Site Selection for External Validity: Theory and an Application to Mobile Money in South Asia[*]

Michael Gechter[†], Keisuke Hirano[‡], Jean Lee[§], Mahreen Mahmud[¶], Orville Mondal[‖], Jonathan Morduch[**], Saravana Ravindran[††] and Abu S. Shonchoy[‡‡]

September 28, 2023

Abstract

Choosing *where* to conduct field experiments is a crucial aspect of experimental design that may influence not only the estimates obtained at the chosen sites, but also the extent to which the empirical findings can be generalized to other settings. We develop a decision-theoretic approach to experimental site selection that frames external validity through a policy lens. A decision-maker must eventually make policy recommendations across a wide set of locations on the basis of data from a limited set of experimental sites. We evaluate potential choices of experimental sites by the degree to which they will lead to better policy recommendations across *all* sites, not only the ones selected for experimentation. Taking a Bayesian approach, we develop a prior specification for the joint distribution of site-level treatment effects based on a microeconometric structural model, while also allowing for additional sources of heterogeneity. We apply our approach to select sites for an experiment on mobile money training from 41 migration corridors (destination-origin district pairs) in Bangladesh, 60 in Pakistan, and 740 in India. Learning is optimized when the experiments take place in sites informative for the largest number of other sites for which suitability of the policy is a priori uncertain. Conventional rules of thumb for choosing experimental sites, such as simple randomization, can entail large efficiency losses.

[*]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 World Bank, Yale, and Y-RISE for helpful comments.

[†]Pennsylvania State University

[‡]Pennsylvania State University

[§]The World Bank

[¶]University of Exeter

[‖]Pennsylvania State University

[**]New York University

[††]National University of Singapore

[‡‡]Florida International University

1 Introduction

The increasing use of field experiments in economics has allowed researchers to evaluate social interventions using data designs with strong internal validity. However, the goal of such analyses is often to provide useful policy advice for other places and times. As a result, there is also increasing concern about the external validity of randomized controlled trials (RCTs) in the field (see e.g. Deaton and Cartwright 2018, Bates and Glennerster 2017, Vivalt 2020, Allcott 2015). In this paper, we tackle the concern directly, by considering a social planner who must use the results from a single experiment, or a limited set of coordinated experiments, to make policy decisions across a broader range of sites. Our notion of external validity departs from the narrow question of whether a treatment effect from one site is expected to be similar to treatment effects elsewhere. Instead, we ask how informative is a treatment effect estimated in a given site about the expected treatment effects in other sites, taking into account knowledge about the structure of economic relationships, the broad range of economic contexts, and previous estimates. With this as our notion of external validity, we consider the experimental design problem of choosing the sites in which to run an experiment. We develop a novel method to choose future 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. We illustrate and apply our general approach to the design of a new multi-site experiment underway in South Asia.

With few exceptions, RCTs in economics are limited to one or a small number of sites, but researchers and policymakers seek to draw broader conclusions from such studies.[1] In our application, we choose a small set of migration corridors from hundreds of candidates across South Asia in which to experimentally evaluate a program that teaches migrant workers to send and receive remittances via mobile money. Lee et al. (2021) report on an intervention, implemented in a single migration corridor in Bangladesh, that increased the volume of urban-to-rural remittances and reduced rural poverty. The goal of the new experiments is to derive policy recommendations for all the candidate sites. Specifically, the experiments are chosen to inform a rule for each of the hundreds of migration corridors. The rule maps each corridor's characteristics to an up-or-down recommendation for whether, ultimately, the training program should be implemented there. 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]

Each new experiment takes place in a "site," which in our context is a migration corridor consisting of an origin and destination district (a geographic area roughly equivalent to US county). We anticipate that

[1] For example, Banerjee et al. (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 et al. 2017). After the study, the program was scaled to more than 2.1 million households in Bangladesh and over 1.1 million households in 14 other countries (JPAL, 2022).

[2] Muralidharan and Niehaus (2017) discuss the scaleup from experimentation in a pilot experiment in one site to several sites, raising additional 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.

the average effect 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. While the experiments will provide direct estimates of the average treatment effects (ATEs) for the chosen sites, we wish to be able to extrapolate those estimates to gain knowledge about the overall distribution of site-level ATEs across Bangladesh, India, and Pakistan. In our analysis, we take advantage of the prior experimental study that was conducted in Bangladesh (Lee et al., 2021). Given the large number of possible sites and the limited scope for experimentation, it will be crucial to develop a framework that allows us to link this prior experimental evidence and the effects at the chosen experimental sites to the effects at the remaining sites. 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.

We propose a quasi-Bayesian framework for this experimental design problem that allows for various prior assumptions about the heterogeneity in site-level effects. A randomized experiment at a particular site will provide a noisy measure of the treatment effect at that site, and will also be informative about the treatment effects at other sites that are related to the experimental site through the prior distribution over *all* site effects. The optimal design chooses sites to maximize the overall information gained, as measured by the improvement in policy decisions that they enable.

One version of this approach uses a "smoothing prior" that makes treatment effects in different sites similar when they are close to each other in certain site-level characteristics such as origin–destination distance and average household income. This encodes the intuition that sites with similar characteristics should have similar effects. If there are many site characteristics, however, the analysis can be sensitive to the inclusion and weighting of different characteristics.

An alternative, structural approach is to introduce a model for migrant remittance decisions that provides a microeconomic foundation for site-level effect heterogeneity and allows us to link the original experimental data to the experimental design problem. In particular, we specify and estimate a structural model using the data from Lee et al. (2021), which we regard as a pilot experiment, and we then use this 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 found structural modeling methods to be useful for analyzing 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 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 a "smoothing" prior that allows sites with similar characteristics to have similar predicted treatment effects. Despite all these features, the model is computationally tractable.

We use the empirical setting to demonstrate how our procedure works, and to compare it to alternatives. 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 et al. 2019) for home household members, where we proxy individual utility by per capita expenditure.

The data show that a rule-of-thumb procedure for selecting sites can be very inefficient relative to our welfare-optimizing approach. To demonstrate this, we implement a heuristic procedure (loosely modeled after a recommendation in Allcott 2015) that selects sites randomly, computes the average of ATEs across sites, and makes a uniform recommendation to implement the program in all sites or none based on whether the average of ATEs exceeds the cost of the program. Experimenting in a single randomly selected sites can result in zero welfare and even with more sites welfare is always well below the optimum.

The sites selected through the welfare-optimizing approach are relatively centrally-located in the space of characteristics of all sites. 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. Finally, while it may be tempting to implement interventions in places where they are most likely to succeed, the approach also highlights that welfare gains are concentrated among sites whose *a priori* optimal assignment to receive the treatment is fundamentally uncertain.

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. Other work, such as Stuart et al. (2011), Tipton (2013, 2014), and O'Muircheartaigh and Hedges (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. Dehejia et al. (2021) examine the performance of such forms of covariate-based extrapolation and their implications for policy choice in an empirical replication setting, and Gechter et al. (2019) develop formal methods for evaluating the performance of various extrapolation methods for choosing policy in a target context. See also Hotz et al. (2005), Bareinboim et al. (2013), Gechter (2023), Andrews et al. (2023), Andrews and Oster (2019), and Menzel (2023) for discussions of identification issues and strategies for transporting causal estimates across contexts.

A closely related recent literature in machine learning and econometrics has focused on constructing forecasts and policy recommendations that are robust under bounded shifts in the underlying distributions of covariates, potential outcomes, or both, across contexts. Typically, this variation across sites must be relatively small to obtain useful robustness properties of the type sought in this work. See for example Kuang et al. (2018), Duchi and Namkoong (2021), Si et al. (2022), Adjaho and Christensen (2022), Kallus and Zhou (2021), Lei et al. (2023), and Chen et al. (2023). We do not impose bounds on the differences between site-level populations as in this literature, but focus on crafting a prior that allows us to model varying degrees of similarity across sites.

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 et al. (2023) and Angrist et al. (2022) report on six RCTs of remote instruction for primary school children during the school shut-downs caused by COVID-19 and Cusolito et al. (2023) evaluate an intervention in six Balkan countries that provides skills to help small businesses compete in the export sector.

Another key contribution of our paper is to consider how the concern for external generalizability should inform experimental design. Glennerster (2017) and Chassang and Kapon (2021), among others, discuss the importance of considering external validity when designing experiments. Under the assumption that all relevant predictors for treatment effects are measured, Tipton (2014) and O'Muircheartaigh and Hedges (2014) discuss how experiments should be designed to ensure that the individual-level covariate distribution in the experiment is sufficiently rich for covariate-based extrapolation to other contexts. However, there has been relatively little work developing analytical approaches to experimental design to account for more more general forms of heterogeneity across sites. Finan and Pouzo (2023) consider sequential experimentation in one site or context, but using various sources of prior information including past data from other contexts to inform the experimental stopping rule and policy recommendation. We similarly seek to use prior information to inform the experimental design, but our key design choice is in the selection of experimental sites.[3]

Our framework is Bayesian, and sequential in that we use data from a pilot experiment to inform our choice of future experimental sites. Bayesian approaches to experimental design (surveyed in Spiegelhalter et al. 1994 and Chaloner and Verdinelli 1995) have become more practical for complex problems with the rise of efficient simulation algorithms and computing power. There has also been a recent revival of interest in adaptive or sequential experiments, which offer the prospect of improving efficiency and lowering costs of randomized experiments in the social sciences. See, for example, Hu and Rosenberger (2006), Hahn et al. (2011), Tabord-Meehan (2018), Kasy and Sautmann (2021), Caria et al. (2021), Athey et al. (2021), Athey et al. (2022), Cytrynbaum (2022), and Xiong et al. (2023). To link experimental design to the problem of external validity, we embed a statistical treatment choice problem into the de-

[3] Site selection is an important concern for clinical trials in medicine; see for example Potter et al. (2011) and Hurtado-Chong et al. (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).

sign problem. Thus our work also relates to a recent literature on empirical treatment choice, including Manski (2004), Dehejia (2005), Stoye (2009), Hirano and Porter (2009), Chamberlain (2011), and Kitagawa and Tetenov (2018).

