in the rest of the section in Appendix K.

prior, a point in the center of one cluster may then not generalize easily to the other cluster. The observation motivates the next questions: When is it valuable to experiment in more than one site? Where should sites be located when experimenting in two or more sites, and how much more informative is it for policymakers?

The questions are not hypothetical: the funder of the research in South Asia in fact provided funding for experiments in two sites per country with a sample of 2,000 households in each site. Here, we analyze the problem of selecting two migration corridors in each country, then in 5.3.2 we quantify the welfare gains from experimenting in various numbers of sites weighed against the added costs.[28]

Impact of Optimal Site Selection Table 3 shows the range of welfare values possible under different choices of experimental sites in each country, with average welfare per migrant at the site level evaluated under our preferred mixed prior. The impacts of optimal site selection emerge as quite substantial in Bangladesh, with the worst site combination leading to a maximized welfare which is about 13% of the welfare in the best site combination. In both Pakistan and India, the ratio of welfare achieved by choosing the best sites compared to the worst is smaller than in Bangladesh. We return to this point in Section 5.2.3.

Table 3: Average Welfare from Best and Worst Site Combinations by Country

Country	Highest Average Welfare	Lowest Average Welfare
Bangladesh	0.030	0.002
Pakistan	0.111	0.103
India	0.209	0.197

Notes: Average Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

Appendix Table A3 goes into more specific detail, showing the five best and worst migration corridor pairs in each all three country. In Pakistan three of the worst combinations involve migration from the district containing the capital, Islamabad. In India again corridors involving origin districts with large cities appear in many of the worst site combinations. These corridors do have high migrant densities but the home households are outliers—they have the highest incomes among candidate sites.

[28] Throughout our discussion of site selections in this application, we choose sites from each country independently to ease computational costs. To ensure we are not leaving gains from joint site selection on the table, we select sites for Bangladesh and Pakistan jointly and find that the top site combinations from selecting sites independently achieve 97% of the welfare of the top combinations from joint selection.

Appendix Figures A3 – A5 show where the best and worst site combinations for each country lie in the space of characteristics considered by the smoothing prior. The best choices for the new experiment lie within clusters and the worst choices are always outliers in the feature space. We will see that this is due to the influence of the smoothing prior, which, roughly speaking, tends to prefer combinations of sites such that all candidate corridors are close to at least one experimental site.[29] Under the smoothing prior conducting an experiment on a migration corridor that is very different from most other sites will yield little information about how the treatment might impact other sites. [30]

Top Site Choices by Prior We now investigate how the choice of top site combinations varies with the prior. Figure 6 shows the best destination–origin pairs under the different priors discussed in the previous subsection. We again mark pairs of corridors selected under the pure smoothing prior, the pure structural prior (mixed prior with 0 weight on smoothing), and mixed priors with differing weights including our preferred 0.5.[31] We also mark the pair of corridors with the largest sum of average predicted treatment effects in preparation for considering the effect of site selection bias, where the most promising sites in terms of treatment effects are also chosen for earlier experimentation (Allcott, 2015).

The pure structural prior tends to choose sites that are closer in average home household income and corridor distance than priors involving smoothing. This is because many more site characteristics enter the structural prior, in nonlinear and interactive ways. We show this in Appendix O where we linearly regress correlations between site effects in the structural prior on differences in site characteristics and find (1) the relationships between characteristic differences and ATE correlations differ substantially across countries despite the common structural model and (2) the vector of simple differences between the characteristics of two sites imperfectly explains their predicted treatment effect correlation even within-country. The pure smoothing prior, as discussed previously, tends to want all candidate sites to be close to at least one experimental site. This feature largely carries over to the mixed priors, including the mixed prior placing only 0.25 weight on smoothing.[32]

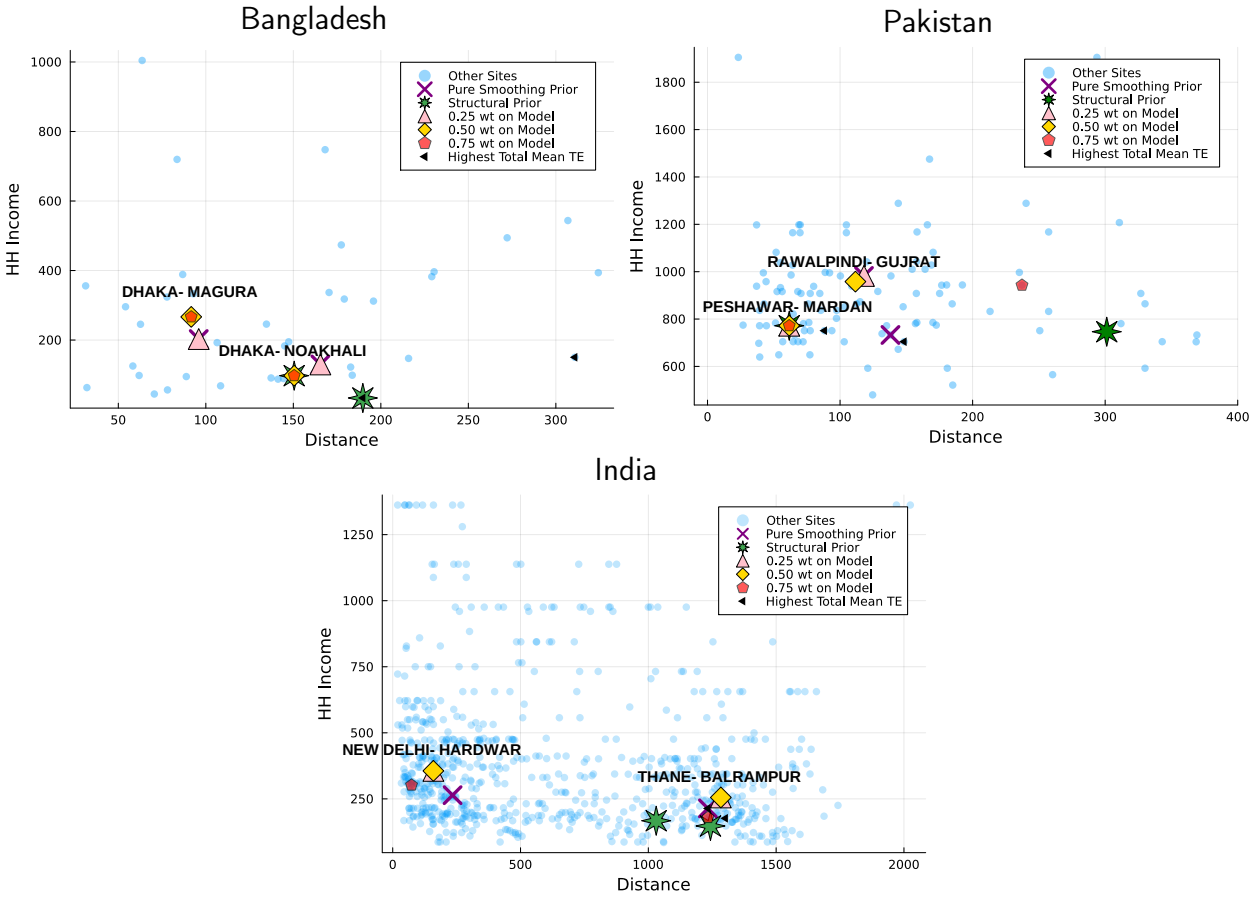
[29] We could consider an alternative heuristic rule that minimizes the sum of the distances in characteristic space between each site and its closest experimental site. However, this would not account for the fact that for some sites, there is little uncertainty about whether the treatment will be effective there or not.

[30] Appendix Figures A4 and A5 show the same pattern in Pakistan and India, respectively.

[31] Our choice of 0.5 weight in the mixing prior should not be taken to represent agnosticism between the pure smoothing and structural approaches. We recommend that practitioners check for robustness of selected sites and welfare to a variety of mixing weights, as we do in Appendix L.

[32] In Appendix P we show that the set of selected sites is robust to variation in ψ, the parameter capturing the effect of the training program on the effective price of mobile money, which helps us be confident in our choice of sites even if the treatment effect may be moderated by unmodeled environmental changes taking place over time and space since the original Lee et al. (2021) study, such as the extent of mobile money penetration.

Figure 6: Optimal Experimental Sites Under Alternative Decision Rules, Two Sites Per Country



Notes: Light dots represent candidate sites. HH Income refers to the average income in 2015 Bangladesh taka (1 2015 USD = 78 BDT), excluding remittances, of a household at origin reporting having sent an internal migrant. Distance is km between corridor origin and destination as the crow flies. See Appendix J.1 for details. Site names are in destination-origin format and refer to the sites selected under our preferred mixed prior with 50% weight on the structural model.

Since the pure structural prior considers more site characteristics than the pure smoothing prior, it may select sites that are more central on dimensions not targeted by the smoothing prior. We show an example of this in Appendix Figure A6, where the smoothing prior picks an outlying site in terms of both average household size at origin and the the ratio of origin prices to destination prices. On the latter characteristic, in particular, the smoothing prior picks a quite unusual site where prices in the origin district are higher than they are at destination.

5.2.3 Welfare Gains When Policy Choices Are *A Priori* Uncertain

In the Bayesian framework, evidence from experiments is especially valuable for decisions in sites where it is not already clear from the prior whether the treatment is likely to be effective. These "marginal" sites play a particularly important role in optimal site selection.

We consider a migration corridor to be marginal if its treatment effect under a given prior has a non-trivial probability of exceeding or falling short of the cost of treatment there. Specifically, given a value $\kappa \in (0,1)$, a site s with treatment effect τ_s and cost of treatment cost_s is said to be marginal if the following conditions hold:

$$\mathbb{P}(\tau_s \geq \text{cost}_s) \geq \kappa \quad \text{and} \quad \mathbb{P}(\tau_s \leq \text{cost}_s) \geq \kappa,$$

where the probabilities are calculated under the relevant prior for τ_s. In our case with normal priors, these probabilities can be evaluated easily for any value of κ.

Sites that are not marginal are those which will almost always be treated, or will almost never be treated, regardless of the choice of experimental sites, because there is little chance that the experimental data will revise beliefs enough to change the sign of $(\tau_s - \text{cost}_s)$. From the perspective of the policy-maker, it is more useful to focus instead on the potential gains in corridors for which the treatment assignment rule may be influenced in a nontrivial manner by experimentation. Welfare from sites that are always treated may also inflate the aggregate welfare values shown in Table 3.

Table 4: Average Welfare from Best and Worst Site Combinations by Country for Marginal Sites, κ =0.25

Country	Highest Average Welfare	Lowest Average Welfare
Pakistan	0.048	0.034
India	0.101	0.076

Notes: Average Welfare refers to the average across marginal sites of the welfare increase from site-level implementation recommendations based on the welfare-maximizing and minimizing sites from Table A3. Average welfare is predicted under our preferred mixed prior with 0.5 weight on the structural model. Site s is considered marginal if $\mathbb{P}(\tau_s \geq \text{cost}_s) \geq \kappa$ and $\mathbb{P}(\tau_s \leq \text{cost}_s) \geq \kappa$, where τ_s and cost_s are the welfare benefit and cost from implementing the mobile money training program there.

In Table 3 the best pair of sites in Pakistan is predicted to achieve only 8% more welfare than the worst pair.[33] Table 4 instead shows that the gain is largers when welfare is averaged over marginal sites only (showing the per site average welfare from the same best and worst site combinations as underlie Table 3). Focusing on sites with at least 25% probability of an

[33]To understand why the percentage welfare gains from using the best site combination instead of the worst are smaller in Pakistan and India than Bangladesh, recall from Equation (2) that welfare from site selection strategies will differ only to the extent that they generate different recommendations for where to implement the mobile money training program. If the mean predicted treatment effect in a given migration corridor is sufficiently high or low under the prior, or there is little uncertainty about the underlying treatment effect, it is considered very unlikely that subsequent experimentation will change the policy recommendation. All treatment assignment rules will benefit from the treatment effect in sites receiving the training program regardless of which sites are selected for experimentation.

opposite-signed net benefit of program implementation ($\kappa = 0.25$) in Table 4, the best pair in Pakistan is now predicted to achieve 41% higher welfare than the worst pair. For India, the corresponding numbers are 6% without restricting to marginal sites and 33% when doing so. In Bangladesh all sites are marginal when $\kappa = 0.25$. Appendix Tables A4 and A5 go into more detail, giving the identities of the specific sites in Pakistan and India, and showing how results change with κ in Pakistan.

5.3 The Gains from Experimenting, With and Without Bayesian Learning

5.3.1 The Welfare Gains from Experimenting

Experiments are undertaken to generate evidence that informs policy choices, and those policy choices are expected to lead to higher welfare. Figure 7 shows that this is not necessarily so. Gains in expected welfare depend on how sites are chosen and how findings are extrapolated.

The vertical axis of the figure shows the welfare from experiments, normalized so that welfare without experimentation is zero. The no-experiment scenario is not a zero-knowledge scenario: decision-makers can draw on data about the settings, the relationships in the structural model, and evidence from the pilot study (Lee et al., 2021). Treatment assignment is based on the prior itself, i.e. a site is treated if the prior mean for that site exceeds the cost of treatment.

The question is how further experimentation informs subsequent policy choices. The horizontal axis in Figure 7 gives the the number of experimental sites. We start by considering welfare when only one experiment is run, as in Section 5.2.1, and we predict welfare using the 0.5-weight mixed prior. By construction, site selection using that prior will attain optimal welfare; this is represented by the top line with star markers. The lowest line in the figure, marked by square boxes, shows welfare when the experimental site is instead the one that *a priori* is expected to deliver the largest treatment effect. As Figure 5 shows, the site's context is atypical relative to others. This is an extreme version of a tendency for positive site-selection bias described in a range of contexts by Allcott (2015), Wang and Yang (2025), and Mentges et al. (2021).[34]

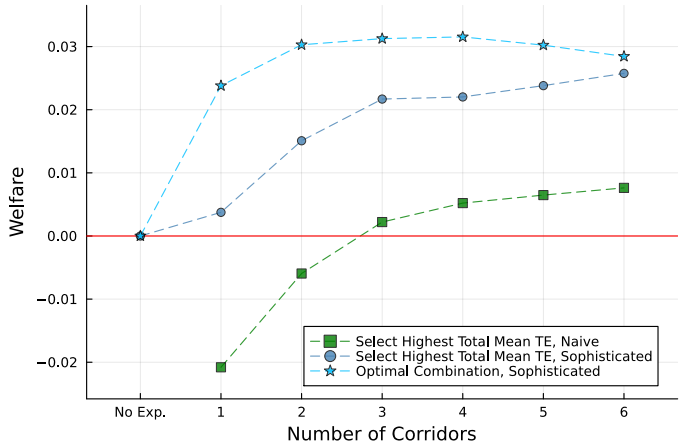
In this scenario, we assume that decision-makers in other sites directly apply the experimental finding to their own sites without adjusting for differences in context. We say that a "naive" policymaker assigns the treatment to all sites if the average of experimental ATEs exceeds the cost of treatment, irrespective of their contexts. This behavior arises as a limiting case of the smoothing prior. We show in Appendix D.3 that as c_1 and c_0 grow large, the Bayesian policy rule converges to the naive one. Hjort, Moreira, Rao, Santini, and Rao (2021) provide

[34] Stuart et al. (2017) provide an example of negative site-selection in the context of the evaluation of a reading intervention. Negative site-selection leads to similar concerns as those described in our analysis.

evidence that policymakers may extrapolate in this way. In their study, Brazilian municipal officials displayed a concern for the sample size of prototypical experiments, but not the contexts in which they were performed as measured by geographic location and average income of the country where the experiment was conducted.

In our results, naive policymakers fare very poorly when there is extreme site selection bias; that is, when experimental sites are chosen for their high predicted treatment effects. With a single experiment the change in welfare is negative; the resulting normalized value of welfare is -0.02. In other words, welfare would have been higher if no experiment had been run. In the context of our framework, the evidence is from the wrong place and applied inappropriately.

Figure 7: Welfare Gains from Experimenting, Bangladesh



Notes: Average Per-Migrant Welfare refers to the average across all candidate sites of the welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Surveying budget is held fixed across different numbers of sites. See Appendix M.2 for details.

The middle line, marked by circles, shows welfare when policymakers again use evidence from the settings with the highest expected treatment effects. Here, however, policymakers are Bayesians, and they process the data using the framework developed above. Because some of the sites with high predicted effects are informative, the welfare loss can be moderated by a sophisticated policymaker who fully accounts for the biased site selection. The sophisticated extrapolation of evidence thus improves policy choices and increases welfare. Welfare is no longer negative, but the increase in welfare from one experiment is small relative to having no experiment.

In contrast, the top line, marked by stars, shows that a single experiment in a site optimized for external validity can substantially increase welfare. This relies on decision-makers being sophisticated in their use of evidence (as defined by the Bayesian framework above) and researchers choosing sites with those decision-makers in mind.

5.3.2 How Many Experimental Sites?

In Sections 5.2.1 and 5.2.2, we analyzed the optimal selection of up to two migration corridors in each country. For the same budget, however, it may be possible to experiment in additional sites and thus estimate impacts in a wider range of different settings. An important cost is the loss of statistical power in each experiment, since the total sample is now split over multiple experiments, as we assume the budget won't be expanded to adjust for the new sites. We maintain a constant budget based on our two-sites-per-country, 2,000-households-per-site mandate, and because experimenting in each site involves fixed costs, the total number of households we can sample diminishes as sites are added. Still, the benefit of being able to learn from more settings may outweigh the cost of reduced precision in each setting.

Our framework allows us to explore the question by modifying the sampling error standard deviation σ_ϵ to reflect the corresponding change in the sample size per site and specifying the appropriate choice set for site combinations. Figure 7 shows per-site average welfare in Bangladesh as we further vary the number of experimental sites, based on cost calculations with a Bangladeshi survey firm reported in Appendix M.2.[35]

As noted in Section 5.3.1, we see large gains in welfare from the first experimental site when sites are chosen optimally. Adding a second site brings evidence from a different setting, and further increases welfare. Welfare gains diminish for subsequent sites until welfare begins to decrease as experimental site effects become noisier due to smaller sample sizes with 5 or 6 sites. In Pakistan, the pattern is similar, with welfare beginning to decrease with the 3rd site (Appendix Figure A11).