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 will then extend this framework to incorporate prior experimental data, turning this into an adaptive experimental design problem, adopt a Bayesian solution concept, and develop tractable prior specifications that will facilitate its implementation.

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 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.[4]

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

[4] 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.

for $s \in \mathscr{S}$. As a notational shorthand, 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. More specifically, we operationalize our goal of designing the experiment for external validity or generalizability 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 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.[5]

The basic decision problem outlined in Figure 1 is difficult in part because the decision-relevant param-

[5] 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.

eter 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.

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 $W(\tau, T)$.

2.2 Bayesian Preposterior 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 for the S-dimensional vector τ:

$$\tau \sim \Pi.$$

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, 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 $\Pi(\tau \mid \hat{\tau}_{\mathscr{S}})$. Then, 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 many times and averaging over the draws will approximate the *ex ante* or "preposterior"[6] expected welfare that the policymaker can anticipate for a particular choice of experimental sites $\mathscr{S}$, incorporating both the prior uncertainty about the true τ and the sampling variability of the experimental estimates $\hat{\tau}_{\mathscr{S}}$.[7] [8] 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.

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

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

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

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.

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

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

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 $c > 0$ and $d > 0$ are tuning parameters for the prior; c controls the amount of local smoothing, and d controls the overall scale of the variance matrix.

2.3.2 Structural Prior

The smoothing prior above uses a specification for the covariance matrix that is well known in the literature on Gaussian smoothing methods.[9] However, this conventional approach may be sensitive to the choice of the function $h(V_s)$ and the tuning parameters, especially when the dimension of V_s is high (as it is in our application). Moreover, we would like to take full advantage of the rich individual-level data in the pilot RCT, and insights from microeconomic modeling, to bring as much prior information as pos-

[9] See, e.g., Rasmussen and Williams (2006).

sible to this difficult experimental design problem. 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 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[10] 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

$$\tau \sim N_S(\mu_\tau, \Sigma_\tau).$$

[10] 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.

Recall that every $\hat{\tau}_s$ is also normally distributed and centered at τ_s, so 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

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

and variance

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

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. In Appendix A, we illustrate these posterior updating formulas in a simple example with two potential sites and one experimental site.

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:

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

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

$$T^*_s = \mathbf{1}(\bar{\tau}_s \geq \texttt{cost}_s). \tag{3}$$

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, 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 greatly 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 $p_h \leq 1$.

This model can be solved to obtain the optimal remittance amount R^*. Depending on the values of the parameters, optimal remittances may be zero, or strictly positive as indicated in the following equation.

$$R^* = \begin{cases} 0 & \text{if} \quad p_r \geq \frac{24\lambda a_h \eta w_m}{p_h y_h} \\ p_h \left\{ \frac{\lambda a_h \eta (24 w_m + p_r p_h y_h)}{p_h p_r (1 + \lambda a_h \eta)} - y_h \right\} & \text{if} \quad p_r < \frac{24\lambda a_h \eta w_m}{p_h y_h} \end{cases}. \tag{4}$$

The optimal values of migrant consumption and leisure are:

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

The mapping $\lambda \mapsto R^*$ is monotonic and continuous at the boundary points. Below, we will 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 et al. 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.[11] 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 $\bar{\epsilon}$:

$$\epsilon \sim \text{Exponential}(\bar{\epsilon}).$$

This ensures that ϵ is nonnegative in accordance with its interpretation as the increase in remittance cost with an 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 monetary per-taka-sent cost of remitting using mobile money, which we obtain from the actual price charged by bKash, the largest mobile money service in Bangladesh. We describe how we determine the value of γ in Appendix E.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

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

may be implemented. Some mobile money services require a physical agent to complete or record any transactions made using the platform. 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

$$\xi \sim \text{Exponential}(\bar{\xi}),$$

where ξ is independent of ϵ.[12] The stochastic terms ϵ and ξ generate variation across individuals in their choice of remittance mode, as well as amount remitted through the optimality condition in Equation (6).

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

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

Let M denote the remittance mode choice, with $M = 1$ for mobile money and $M = 0$ for traditional remittance. Then

$$\begin{aligned} M^* &= \mathbf{1}\left\{p_{r,\text{trad}} \geq p_{r,\text{mm}}(t)\right\} \\ &= \mathbf{1}\left\{1 + d\epsilon \geq 1 + \gamma + \exp(\delta o + \psi t)\xi\right\}. \end{aligned} \tag{5}$$

We can think of the migrant first choosing the remittance mode, which determines the price of remittance, and then selecting the level of remittance R^* according to the utility-maximization problem described above.

3.3 Distribution of Optimal Remittances

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).

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

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

[12] 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.

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

$$\mathbb{P}(R^* \leq r \mid X) = \mathbb{P}\left(p_r \geq \frac{24\lambda a_h w_m \eta}{r(1+\lambda a_h \eta) + p_h y_h} \;\middle|\; X\right). \tag{6}$$

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

$$\begin{aligned}\mathbb{P}(M^* = 1, R^* > 0 | X) = \exp\left(\frac{-\gamma}{\bar{\epsilon}d}\right)\cdot\frac{\bar{\epsilon}d}{\bar{\epsilon}d + \bar{\xi}e^{\psi t+\delta o}} \\ \times \mathbf{1}\left\{w_m \geq \frac{(1+\gamma)\cdot p_h y_h}{24\lambda a_h \eta}\right\}\cdot\left(1 - \exp\left(-\left(\frac{1}{\bar{\xi}e^{\psi\cdot t+\delta o}} + \frac{1}{\bar{\epsilon}d}\right)\cdot\left(\frac{24\lambda a_h \eta w_m}{p_h y_h} - 1 - \gamma\right)\right)\right).\end{aligned} \tag{7}$$

The indicator variable in Equation (7) reflects the necessary condition for mobile money to be the preferred mode of remittance. The distribution of optimal remittances, and the probability of remitting a positive amount using mobile money, 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.

4 Fitting the Structural Model to the Pilot Experiment

Lee et al. (2021) conducted an experiment in Bangladesh in 2014–2016 to measure the impacts of an intervention that trained migrant workers in the city of Dhaka and their family members at home to use a digital money transfer service. The study measured multiple outcomes of interest for both the migrants and their family members. The experiment tracked 813 migrant-home family pairs whose home families were located in the district of Gaibandha. 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 who report 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. Table 1 displays summary statistics for the sample used for model estimation.

Table 1: Summary Statistics for Estimation Sample from bKash Experiment

	Treated				Control			
	Male		Female		Male		Female	
Variable	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev.	Mean	Std. Dev
Daily Remittances Sent	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 Daily Wage	32.77	9.07	25.49	5.68	32.01	8.89	25.68	5.14
Home Daily Income	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: Lee et al. (2021) recorded remittance behavior over multiple months which we aggregated up to household level; mobile money usage conditional on remitting is the fraction of remittance transactions reported to have been executed via mobile money.

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 C. 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 on outcomes Y_i through the structural model $F_{Y|X}(Y_i|X_i, \theta)$.[13] At the

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

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 classical minimum distance, following Newey and McFadden (1994).[14] Let $q(\theta)$ be a K-dimensional vector of quantities implied by model parameter θ. These quantities will generally be conditional expectations of functions of Y given X. Let $\hat{q}(\theta)$ be the sample versions of these model-implied conditional expectations. We will specify $q(\theta)$ and $\hat{q}(\theta)$ in detail below.

Let $\hat{\pi}$ be sample estimators of the corresponding conditional expectations (which will typically have the form of conditional sample averages). Suppose that $\hat{\pi} \xrightarrow{p} \pi_0$ for some π_0 and that

$$\sqrt{n}(\hat{\pi} - \pi_0) \xrightarrow{d} N(0, \Sigma_\pi).$$

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

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

where $\hat{W}_n$ is a weighting matrix with $\hat{W}_n \xrightarrow{p} W$ for some W. Suppose there is a unique solution θ_0 to the population minimum distance problem:

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

We will take θ_0 to be the pseudo-true value of the parameter.[15] Under suitable conditions we will have $\hat{\theta} \xrightarrow{p} \theta_0$, and

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

where

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

and $G = \frac{\partial}{\partial\theta} q(\theta_0)$ is the Jacobian of $q(\theta)$ evaluated at θ_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 or numerically. We also estimate the variance matrix Σ_π associated with $\hat{\pi}$ (see Appendix B.1 for the derivations).