With optimal site selection, losses of power as sites are added suggest that running experiments in 2 sites is roughly optimal given a fixed budget and costs of adding experiments. Comparing the scenarios in Figure 7, the work to select optimal sites may thus be compensated by the reduced need to run as many experiments and the greater statistical power in each experiment.

With sub-optimal sites, more experiments bring a broader range of information, and welfare gains increase substantially in moving from a single experimental site. As we noted, in Bangladesh the site with the highest single predicted treatment effect is very unusual and contributes little to learning about other sites. However, the site with the second-highest predicted treatment effect happens to be less outlying and is even selected by the structural prior in the two-site case. All the same, it is necessary to run roughly 6 experiments in the non-optimal, high-treatment effect settings to provide evidence that generates the level of welfare that can

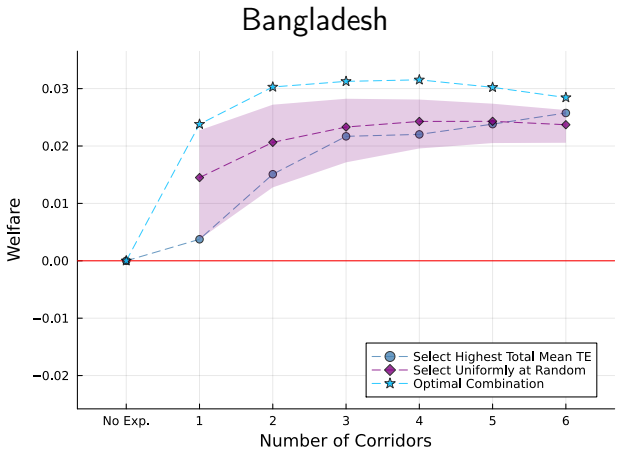
[35] The calculations are computationally intensive, and we illustrate ideas with results from Bangladesh and Pakistan (in Appendix M.3), where computation is easier due to the much smaller number of candidate sites than in India.

be generated by evidence from just a single optimally-selected site.

5.3.3 Selecting Sites Randomly

The analyses above show the welfare gains to optimal site selection relative to sites selected to maximize expected treatment effects. In response to this kind of positively biased selection, Allcott (2015) proposes that sites be chosen randomly instead. With a large number of experiments, randomization can eliminate selection bias. Here, we consider the more typical case where it is possible to run just one experiment or a small number.

Figure 8: Randomized Site Selection with Sophisticated Extrapolation, Bangladesh



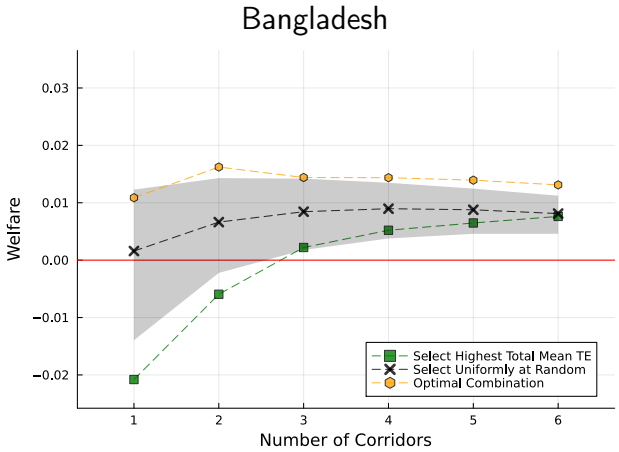
Notes: Average Per-Migrant Welfare refers to the average across all candidate sites of the welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Surveying budget is held fixed across different numbers of sites. See Appendix M.2 for details.

The top line (marked by stars) and bottom line (marked by circles) in Figure 8 correspond to the two top lines in Figure 7. They reflect welfare when a sophisticated, Bayesian decision-maker takes evidence from an optimal site (top line) or sites with the highest expected treatment effect (bottom line). The middle line (marked by diamonds) reflects evidence generated by experiments in randomly-selected places, with the shaded band showing the 10th to 90th percentile range of welfare values across repeated random choices. As Allcott (2015) suggests, welfare is higher over most of the range, relative to experiments in positively-selected sites. Sophisticated decision-makers benefit from the greater range of information provided by the randomly-chosen sites.

Still, if there can only be a small number of experiments, this strategy fares relatively poorly, with a range of values that can provide even worse welfare than when there is extreme site selection. While this may be surprising at first glance, randomization has a clear shortcoming in our setting. If a very small number of sites are chosen at random, there is a reasonably high prob-

ability that the chosen sites will be very unrepresentative of the entire country. In this context, welfare from optimally-selected sites is always greater than that from random sites.

Figure 9: Randomized Site Selection with Naive Extrapolation, Bangladesh



Notes: Average Per-Migrant Welfare refers to the average across all candidate sites of the welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Surveying budget is held fixed across different numbers of sites. See Appendix M.2 for details.

Figure 9 shows expected welfare when decision-makers naively extrapolate evidence. Under the same methods for selecting sites (highest treatment effects, random, and optimized), welfare is uniformly lower than when decision-makers are sophisticated. The ordering of the lines shows that optimal site selection improves welfare even if decision-makers are naive.

6 Conclusion

The social sciences have become increasingly empirical, driven by new methods for causal identification and the promotion of evidence-based decision-making (Angrist and Pischke, 2010). Field experiments play an important role, especially in development economics (Banerjee and Duflo, 2025). Even with the growing number of high-quality empirical studies, however, decision-makers usually need to extrapolate results from other contexts to make choices in their own contexts (Bates and Glennerster, 2017). Concern with the internal validity of research designs is then joined by the need for external validity and the ability to generalize results (Al-Ubaydli et al. 2017b; Vivalt 2020).

We have developed a general framework for experimental site selection that incorporates external validity as its key aim. Our approach is flexible, allowing for various specifications of the decision-maker's social welfare function, various constraints on the choice set of feasible combinations of sites, and a range of prior specifications that incorporate different sources of information about the heterogeneity of effects across sites.

In our application, we developed a structural model of individual-level responses to the intervention, which allowed us to leverage the rich data available in a pilot experiment to build an informative prior for the treatment effects at other sites. Then, to allow the possibility that the structural model may not capture all the underlying causal mechanisms and sources of heterogeneity, we mixed the structural prior with a smoothing prior based on a set of aggregate site-level characteristics. Working with quasi-posteriors associated with standard estimation methods makes it convenient to incorporate standard structural estimation strategies without needing to invoke the full machinery of Bayesian posterior inference on the structural model.

This approach to incorporating structural modeling in experimental analyses while allowing for departures from the model may be useful in other problems. In our application to an ongoing experimental design in Bangladesh, Pakistan, and India, our method is computationally feasible and leads to substantial prospective gains in expected welfare relative to *ad hoc* approaches that approximate common practice.

The approach is data-intensive and requires building computational machinery, but the costs incurred in the research design can reduce the costs of running experiments. For example, in our application we find that running fewer experiments in optimally-chosen sites outperforms running a greater number of experiments in sub-optimal sites.

Even if researchers do not incorporate the full Bayesian decision-theoretic approach into research designs, the approach generates broad insights for selecting experimental sites. First, external validity is often framed as a question about whether a given treatment effect is likely to be similar in other sites. What matters for evidence-based policy, however, is not similarity but how the evidence from the given experimental site informs policy decisions in other sites, when coupled with contextual information and knowledge of economic relationships. Second, sites chosen to maximize expected treatment effects can be valuable in demonstrating the plausibility of ideas, but evidence may not travel far and, if applied in other sites without adjusting for contexts, can lead to policy choices that *reduce* expected welfare. Third, in contrast to sites where treatment effects are expected to be large, experiments from sites where treatment effects are *a priori* uncertain provide particularly valuable evidence. Fourth, while the random selection of experimental sites is a common benchmark against which to judge selection bias, when only a small number of experiments is run, evidence from sites that are randomly-selected is generally not as valuable as evidence from sites that are deliberately chosen to be most informative for policymakers. Fifth, while we have focused on site selection, the nature of learning matters in parallel. Extrapolation that is sensitive to the heterogeneity of settings can mitigate welfare losses associated with taking evidence from sites that are selected sub-optimally. Together with optimal site-selection, careful extrapolation leads to the highest welfare gains.

Further extensions of our approach are possible. Our two-stage adaptive design setting

could be extended to multiple rounds, where at each round, the available information can be used to inform future experimental design choices. Additional dimensions of choice beyond the selection of sites, such as the joint choice of sites and sample sizes, or more sophisticated choices of conditional treatment randomization probabilities, could also be incorporated into the approach.

References

ABADIE, A. AND M. D. CATTANEO (2018): "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

ADJAHO, C. AND T. CHRISTENSEN (2022): "Externally Valid Treatment Choice," Working paper.

AL-UBAYDLI, O., J. LIST, D. LORE, AND D. SUSKIND (2017a): "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature," *Journal of Economic Perspectives*, 31, 125–44.

AL-UBAYDLI, O., J. LIST, AND D. SUSKIND (2017b): "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results," *American Economic Association Papers and Proceedings*, 107, 282–286.

ALDERMAN, H., J. BEHRMAN, AND A. TASNEEM (2019): "The Contribution of Increased Equity to the Estimated Social Benefits from a Transfer Program: An Illustration from PROGRESA/Oportunidades," *The World Bank Economic Review*, 33, 535–550.

ALLCOTT, H. (2015): "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130, 1117–1165.

ANDREWS, I., D. FUDENBERG, L. LEI, A. LIANG, AND C. WU (2023): "The Transfer Performance of Economic Models," ArXiv:2202.04796.

ANGRIST, J. D. AND J.-S. PISCHKE (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.

ANGRIST, N., M. AINOMUGISHA, S. P. BATHENA, P. BERGMAN, C. CROSSLEY, C. CULLEN, T. LETSOMO, M. MATSHENG, R. MARLON PANTI, S. SABARWAL, AND T. SULLIVAN (2023): "Building Resilient Education Systems: Evidence from Large-Scale Randomized Trials in Five Countries," Working Paper 31208, National Bureau of Economic Research.

ANGRIST, N., P. BERGMAN, AND M. MATSHENG (2022): "Experimental Evidence on Learning Using Low-Tech when School is Out," *Nature Human Behaviour*, 6, 941–950.

ARADILLAS FERNANDEZ, A., J. BLANCHET, J. L. MONTIEL OLEA, C. QIU, J. STOYE, AND L. TAN (2024): "ϵ-Minimax Solutions of Statistical Decision Problems via the Hedge Algorithm," 1–53.

ATHEY, S., K. BERGSTROM, V. HADAD, J. C. JAMISON, B. OZLER, L. PARISOTTO, AND J. D. SAMA (2021): "Shared Decision-Making: Can Improved Counseling Increase Willingness to Pay for Modern Contraceptives?" Policy Research Working Paper 9777, World Bank.

BANDIERA, O., R. BURGESS, N. DAS, S. GULESCI, I. RASUL, AND M. SULAIMAN (2017): "Labor Markets and Poverty in Village Economies," *Quarterly Journal of Economics*, 132, 811–870.

BANERJEE, A., E. DUFLO, N. GOLDBERG, D. KARLAN, R. OSEI, W. PARIENTE, J. SHAPIRO, B. THUYSBAERT, AND C. UDRY (2015): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348, 1260799–1260799.

BANERJEE, A. V. AND E. DUFLO (2025): "Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty (Revised and Updated)," .

BATES, M. A. AND R. GLENNERSTER (2017): "The Generalizability Puzzle," *Stanford Social Innovation Review*, 15, 50–54.

BERGER, J. O. (1993): *Statistical Decision Theory and Bayesian Analysis*, New York: Springer, 2nd ed.

BREZA, E., A. G. CHANDRASEKHAR, AND D. VIVIANO (2026): "Evidence aggregation with ignorance in mind: learning what we do (not) know for archetypes discovery," Working paper arXiv:2501.13355.

CAMPBELL, D. AND J. STANLEY (1966): *Experimental and Quasi-experimental Designs for Research*, Chicago: Rand McNally.

CAMPBELL, D. T. (1986): "Relabeling Internal and External Validity for Applied Social Scientists," *New Directions for Program Evaluation*, 1986, 67–77.

CARIA, S., G. GORDON, M. KASY, S. QUINN, S. SHAMI, AND A. TEYTELBOYM (2021): "An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan," Working paper.

CHASSANG, S. AND S. KAPON (2021): "Designing Randomized Controlled Trials with External Validity in Mind," Working paper.

COOK, T. AND W. SHADISH (1994): "Social Experiments: Some Developments over the Past Fifteen Years," *Annual Review of Psychology*, 45, 45–580.

CUSOLITO, A., O. DAROVA, AND D. MCKENZIE (2023): "Capacity Building as a Route to Export Market Expansion: A Six-Country Experiment in the Western Balkans," CEPR Discussion Paper 17789, Center for Economic and Policy Research.

DEATON, A. (1997): *The Analysis of Household Surveys*, The World Bank.

DEATON, A. AND N. CARTWRIGHT (2018): "Understanding and Misunderstanding Randomized Controlled Trials," *Social Science and Medicine*, 210, 2–21.

DEHEJIA, R., C. POP-ELECHES, AND C. SAMII (2021): "From Local to Global: External Validity in a Fertility Natural Experiment," *Journal of Business and Economic Statistics*, 39, 217–243.

EGAMI, N. AND D. D. I. LEE (Forthcoming): "Designing Multi-Site Studies for External Validity: Site Selection via Synthetic Purposive Sampling," *American Journal of Political Science*.

EPANOMERITAKIS, A. AND D. VIVIANO (2025): "Learning What to Learn: Experimental Design when Combining Experimental with Observational Evidence," 1–58.

FINAN, F. AND D. POUZO (2023): "Reinforcing RCTs with Multiple Priors while Learning about External Validity," ArXiv:2112.09170.

FINKELSTEIN, A. AND N. HENDREN (2020): "Welfare Analysis Meets Causal Inference," *Journal of Economic Perspectives*, 34, 146–67.

GECHTER, M. (2023): "Generalizing the Results from Social Experiments: Theory and Evidence from India," Working paper.

GECHTER, M., C. SAMII, R. DEHEJIA, AND C. POP-ELECHES (2019): "Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference," Working paper.

GLENNERSTER, R. (2017): "The Practicalities of Running Randomized Evaluations," in *Handbook of Economic Field Experiments*, Elsevier, vol. 1, 175–243.

HAHN, J., K. HIRANO, AND D. KARLAN (2011): "Adaptive Experimental Design Using the Propensity Score," *Journal of Business and Economic Statistics*, 29, 96–108.

HASHEMI, S. AND A. DE MONTESQUIOU (2011): "Reaching the Poorest: Lessons from the Graduation Model," Tech. Rep. Focus Note 69, CGAP, Washington, DC.

HJORT, J., D. MOREIRA, G. RAO, J. F. SANTINI, AND G. RAO (2021): "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities," *American Economic Review*, 111, 1442–1480.

HOTZ, V. J., G. W. IMBENS, AND J. H. MORTIMER (2005): "Predicting the Efficacy of Future Training Programs using Past Experiences at Other Locations," *Journal of Econometrics*, 125, 241–270.

HURTADO-CHONG, A., A. JOERIS, D. HESS, AND M. BLAUTH (2017): "Improving Site Selection in Clinical Studies: a Standardised, Objective, Multistep Method and First Experience Results," *BMJ Open*, 7, e014796.

IMBERT, C. AND J. PAPP (2020): "Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India," *Journal of the European Economic Association*, 18, 927–963.

ISHIHARA, T. AND T. KITAGAWA (2021): "Evidence Aggregation for Treatment Choice," ArXiv:2108.06473.

JPAL (2022): "Targeting the Ultra-Poor to Improve Livelihoods," https://www.povertyactionlab.org/case-study/targeting-ultra-poor-improve-livelihoods, updated: November 2022.

KALLUS, N. AND A. ZHOU (2021): "Minimax-Optimal Policy Learning Under Unobserved Confounding," *Management Science*, 67, 2870–2890.

KASY, M. AND A. SAUTMANN (2021): "Adaptive Treatment Assignment in Experiments for Policy Choice," *Econometrica*, 89, 113–132.

KUANG, K., R. XIONG, P. CUI, S. ATHEY, AND B. LI (2018): "Stable Prediction across Unknown Environments," ArXiv:1806.06270.

KYDLAND, F. E. AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–

1370.

LEE, J. N., J. MORDUCH, S. RAVINDRAN, AND A. S. SHONCHOY (2022): "Narrowing the Gender Gap in Mobile Banking," *Journal of Economic Behavior and Organization*, 193, 276–293.

LEE, J. N., J. MORDUCH, S. RAVINDRAN, A. S. SHONCHOY, AND H. ZAMAN (2021): "Poverty and Migration in the Digital Age: Experimental Evidence on Mobile Banking in Bangladesh," *American Economic Journal: Applied Economics*, 13, 38–71.

MEAGER, R. (2019): "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 11, 57–91.

MENTGES, A., S. BLOWES, D. HODAPP, H. HILLEBRAND, AND J. CHASE (2021): "Effects of site-selection bias on estimates of biodiversity change," *Conservation Biology*, 35, 688–698.

MENZEL, K. (2023): "Transfer Estimates for Causal Effects across Heterogeneous Sites," ArXiv:2305.01435.

MONTIEL OLEA, J. L., C. QIU, AND J. STOYE (2023): "Decision Theory for Treatment Choice Problems with Partial Identification," ArXiv:2312.17623.

MURALIDHARAN, K. AND P. NIEHAUS (2017): "Experimentation at Scale," *Journal of Economic Perspectives*, 31, 103–124.

NEWEY, W. AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics 4*, ed. by R. Engle and D. McFadden, Elsevier, 2111–2245.

O'MUIRCHEARTAIGH, C. AND L. V. HEDGES (2014): "Generalizing from Unrepresentative Experiments: A Stratified Propensity Score Approach," *Journal of the Royal Statistical Society Series C: Applied Statistics*, 63, 195–210.

POTTER, J. S., D. M. DONOVAN, R. D. WEISS, J. GARDIN, R. LINDBLAD, P. WAKIM, AND D. DODD (2011): "Site Selection in Community-Based Clinical Trials for Substance Use Disorders: Strategies for Effective Site Selection," *The American Journal of Drug and Alcohol Abuse*, 37, 400–407.

PRITCHETT, L. AND J. SANDEFUR (2015): "Learning from Experiments when Context Matters," *American Economic Review*, 105, 471–475.

RASMUSSEN, C. E. AND C. K. I. WILLIAMS (2006): *Gaussian Processes for Machine Learning*, Cambridge, Massachusetts: The MIT Press.

SANCHEZ-BECERRA, A., M. GECHTER, R. DEHEJIA, C. POP-ELECHES, J. N. LEE, M. MAHMUD, J. MORDUCH, S. RAVINDRAN, C. SAMII, AND A. S. SHONCHOY (2026): "Evaluating Expert Predictions for the Impact of Mobile Money on Poverty: Evidence from a Multi-Site Experiment," Working paper.

STUART, E., S. BELL, C. EBNESAJJAD, R. OLSEN, AND L. ORR (2017): "Characteristics of School Districts That Participate in Rigorous National Educational Evaluations," *Journal of Research on Educational Effectiveness*, 10, 168–206.

STUART, E. A., S. R. COLE, C. P. BRADSHAW, AND P. J. LEAF (2011): "The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 174, 369–386.

TABORD-MEEHAN, M. (2018): "Stratification Trees for Adaptive Randomization in Randomized Controlled Trials," Working paper.

TIPTON, E. (2013): "Improving Generalizations From Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts," *Journal of Educational and Behavioral Statistics*, 38, 239–266.

——— (2014): "How Generalizable Is Your Experiment? An Index for Comparing Experimental Samples and Populations," *Journal of Educational and Behavioral Statistics*, 39, 478–501.

TODD, P. E. AND K. I. WOLPIN (2010): "Structural Estimation and Policy Evaluation in Developing Countries," *Annual Review of Economics*, 2, 21–50.

——— (2023): "The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling," *Journal of Economic Literature*, 61, 41–85.

VIVALT, E. (2020): "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association*, 18, 3045–3089.

WANG, S. AND D. Y. YANG (2025): "Policy Experimentation in China: The Political Economy of Policy Learning," *Journal of Political Economy*, 133, 2180–2228.

Appendices, for Online Publication Only

A Additional Tables Referenced in the Main Text

Table A1: Summary Statistics for Estimation Sample from the Lee et al. (2021) Experiment by Migrant Gender

	Treated				Control			
	Male		Female		Male		Female	
Variable	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev.	Mean	Std. Dev
Average Remittance Amount per Day	92.49	81.00	49.16	48.83	83.12	68.87	51.12	56.52
Mobile Money Usage (cond. on remitting)	0.54	0.45	0.46	0.47	0.43	0.46	0.34	0.46
Migrant Hourly Wage	32.77	9.07	25.49	5.68	32.01	8.89	25.68	5.14
Home Income Per Day	362.02	260.36	317.47	216.01	369.06	247.75	352.18	239.63
Home Household Size	4.09	1.57	4.10	1.41	4.27	1.64	4.18	1.51
N	241		83		233		100	

Notes: currency amounts are in Bangladesh taka (1 2015 USD = 78 BDT). Mobile money usage conditional on remitting is the fraction of remittance transactions reported to have been executed via mobile money.

Table A2: Site Characteristics, Summary Statistics by Country

Characteristic		Bangladesh	Pakistan	India
Mean Household Income (Daily)	Mean	267.05	897.51	378.94
	SD	211.38	262.3	261.15
SD Household Income (Daily)	Mean	239.57	836.32	581.02
	SD	230.49	276.15	390.39
Mean Household Size	Mean	4.03	5.31	6.37
	SD	0.89	1.03	1.04
SD Household Size	Mean	2.35	2.73	2.97
	SD	1.83	0.72	0.75
Mean Male Migrant Wage (Hourly)	Mean	55.55	125.49	77.03
	SD	5.53	61.54	29.16
SD Male Migrant Wage (Hourly)	Mean	72.96	75.37	65.19
	SD	8.26	63.54	34.87
Mean Female Migrant Wage (Hourly)	Mean	33.07	44.17	50.40
	SD	25.52	53.52	93.10
SD Female Migrant Wage (Hourly)	Mean	43.03	26.04	27.53
	SD	33.67	54.21	83.47
Mean Remittance (Daily)	Mean	178.08	459.41	227.46
	SD	71.05	152.03	215.361
Operator Density	Mean	0.07	0.04	0.05
	SD	0.02	0.01	0.05
Missing Operator Density (District Level)	Percent	41.5	48.3	48.8
Price Index	Mean	0.88	0.99	0.81
	SD	0.145	0.02	0.17
Distance between Sites in Corridor	Mean	142.63	227.83	652.70
	SD	75.85	315.12	502.16
Migrant Density	Mean	0.03	0.01	0.17
	SD	0.022	0.004	0.22
Number of Sites	Count	41	60	740

Notes: Site characteristics are themselves means and standard deviations across households within each site. The means and standard deviations reported in this table are the mean of means and standard deviations across sites in each country, and the standard deviations of site-level means and standard deviations. Wages and incomes are in Bangladesh taka (1 USD = 78 BDT as of December 2015). Sites missing operator density have operator density replaced with the mean operator density at the next-highest administrative level (e.g. the state mean in India). Sites come from a set restricted to ensure feasibility of treatment delivery and robustness to prior misspecification. See Appendix J.1 for variable construction and site restriction details.

Table A3: Welfare-Maximizing and Minimizing Experimental Site Combinations

Bangladesh					
Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.030	Dhaka-Madaripur	Gazipur-Panchagarh	0.004
Dhaka-Kishoregonj	Dhaka-Noakhali	0.030	Dhaka-Gopalganj	Gazipur-Panchagarh	0.004
Dhaka-Barguna	Dhaka-Kishoregonj	0.030	Dhaka-Bhola	Gazipur-Panchagarh	0.003
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.030	Dhaka-Bhola	Dhaka-Gopalganj	0.002
Dhaka-Feni	Dhaka-Narsingdi	0.029	Dhaka-Bhola	Dhaka-Madaripur	0.002
Pakistan					
Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Rawalpindi-Gujrat	Peshawar-Mardan	0.111	Karachi-Mianwali	Karachi-Sukkur	0.103
Rawalpindi-Mandi Bahauddin	Peshawar-Mardan	0.111	Karachi-Bannu	Rawalpindi-Islamabad	0.103
Rawalpindi-Gujrat	Mardan-Swabi	0.111	Karachi-Bannu	Chitral-Islamabad	0.103
Rawalpindi-Gujrat	Hangu-Kohat	0.111	Rawalpindi-Islamabad	Karachi-Sukkur	0.103
Faisalabad-Bahawalnagar	Peshawar-Mardan	0.111	Chitral-Islamabad	Karachi-Sukkur	0.103
India					
Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Thane-Balrampur	New Delhi-Hardwar	0.209	New Delhi-Bangalore	Kancheepuram-Chandigarh	0.197
New Delhi-Hardwar	Mumbai-Sultanpur	0.209	Bangalore-Chandigarh	Sitamarhi-Mumbai	0.197
New Delhi-Hardwar	Thane-Sultanpur	0.209	Bangalore-Chandigarh	New Delhi-Chennai	0.197
New Delhi-Hardwar	Thane-Sant Kabir Nagar	0.209	Kancheepuram-Chandigarh	New Delhi-Chennai	0.197
Thane-Basti	New Delhi-Hardwar	0.209	Kancheepuram-Chandigarh	Sitamarhi-Mumbai	0.197

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Site names are in destination-origin format.

Table A4: Contributions of Marginal Sites to Average Per-Migrant Welfare in Pakistan

Best Combinations		κ			Worst Combinations		κ		
Corridor 1	Corridor 2	0.05	0.15	0.25	Corridor 1	Corridor 2	0.05	0.15	0.25
Rawalpindi-Gujrat	Peshawar-Mardan	0.109	0.077	0.048	Karachi-Mianwali	Karachi-Sukkur	0.100	0.066	0.035
Rawalpindi-Mandi Bahauddin	Peshawar-Mardan	0.109	0.077	0.048	Karachi-Bannu	Rawalpindi-Islamabad	0.100	0.066	0.034
Rawalpindi-Gujrat	Mardan-Swabi	0.109	0.077	0.048	Karachi-Bannu	Chitral-Islamabad	0.100	0.066	0.034
Rawalpindi-Gujrat	Hangu-Kohat	0.109	0.076	0.048	Rawalpindi-Islamabad	Karachi-Sukkur	0.100	0.066	0.034
Faisalabad-Bahawalnagar	Peshawar-Mardan	0.109	0.076	0.048	Chitral-Islamabad	Karachi-Sukkur	0.100	0.066	0.034

Table A5: Contributions of Marginal Sites to Average Per-Migrant Welfare in India

Best Combinations		κ	Worst Combinations		κ
Corridor 1	Corridor 2	0.25	Corridor 1	Corridor 2	0.25
Thane-Balrampur	New Delhi-Hardwar	0.100	New Delhi-Bangalore	Kancheepuram-Chandigarh	0.078
New Delhi-Hardwar	Mumbai-Sultanpur	0.100	Bangalore-Chandigarh	Sitamarhi-Mumbai	0.077
New Delhi-Hardwar	Thane-Sultanpur	0.101	Bangalore-Chandigarh	New Delhi-Chennai	0.077
New Delhi-Hardwar	Thane-Sant Kabir Nagar	0.101	Kancheepuram-Chandigarh	New Delhi-Chennai	0.076
Thane-Basti	New Delhi-Hardwar	0.099	Kancheepuram-Chandigarh	Sitamarhi-Mumbai	0.077

Notes: Average Per-Migrant Welfare refers to the average across all candidate sites of the welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Site s is considered marginal if $\mathbb{P}(\tau_s \geq \text{cost}_s) \geq \kappa$ and $\mathbb{P}(\tau_s \leq \text{cost}_s) \geq \kappa$, where τ_s and cost_s are the welfare benefit and cost from implementing the mobile money training program there. Site names are in destination-origin format.

B Additional Figures Referenced in the Main Text

B.1 Prior Implied Conditional Mean Welfare Treatment Effects for Pakistan

Figures A1 and A2 replicate Figures 3 and 4 for potential experimental sites in Pakistan. Mean welfare treatment effects are shown, conditional on the welfare treatment effect in the corridor originating in Mardan and going to Peshawar.

Figure A1: Expected Welfare TE (Structural Prior)

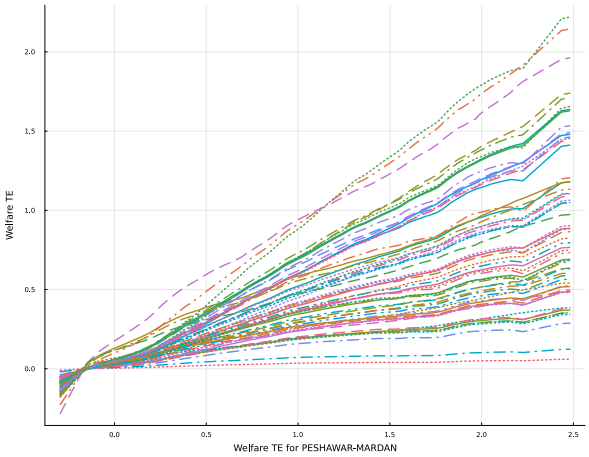
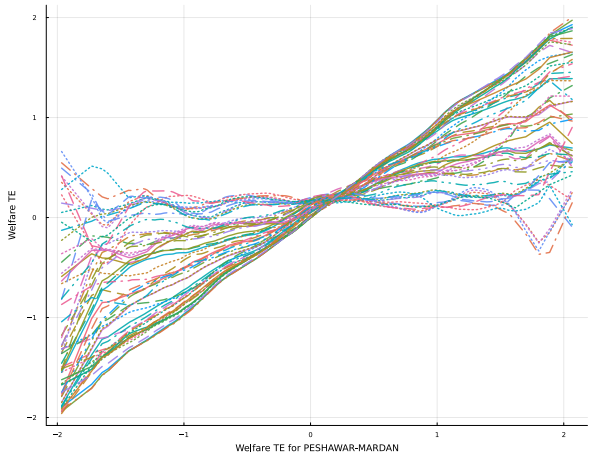


Figure A2: Expected Welfare TE (Smoothing Prior)

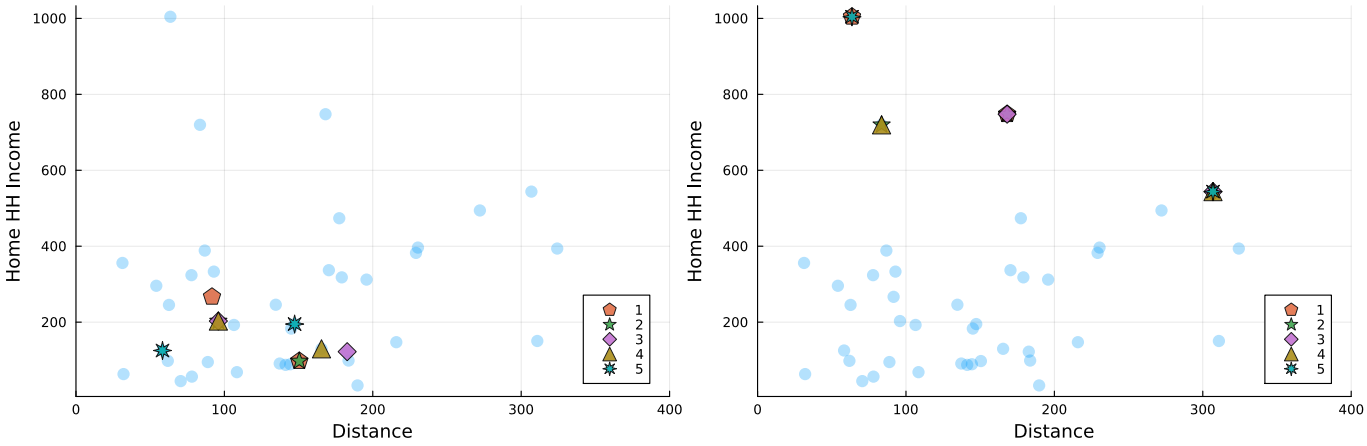


Notes: Each line in the figures shows the expected welfare treatment effect in a Pakistani migration corridor conditional on knowing the welfare treatment effect in the Peshawar-Mardan corridor is equal to the value on the X-axis.

B.2 Top 5 Best Corridor Combinations in Characteristic Space

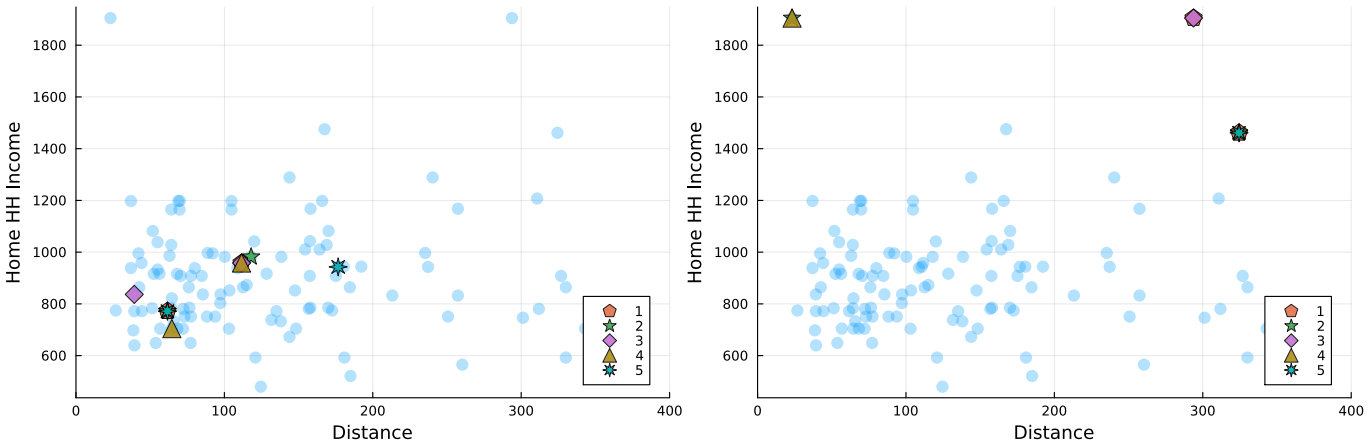
Figures A3 - A5 show where the best and worst combinations of two sites according to our preferred mixed prior lie in the space of characteristics considered by the smoothing prior. The best site combinations are located in the middle of the distribution of characteristics while the worst site combinations are outliers.

Figure A3: Welfare Maximizing (Left) and Minimizing (Right) Experimental Site Combinations in Bangladesh



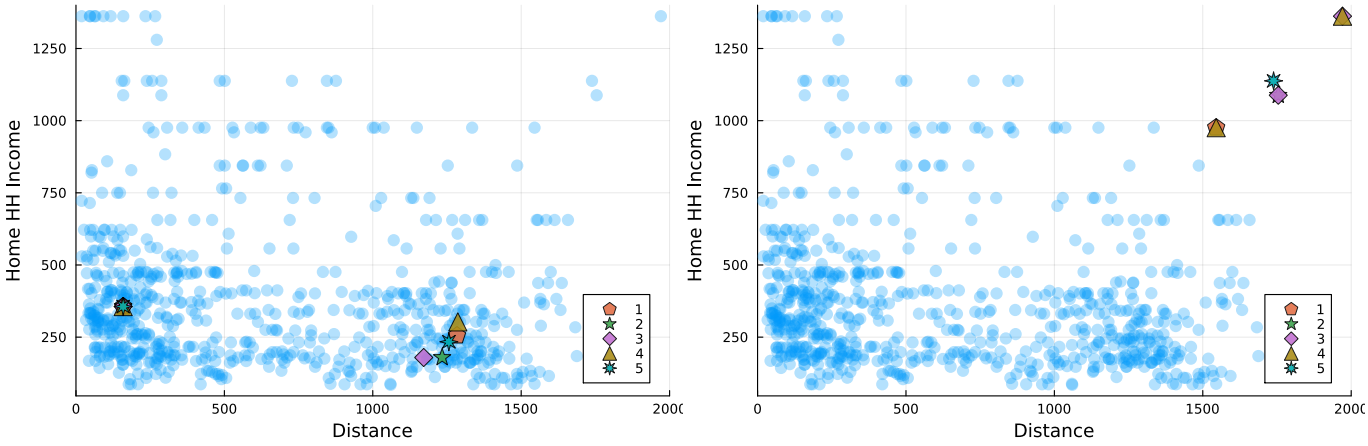
Notes: Light dots represent candidate sites. Numbers in the legend refer to rank among the top or bottom sites. E.g. 1 refers to the best site pair in the left panel and the worst pair in the right panel. Ranks are computed using our preferred mixed prior, with 0.5 weight on the structural model.