[14] 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 the exact Bayesian posterior expectation. 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.

[15] 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.

4.2 Specification of Estimator

Here we specify the function $q(\theta)$ used for our model. First, we define the components of π and $q(\theta)$ that relate to the quantity remitted by the migrant. We fix a set of remittance values

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

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

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

These conditional probabilities may be estimated directly by their sample analogs, or derived from the structural model given a parameter vector. The sample analogs will constitute the vector $\hat{\pi}$ in the minimum distance estimation routine, while the model implied versions will constitute the vector $h(\cdot)$. The experiment of Lee et al. (2022) 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 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). The definitions of the components in $\hat{\pi}$ are modified accordingly. Specifically, for any value $r \in \mathscr{R}$, the corresponding component of $\hat{\pi}$ is the following estimated conditional probability:

$$\hat{\pi}_l = \widehat{\mathbb{P}}\left(R \leq r \mid \text{male} = m, T = t\right) = \frac{\sum_{i=1}^n \mathbf{1}\{R_i \leq r, \text{male}_i = m, T_i = t\}}{\sum_{i=1}^n \mathbf{1}\{\text{male}_i = m, T_i = t\}}.$$

The model-implied quantity is:

$$\hat{q}_l(\theta) = \frac{\sum_{i=1}^n \mathbb{P}(R^* \leq r \mid X_i, \theta) \cdot \mathbf{1}\{\text{male}_i = m, T_i = t\}}{\sum_{i=1}^n \mathbf{1}\{\text{male}_i = m, T_i = t\}},$$

where $\mathbb{P}(R^* \leq r \mid X_i, \theta)$ is as defined in Equation (6). Similarly, for the probability of remitting a positive amount through mobile money, conditional on migrant gender and treatment status we have:

$$\hat{\pi}_k = \widehat{\mathbb{P}}(M_i = 1, R_i > 0 \mid \text{male}_i = m, T_i = t) = \frac{\sum_{i=1}^n \hat{\mathbb{P}}\left(M_i = 1, R_i > 0\right) \cdot \mathbf{1}\{\text{male}_i = m, T_i = t\}}{\sum_{i=1}^n \mathbf{1}\{\text{male}_i = m, T_i = t\}},$$

where $\hat{\mathbb{P}}\left(M_i = 1, R_i > 0\right)$ for household i represents their rate of using mobile money across all remittance transactions they report making in the experimental data. The corresponding component of $\hat{q}(\theta)$ is:

$$\hat{q}_k(\theta) = \frac{\sum_{i=1}^n \mathbb{P}\left(M_i = 1, R_i > 0 \mid X_i, \theta\right) \cdot \mathbf{1}\{\text{male}_i = m, T_i = t\}}{\sum_{i=1}^n \mathbf{1}\{\text{male}_i = m, T_i = t\}},$$

where $\mathbb{P}\left(M_i = 1, R_i > 0 \mid X_i, \theta\right)$ is given in Equation (7). We stack these conditional probabilities across groups based on the four possible values of the vector (male, T) to define the first $4 \cdot (|\mathscr{R}| + 1)$ elements of $\hat{\pi}$ and $h(\cdot)$, respectively.

We also include a component of $\hat{\pi}$ and h 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:

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

Substituting E_m in the optimality condition and rearranging terms results in the following expression for the parameter α:

$$\alpha = \frac{\mathbb{E}[E_m]}{\mathbb{E}[E_m] + \mathbb{E}[L_m \cdot w_m]} \tag{8}$$

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

Finally, to complete the specification of the minimum distance objective function, we set $\hat{W}_n$ equal to the identity matrix. The asymptotic variance of the estimated parameter vector depends on the limiting gradient matrix for the estimation moment. Expressions for the derivatives of $q(\cdot)$ can be found in Appendix B.2. Table 2 displays estimated parameters, with standard errors in parentheses. Appendix D 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 2: 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)

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).

4.3 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.[16]

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 means depending on the draw for θ). 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 Equations (4) and (5).

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 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 4.1-4.2 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 ma-

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

trix 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 $\tau_1, \dots, \tau_S$. For each draw of θ, we employ the same method to generate the implied value of τ_s across all sites $s = 1, \dots, S$. 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, we do not have data with this level of granularity. Specifically, data linking households in rural areas to migrant members is not available. We use nationally representative data sets—which we will hereafter refer to as administrative data—to construct site-level summary statistics. In these data, the only site-level statistics that we regard as being of sufficiently high quality for our purposes are the marginal distribution of gender among migrants, and the marginal sample means and variances of migrant wage, home household size, and home household income conditional on migrant gender. 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 3 demonstrates the impact of the data limitations described above. It shows the model implied treatment effect on daily remittances sent by migrants when using the full data of Lee et al. (2022) and compares it to the model predicted value when restricted to only summary statistics for the migrant and home household characteristics.

Table 3: Model Implied Treatment Effects for Migrants in bKash Experiment.

Outcome	Model	BKash Experiment
Daily Remittance (Complete Microdata)	5.94	6.41
Daily Remittance (Summary Statistics Only)	6.44	6.41

5 Application: Designing Experiments in South Asia

We apply the experimental site selection methodology developed in the previous sections to determine where to locate new randomized trials in Bangladesh, India, and Pakistan. The experimental intervention will again be aimed at reducing the costs a migrant may face when remitting to their family in distant home districts. A migration corridor (site) is defined as an origin district and destination district pair in any country. "District" is the administrative unit we use as the end point of a migration corridor. Approximately equivalent to a county in the US, they are referred to as districts 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.

We first describe the social welfare function and treatment effects we will target and translate costs into welfare terms, and specify the site-level data and prior distributions we will use for the analyses. Then, we begin our exploration of experimental designs by applying the method developed in Section 2 to the problem of choosing a single experimental site in each country. We then consider site selection when it is possible to target two migration corridors in each country, which is the case we can actually take to the field. We evaluate the welfare gains and discuss the potential gains (and costs) of experimenting on multiple sites.

5.1 Specifying Social Welfare

5.1.1 Benefit of Treating a Site

The household-level outcome of interest in this application is the sum of additive log social welfare associated with each of its home members' utility,[17]

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

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:

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

[17] 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 a 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.

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 detailed in Section 3.3. 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 4.3. The average treatment effect on this measure can be interpreted as approximating the home-household size-weighted average percentage change in per-capita consumption due 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.

5.1.2 Cost of Treating a Site

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 therefore now turn to putting the per-migrant cost of the program, given in Table 3 of Lee et al. (2021) as 885.84 taka, in the social welfare units described above. We suppose that the cost is constant across sites within a country and the program is financed by a uniform consumption tax in the country. 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.[18]

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

[18] 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.

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,

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

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 World Development Indicators 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. This procedure yields welfare-unit treatment costs of 0.183, 0.145, and 0.153 for Bangladesh, Pakistan, and India, respectively.

5.2 Candidate Migration Corridors

While data is available for a large number of migration corridors in each country (Appendix E.1 discusses the construction of our dataset of site characteristics), we focus on a smaller subset for the purposes of experimental site selection. 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 is survey and program administration practicality: origin 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 E.2 provides details on the migrant density restrictions we use, and Table 4 lists summary statistics on the characteristics of sites matching our criteria across the three countries.

Table 4: Summary Statistics for Site Characteristics Across Corridors by Country

Variables		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 the 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. 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 set restricted to ensure feasibility of treatment delivery and robustness to prior misspecification, see Appendix E.2 for details.

5.3 Description of 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 standard 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 mean zero 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, including the measure of distance between site characteristics.[19] This is 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 the tuning parameter c to 1 while the parameter d is chosen such that the scale of covariance values implied by the smoothing prior are similar to those of the structural prior. In particular, d is chosen to approximately minimize the maximum absolute difference in variance values as defined by the structural and smoothing priors.[20]

While most empirical results shown below are with reference 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. (2022). 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 d 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 $d \approx 0.364$, and we set c to a benchmark value of 1.

Distribution of $\hat{\tau}$ given τ Recall that the distribution of $\hat{\tau}$ is related to the true outcome τ is

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

[19] In Pakistan and India there are more destinations and wage at destination varies more across corridors (see Table 4). We keep the number of characteristics at two for consistency with Bangladesh and because of the aforementioned concerns about dimensionality in the smoothing prior.

[20] We set $d = 0.1$ for Bangladesh and Pakistan, while for India we use $d = 0.073$.

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 treatment effect as observed in the experiment of Lee et al. (2021) 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 H.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.[21] 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 O.1.