Figure A4: Welfare Maximizing (Left) and Minimizing (Right) Experimental Site Combinations in Pakistan



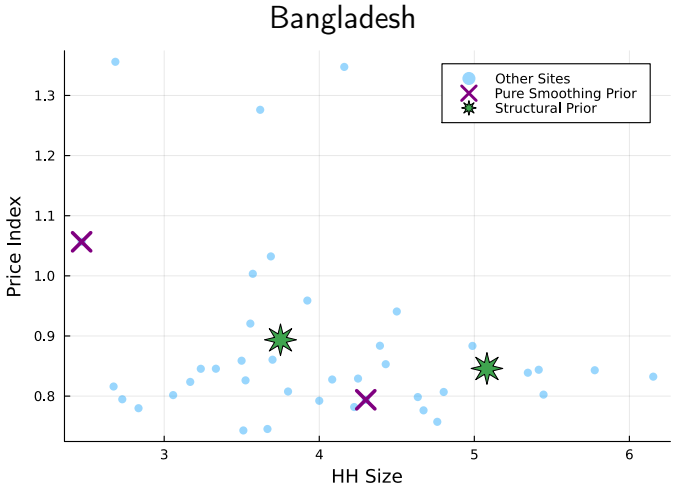
Notes: Light dots represent candidate sites. Numbers in the legend refer to rank among the top or bottom sites. E.g. 1 refers to the best site pair in the left panel and the worst pair in the right panel. Ranks are computed using our preferred mixed prior, with 0.5 weight on the structural model.

Figure A5: Welfare Maximizing (Left) and Minimizing (Right) Experimental Site Combinations in India



Notes: Light dots represent candidate sites. Numbers in the legend refer to rank among the top or bottom sites. E.g. 1 refers to the best site pair in the left panel and the worst pair in the right panel. Ranks are computed using our preferred mixed prior, with 0.5 weight on the structural model.

Figure A6: Sites Selected by the Smoothing Prior Need Not Be Central on Non-Targeted Dimensions



Notes: Light dots represent candidate sites. Y-axis is the ratio of price level in the origin district to that in the destination district, X-axis is the average household size in the origin district.

C Posterior Updating Rule

Let the prior for τ be $\tau \sim N(\mu_\tau, \Sigma_\tau)$. Since each $\hat{\tau}_s$ is distributed $N(\tau_s, \sigma^2_{\epsilon,s})$ conditional on τ, we can write the full potential vector of estimates $\hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_S)$ as

$$\hat{\tau} \mid \tau \sim N(\tau, \Sigma_\epsilon),$$

where $\Sigma_\epsilon = \text{diag}\left\{\sigma^2_{\epsilon,1}, \ldots, \sigma^2_{\epsilon,S}\right\}$ is the diagonal matrix of sampling variances. Combining these, we have

$$\begin{pmatrix} \tau \\ \hat{\tau} \end{pmatrix} \sim N_{2S}\left(\begin{pmatrix} \mu_\tau \\ \mu_\tau \end{pmatrix}, \begin{pmatrix} \Sigma_\tau & \Sigma_\tau \\ \Sigma'_\tau & \Sigma_\tau + \Sigma_\epsilon \end{pmatrix}\right).$$

However, given a site-selection subset $\mathscr{S}$, we will only observe $\hat{\tau}_s$ for $s \in \mathscr{S}$. Let $\hat{\tau}[\mathscr{S}]$ denote the subvector of $\hat{\tau}$ selected by $\mathscr{S}$, and define $\hat{\mu}_\tau[\mathscr{S}]$ analogously. For any matrix Σ, let $\Sigma[:, \mathscr{S}]$ denote the submatrix with columns selected by $\mathscr{S}$ and all rows, and let $\Sigma[\mathscr{S}, \mathscr{S}]$ denote the square submatrix with rows and columns selected by $\mathscr{S}$. Then the marginal distribution of $(\tau, \hat{\tau}[\mathscr{S}])$ is given by:

$$\begin{pmatrix} \tau \\ \hat{\tau}[\mathscr{S}] \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\tau \\ \mu_\tau[\mathscr{S}] \end{pmatrix}, \begin{pmatrix} \Sigma_\tau & \Sigma_\tau[:, \mathscr{S}] \\ \Sigma_\tau[:, \mathscr{S}]' & (\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}] \end{pmatrix}\right).$$

By standard calculations, the conditional distribution of τ given $\hat{\tau}[\mathscr{S}]$ is multivariate normal with mean

$$\mathbb{E}[\tau \mid \hat{\tau}[\mathscr{S}]] = \mu_\tau + \Sigma_\tau[:, \mathscr{S}] \left\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}]\right\}^{-1} (\hat{\tau}[\mathscr{S}] - \mu_\tau[\mathscr{S}]),$$

and variance

$$\mathbb{V}[\tau \mid \hat{\tau}[\mathscr{S}]] = \Sigma_\tau - \Sigma_\tau[:, \mathscr{S}] \left\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}]\right\}^{-1} \Sigma_\tau[:, \mathscr{S}]'.$$

With these formulas in hand, the integrals required to calculate the preposterior welfare for any choice of $\mathscr{S}$ can be approximated efficiently by numerical simulation.

To illustrate the posterior updating rule under multivariate normal priors, suppose there are two potential sites $s = 1, 2$, and under the prior their treatment effects $\tau = (\tau_1, \tau_2)'$ have distribution

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2_1 & \sigma_{12} \\ \sigma_{12} & \sigma^2_2 \end{pmatrix}\right).$$

Suppose we are considering experimenting only on site $s = 1$, so that the site selection vector is $\mathscr{S} = [1]$. As in the main text, suppose that the estimation error if an experiment is conducted in a site is independent of the site, and determined by the sample size of the prospective experiment so that the variance matrix for the estimation errors is $\Sigma_\epsilon = \sigma^2_\epsilon I$. Then we will observe $\hat{\tau}[\mathscr{S}] = \hat{\tau}_1 = \tau_1 + \epsilon_s$, which has

conditional distribution

$$\hat{\tau}_1 \mid \tau \sim N\left(\tau_1, \sigma_\epsilon^2\right).$$

If we choose to experiment in site 1 and observe $\hat{\tau}_1$, we will be able to update our beliefs about $\tau = (\tau_1, \tau_2)$ to its posterior distribution given $\hat{\tau}_1$. We apply the formulas above with

$$\hat{\tau}[\mathscr{S}] = \hat{\tau}_1, \quad \mu[\mathscr{S}] = \mu_1, \quad \Sigma_\tau[:, \mathscr{S}] = \begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \end{bmatrix}, \quad (\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}] = \left[\sigma_1^2 + \sigma_\epsilon^2\right].$$

Then, the posterior mean for τ given $\hat{\tau}_1$ is

$$\begin{aligned}
\mathbb{E}\left[\tau \mid \hat{\tau}_1\right] &= \mu_\tau + \Sigma_\tau[:, \mathscr{S}] \left\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}]\right\}^{-1} \left(\hat{\tau}[\mathscr{S}] - \mu_\tau[\mathscr{S}]\right) \\
&= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \end{bmatrix} \frac{1}{\sigma_1^2 + \sigma_\epsilon^2} \left(\hat{\tau}_1 - \mu_1\right) \\
&= \begin{pmatrix} \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}(\hat{\tau}_1 - \mu_1) \\ \mu_2 + \frac{\sigma_{12}}{\sigma_1^2 + \sigma_\epsilon^2}(\hat{\tau}_1 - \mu_1) \end{pmatrix}.
\end{aligned}$$

The posterior variance is

$$\begin{aligned}
\mathbb{V}\left[\tau \mid \hat{\tau}_1\right] &= \Sigma_\tau - \Sigma_\tau[:, \mathscr{S}] \left\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}]\right\}^{-1} \Sigma_\tau[:, \mathscr{S}]' \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} - \begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \end{bmatrix} \frac{1}{\sigma_1^2 + \sigma_\epsilon^2} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 \left(1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right) & \sigma_{12} \left(1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right) \\ \sigma_{12} \left(1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right) & \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2 + \sigma_\epsilon^2} \end{bmatrix}.
\end{aligned}$$

To interpret the posterior distribution, first consider the posterior mean of τ_1. This is

$$\mathbb{E}[\tau_1 \mid \hat{\tau}_1] = \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2} \left(\hat{\tau}_1 - \mu_1\right).$$

The second term on the right gives the incremental update to the prior mean towards the estimate $\hat{\tau}_1$. If we observe $\hat{\tau}_1 > \mu_1$, then the posterior mean will be greater than the prior mean μ_1. We can also rearrange the expression to write

$$\mathbb{E}[\tau_1 \mid \hat{\tau}_1] = \frac{\sigma_\epsilon^2}{\sigma_1^2 + \sigma_\epsilon^2} \cdot \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2} \cdot \hat{\tau}_1.$$

The posterior mean of τ_1 will be a weighted average of the prior mean μ_1 and the experimental estimate

$\hat{\tau}_1$, with weights depending on the relative uncertainty in the prior (measured by σ_1^2) and the noisiness of the estimate (through σ_ϵ^2). For example, if σ_ϵ^2 is close to zero, so that the estimate is known to be very precise, then the posterior mean will be close to the estimate $\hat{\tau}_1$.

Next, consider the posterior mean of τ_2 given $\hat{\tau}_1$:

$$\mathbb{E}[\tau_2 \mid \hat{\tau}_1] = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2 + \sigma_\epsilon^2}\left(\hat{\tau}_1 - \mu_1\right).$$

If $\sigma_{12} > 0$, so that according to the prior τ_1 and τ_2 are positively correlated because we consider them similar in some way, then observing $\hat{\tau}_1 > \mu_1$ implies that the posterior mean for τ_2 will be greater than its prior mean. On the other hand, if $\sigma_{12} < 0$, then the mean for τ_2 will be updated in the opposite direction to τ_1.

To interpret the posterior variances, note that

$$\mathbb{V}[\tau_1 \mid \hat{\tau}_1] = \sigma_1^2\left(1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right).$$

The marginal posterior variance of τ_1 will be smaller than the prior variance, unless $\sigma_1^2 = 0$ to begin with (meaning there was no prior uncertainty about τ_1), or $\sigma_\epsilon^2 = \infty$ (which implies that the estimator $\hat{\tau}_1$ is completely uninformative).

To interpret the posterior variance of τ_2, let $\rho = \sigma_{12}/(\sigma_1\sigma_2)$ be the correlation of (τ_1, τ_2) under the prior. The posterior variance can be written as

$$\mathbb{V}[\tau_2 \mid \hat{\tau}_1] = \sigma_2^2\left(1 - \frac{\rho^2\sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right).$$

If $\sigma_1^2 > 0$ and $\sigma_\epsilon^2 < \infty$, then the posterior variance will be smaller than the prior variance (σ_2^2) by a factor that depends on the correlation between τ_1 and τ_2. If $\rho = 0$ then there is no information about τ_2 from $\hat{\tau}_1$, and the posterior variance will equal the prior variance.

D Smoothing Prior Parameters

Recall that the smoothing prior has $\tau \mid V \sim N_S\left(\mu, \Sigma(V)\right)$, where $\Sigma(V)$ is a covariance matrix with elements

$$\mathbb{C}(\tau_s, \tau_{s'}) = c_0 \cdot \exp\left(-c_1^{-1}\left\|h(V_s) - h(V_{s'})\right\|^2\right),$$

$h(V_s)$ is some low-dimensional function of the characteristics V_s, and $c_0 > 0$ and $c_1 > 0$ control the overall scale of the variance matrix and the amount of local smoothing, respectively. Write $d_{ss'} := \|h(V_s) - h(V_{s'})\|$ for the distance between sites s and s' in transformed characteristic space, so that $\mathbb{C}(\tau_s, \tau_{s'}) = c_0 \exp\left(-c_1^{-1} d_{ss'}^2\right)$.

This appendix first illustrates in Section D.1 how c_1 controls how local the learning from experimental results is. We next show that c_1 can alternatively be identified from expert forecasts (Section D.2). Finally, we derive the naive extrapolation rule of Section 5.3.1 as the limit of the posterior-optimal treatment rule when c_1 and c_0 grow large (Section D.3).

D.1 How c_1 Governs Updating

Figure A7 illustrates how c_1 changes updating behavior in our Bangladesh application. Sites are migration corridors and the smoothing prior uses two site characteristics: the distance between a corridor's destination and origin districts and household income at the origin (see Section 5.1.3). There we take $h(\cdot)$ to be the identity function applied to the two characteristics, each standardized to have zero mean and unit standard deviation across sites. Each panel plots the candidate sites by their two characteristics in raw units. The prior measures the distance between sites after standardization. Stars mark the two corridors treated as experimental sites for the illustration, Dhaka–Bagerhat and Dhaka–Kishoregonj.

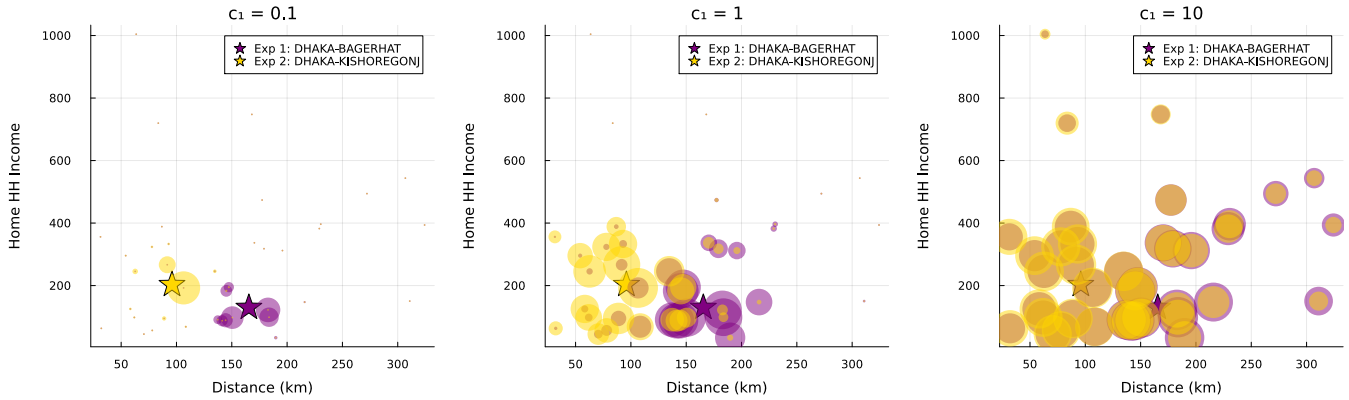
To measure how much we learn about a site's ATE from an experiment run in another site, consider the two-site example from Appendix C: if site s' is the only experimental site, then by Gaussian conditioning the posterior mean for the other site s moves with the estimate $\hat{\tau}_{s'}$ according to

$$\mathbb{E}[\tau_s \mid \hat{\tau}_{s'}] = \mu_s + \beta_{ss'}\left(\hat{\tau}_{s'} - \mu_{s'}\right), \qquad \beta_{ss'} := \frac{\mathbb{C}(\tau_s, \tau_{s'})}{\mathbb{V}(\tau_{s'}) + \sigma_\epsilon^2},$$

where μ_s is element s of the prior mean vector μ and σ_ϵ^2 is the sampling variance of the experimental estimate. We call $\beta_{ss'}$ the updating weight of site s on experimental site s': it multiplies the deviation of the estimate in site s' from its prior mean in the posterior mean for τ_s. In Figure A7, each site is drawn with one circle per experimental site, colored to match that experimental site's star, and a circle's size increases with the site's updating weight $\beta_{ss'}$ on that experiment, computed one experimental site at a time.

The panels vary c_1. When c_1 is small the exponent $-c_1^{-1} d_{ss'}^2$ is far below zero for all but the closest pairs of sites so the prior covariance, and with it the weight $\beta_{ss'}$, falls quickly with characteristic distance. In the left panel, with $c_1 = 0.1$, learning is hyper-local: only sites close to an experimental site in characteristic space update meaningfully. As c_1 grows, the exponent rises toward zero for every pair of sites, and prior correlation extends across the whole characteristic space. The middle panel shows $c_1 = 1$, the value we use in the application (Section 5.1.3). As noted there, $c_1 = 1$ implies that two sites one cross-site standard deviation apart in characteristic space have prior correlation $\exp(-1) \approx 0.37$. In the right panel, with $c_1 = 10$, the prior extrapolates widely: even sites far from the experimental sites update substantially. The limiting case $c_1 \to \infty$, in which distance stops mattering entirely, is the starting point of Appendix D.3.

Figure A7: The Smoothing Prior Parameter c_1 Controls How Local Learning Is



Notes: Each panel plots Bangladesh's candidate sites by destination–origin distance and home household income, the two characteristics we use in the smoothing prior (Equation (3)) in our application. Stars mark two hypothetical experimental sites. Each site is drawn with one circle per experimental site, colored to match that site's star. A circle's size is proportional to the weight the site's posterior mean would place on the estimate from that experimental site, were it the only experiment run (the updating weight $\beta_{ss'}$ defined in Appendix D.1). Axes are in raw units, the prior computes distances on the standardized characteristics.

D.2 Identifying c_1 from Forecasts

In a companion paper (Sanchez-Becerra et al., 2026), we ask experts to predict the effects of the mobile money intervention in the six selected migration corridors, two in each of the three countries of our application. Viewing an expert's beliefs through the lens of the smoothing prior, their point and probabilistic predictions for site-level ATEs pin down all the prior hyperparameters, except c_1. Point predictions identify the prior means of the surveyed corridors. Uncertainty about any single corridor's effect identifies c_0: setting $s' = s$ in the smoothing prior covariance function gives $d_{ss} = 0$ and hence $\mathbb{V}(\tau_s) = c_0$, so every corridor's prior variance equals c_0. The parameter c_1, by contrast, enters only the covariance between distinct sites, so no question about a single site's effect in isolation can identify it.

We can, however, identify c_1 by asking how an expert would revise their prediction for one site after seeing a result from another. We present the expert with a hypothetical experimental result y for site s' and ask for their revised prediction for site s. Treating the expert's beliefs as the smoothing prior and the hypothetical result as known without estimation error, Gaussian conditioning as in Appendix C gives the revised prediction

$$\mathbb{E}\left[\tau_s \mid \tau_{s'} = y\right] = \mu_s + \frac{\mathbb{C}(\tau_s, \tau_{s'})}{\mathbb{V}(\tau_{s'})}\left(y - \mu_{s'}\right).$$

The revision in s, relative to the revision in s', therefore reveals the ratio

$$\frac{\mathbb{E}\left[\tau_s \mid \tau_{s'} = y\right] - \mu_s}{y - \mu_{s'}} = \frac{\mathbb{C}(\tau_s, \tau_{s'})}{\mathbb{V}(\tau_{s'})} = \frac{c_0 \exp\left(-c_1^{-1} d_{ss'}^2\right)}{c_0} = \exp\left(-c_1^{-1} d_{ss'}^2\right),$$

where the second equality substitutes in the smoothing prior covariance function. This ratio equals

the updating weight $\beta_{ss'}$ defined above, with $\sigma_\epsilon^2 = 0$ in the updating weight, because we present the hypothetical result as known rather than estimated. Because the distance $d_{ss'}$ is computable directly from the site characteristics, inverting the expression identifies c_1. We can solve for c_1 in

$$\frac{\mathbb{E}\left[\tau_s \mid \tau_{s'} = y\right] - \mu_s}{y - \mu_{s'}} = \exp\left(-c_1^{-1} d_{ss'}^2\right),$$

The more heavily an expert discounts revisions from site s' when revising their prediction for site s, the smaller is $\beta_{ss'}$ and the smaller the implied c_1.

D.3 Naive Extrapolation as a Limiting Case of the Smoothing Prior

In Section 5.3.1 we describe a "naive" policymaker who assigns the treatment to all sites if the average of the experimental ATEs exceeds the cost of treatment. Here we show that this policy rule arises as a limiting case of the smoothing prior as its tuning parameters c_1 and c_0 grow large.

As in the pure smoothing prior of Section 5.1.3, we take a mean that is constant across sites: $\mu_s = \bar{\mu}$ for all s. Experiments are run in the K sites in $\mathscr{S}$, yielding estimates $\hat{\tau}_s = \tau_s + \epsilon_s$ for $s \in \mathscr{S}$, with the ϵ_s independent $N(0, \sigma_\epsilon^2)$ errors, and the cost of treatment is homogeneous across sites, $\text{cost}_s = \text{cost}$, as we assume within each country in our application (Section 5.1.2). Because the welfare function is additive across sites, the posterior-optimal treatment rule treats site s if and only if $\mathbb{E}[\tau_s \mid \hat{\tau}_{\mathscr{S}}] \geq \text{cost}_s$, as in Equation (4).

First let $c_1 \to \infty$. Fix a pair of sites (s, s'). The sites, and hence the distance $d_{ss'}$ between their transformed characteristics, are fixed as c_1 varies, so $c_1^{-1} d_{ss'}^2 \to 0$ and, by continuity of the exponential, $\mathbb{C}(\tau_s, \tau_{s'}) = c_0 \exp(-c_1^{-1} d_{ss'}^2) \to c_0$. Since $\Sigma(V)$ has finitely many elements, entrywise convergence is convergence of the matrix: the prior covariance matrix converges to $c_0 \mathbf{1}_S \mathbf{1}_S'$, the $S \times S$ matrix with c_0 in every cell, which is the outer product of $\sqrt{c_0}\,\mathbf{1}_S$ with itself and hence has rank one. In words, c_1 governs how quickly prior correlation decays with distance in characteristics. With infinite c_1, distance stops mattering, and every pair of sites is as correlated as a site is with itself.

Under the limiting covariance matrix, every variance and every covariance equals c_0, so for any pair of sites

$$\mathbb{V}(\tau_s - \tau_{s'}) = \mathbb{V}(\tau_s) + \mathbb{V}(\tau_{s'}) - 2\,\mathbb{C}(\tau_s, \tau_{s'}) = c_0 + c_0 - 2c_0 = 0,$$

and since $\mathbb{E}[\tau_s - \tau_{s'}] = \bar{\mu} - \bar{\mu} = 0$, all site effects are equal. Define $\delta := \tau_s - \bar{\mu}$, the same random variable for every s. As a linear function of the Gaussian vector τ, δ is Gaussian, with mean zero and variance c_0, so under the limiting prior we can write, exactly,

$$\tau_s = \bar{\mu} + \delta \text{ for all } s, \qquad \delta \sim N(0, c_0).$$

Substituting this representation into the sampling model gives $\hat{\tau}_s = \bar{\mu} + \delta + \epsilon_s$ for $s \in \mathscr{S}$: conditional on δ, the experimental estimates are K independent noisy observations of the common shift δ, with the independence carried entirely by the errors ϵ_s. Standard normal updating, as in Appendix C, gives

$$\mathbb{E}[\delta \mid \hat{\tau}_{\mathscr{S}}] = w_K(\bar{\hat{\tau}} - \bar{\mu}), \qquad \bar{\hat{\tau}} := \frac{1}{K}\sum_{s\in\mathscr{S}} \hat{\tau}_s, \qquad w_K := \frac{Kc_0}{Kc_0 + \sigma_\epsilon^2} \in (0,1),$$

and hence, for every site s, experimental or not,

$$\mathbb{E}[\tau_s \mid \hat{\tau}_{\mathscr{S}}] = \bar{\mu} + w_K(\bar{\hat{\tau}} - \bar{\mu}).$$

All sites share a common posterior mean: the policymaker pools the experimental estimates and applies them uniformly, without adjusting for how the experimental sites compare to any target site. With homogeneous costs, the treatment decision is already an all-or-nothing rule, but the pooled average is still shrunk toward the prior mean $\bar{\mu}$.

Now let $c_0 \to \infty$, so that the limiting prior is diffuse. Then $w_K \to 1$ and $\mathbb{E}[\tau_s \mid \hat{\tau}_{\mathscr{S}}] \to \bar{\hat{\tau}}$ for every site s. The posterior-optimal rule (4) becomes: treat all sites if $\bar{\hat{\tau}} \geq$ cost, and no sites otherwise. This is exactly the naive policy rule.

E Model Solutions

E.1 Optimal Choices Given Remittance Price

The model laid out in Section E.1 can be solved to obtain the optimal remittance amount R^*in closed form. Depending on the values of the parameters, optimal remittances may be zero, or strictly positive as indicated in the following equation:

$$R^* = \begin{cases} 0 & \text{if} \quad p_r \geq \frac{24\lambda a_h \eta w_m}{p_h y_h} \\ p_h\left\{\frac{\lambda a_h \eta(24w_m + p_r p_h y_h)}{p_h p_r(1+\lambda a_h \eta)} - y_h\right\} & \text{if} \quad p_r < \frac{24\lambda a_h \eta w_m}{p_h y_h} \end{cases}. \tag{6}$$

The mapping $\lambda \mapsto R^*$ is monotonic and continuous at the boundary points. The optimal values of migrant consumption and leisure are:

$$C_m^* = \frac{(24w_m + p_r p_h y_h)\cdot\alpha}{1+\lambda a_h \eta}; \quad L_m^* = \frac{(24w_m + p_r p_h y_h)\cdot(1-\alpha)}{w_m\left(1+\lambda a_h \eta\right)}.$$

E.2 Optimal Distribution of Remittances with Stochastic Remittance Price

Let X denote the vector of household characteristics, including treatment status t:

$$X = (p_h, a_h, w_m, y_h, \text{male}, o, d, t).$$

Let the outcome variables be

$$Y = (R^*, M^* \cdot \mathbf{1}(R^* > 0)).$$

We first derive the distribution of the price to remit p_r. Given the distributions of ϵ, ξ, we have:

$$\begin{aligned}
\mathbb{P}\left(p_r \geq \tilde{p} \mid X\right) &= \exp\left(\min\left\{-\frac{\tilde{p}-1}{\bar{\epsilon}\cdot d}, 0\right\}\right) \cdot \exp\left(\min\left\{-\frac{\tilde{p}-1-\gamma}{\bar{\xi}\cdot e^{\psi\cdot t+\delta\cdot o}}, 0\right\}\right) \\
&= \begin{cases} 1 & \text{if} \quad \tilde{p} < 1, \\ \exp\left(-\frac{\tilde{p}-1}{\bar{\epsilon}\cdot d}\right) & \text{if} \quad 1 \leq \tilde{p} < 1+\gamma, \\ \exp\left(-\frac{\tilde{p}-1}{\bar{\epsilon}\cdot d} - \frac{\tilde{p}-1-\gamma}{\bar{\xi}\cdot e^{\psi\cdot t+\delta\cdot o}}\right) & \text{if} \quad 1+\gamma \leq \tilde{p}. \end{cases}
\end{aligned}$$

Given the expression for optimal remittances, for any value $r \geq 0$ we have

$$\mathbb{P}(R^* \leq r \mid X) = \mathbb{P}\left(p_r \geq \frac{24\lambda a_h w_m \eta}{r(1+\lambda a_h \eta) + p_h y_h} \;\middle|\; X\right). \tag{7}$$

The optimal distribution of remittances now follows from the expression for $\Pr(p_r \geq \tilde{p} \mid X)$ for any $\tilde{p} \in \mathbb{R}_+$.

E.3 Probability of Remitting Using Mobile Money

Another quantity of interest is the probability that a migrant will remit a positive amount through mobile money. This probability, conditional on migrant characteristics, is

$$\begin{aligned}
\mathbb{P}(M^* = 1, R^* > 0 | X) = \exp\left(\frac{-\gamma}{\bar{\epsilon}d}\right) \cdot \frac{\bar{\epsilon}d}{\bar{\epsilon}d + \bar{\xi}e^{\psi t+\delta o}} \\
\times \mathbf{1}\left\{w_m \geq \frac{(1+\gamma)\cdot p_h y_h}{24\lambda a_h \eta}\right\} \cdot \left(1 - \exp\left(-\left(\frac{1}{\bar{\xi}e^{\psi\cdot t+\delta o}} + \frac{1}{\bar{\epsilon}d}\right) \cdot \left(\frac{24\lambda a_h \eta w_m}{p_h y_h} - 1 - \gamma\right)\right)\right).
\end{aligned} \tag{8}$$

The indicator variable in Equation (8) reflects the necessary condition for mobile money to be the preferred mode of remittance.

F Asymptotic Distribution of the Structural Parameter Estimator

The minimum distance estimator $\hat{\theta}$ solves:

$$\min_{\theta} \left(\hat{\pi} - q(\theta)\right)' \hat{W}_n \left(\hat{\pi} - q(\theta)\right),$$

where $\hat{W}_n \xrightarrow{p} W$ for some W. We assume there is a unique solution θ_0 to the population minimum distance problem

$$\min_{\theta} (\pi_0 - q(\theta))' W (\pi_0 - q(\theta)).$$

Under suitable conditions, we will have

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta),$$

where

$$\Sigma_\theta = (G'WG)^{-1} G'W\Sigma_\pi WG(G'WG)^{-1},$$

and $G = \frac{\partial}{\partial \theta} q(\theta_0)$. In order to estimate the sampling variance of $\hat{\theta}$, we estimate G by evaluating the Jacobian at $\theta = \hat{\theta}$, taking the derivative analytically (see Appendix F.2 for the specific expressions). The matrix Σ_π is the asymptotic variance matrix of $\hat{\pi}$, which can also be estimated using the expressions in Appendix F.1.

F.1 Asymptotic Distribution of Reduced Form Estimator

The vector π in the minimum distance estimator consists of elements of the form

$$\mathbb{P}(R \leq r \mid X \in A), \quad \mathbb{P}(M = 1 \mid R > 0, X \in A),$$

for suitable choices of r, X, A. In addition, π contains one component equal to the expression for α in Equation (5). The reduced form estimator $\hat{\pi}$ consists of sample analogs of the population quantities in π. Under i.i.d. sampling and ruling out degenerate cases, we will have

$$\sqrt{n}(\hat{\pi} - \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i + o_P(1) \xrightarrow{d} N(0, \Sigma_\pi),$$

where $\Sigma_\pi = \mathbb{E}\left[\zeta_i \zeta_i'\right]$. The asymptotic variance can be estimated by the bootstrap, or by a sample analog

$$\hat{\Sigma}_\pi = \frac{1}{n} \sum_{i=1}^{n} \hat{\zeta}_i \hat{\zeta}_i,$$

where the $\hat{\zeta}_i$ replace unknown population quantities in the formulas for ζ_i with sample estimates. For completeness, we give specific expressions for the influence function ζ_i next.

For components of π containing quantities of the form $\mathbb{P}(R \le r \mid X \in A)$, the corresponding component of ζ_i is

$$\frac{\mathbf{1}\{R_i \le r, X_i \in A\} - \mathbb{P}(R \le r, X \in A)}{\mathbb{P}(X \in A)} - \frac{\mathbb{P}(R \le r \mid X \in A)}{\mathbb{P}(X \in A)}\Big(\mathbf{1}\{X_i \in A\} - \mathbb{P}(X \in A)\Big).$$

For components of π containing quantities of the form $\mathbb{P}(M = 1 \mid R > 0, X \in A)$, the corresponding component of ζ_i is

$$\frac{\mathbf{1}\{M_i = 1, R_i > 0, X_i \in A\} - \mathbb{P}(M = 1, R > 0, X \in A)}{\mathbb{P}(R > 0, X \in A)} - \frac{\mathbb{P}(M = 1 \mid R > 0, X \in A)}{\mathbb{P}(R > 0, X \in A)}\Big(\mathbf{1}\{R_i > 0, X_i \in A\} - \mathbb{P}(R > 0, X \in A)\Big)$$

Finally, for the element of π equal to α, we have the following expression for the corresponding element of ζ_i:

$$\frac{\mathbb{E}[E_m]^2}{(\mathbb{E}[E_m] + \mathbb{E}[W_m] \cdot \mathbb{E}[L_m])^2}\left\{\frac{\mathbb{E}[W_m] \cdot \mathbb{E}[L_m]}{\mathbb{E}[E_m]^2} \cdot \big(E_{m,i} - \mathbb{E}[E_m]\big) - \frac{\mathbb{E}[W_m]}{\mathbb{E}[E_m]}\big(L_{m,i} - \mathbb{E}[L_m]\big) - \frac{\mathbb{E}[L_m]}{\mathbb{E}[E_m]}\big(W_{m,i} - \mathbb{E}[W_m]\big)\right\}.$$

F.2 Derivatives of Estimation Moments

Consider the estimation moments,

$$\mathbb{P}(R^*(\theta) \le r \mid T = t, \text{male} = m).$$

Fixing t and m, and denoting $\lambda = \exp(\phi_0 + \phi_1 \cdot m)$, the model (refer to Equation (7)) implies the following value for the joint probability $\mathbb{P}(R^*(\theta) \le r, T = t, \text{male} = m)$:

$$\sum_a \int_0^\infty \int_0^{\frac{r(1+\lambda a\eta)+p_h y}{24\lambda a\eta}} f_X(w, y, a)\, dw dy$$

$$+ \sum_a \int_0^\infty \int_{\frac{r(1+\lambda a\eta)+p_h y}{24\lambda a\eta}}^{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}} \exp\left(-\frac{1}{\bar{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y} - 1\right)\right) f_X(w, y, a)\, dw dy$$

$$+ \sum_a \int_0^\infty \int_{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}}^\infty \exp\left(-\frac{1}{\bar{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y} - 1\right) - \frac{1}{\bar{\xi} e^{\psi t + \delta o}}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y} - 1 - \gamma\right)\right) f_X(w, y, a)\, dw dy,$$

where f_X denotes the joint density function for (w_m, y_h, a_h). The derivatives of this object with respect to $(\psi, \lambda, \overline{\epsilon}, \overline{\xi})$ are displayed below.