Figure 3: Mean Welfare TE (Structural Prior)

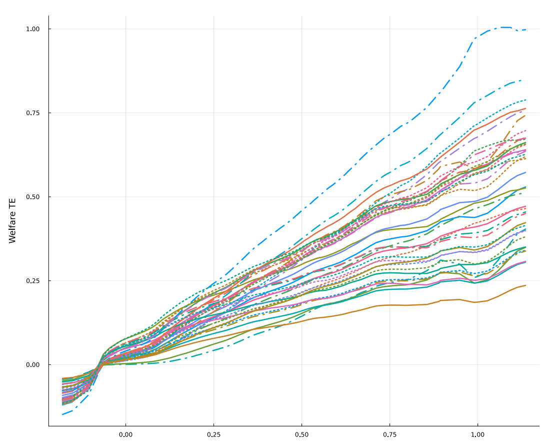
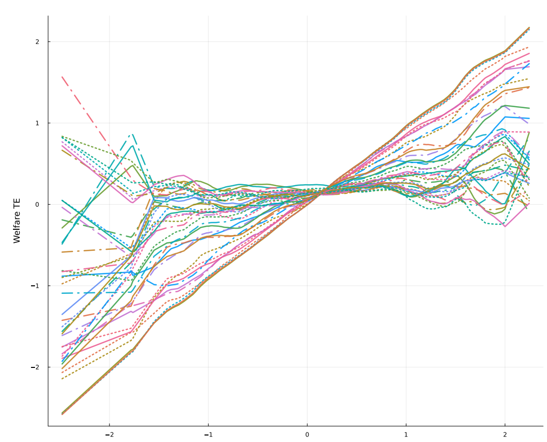


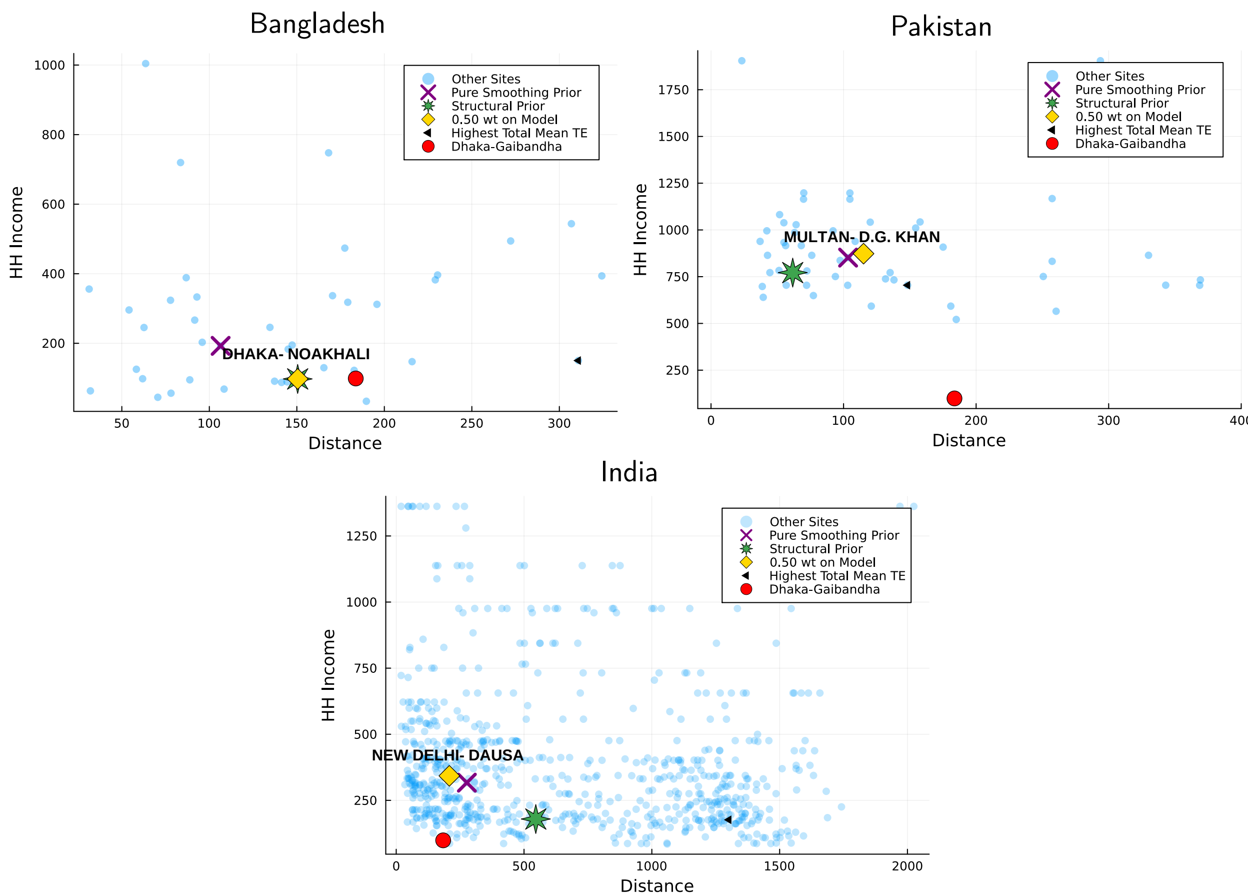
Figure 4: Mean Welfare TE (Smoothing Prior)



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

5.4 Optimal Choice of a Single Experimental Site

Figure 5: Optimal Sites under Different Rules When Choosing a Single Site in Each Country



Notes: Light dots represent candidate sites. HH Income refers to the average income in 2015 Bangladesh Taka, 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 E.1 for details. Site names refer to the sites selected under our preferred mixed prior with 50% weight on the structural model.

We now 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. The top left panel of Figure 5 shows other migration corridors in Bangladesh. 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. The red circle which locates the Dhaka–Gaibandha corridor is thus on the periphery of the main cluster of sites in the figure.

A black triangle indicates the corridor where the highest total mean treatment effect from the intervention is expected, according to the mixed prior. It also involves a relatively poor district like Gaibandha, and it is much further from Dhaka. Choosing a site for additional experimentation which is similar to Dhaka–Gaibandha—or a site where the treatment effect is expected to be particularly large (shown by the black triangle)—would lead to choices that are unlike most other migration corridors in Bangladesh.

The optimally chosen sites for experimentation under the mixed, pure structural, and pure smoothing prior are given by a diamond, 8-pointed star, and X, respectively.[22] These, in contrast, are more similar to the other candidate sites, all of which are of interest to the policymaker. The optimally-chosen corridors are shorter, and the populations are somewhat better off economically under the smoothing prior. The optimal location for experimentation reflects the full decision problem faced by the policy maker. For some sites, the choice of adopting the policy is relatively clear, even without the new experimental data. But policy choices in other sites will be *a priori* uncertain. The policymaker's choices benefit less from experimenting in sites informative for locations where it is already clear from the prior whether or not the treatment is likely to be effective. Experimentation in sites that can inform policy choices where decisions are uncertain thus get more weight in the experimental design because the results are more likely to affect overall welfare.

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 figure shows that the chosen site is far more typical than the Dhaka–Gaibandha corridor in these two dimensions.

As described in Section 2.3, other dimensions can matter as well. 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. Site selection with the structural prior thus reflects dimensions beyond just household income and corridor distance. The richer information yields a site selected 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 longer 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).

The two other panels in Figure 5 show optimal single experimental site selections in Pakistan and India.

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

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.

The bottom panel of Figure 5 shows site selection in India, and it illustrates 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 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? In Sections 5.5–5.6 we analyze the problem of selecting two migration corridors in each country, then in 5.7 we quantify the welfare gains from experimenting in various numbers of sites weighed against the added costs.

5.5 Optimally Choosing Two Sites in Each Country

In Figure 5 for India, there appeared to be two clusters of sites based on distance and household income. The single-site analysis recommended sites relatively central to one of the clusters, but far from the other cluster. Experimenting in two sites, instead of just one, could allow us to learn more about sites in the other cluster. Fortunately, in the actual experiment we are designing, we have the ability (in fact a mandate) to select two sites in each country. We turn to this case now.

Range of Welfare Values Table 5 displays the five best and worst migration corridor pairs for all three countries with average welfare per migrant at the site level evaluated under our preferred mixed prior. The table shows the range of welfare values possible under different choices of experimental sites. As mentioned, there are few destination districts among high-density corridors in Bangladesh, with only two destinations represented in the table. The impacts of optimal site selection are substantial here, with the worst site combination leading to an maximized welfare which is about 7% of the welfare in the best site combination.

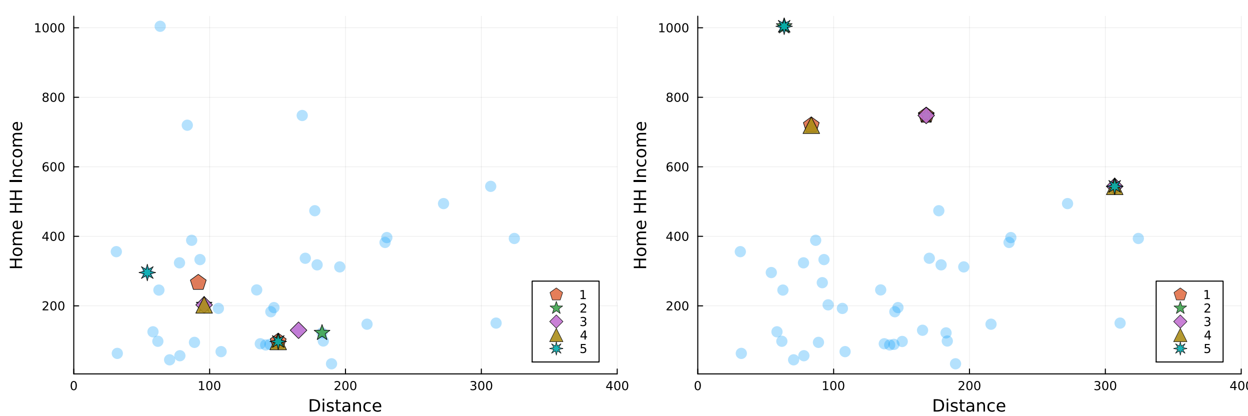
We see that there are more destinations among the high-density corridors in Pakistan than in Bangladesh. Several of the worst combinations involve migration from the district containing Islamabad. In India again corridors involving origin districts with large cities appear in many of the worst site combinations. These corridors do have high migrant densities but the home households are outliers—they have the highest incomes among candidate sites. 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