$$\frac{\partial \mathbb{P}(R^*(\theta) \leq r, T = t, \text{male} = m)}{\partial \lambda} =$$

$$-\sum_a \int_0^\infty \int_{\frac{r(1+\lambda a\eta)+p_h y}{24\lambda a\eta}}^{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}} \exp\left(-\frac{1}{\overline{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)\right)\cdot\left(\frac{(r+p_h y)24a\eta w}{(r(1+\lambda a\eta)+p_h y)^2}\cdot\frac{1}{\overline{\epsilon} d}\right) f_X(w,y,a)\,dw dy$$

$$-\sum_a \int_0^\infty \int_{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}}^{\infty} \exp\left(-\frac{1}{\overline{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)-\frac{1}{\overline{\xi} e^{\psi t+\delta o}}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1-\gamma\right)\right)$$

$$\times\left(\frac{(r+p_h y)24a\eta w}{(r(1+\lambda a\eta)+p_h y)^2}\cdot\left(\frac{1}{\overline{\epsilon} d}+\frac{1}{\overline{\xi} e^{\psi t+\delta o}}\right)\right) f_X(w,y,a)\,dw dy.$$

$$\frac{\partial \mathbb{P}(R^*(\theta) \leq r, T = t, \text{male} = m)}{\partial \overline{\epsilon}} =$$

$$\sum_a \int_0^\infty \int_{\frac{r(1+\lambda a\eta)+p_h y}{24\lambda a\eta}}^{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}} \exp\left(-\frac{1}{\overline{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)\right)\cdot\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)\cdot\frac{1}{\overline{\epsilon}^2 d} f_X(w,y,a)\,dw dy$$

$$+\sum_a \int_0^\infty \int_{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}}^{\infty} \exp\left(-\frac{1}{\overline{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)-\frac{1}{\overline{\xi} e^{\psi t+\delta o}}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1-\gamma\right)\right)$$

$$\times\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)\cdot\frac{1}{\overline{\epsilon}^2 d} f_X(w,y,a)\,dw dy.$$

$$\frac{\partial \mathbb{P}(R^*(\theta) \leq r, T = t, \text{male} = m)}{\partial \overline{\xi}} =$$

$$\sum_a \int_0^\infty \int_{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}}^{\infty} \exp\left(-\frac{1}{\overline{\epsilon} d}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1\right)-\frac{1}{\overline{\xi} e^{\psi t+\delta o}}\left(\frac{24\lambda a w\eta}{r(1+\lambda a\eta)+p_h y}-1-\gamma\right)\right)$$

$$\times \left(\frac{24\lambda a w \eta}{r(1+\lambda a \eta) + p_h y} - 1 - \gamma \right) \cdot \frac{1}{\overline{\xi}^2 e^{\psi t + \delta o}} \, f_X(w, y, a) \, dw dy.$$

$$\frac{\partial \mathbb{P}(R^*(\theta) \leq r, T = t, \text{male} = m)}{\partial \psi} =$$

$$\sum_a \int_0^{\infty} \int_{\frac{(1+\gamma)\cdot(r(1+\lambda a\eta)+p_h y)}{24\lambda a\eta}}^{\infty} \exp\left(-\frac{1}{\overline{\epsilon} d} \left(\frac{24\lambda a w \eta}{r(1+\lambda a\eta) + p_h y} - 1 \right) - \frac{1}{\overline{\xi} e^{\psi t + \delta o}} \left(\frac{24\lambda a w \eta}{r(1+\lambda a\eta) + p_h y} - 1 - \gamma \right) \right)$$

$$\times \left(\frac{24\lambda a w \eta}{r(1+\lambda a\eta) + p_h y} - 1 - \gamma \right) \cdot \frac{t}{\overline{\xi} e^{\psi t + \delta o}} \, f_X(w, y, a) \, dw dy.$$

Estimation also involves the probability of using mobile money, conditional on remitting a positive amount. Given characteristics X, this is the probability assigned to the event:

$$\gamma + \xi e^{\psi \cdot t + \delta \cdot o} \leq \epsilon \cdot d \quad \text{and} \quad p_r \leq \frac{24\lambda a_h \eta w_m}{p_h \cdot y_h}.$$

The above event is non-empty only if,

$$w_m \geq (1+\gamma) \cdot \frac{p_h \cdot y_h}{24\lambda a_h \eta}.$$

Once again, fixing t, m and denoting $\lambda = \exp(\phi_0 + \phi_1 \cdot m)$, recall that the probability $\mathbb{P}(M^* = 1, R^* > 0 \mid T = t, \text{male} = m)$ equals

$$\exp\left(\frac{-\gamma}{\overline{\epsilon} d}\right) \cdot \frac{\overline{\epsilon} d}{\overline{\epsilon} d + \overline{\xi} e^{\psi t + \delta o}}$$
$$\times \sum_a \int_0^{\infty} \int_{\frac{(1+\gamma)\cdot p_h y}{24\lambda a\eta}}^{\infty} \left(1 - \exp\left(- \left(\frac{1}{\overline{\xi} e^{\psi \cdot t + \delta o}} + \frac{1}{\overline{\epsilon} d} \right) \cdot \left(\frac{24\lambda a \eta w}{p_h y} - 1 - \gamma \right) \right) \right) f_X(w, y, a) \; dw dy.$$

Denote the integral shown above as $I(\theta)$, and the function multiplying it $C(\theta)$. Partial derivatives are displayed below.

$$\frac{\partial \mathbb{P}(M^* = 1, R^* > 0, T = t, \text{male} = m)}{\partial \lambda} =$$

$$C(\theta) \cdot \sum_a \int_0^{\infty} \int_{\frac{(1+\gamma)\cdot p_h y}{24\lambda a\eta}}^{\infty} \exp\left(- \left(\frac{1}{\overline{\xi} e^{\psi \cdot t + \delta o}} + \frac{1}{\overline{\epsilon} d} \right) \cdot \left(\frac{24\lambda a \eta w}{p_h y} - 1 - \gamma \right) \right) \left(\frac{1}{\overline{\xi} e^{\psi \cdot t + \delta o}} + \frac{1}{\overline{\epsilon} d} \right) \cdot \frac{24\eta w a}{p_h \cdot y} \, f_X(w, y, a) \, dw dy.$$

$$\frac{\partial \mathbb{P}(M^*=1, R^*>0, T=t, \text{male}=m)}{\partial \bar{\epsilon}} =$$

$$-C(\theta)\cdot\sum_a \int_0^{\infty}\int_{\frac{(1+\gamma)\cdot p_h y}{24\lambda a\eta}}^{\infty} \exp\left(-\left(\frac{1}{\bar{\xi}e^{\psi\cdot t+\delta o}}+\frac{1}{\bar{\epsilon}d}\right)\cdot\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\right)\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\frac{1}{\bar{\epsilon}^2 d}\, f_X(w,y,a)\,dw dy$$

$$+\exp\left(\frac{-\gamma}{\bar{\epsilon}d}\right)\cdot\left[\frac{\gamma}{\bar{\epsilon}\cdot\left(\bar{\epsilon}d+\bar{\xi}e^{\psi t+\delta o}\right)}+\frac{\bar{\xi}e^{\psi t+\delta o}d}{\left(\bar{\epsilon}d+\bar{\xi}e^{\psi t+\delta o}\right)^2}\right]\cdot I(\theta).$$

$$\frac{\partial \mathbb{P}(M^*=1, R^*>0, T=t, \text{male}=m)}{\partial \bar{\xi}} =$$

$$-C(\theta)\cdot\sum_a \int_0^{\infty}\int_{\frac{(1+\gamma)\cdot p_h y}{24\lambda a\eta}}^{\infty} \exp\left(-\left(\frac{1}{\bar{\xi}e^{\psi\cdot t+\delta o}}+\frac{1}{\bar{\epsilon}d}\right)\cdot\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\right)\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\frac{1}{\bar{\xi}^2 e^{\psi t+\delta o}}\, f_X(w,y,a)\,dw dy$$

$$-\exp\left(\frac{-\gamma}{\bar{\epsilon}d}\right)\cdot\left[\frac{\bar{\epsilon}d e^{\psi t+\delta o}}{\left(\bar{\epsilon}d+\bar{\xi}e^{\psi t+\delta o}\right)^2}\right]\cdot I(\theta).$$

$$\frac{\partial \mathbb{P}(M^*=1, R^*>0, T=t, \text{male}=m)}{\partial \psi} =$$

$$-C(\theta)\cdot\sum_a \int_0^{\infty}\int_{\frac{(1+\gamma)\cdot p_h y}{24\lambda a\eta}}^{\infty} \exp\left(-\left(\frac{1}{\bar{\xi}e^{\psi\cdot t+\delta o}}+\frac{1}{\bar{\epsilon}d}\right)\cdot\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\right)\left(\frac{24\lambda a_h\eta w_m}{p_h y_h}-1-\gamma\right)\frac{t}{\bar{\xi}e^{\psi t+\delta o}}\, f_X(w,y,a)\,dw dy$$

$$-\exp\left(\frac{\gamma}{\bar{\epsilon}d}\right)\cdot\frac{t}{\bar{\xi}\cdot e^{\psi\cdot t+\delta o}}\left(\frac{2}{\bar{\xi}e^{\psi t+\delta o}}+\frac{1}{\bar{\epsilon}d}\right)\cdot I(\theta).$$

G Parameters Normalized and Estimated Separately

The parameter γ, reflecting the price of remittance transfer via mobile money, is obtained directly from the actual price schedules as described in Section J.1. The parameter η will be normalized to equal one, because only the product of η and λ appear in the expression for the distribution of optimal remittances in Equation (7).

A similar issue arises for inferring δ and $\bar{\xi}$. The experimental data has no variation in operator density since all observations came from a single migration corridor. For estimation purposes we set $\delta = 0$ and estimate the parameter $\bar{\xi}$. In Section 4.2 we describe how the model is used to make predictions about

treatment effects in sites not covered in the study of Lee et al. (2021). These sites have different values for operator density, and we can incorporate this information by calibrating a value for δ given our estimate of $\bar{\xi}$. In particular, let o_g denote the operator density in Gaibandha, and let $\hat{\bar{\xi}}$ denote the estimated value of $\bar{\xi}$. We then set:

$$\hat{\delta} = \frac{\ln \hat{\bar{\xi}}}{o_g}.$$

H Model Fit to the Pilot Experiment

Table A6 displays model implied values for the treatment effect on daily remittances sent and the probability of remitting a positive amount via mobile money (bKash).

Table A6: Model-Implied and Estimated Average Treatment Effects

Calculated Using	Daily Remittances	$\mathbb{P}(MM = 1, R > 0)$
Lee et al. (2021) Data	6.41	0.10
Structural Model	5.94	0.11

Figures A8 and A9 display the observed and model implied cumulative distribution functions of average daily remittances when the model of Section 3 is fit to the experimental data of Lee et al. (2021). We use the point estimates in Table 1 to generate the model-implied distributions of daily remittances, and show the empirical and model-implied CDFs separately by gender and treatment status. The parsimonious model largely fits the data well, with the exception of overpredicting the treatment effect on remittances for female migrants, which consitute a small share of the total sample, as shown in Table A1.

Figure A8: Average Daily Remittances for Female Migrants

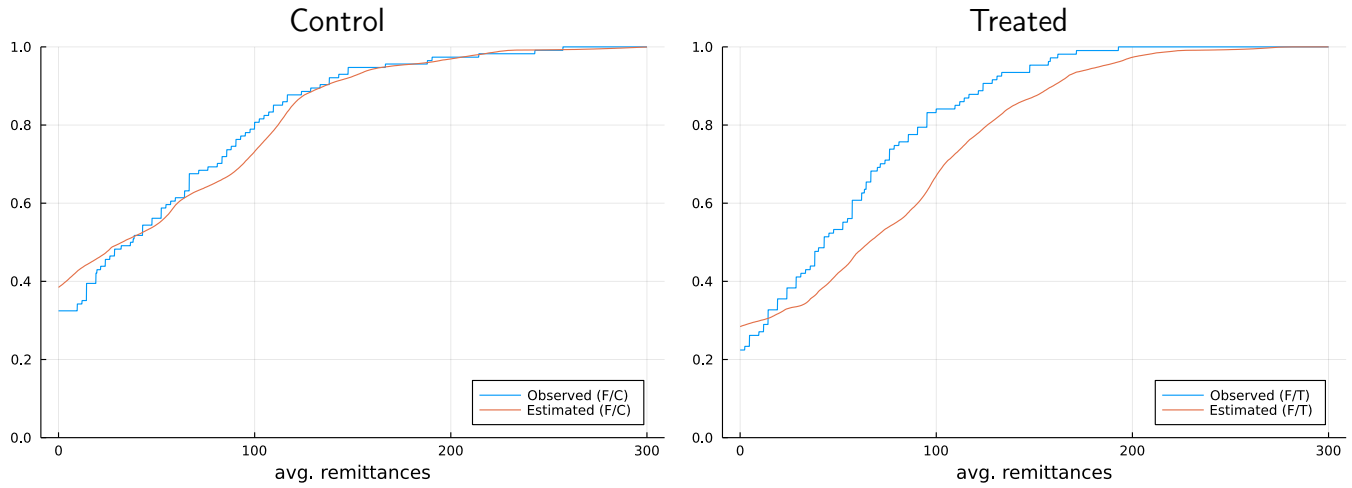
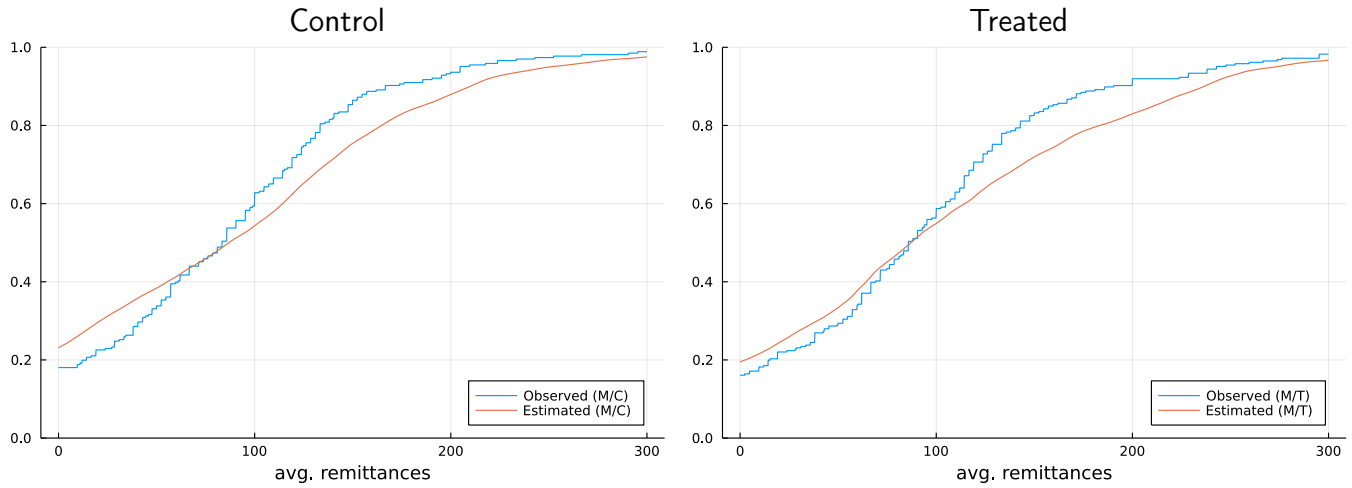


Figure A9: Average Daily Remittances for Male Migrants



Notes: observed refers to the empirical CDF from the Lee et al. (2021) data, estimated to the structural model-implied CDF at the estimated parameters.

I Deriving the Welfare Cost of Treating a Site

Following the standard benchmark in the public finance literature (c.f. Finkelstein and Hendren 2020), we assume taxpayers incur a 30% deadweight loss, so that the cost of the program from their perspective is 1152 taka.[36] Consider the consumption tax rate ω needed to pay for the program on a per-migrant basis. As above, u_j for taxpayer j is per-capita expenditure in j's household. To pay for the program with a uniform consumption tax at rate ω, u_j must go down by

$$\omega u_j = \omega \cdot \frac{\text{household expenditure}_j}{\text{hh size}_j}.$$

We take a first-order approximation to the impact of this change in u_j on social welfare, multiplying the change in u_j by the social marginal utility associated with u_j. Because of the log specification for the social value of u_j, the social marginal utility of a change in u_j is

$$\frac{\text{hh size}_j}{\text{household expenditure}_j}.$$

All members j within the same household have the same social marginal utility and change in u_j so we can sum over households h to get the aggregate effect of the the tax,

$$\sum_h \text{hh size}_h \cdot \omega. \tag{9}$$

[36] This may differ in developing countries — for example because of lower overall tax burdens, a lower labor supply elasticity and the mode of taxes (e.g. VAT vs. labor tax). These factors may result in a different, potentially lower deadweight loss.

It remains to derive the size of ω needed to pay for the program on a per migrant basis. This solves:

$$1152 = \omega \cdot \text{total expenditure in the country.} \tag{10}$$

Combining (9) and (10), we can rewrite the treatment cost in welfare terms as:

$$1152 \cdot \frac{1}{\text{national per-capita expenditure}}.$$

To compute national per-capita expenditure in the three countries, we use the "Final consumption expenditure" series from the World Bank's World Development Indicators database in constant local currency units and divide it by the contemporaneous value of total population, also from World Development Indicators. For compatibility with the per-capita expenditures from Lee et al. (2021), we convert "Final consumption expenditure" to the monthly level and use 2015 values. For Pakistan and India, we convert rupee values to Bangladesh taka using the April 1, 2015 market exchange rate from xe.com.

J Migration Corridor Data and Restrictions

J.1 Data on Corridor Characteristics

As discussed in Section 4.2, to make predictions for future experimental sites or migration corridors, we need site-level data such as the migrant's wage rate and their family's income. This information is taken from multiple sources for each country. Monetary values are adjusted for inflation to ensure that all values accord with the original experiment of Lee et al. (2021). Home households are defined as those who report having sent a domestic migrant or, when this information is not available, receiving remittances from a domestic source.

For all three countries, we measure mobile money operator density in the origin district information from the latest available waves of InterMedia's Financial Inclusion Insights surveys, conducted in collaboration with the Bill and Melinda Gates Foundation. For Bangladesh and India, we use data from 2018, while for Pakistan we use data from 2020. Our operator density measure is based on the time it takes an individual to reach their nearest mobile money agent. A lower average time corresponds to higher operator density in a district. To ease interpretation of this measure as operator density, we invert it so that it is 1 divided by the time taken (in minutes) to the nearest mobile money operator. We measure the length of a corridor as the distance in kilometers as the crow flies between the centroid of the origin district and the destination district.