Table 5: Best and Worst Corridor Combinations in Bangladesh, Pakistan, and India

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-Barguna	Dhaka-Kishoregonj	0.030	Dhaka-Gopalganj	Gazipur-Panchagarh	0.004
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.030	Dhaka-Bhola	Gazipur-Panchagarh	0.004
Dhaka-Kishoregonj	Dhaka-Noakhali	0.030	Dhaka-Bhola	Dhaka-Madaripur	0.002
Dhaka-Faridpur	Dhaka-Noakhali	0.029	Dhaka-Bhola	Dhaka-Gopalganj	0.002
Pakistan					
Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Rawalpindi-Kohat	Okara-Pak Pattan	0.137	Rawalpindi-Islamabad	Karachi-Swat	0.104
Peshawar-Mardan	Okara-Pak Pattan	0.137	Chitral-Islamabad	Karachi-Swat	0.103
Rawalpindi-Mansehra	Okara-Pak Pattan	0.137	Karachi-Lodhran	Karachi-Sukkur	0.102
Rawalpindi-Mardan	Okara-Pak Pattan	0.137	Rawalpindi-Islamabad	Karachi-Sukkur	0.101
Peshawar-Kohat	Bahawalpur-Lodhran	0.137	Chitral-Islamabad	Karachi-Sukkur	0.101
India					
Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Thane-Balrampur	New Delhi-Dausa	0.217	New Delhi-Bangalaore	Bangalore-Chandigarh	0.194
Thane-Balrampur	New Delhi-Hardwar	0.217	New Delhi-Bangalaore	Kancheepuram-Chandigarh	0.194
Thane-Basti	New Delhi-Dausa	0.217	New Delhi-Bangalaore	New Delhi-Chennai	0.194
New Delhi-Dausa	Thane-Sant Kabir Nagar	0.217	Bangalore-Chandigarh	New Delhi-Chennai	0.194
New Delhi-Hardwar	Thane-Sant Kabir Nagar	0.217	Kancheepuram-Chandigarh	New Delhi-Chennai	0.194

the next subsection.

Figure 6: Best (left) and Worst (right) Migration Corridor 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.

Figure 6 shows where the best and worst site combinations from Table 5 for Bangladesh 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 sites such that all candidate corridors are close to at least one experimental site.[23] 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. [24]

Top Site Choices by Prior We now investigate how the choice of top site combinations varies with the prior. Figure 7 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.[25] For Bangladesh and Pakistan 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 to be 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 K 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) differences in characteristics imperfectly explain treatment effect correlations 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.[26] [27]

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 Figure 8, where the smoothing prior picks outlying sites 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 quite unusual sites where prices in the origin district are higher than they are at destination.

[23]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.

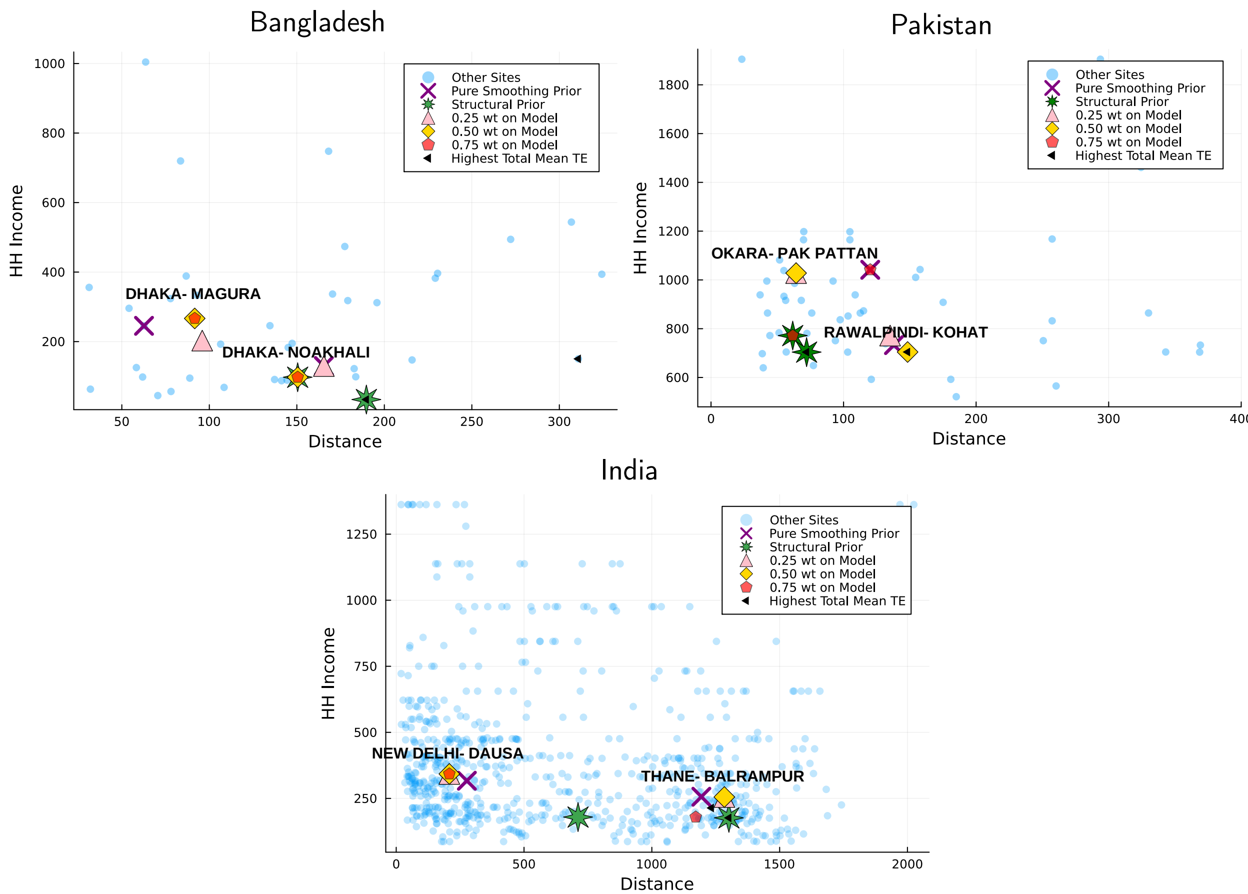
[24]Appendix Figures 20 and 21 show the same pattern in Pakistan and India, respectively.

[25]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 to a variety of mixing weights, as we do in Appendix G.

[26]In Appendix J we compare our decision-theoretic analysis based on the predicting the welfare associated with recommending different sites for treatment to a heuristic approach using the same prior beliefs: experimenting in sites predicted to shrink the uncertainty in beliefs about ATEs across all sites the most.

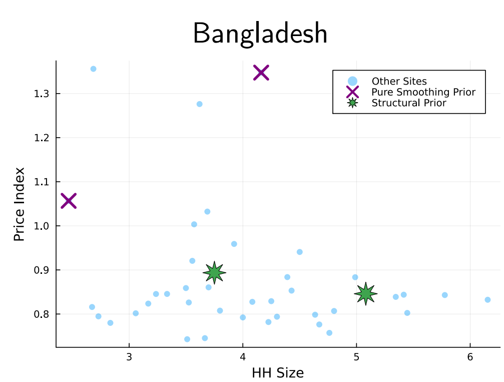
[27]In Appendix L we show that the set of selected sites is robust to variation in ψ, the parameter capturing the effect of the training program on the effective price of mobile money, which helps us be confident in our choice of sites even if the treatment effect may be moderated by unmodeled environmental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

Figure 7: Optimal Sites under Different Priors



Notes: Light dots represent candidate sites. HH Income refers to the average income in Bangladesh taka, 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 E.1 for details. Site names refer to the sites selected under our preferred mixed prior with 50% weight on the structural model.

Figure 8: Smoothing Prior Sites 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.

Sites with the Highest Predicted Effects Figure 7 also shows that the site predicted to have the second-highest effect in Bangladesh coincides with one of the sites selected under the pure structural prior. In Pakistan, the second-highest predicted effect site is also more central than the site with the highest predicted effect. There is no reason to expect this to be so in other applications, but for us it means adding a second site helps with the danger of site selection bias leading to poor policy prescriptions so long as the policymaker makes appropriate use of site characteristics and experimental results to "undo" the site selection bias. If the sites with the highest predicted ATEs were instead among those shown in the right panels of Figure 6, site selection bias could result in up to 93% worse welfare despite a sophisticated policymaker's best efforts to undo it.

5.6 Marginal Sites

A motivation to experiment is to improve policy recommendations in sites where it is not already clear from the prior whether or not 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 nontrivial 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 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 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 5.

To understand why the percentage welfare gains from using best site selection rule instead of the worst is 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 rules will benefit from the treatment effect in sites receiving the training program regardless of which sites are selected for experimentation.

Table 6 shows the per site average welfare from the best and worst site combinations in Pakistan when the welfare values only aggregate over marginal sites. A site pair is optimal precisely because it is informative for marginal sites so the combinations of sites are the same as the best combinations in Table 5. We are interested here in the ratio of predicted welfare when choosing the best vs. the worst site combinations under the 0.5-weight mixed prior. In Table 5 the best pair of sites is predicted to achieve 36% more welfare than the worst pair. Focusing on sites with at least 25% probability of an opposite-signed net benefit of program implementation ($\kappa = 0.25$) in Table 6, however, the best pair is predicted to achieve 70% higher welfare than the worst pair. Table 7 conducts the same exercise for India with similar results.

Table 6: Per Site Welfare from Different Sets of Marginal Sites 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-Kohat	Okara-Pak Pattan	0.137	0.123	0.097	Rawalpindi-Islamabad	Karachi-Swat	0.104	0.089	0.061
Peshawar-Mardan	Okara-Pak Pattan	0.137	0.123	0.097	Chitral-Islamabad	Karachi-Swat	0.103	0.088	0.060
Rawalpindi-Mansehra	Okara-Pak Pattan	0.137	0.123	0.097	Karachi-Lodhran	Karachi-Sukkur	0.102	0.088	0.059
Rawalpindi-Mardan	Okara-Pak Pattan	0.137	0.123	0.097	Rawalpindi-Islamabad	Karachi-Sukkur	0.101	0.087	0.058
Peshawar-Kohat	Bahawalpur-Lodhran	0.137	0.123	0.097	Chitral-Islamabad	Karachi-Sukkur	0.101	0.086	0.057

Table 7: Per Site Welfare from Different Sets of Marginal Sites in India

Best Combinations		κ	Worst Combinations		κ
Corridor 1	Corridor 2	0.25	Corridor 1	Corridor 2	0.25
Thane-Balrampur	New Delhi-Dausa	0.113	New Delhi-Bangalore	Bangalore-Chandigarh	0.080
Thane-Balrampur	New Delhi-Hardwar	0.113	New Delhi-Bangalore	Kancheepuram-Chandigarh	0.080
Thane-Basti	New Delhi-Dausa	0.113	New Delhi-Bangalore	New Delhi-Chennai	0.080
New Delhi-Dausa	Thane-Sant Kabir Nagar	0.114	Bangalore-Chandigarh	New Delhi-Chennai	0.080
New Delhi-Hardwar	Thane-Sant Kabir Nagar	0.114	Kancheepuram-Chandigarh	New Delhi-Chennai	0.080

Figure 9 makes the point visually, plotting the 0.5-weight prior mean and variance for all sites across the three countries and distinguishing marginal and non-marginal sites by color. In Bangladesh all but three sites are marginal, while in Pakistan there is a substantial fraction (about 22%) of sites whose recommendation to receive the training program is unlikely to change based on the experimental results. India has the highest fraction of non-marginal sites (45.5%), with many migration corridors already predicted to benefit a great deal from the program.

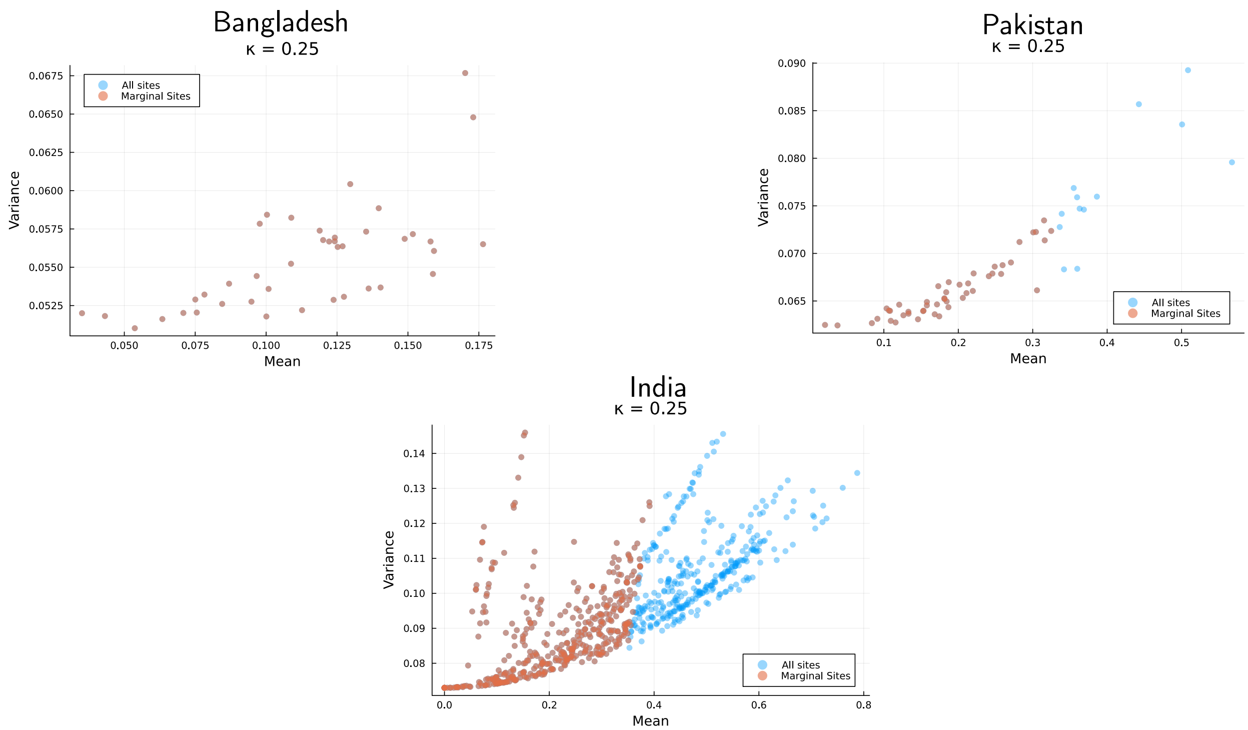
5.7 Choosing the Number of Experimental Sites

In Section 5.5, we analyzed the optimal selection of two migration corridors in each country in which to run experiments. The question is 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.

It is natural to wonder whether we might improve welfare by experimenting in a different number of sites while maintaining the same budget. Our framework allows us to explore such alternatives, by changing the choice set to allow for a different numbers of sites and modifying the sampling error term σ_ϵ to reflect the corresponding change in the sample size per site.

Based on calculations with a Bangladesh survey firm reported in Appendix H.2, in Figure 10 we report

Figure 9: Mean and Variance of Site Level Treatment Effects under the Mixed Prior for Marginal Sites



per-site average welfare in Bangladesh and Pakistan as we vary the number of experimental sites. For Pakistan, we report welfare values over marginal sites with $\kappa = 0.25$ to sharpen comparisons. We maintain a constant budget based on our two-site-per-country, 2,000-households-per-site mandate, which means the number of households we can sample diminishes with more sites since there are fixed costs associated with experimenting in each site. As a baseline, the figure also displays the welfare from no experimentation where 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.

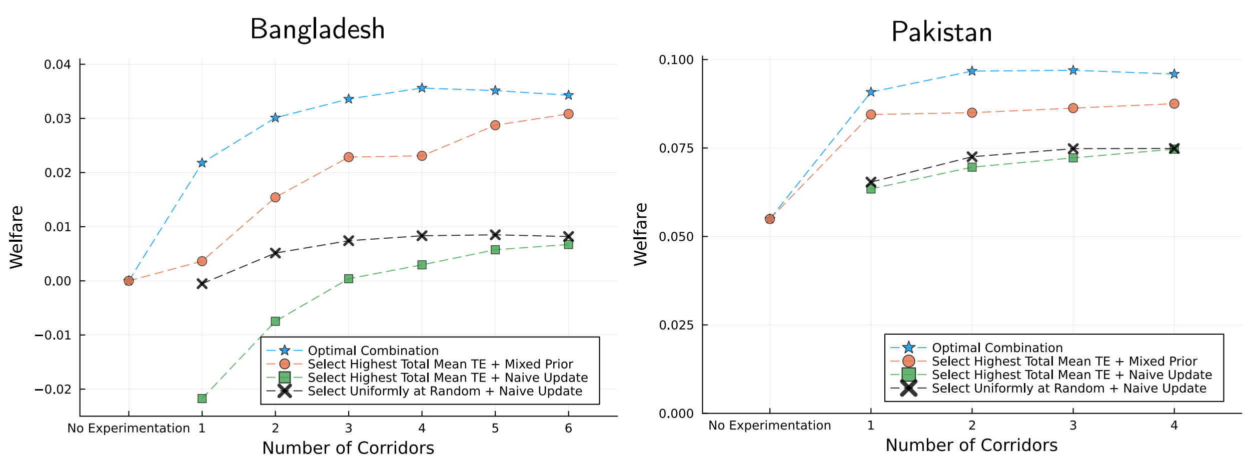
We predict welfare using the 0.5-weight mixed prior so site selection using that prior will attain optimal welfare. This is represented by the line with star markers. In Bangladesh we see large gains in welfare from the first experimental site, and diminishing welfare gains for subsequent sites until 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.