The explicit per-unit monetary cost of remitting via mobile money, parameterized by γ, is calculated separately for each country. For Bangladesh we use the per-unit cost to remit using mobile money as reported in Lee et al. (2021). For Pakistan, we use the cost schedule from a popular mobile money service

(Telenor Microfinance Bank) to construct the parameter. For India, we construct γ using the cost of remitting though common digital transfer methods available to those who have bank accounts at post offices. Specifically, we use the per-unit cost of transferring money from one bank account to another through the National Electronic Funds Transfer (NEFT) protocol.[37]

Bangladesh For Bangladesh, we use the Bangladesh Household Income and Expenditure Survey 2016 (BHIES) for home household incomes and size. We only use data for households which report having sent a migrant at any point in the past, and for which the migrant in question had been working within the country at the time of the survey. A household surveyed in the BHIES reports if any of their members have migrated in the past, where they migrated to, and their age and sex, but does not record their wages which we need for our analysis. To resolve this missing data problem we turn to another data source. The Bangladesh Integrated Household Survey 2015 (BIHS) contains wage and demographic information for migrants, but is a much smaller survey than the BHIES. To leverage the BHIES, we estimate a wage model using data from the BIHS. The model specifies that

$$\ln w_{\text{dist},l} = \beta_0 + \beta_1 \cdot \text{age}_l + \beta_2 \cdot \text{age}_l^2 + \beta_3 \cdot \text{male}_l + \beta_4 \cdot \text{migrant}_l + \alpha_{\text{dist}} + \epsilon_l,$$

where the district-level effects α_{dist} are treated as i.i.d. normal random effects with mean zero and a constant variance. We also specify ϵ_l as i.i.d. normal with mean zero and constant variance, independent of the covariates and random effects. We estimate the β coefficients and the variance parameters by maximum likelihood, and plug in the estimates to obtain empirical Bayes-type estimates of the district effects α_{dist}. This fitted model is used to simulate the distribution of migrant wages in each corridor according to BHIES, and obtain values for the mean and variance of log migrant wage used in the construction of predicted site-level effects outlined in Section 4.2. Price indices are constructed from the 2016–17 Household Income and Expenditure Survey for Bangladesh.

India For India, we use the 10th schedule of the 64th (2007-08) round of the National Sample Survey (NSS) for both home household and migrant data, as well as price indices. We define a home household as one which sent a migrant at any point of time in the past and reports having lived in its current district for at least a year. The 2011 Census provides additional information for long term migrant flows along a corridor.

Pakistan For Pakistan, we obtain data for migrants from the 2014–15 Labour Force Survey (LFS) and use data from the 2019–20 Pakistan Social and Living Standards Measurement Survey (PSLM) for home households. We construct home household summary statistics using only data for households which

[37] In Appendix Q we describe how we approximate the nonlinear price schedules in Pakistan and India with a single price per currency unit of remittance sent in each country.

report having received remittances from a domestic source during the prior year. Price indices are constructed using the Pakistan Household Integrated Economic Survey (PHIES).

J.2 Restrictions on the Set of Candidate Corridors

In the rest of this subsection we describe the density and other restrictions placed on the set of candidate sites in each country, in addition to the requirement that the two corridors have origins in different states (India), provinces (Pakistan), and divisions (Bangladesh), which we mentioned in the main text.

Bangladesh For Bangladesh, we require each site to possess a minimal migrant density of 0.01 in the origin district of that corridor. Migrant density in a district is defined as the number of migrants originating there as a fraction of the total households from that district for which we have data. The density restriction yields 41 possible migration corridors, and 619 possible two-site combinations.

Pakistan Migration flows in Pakistan are much lower in magnitude that in Bangladesh so we require a lower minimal migrant density of 0.003. Our restrictions leave us with 129 possible corridors and 4,763 possible combinations.

India Following Imbert and Papp (2020), for India, we first combine data from the NSS and the 2011 Census to derive a measure of migrant density. This is because while the 10th schedule of the 64th round of the NSS does record how many migrants are sent by a household in a district, it does not record where those migrants move to. The 2011 Census records the number of migrants along each migration corridor but these correspond to long term migrants, while short term migrants are targeted for the training intervention. The migrant density measure we use is defined as,

$$\frac{\text{Migrants along a corridor (Census 2011)}}{\text{Total migrants from an origin district (Census 2011)}} \times \frac{\text{Total migrants from an origin district (NSS 07-08)}}{\text{Total HHs in an origin (NSS 07-08)}}.$$

The above product is meant to approximate the ratio of the total migrants along a corridor to the total households in an origin district. This approach uses long term migration flows to allocate the short term migrants from an origin district to multiple destination districts. This relies on the assumption that conditional on migrating, short term and long term migrants choose similar destinations. Imbert and Papp (2020) provide evidence in favor of this assumption using the REDS survey which measures both short and long term migration.

For India, we impose more stringent restrictions so that the number of site combinations remains computationally tractable. In addition to a minimal migrant density of 0.01 at the origin, as in Bangladesh, we require any site to have at least 0.75 long term migrants (as measured in the 2011 Census) per unit area of the origin. While in Bangladesh and Pakistan there are only a handful of destinations within

high-density migration corridors, in India there are many more destinations so we additionally require site combinations to have destinations in different states. These criteria give us a feasible set of 740 corridors and 223,2000 site-pair combinations in India.

K Simulation Algorithm

This section describes how we use the mixed prior to arrive at the welfare estimates given in e.g. Table A3, and how these estimates determine the sites we select.

K.1 Setup

As described in Section 2.3.3 we compose the mixed prior from a combination of the structural and smoothing priors. The mean predicted site effects, μ_τ, and the model-based variance matrix Σ_{model} are obtained from the structural model as discussed in Section 4.2. The mixed prior with weight $w \in [0,1]$ is normal with mean μ_τ, and variance matrix

$$\Sigma_{\text{mixed}} = w \cdot \Sigma_{\text{model}} + (1-w) \cdot \Sigma_{\text{smooth}}.$$

Sampling uncertainty is independent and identically distributed across sites, with variance σ_ϵ^2 computed as described in Section 5.1.3, so the matrix Σ_ϵ is the diagonal matrix $\Sigma_\epsilon = \sigma_\epsilon^2 I_S$.

K.2 Simulations

Our goal is to approximate by simulation the double integral in Equation (2) for each set of possible experimental sites $\mathscr{S}$. For $b = 1, \ldots, B = 10,000$, we draw a vector of true and (potential) estimated treatment effects for all sites from the joint distribution given in Appendix C:

$$\begin{pmatrix} \tau \\ \hat{\tau} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_\tau \\ \mu_\tau \end{bmatrix}, \begin{bmatrix} \Sigma_{\text{mixed}} & \Sigma_{\text{mixed}} \\ \Sigma_{\text{mixed}} & \Sigma_{\text{mixed}} + \Sigma_\epsilon \end{bmatrix} \right).$$

To evaluate alternative choices for $\mathscr{S}$, we take these simulation draws and estimate the welfare that would obtain from selecting each feasible subset of sites. Given $\mathscr{S}$, we keep sub-vector $\hat{\tau}[\mathscr{S}]$, the estimated treatment effects in the experimental sites,

$$\hat{\tau}[\mathscr{S}] = \{\hat{\tau}[s] \text{ for } s \in \mathscr{S}\},$$

and compute $\mathbb{E}[\tau \mid \hat{\tau}[\mathscr{S}]]$, the posterior mean of the full vector of true site effects τ given $\hat{\tau}[\mathscr{S}]$ using the formula in Appendix C. We then compute the optimal site-level treatment vector T^*, a vector of

indicators for whether each site's posterior mean treatment effect exceeds the cost of implementation:

$$T^* = \mathbb{1}\{\mathbb{E}[\tau \mid \hat{\tau}[\mathscr{S}]] \geq c\}.$$

Finally, we calculate welfare under T^* for simulation b. For site s, welfare is 0 if not treated and $\tau_s - c$ otherwise, giving aggregate welfare

$$W_b[\mathscr{S}] = \sum_{s=1}^{S} (\tau_s - c) T_s^*,$$

where τ_s is an element of the vector we drew at the beginning of simulation b.

We then compute the average welfare over all the simulation draws

$$\overline{W}[\mathscr{S}] = \frac{1}{B} \sum_{b=1}^{B} W_b[\mathscr{S}],$$

which gives us an entry in e.g. Table A3 for experimental site combination $\mathscr{S}$. To select sites, we choose the $\mathscr{S}$ with the highest $\overline{W}[\mathscr{S}]$.

L Site Selection Rule Performance Under Alternative Priors

We now examine the welfare predicted under specific mixed priors when choosing sites using different priors. Figure A10 shows the results. Each line in the graph gives the welfare anticipated under the mixed prior specified on the X-axis when sites are selected according the prior represented by a given line. In other words, sites are selected using the prior indicated by the line but evaluated under the prior on the X-axis. For instance, the line marked with upside-down triangles shows the welfare predicted when selecting sites using the pure structural prior, according to different mixed priors.

It is never optimal to use the pure structural prior to select sites since Figure A10 only predicts welfare using mixed priors placing between 0.1 and 0.95 weight on the structural prior, but it may be close. It *is* optimal to use the 0.5 mixed prior when also predicting welfare from the site combination chosen using the same prior. Matching the mixed prior used to select sites and to predict welfare yields the upper envelope for expected performance.

Looking at the figure, we see that performance of selecting sites using any mixed prior is quite robust to the mixing weight used when predicting welfare. This is in part because the mixed priors often select the same sites. We suggest all practitioners using our method perform a similar robustness check since it is difficult to determine the appropriate weight to place on the structural model *a priori*. One might believe that a 0.5 weight reflects complete ambivalence between the structural and smoothing priors but - in all countries - selecting sites using the pure smoothing prior leads to better predicted performance than selection using the pure structural prior when predictions are computed using the 0.5-weight prior.

Selecting sites using the model only begins to outperform selecting sites using the pure smoothing prior when the weight on the model in the prior used to generate predicted welfare exceeds 0.8, 0.6, and 0.85 in Bangladesh, Pakistan, and India respectively. This also shows the importance of allowing for deviations from the structural model when selecting sites, since the performance of site selections using only the structural model degrades when considering such deviations.

Nevertheless, site selection using the pure structural prior always outperforms by a wide margin the naive approach of selecting sites randomly and recommending the program be implemented in all sites if the average of experimental site ATEs exceeds the implementation cost. We consider this approach in Section 5.3.2 and refer to it as the uniform rule.

M Choosing Different Numbers of Experimental Sites

M.1 Best/Worst Single Experimental Site Choices

Table A7 shows the five best and worst choices when a single corridor is chosen to be experimented on in Bangladesh. Table A8 shows corresponding results for Pakistan, with the welfare values are based on the full set of admissible corridors instead of just marginal corridors.

Table A7: Welfare-Maximizing and Minimizing Experimental Sites, Single Site Case, Bangladesh

Best Choice		Worst Choice	
Corridor	Avg. Welfare	Corridor	Avg. Welfare
Dhaka-Noakhali	0.024	Dhaka-Thakurgaon	0.004
Dhaka-Feni	0.024	Gazipur-Panchagarh	0.003
Dhaka-Kishoregonj	0.023	Dhaka-Gopalganj	0.001
Dhaka-Bogra	0.023	Dhaka-Madaripur	0.001
Chittagong-Comilla	0.023	Dhaka-Bhola	0.001

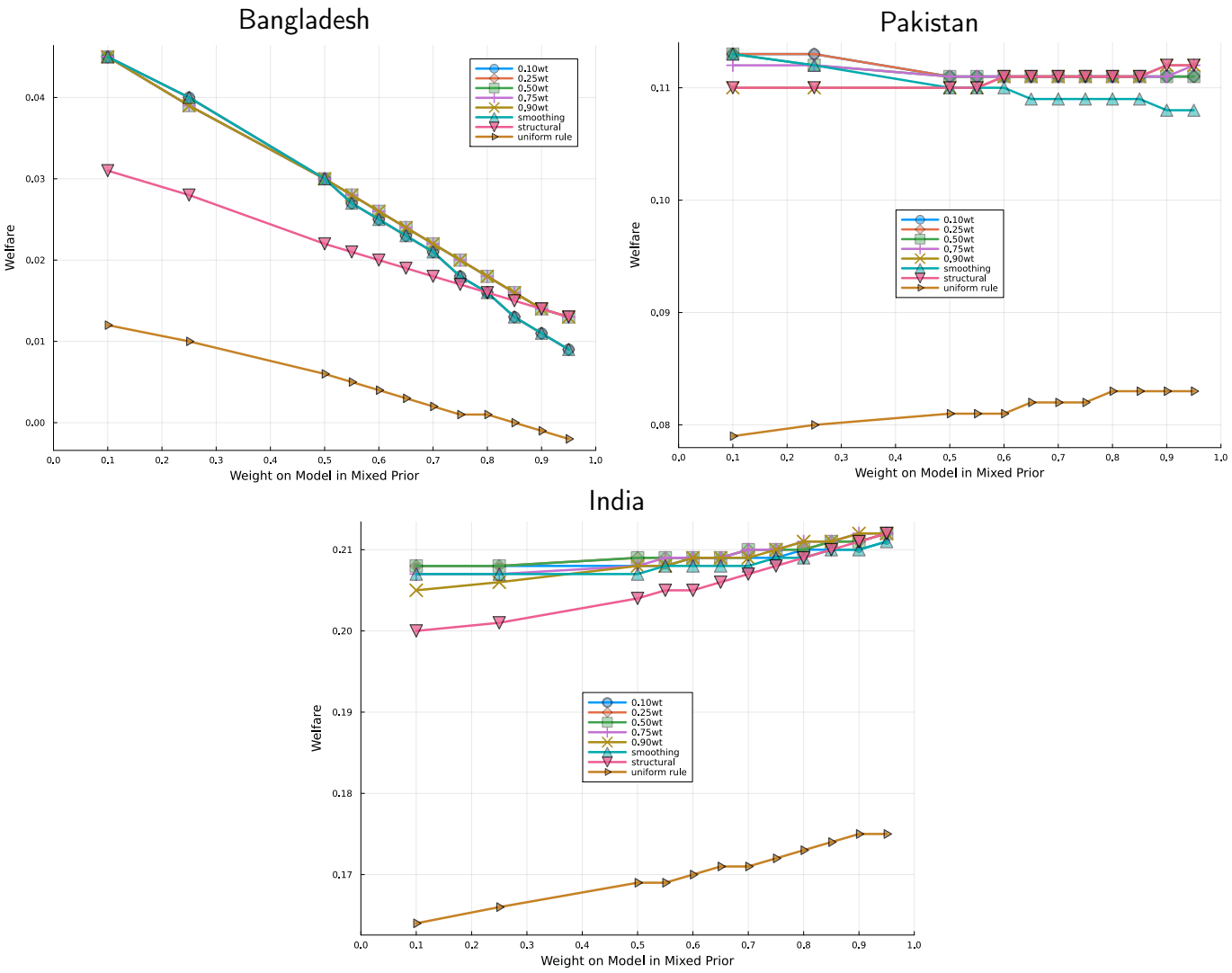
Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

M.2 Calculating the Number of Observations with Differing Numbers of Sites

To calculate the cost of collecting data in different numbers of corridors, we started with the scenario above where 8000 surveys would be collected in 2 corridors, one for each migrant and one for their household at home. We then asked the survey firm that conducted the surveys in Lee et al. (2021) to create a hypothetical budget for survey collection in "typical" sites in Bangladesh.

The firm's work led to an estimate of the total cost of the 8000 surveys in the 2 corridors of $175,415 or $21.93 per survey (converted to dollars from taka on July 18, 2023). We then took that budget as fixed

Figure A10: Average Per-Migrant Welfare of Experimental Sites by Site Selection Method and Evaluating Mixed Prior



Notes: Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under the mixed prior with model weight given by the X-coordinate. Lines index different site selection methods, with $p$wt indicating a mixed prior with p weight on the structural model.

and calculated how many surveys would be possible with 1, 3, 4, 5, and 6 corridors, using the survey firm's assumptions about fixed and variable costs. Variable costs are not linear in the number of surveys since some costs increase with the number of sites and some increase with the number of surveys.

Before collecting surveys, the firm completes a process of "listing" which involves collecting basic data on the households. The firm assumes that there would be 15 enumerators for this part of the work, each conducting 20 listing surveys per day (with buffer days added). They collect data for 110% of the eventual surveys to be completed. Other staff members include 3 supervisors, 2 research assistants (who assist with training, data collection, and data cleaning), and one research manager (who monitors and

Table A8: Welfare-Maximizing and Minimizing Experimental Sites, Single Site Case, Pakistan

Best Choice		Worst Choice	
Corridor	Avg. Welfare	Corridor	Avg. Welfare
Mardan-Swabi	0.111	Karachi-Bannu	0.103
Peshawar-Swabi	0.111	Karachi-Mianwali	0.103
Peshawar-Mardan	0.111	Karachi-Sukkur	0.103
Kohat-Hangu	0.111	Rawalpindi-Islamabad	0.103
Swat-Shangla	0.111	Chitral-Islamabad	0.103

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

Table A9: Welfare-Maximizing and Minimizing Experimental Sites, Single Site Case, India

Best Choice		Worst Choice	
Corridor	Avg. Welfare	Corridor	Avg. Welfare
New Delhi-Dausa	0.207	Patiala-Chandigarh	0.197
New Delhi-Hardwar	0.207	New Delhi-Chennai	0.197
New Delhi-Jaipur	0.207	Sitamarhi-Mumbai	0.197
New Delhi-Agra	0.206	Bangalore-Chandigarh	0.197
New Delhi-Pilibhit	0.206	Kancheepuram-Chandigarh	0.197

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

supervises the complete process). For the main household survey, the budget covers 30 enumerators who each can collect 3 surveys per day (with buffer days added). Again there will also be 2 research assistants and a research manager. This phase of the work includes additional staff to ensure data quality: 2 "scutinizers" and 6 "back-checkers."

The costs also include subscriptions of survey software and the costs of training the enumerators (including space rental, food, and transportation). Travel costs for visiting the sites by supervisors and research assistants are included.

The costs were constructed for "typical" locations several hours from Dhaka. Survey costs in Dhaka itself would be higher in general, but there would be savings on travel costs and related expenses.

The cost calculations allow us to determine how many surveys could be collected in each corridor if we kept the budget capped at $175,415. As noted, with the 2 corridor benchmark, 8,000 surveys could be collected. Table A10 shows that with only 1 corridor, it is possible to increase the number to 8,673 surveys in total. With 6 corridors, however, only 5,305 can be collected with the same budget.