To compare with optimal site selection, we consider two cases of site selection bias, where the sites chosen have the maximal sum of ATEs according to the prior. In the first, represented by the line with circular markers, we allow the policymaker to make optimal use of the treatment effect estimates from the selected sites, updating the predicted treatment effects for other sites according to the informativeness of the highest-treatment-effect sites for each one. This represents a sophisticated policymaker who uses all of our data and machinery to account for the non-random selection of sites. Since we saw in the Bangladesh and Pakistan panels of Figure 7 that some of the sites with the highest predicted treatment effects happen to be chosen under some mixed priors, the welfare cost of site selection bias can be muted to some degree in our application as long as the policymaker is very sophisticated.

We compare these sophisticated scenarios with a naive policymaker who assigns treatment to all sites if the average of estimated ATEs exceeds the cost of treatment, a case considered in Allcott (2015), which we call the uniform rule. Hjort et al. (2021) provide evidence that policymakers may extrapolate experimental results naively. Their sample of Brazilian municipal officials shows a concern for the sample size of prototypical experiments, but not the context where they were performed. In our numerical experiment, naive policymakers fare very poorly when there is extreme site selection bias. Selecting sites with the highest predicted effects and applying the naive policy rule lead to the uniformly lowest expected welfare values.

We also consider the performance of the uniform rule where sites are selected at random, which roughly approximates Allcott (2015)'s recommendation of choosing sites at random to combat site selection bias when making uniform implementation recommendations on the basis of a set of experiments. These are indicated in Figure 10 by the points marked by an "X." This strategy also fares quite poorly, in some cases little better 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 probability that the chosen sites will be very unrepresentative of the entire country.

Figure 10: Gains in Average Welfare When Choosing Sites Optimally vs. Picking Sites with Highest Total Mean Treatment Effect



6 Conclusion

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.

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 our approach.

References

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

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.

ANDREWS, I. AND E. OSTER (2019): "A Simple Approximation for Evaluating External Validity Bias," *Economics Letters*, 178, 58–62.

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.

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.

ATHEY, S., U. BYAMBADALAI, V. HADAD, S. K. KRISHNAMURTHY, W. LEUNG, AND J. J. WILLIAMS (2022): "Contextual Bandits in a Survey Experiment on Charitable Giving: Within-Experiment Outcomes versus Policy Learning," ArXiv:2211.12004.

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.

BAREINBOIM, E., S. LEE, V. HONAVAR, AND J. PEARL (2013): "Transportability from Multiple Environments with Limited Experiments," in *Advances in Neural Information Processing Systems*, ed. by C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Curran Associates, Inc., vol. 26.

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.

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.

CHALONER, K. AND I. VERDINELLI (1995): "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 273–304.

CHAMBERLAIN, G. (2011): "Bayesian Aspects of Treatment Choice," in *The Oxford Handbook of Bayesian Econometrics*, ed. by G. Geweke, J.and Koop and H. van Dijk, Oxford University Press, 11–39.

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

CHEN, X., Z. QI, AND R. WAN (2023): "STEEL: Singularity-Aware Reinforcement Learning," Working paper.

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.

CYTRYNBAUM, M. (2022): "Designing Representative and Balanced Experiments by Local Randomization," Working paper.

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.

DEHEJIA, R. H. (2005): "Program Evaluation as a Decision Problem," *Journal of Econometrics*, 125, 141–173.

DUCHI, J. C. AND H. NAMKOONG (2021): "Learning Models with Uniform Performance via Distributionally Robust Optimization," *The Annals of Statistics*, 49.

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.

HIRANO, K. AND J. R. PORTER (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77, 1683–1701.

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.

HU, F. AND W. F. ROSENBERGER (2006): *The Theory of Response-Adaptive Randomization in Clinical Trials*, New York: Wiley.

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.

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.

KITAGAWA, T. AND A. TETENOV (2018): "Who Should be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

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.

LEI, L., R. SAHOO, AND S. WAGER (2023): "Policy Learning under Biased Sample Selection," ArXiv:2304.11735.

MANSKI, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

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.

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

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.

SI, N., F. ZHANG, Z. ZHOU, AND J. BLANCHET (2022): "Distributionally Robust Batch Contextual Bandits," ArXiv:2006.05630.

SPIEGELHALTER, D. J., L. S. FREEDMAN, AND M. K. B. PARMAR (1994): "Bayesian Approaches to Randomized Trials," *Journal of the Royal Statistical Society, Series A*, 157, 357–416.

STOYE, J. (2009): "Minimax Regret Treatment Choice With Finite Samples," *Journal of Econometrics*, 151, 70–81.

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.

XIONG, R., S. ATHEY, M. BAYATI, AND G. IMBENS (2023): "Optimal Experimental Design for Staggered Rollouts," ArXiv:1911.03764.

Appendices

A Example of Posterior Updating Rule

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

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right).$$

In the notation of Section 2.3.4, we have

$$\mu_\tau = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma_\tau = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

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 $\Sigma_\epsilon = \sigma_\epsilon^2 I$. 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 $\hat{\tau}_1$. The specific form of this posterior distribution can be derived from the results in Section 2.3.4. In this case, we would have

$$\hat{\tau}[\mathscr{S}] = \hat{\tau}_1, \quad \mu[\mathscr{S}] = \mu_1.$$

Since estimation error is uncorrelated with underlying treatment effects the unconditional covariance of $\hat{\tau}_1$ with τ_1 is simply the variance of τ_1 under the prior, while its covariance with τ_2 is the covariance of τ_1 with τ_2, both of which are summarized in the following vector

$$\Sigma_\tau[:, \mathscr{S}] = \begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \end{bmatrix},$$

We also need the following expression for the total variance of $\hat{\tau}_1$, which equals the prior variance of τ_1 and the variance of the estimation error:

$$(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S}, \mathscr{S}] = \left[\sigma_1^2 + \sigma_\epsilon^2\right].$$

With these objects in hand we can now apply the formulas from Section 2.3.4. The posterior mean for τ

given $\hat{\tau}_1$ is

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

The posterior variance is

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

So the posterior distribution of $\tau = (\tau_1, \tau_2)'$ given $\hat{\tau}_1$ can be written as

$$\tau \mid \hat{\tau}_1 \sim N\left(\begin{pmatrix}\mu_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_\epsilon^2}(\hat{\tau}_1 - \mu_1) \\ \mu_2 + \frac{\sigma_{12}}{\sigma_1^2+\sigma_\epsilon^2}(\hat{\tau}_1 - \mu_1)\end{pmatrix}, \begin{bmatrix}\sigma_1^2\left(1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_\epsilon^2}\right) & \sigma_{12}\left(1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_\epsilon^2}\right) \\ \sigma_{12}\left(1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_\epsilon^2}\right) & \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2+\sigma_\epsilon^2}\end{bmatrix}\right).$$

To intepret this expression, 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 $\hat{\tau}_1$. From the expression above, this is

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

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

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

So 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. Then we can write

$$\begin{aligned}\mathbb{V}[\tau_2 \mid \hat{\tau}_1] &= \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2 + \sigma_\epsilon^2} \\ &= \sigma_2^2 \left(1 - \frac{\rho^2 \sigma_1^2}{\sigma_1^2 + \sigma_\epsilon^2}\right).\end{aligned}$$

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.

B Asymptotic Distribution of the Structural Parameter Estimator

B.1 Asymptotics for the Vector of Summary Statistics

Under standard conditions $\hat{\pi}$ is asymptotically linear, i.e.

$$\hat{\pi} = \pi_0 + \frac{1}{n}\sum_{i=1}^{n} \zeta_i + o_p(1/\sqrt{n}),$$

and a central limit theorem applies, so that

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

where $\Sigma_\pi = \mathbb{E}\left[\zeta_i \zeta_i'\right]$. The matrix Σ_π can be estimated as $(Z'Z)/n$ where the matrix Z has rows $\{\zeta_i'\}_{i=1}^n$. Alternatively, Σ_π may be estimated by the bootstrap. The appropriate asymptotic linear forms for each component of $\hat{\pi}$, excluding $o_p(1/\sqrt{n})$ terms, are listed below.

$$\widehat{\mathbb{P}}(R \le r \mid X \in A) = \mathbb{P}(R \le r \mid X \in A) + \frac{1}{n}\left\{\sum_{i=1}^{n} \frac{\mathbf{1}\{R_i \le r, X_i \in A\} - \mathbb{P}(R_i \le r, X_i \in A)}{\mathbb{P}(X \in A)} - \frac{\mathbb{P}(R \le r \mid X \in A)}{\mathbb{P}(X \in A)}\Big(\mathbf{1}\{X_i \in A\} - \mathbb{P}(X_i \in A)\Big)\right\},$$

$$\widehat{\mathbb{P}}(M = 1 \mid R > 0, X \in A) = \mathbb{P}(M = 1 \mid R > 0, X \in A) + \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbf{1}\{M_i = 1, R > 0, X_i \in A\} - \mathbb{P}(M_i = 1, R > 0, X_i \in A)}{\mathbb{P}(R > 0, X \in A)} - \frac{\mathbb{P}(M = 1 \mid R > 0, X \in A)}{\mathbb{P}(R > 0, X \in A)}\Big(\mathbf{1}\{R > 0, X_i \in A\} - \mathbb{P}(R > 0, X_i \in A)\Big)\right\},$$

$$\hat{\alpha} = \alpha + \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{E}[E_m]^2}{(\mathbb{E}[E_m] + \mathbb{E}[W_m]\cdot\mathbb{E}[L_m])^2}\left\{\frac{\mathbb{E}[W_m]\cdot\mathbb{E}[L_m]}{\mathbb{E}[E_m]^2}\cdot\big(E_{m,i} - \mathbb{E}[E_m]\big) - \frac{\mathbb{E}[W_m]}{\mathbb{E}[E_m]}\big(L_{m,i} - \mathbb{E}[L_m]\big) - \frac{\mathbb{E}[L_m]}{\mathbb{E}[E_m]}\big(W_{m,i} - \mathbb{E}[W_m]\big)\right\}. \tag{11}$$