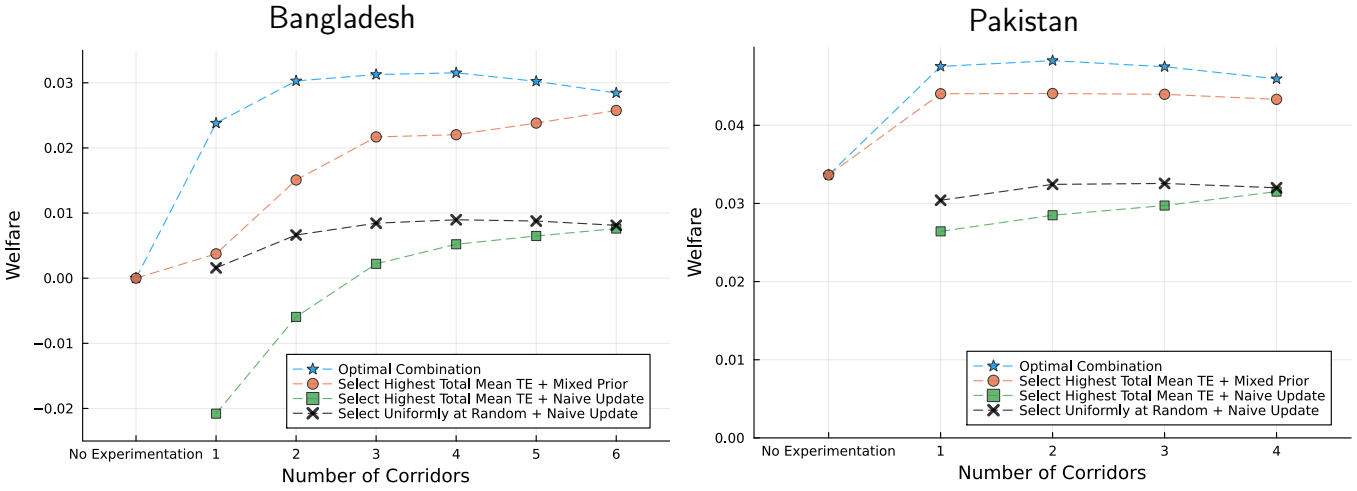
Table A10: Maximum Number of Surveys Possible for a Fixed Budget, By Number of Corridors (Costs in USD)

	1	2	3	4	5	6
Fixed cost	18.308	18,308	18,308	18,308	18,308	18,308
Total variable costs	134,219	134,227	134,220	134,213	134,220	134,213
Indirect costs	22,879	22,880	22,879	22,878	22,879	22,878
Total cost	175,406	175,415	175,407	175,399	175,407	175,400
Fixed cost per survey	2.11	2.29	2.50	2.75	3.06	3.45
Variable cost per survey	15.48	16.78	18.32	20.18	22.45	23.50
Indirect cost per survey	2.64	2.86	3.12	3.44	3.83	4.31
Total cost per survey	20.22	21.93	23.94	26.37	29.34	33.06
Maximum survey number	8,673	8,000	7,326	6,652	5,979	5,305
Surveys per corridor	8,673	4,000	2,442	1,663	1,196	884

M.3 Comparing Optimal Sites in Bangladesh and Pakistan

Figure A11 compares average per-migrant welfare by number of experimental sites and site selection method in Bangladesh and Pakistan. For Pakistan, we report welfare values over marginal sites with $\kappa = 0.25$ to sharpen comparisons.

Figure A11: Average Per-Migrant Welfare by Number of Experimental Sites and Site Selection Method



Notes: Average Per-Migrant Welfare refers to the average across all candidate sites of the welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model. Pakistan plot is restricted to marginal sites, where site s is considered marginal if $\mathbb{P}(\tau_s \geq \text{cost}_s) \geq \kappa$ and $\mathbb{P}(\tau_s \leq \text{cost}_s) \geq \kappa$, where τ_s and cost_s are the welfare benefit and cost from implementing the mobile money training program there. Surveying budget is held fixed across different numbers of sites. See Appendix M.2 for details.

N Choosing Sites Jointly Over Multiple Countries

Instead of choosing a pair of experimental sites in each country separately, we could choose migration corridors jointly across some or all of the countries. Table A11 shows the best choices for four migration corridors with two each from Bangladesh and Pakistan, respectively, and Table A12 shows the worst performing sets. Sites in India are not considered for reasons of computational tractability. As before, we assume a total of 4000 sampled migrant-household pairs in each country.

Table A11: Best Site Choices when Choosing Jointly across Bangladesh and Pakistan

Bangladesh		Pakistan		
Corridor 1	Corridor 2	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Barguna	Dhaka-Kishoregonj	Multan-Khanewal	Peshawar-Swat	0.1064
Dhaka-Kishoregonj	Dhaka-Noakhali	Multan-Muzaffar Garh	Peshawar-Swat	0.1063
Dhaka-Comilla	Dhaka-Pirojpur	Rawalpindi-Gujrat	Peshawar-Swat	0.1063
Dhaka-Comilla	Dhaka-Pirojpur	Gujranwala-Gujrat	Peshawar-Swat	0.1063
Dhaka-Comilla	Dhaka-Pirojpur	Rawalpindi-Gujrat	Rawalpindi-Peshawar	0.1062

Table A12: Worst Site Choices when Choosing Jointly across Bangladesh and Pakistan

Bangladesh		Pakistan		
Corridor 1	Corridor 2	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Bhola	Dhaka-Madaripur	Chitral-Islamabad	Peshawar-Lakki Marwat	0.0708
Dhaka-Bhola	Dhaka-Madaripur	Rawalpindi-Islamabad	Peshawar-Lakki Marwat	0.0708
Dhaka-Bhola	Dhaka-Madaripur	Peshawar-Lakki Marwat	Karachi-Sukkur	0.0707
Dhaka-Bhola	Dhaka-Madaripur	Chitral-Islamabad	Karachi-Sukkur	0.0702
Dhaka-Bhola	Dhaka-Madaripur	Rawalpindi-Islamabad	Karachi-Sukkur	0.0701

O Structural Correlations and Site Characteristics

We investigate how the structural model generates correlations across site-level treatment effects in the structural prior. We fit a linear regression model that relates the structural prior correlations to the absolute difference in site characteristics:

$$\rho_{s,s'} = \beta_0 + |X_s - X_{s'}|'\beta + \varepsilon_{s,s'},$$

where $\rho_{s,s'}$ denotes the correlation in outcomes between sites s and s' in the structural prior, and X_s denotes characteristics for site s. All differences in site level characteristics are standardised to have mean zero and unit variance.

The regression estimates are reported in Tables A13–A15. In all the estimated regressions, the constant term is large. This shows that the structural prior implies high correlations in ATEs across sites,

which reaffirms the message from Figure 3 that the parametric model allows quite different sites to be informative for one another. Some of the site-level characteristics have explanatory power for the prior correlations, but these patterns differ substantially across the three countries. This suggests that the true relationship between site-level characteristics and prior correlations may be highly nonlinear or involve interactions that our simple linear approximation does not capture.

Table A13: Regressing Structural Correlations on Differences in Site Characteristics in Bangladesh

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9908	0.0002	0.9903	0.9913
Distance	-0.0024	0.0002	-0.0029	-0.0019
Home Income	-0.0011	0.0004	-0.0020	-0.0003
Migrant Wage	0.0016	0.0005	0.0006	0.0026
Home HH Size	0.0005	0.0003	-0.0001	0.0010
Operator Density	-0.0102	0.0003	-0.0108	-0.0096
Home Income SD	0.0020	0.0004	0.0012	0.0028
Migrant Wage SD	-0.0021	0.0005	-0.0031	-0.0011
Home HH Size SD	0.0026	0.0003	0.0019	0.0032
Number of Corrs.	820		F	200.7426
SSR	0.0395		(Prob. >F)	0.0000
R^2	0.6645			

Notes: results from regressing correlations in welfare ATEs according to the structural prior across all site pairs on the vector of differences in site characteristics for the corresponding pair. Numbers are rounded to the fourth decimal place.

Table A14: Regressing Structural Correlations on Differences in Site Characteristics in Pakistan

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9447	0.0022	0.9404	0.9490
Distance	0.0052	0.0022	0.0009	0.0095
Home income	0.0046	0.0028	-0.0008	0.0100
Migrant Wage	-0.0758	0.0028	-0.0813	-0.0703
Home HH size	0.0142	0.0033	0.0077	0.0208
Operator Density	-0.0047	0.0022	-0.0090	-0.0004
Home income std. dev.	-0.0038	0.0028	-0.0092	0.0016
Migrant wage std. dev.	0.0480	0.0028	0.0425	0.0536
Home HH size std. dev.	-0.0012	0.0036	-0.0077	0.0054
Number of Corrs.	8256		F	98.58
SSR	326.3269		(Prob. >F)	0.0000
R^2	0.0873			

Notes: results from regressing correlations in welfare ATEs according to the structural prior across all site pairs on the vector of differences in site characteristics for the corresponding pair. Numbers are rounded to the fourth decimal place.

Table A15: Regressing Structural Correlations on Differences in Site Characteristics in India

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9225	0.0003	0.9220	0.9231
Distance	0.0017	0.0003	0.0011	0.0023
Home income	0.0016	0.0004	0.0008	0.0023
Migrant Wage	0.0004	0.0003	-0.0003	0.0010
Home HH size	0.0012	0.0003	0.0005	0.0018
Operator Density	-0.0802	0.0003	-0.0808	-0.0796
Home income std. dev.	0.0060	0.0004	0.0052	0.0067
Migrant wage std. dev.	-0.0024	0.0003	-0.0031	-0.0018
Home HH size std. dev.	0.0011	0.0003	0.0004	0.0017
Number of Corrs.	273,430		F	9525.6627
SSR	6402.9309		(Prob. >F)	0.0000
R^2	0.2180			

Notes: results from regressing correlations in welfare ATEs according to the structural prior across all site pairs on the vector of differences in site characteristics for the corresponding pair. Numbers are rounded to the fourth decimal place.

P Robustness to Alternate Treatment Effects

As a robustness check, we would like to see how the ranking of optimal corridor combinations changes in response to a more or less effective intervention. Since the environment in Bangladesh is different now from how it was in 2015 when the Lee et al. (2021) experiment was carried out (for example, mobile money usage has increased nationally), we would like to see that our site selections are not too sensitive to the specific value of ψ we estimated from the experimental microdata. To do this, we suppose the point estimate for the parameter ψ is larger or smaller than is estimated from the experimental data from Lee et al. (2021) while its estimated variance remains the same. This roughly assumes that the effectiveness of the experimental intervention was greater or lesser while sampling variability remained the same. Changing the structural parameter estimates in this way will lead to different specifications of the mean vector and variance matrix of the structural prior used for the site-selection algorithm, so we can view these results as a form of prior sensitivity or prior robustness exercise.

Tables A16–A19 display the best and worst combinations of two migration corridors in Bangladesh under the mixed prior with 0.5 weight on the model as the estimated treatment effect is magnified or diminished. The top site combination with our estimated value of ψ always appears among the top five highest-welfare site combinations when ψ is altered, along with between one and three other top sites from Table A3.

Table A16: Welfare-Maximizing and Minimizing Experimental Site Combinations in Bangladesh; ψ Scaled by 0.1

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Kishoregonj	Dhaka-Noakhali	0.007	Dhaka-Bhola	Dhaka-Panchagarh	0.001
Dhaka-Feni	Dhaka-Shariatpur	0.007	Dhaka-Gopalganj	Dhaka-Thakurgaon	0.001
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.007	Dhaka-Bhola	Dhaka-Thakurgaon	0.001
Dhaka-Magura	Dhaka-Noakhali	0.007	Dhaka-Bhola	Dhaka-Madaripur	0.001
Dhaka-Noakhali	Dhaka-Shariatpur	0.007	Dhaka-Bhola	Dhaka-Gopalganj	0.001

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

Table A17: Welfare-Maximizing and Minimizing Experimental Site Combinations in Bangladesh; ψ Scaled by 0.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Kishoregonj	Dhaka-Noakhali	0.014	Dhaka-Gopalganj	Dhaka-Thakurgaon	0.002
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.014	Dhaka-Bhola	Dhaka-Thakurgaon	0.002
Dhaka-Magura	Dhaka-Noakhali	0.013	Dhaka-Bhola	Dhaka-Panchagarh	0.002
Dhaka-Narsingdi	Dhaka-Noakhali	0.013	Dhaka-Bhola	Dhaka-Madaripur	0.001
Dhaka-Barguna	Dhaka-Kishoregonj	0.013	Dhaka-Bhola	Dhaka-Gopalganj	0.001

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

Table A18: Welfare-Maximizing and Minimizing Experimental Site Combinations in Bangladesh; ψ Scaled by 1.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.053	Dhaka-Gopalganj	Gazipur-Panchagarh	0.018
Dhaka-Kishoregonj	Dhaka-Noakhali	0.052	Dhaka-Madaripur	Gazipur-Panchagarh	0.018
Dhaka-Faridpur	Dhaka-Noakhali	0.052	Dhaka-Bhola	Gazipur-Panchagarh	0.018
Dhaka-Barguna	Dhaka-Kishoregonj	0.052	Dhaka-Bhola	Dhaka-Gopalganj	0.016
Dhaka-Feni	Dhaka-Narsingdi	0.051	Dhaka-Bhola	Dhaka-Madaripur	0.015

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

Q Specifying the Cost of Remitting Through Mobile Money

The structural model laid out in Section 3 requires a cost of remitting through mobile money γ. For Bangladesh we use the figure from Lee et al. (2021). For India and Pakistan we use cost schedules for digital money transfers using local services to calibrate a value for γ. For India we use the fee schedule for electronic fund transfers between savings accounts at India Post Offices and for Pakistan we use the

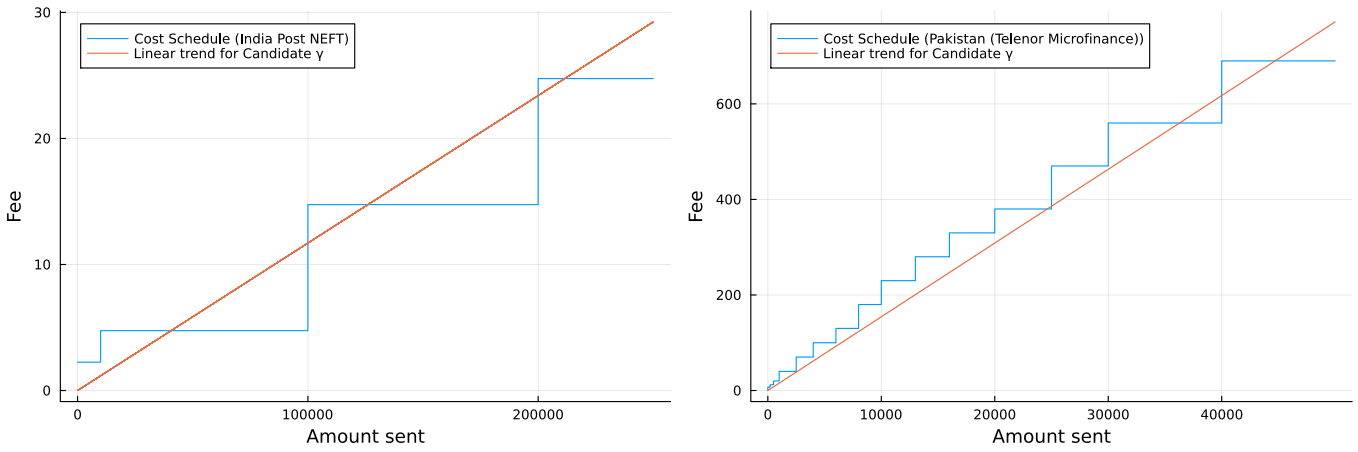
Table A19: Welfare-Maximizing and Minimizing Experimental Site Combinations in Bangladesh; ψ Scaled by 2.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.110	Dhaka-Gopalganj	Gazipur-Panchagarh	0.079
Dhaka-Faridpur	Dhaka-Noakhali	0.110	Dhaka-Bhola	Gazipur-Panchagarh	0.079
Dhaka-Kishoregonj	Dhaka-Noakhali	0.110	Dhaka-Madaripur	Gazipur-Panchagarh	0.078
Dhaka-Narsingdi	Dhaka-Noakhali	0.109	Dhaka-Bhola	Dhaka-Madaripur	0.077
Dhaka-Noakhali	Dhaka-Shariatpur	0.109	Dhaka-Bhola	Dhaka-Gopalganj	0.077

Notes: Avg. Welfare refers to the average across all candidate sites of the per-migrant welfare increase from site-level implementation recommendations, predicted under our preferred mixed prior with 0.5 weight on the structural model.

cost of transferring funds digitally using Telenor Microfinance Bank services. The variable γ is a scalar but the cost schedules above do not charge a flat fee for all transfers. We calibrate γ as an approximate linear trend which represents the respective cost schedules. Figure A12 shows the cost schedules for local digital money transfer options and superimposes a linear trend corresponding to the country specific γ we use for the structural model. The value of γ for Bangladesh, India and Pakistan are 0.02, 0.001 and 0.015, respectively.

Figure A12: Cost Schedules for Local Digital Money Transfer Services



R Price Index Computation

For each locality l and a food product $j \in \{1,\ldots,J\}$, each country's expenditure survey dataset provides two key pieces of information for each household i surveyed in this (l,j) pair: quantity q_{ljl} and expenditure $\exp_{lji}$. Let N_l be the total number of households surveyed in locality l. Then, for each product in each locality, we can obtain the average price $avg_p_{lj} = \frac{\sum_{i\in l} exp_{lji}}{q_{lj}}$ and average consumption

$avg_q_{lj} = \frac{\sum_{i \epsilon l} q_{lji.}}{N_l}$.

For any migration corridor defined by an origin locality $l = o$ and destination locality $l = d$, define the food consumption basket at origin o as the vector $q_o \equiv \left(avg_q_{o1}, \ldots, avg_q_{oJ}\right)$. The corresponding price vector at origin is $p_o \equiv \left(avg_p_{o1}, \ldots, avg_p_{oJ}\right)$ and the price vector at the destination d is $p_d \equiv \left(avg_p_{d1}, \ldots, avg_p_{dJ}\right)$. The Laspeyres price index we use is simply the the expenditure on the origin basket at origin prices (given by $p_o' q_o$) divided by expenditure on the origin basket at destination prices (given by $p_d' q_o$).