Sample analogs of the expressions above can be used to estimate Σ_π. For reference, the asymptotic variances for each of the above are displayed below.

$$n \cdot V\left(\widehat{\mathbb{P}}(R \le r \mid X \in A)\right) \to \frac{\mathbb{P}(R \le r, X \in A)(1 - \mathbb{P}(R \le r, X \in A)) - \mathbb{P}(X \in A)(1 - \mathbb{P}(X \in A)\mathbb{P}(R \le r \mid X \in A)^2}{\mathbb{P}(X \in A)^2}.$$

$$n \cdot V\left(\widehat{\mathbb{P}}(M = 1 \mid R > 0, X \in A)\right) \to \frac{\mathbb{P}(M = 1, R > 0, X \in A)(1 - \mathbb{P}(M = 1, R > 0, X \in A)) - \mathbb{P}(X \in A)(1 - \mathbb{P}(X \in A)\mathbb{P}(M = 1 \mid R > 0, X \in A)^2}{\mathbb{P}(R > 0, X \in A)^2}.$$

$$\begin{aligned} n \cdot V(\hat{\alpha}) \to \frac{1}{(\mathbb{E}[E_m] + \mathbb{E}[W_m]\cdot\mathbb{E}[L_m])^4}\Big\{ & (\mathbb{E}[W_m]\mathbb{E}[L_m])^2 V(E_m) + (\mathbb{E}[E_m]\mathbb{E}[W_m])^2 V(L_m) \\ & + (\mathbb{E}[E_m]\mathbb{E}[L_m])^2 V(W_m) - 2(\mathbb{E}[W_m])^2 \mathbb{E}[L_m]\mathbb{E}[E_m]\mathrm{Cov}(E_m, L_m) \\ & + 2(\mathbb{E}[E_m])^2 \mathbb{E}[W_m]\mathbb{E}[L_m]\mathrm{Cov}(W_m, L_m) \\ & - 2(\mathbb{E}[L_m])^2 \mathbb{E}[E_m]\mathbb{E}[W_m]\mathrm{Cov}(W_m, E_m)\Big\}. \end{aligned}$$

B.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 (6)) 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 $(\psi, \lambda, \bar{\epsilon}, \bar{\xi})$ are displayed below.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

The above event is non-empty only if,

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

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

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

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

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

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

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

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

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

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

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

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

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

C 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 E.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 (6).

A similar issue arises for inferring δ and $\overline{\xi}$. 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 $\overline{\xi}$. In Section 4.3 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 $\overline{\xi}$. In particular, let o_g denote the operator density in Gaibandha, and let $\hat{\overline{\xi}}$ denote the estimated value of $\overline{\xi}$. We then set:

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

D Model Fit to the Pilot Experiment

Table 8 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 8: Model-Implied and Estimated Average Treatment Effects

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

Figures 11 and 12 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 2 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.

Figure 11: Average Daily Remittances for Female Migrants

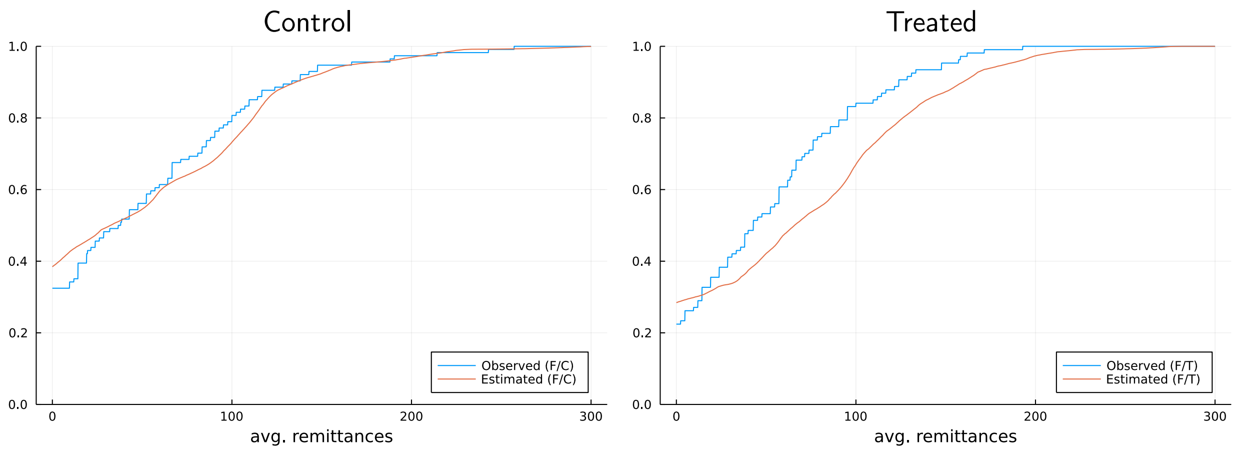
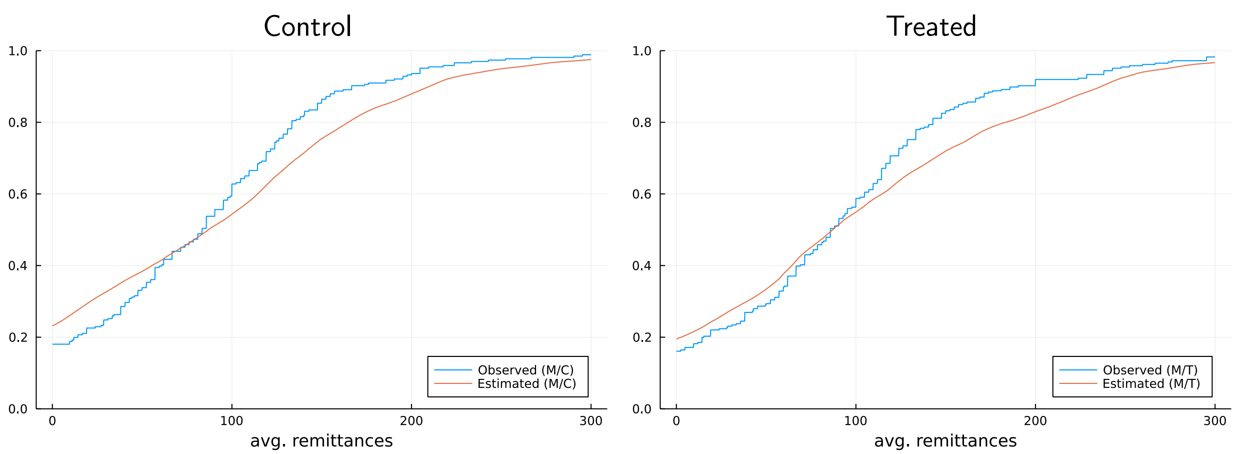


Figure 12: Average Daily Remittances for Male Migrants



E Migration Corridor Data and Restrictions

E.1 Data on Corridor Characteristics

As discussed in Section 4.3, 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). Summary statistics for home households are calculated using only data for households who report sending a migrant or 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.[28]

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 4.3. 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

[28] In Appendix N 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.

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

E.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. In Appendix M we show that the identities of the site combinations chosen are robust to alternative density restrictions.

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 0.005. Our restrictions leave us with 60 possible corridors and 1,125 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.

F Simulation Algorithm

This appendix describes how we use a mixed prior to arrive at the kind of welfare estimates given in e.g. Table 5, and how these estimates determine the sites we select.

F.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, are taken from the structural model. The covariance matrix over predicted site effects, Σ_{mixed}, is a convex combination of the covariance matrix obtained from the structural model and that obtained from the smoothing prior, with weight $w \in (0,1]$ on the structural model:

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

Under the prior the mean of the estimated treatment effects is also equal to the prior mean of site effects so that the joint mean is

$$\begin{bmatrix} \mu_{\tau} \\ \mu_{\hat{\tau}} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix}.$$

Sampling uncertainty is independent and identically distributed across sites, with variance σ_{ϵ}^2 computed as described in Section 5.3:

$$\Sigma_{\epsilon} = \text{Diagonal}\left(\sigma_{\epsilon}^2\right).$$

F.2 Simulations

Our goal is to approximate by simulation the double integral in Equation (2) for each set of possible experimental sites $\mathscr{S}$. Given a value for the vector of true site effects we (1) compute the welfare from observing estimated treatment effects in sites $\mathscr{S}$ then assigning sites to receive treatment following Equation (1), averaging over the distribution of sampling error in the effect estimates (the inner integral).

We then (2) average (1) over the prior belief distribution over possible true site effect vectors (the outer integral).

For each $\mathscr{S}$, then, we run B (10,000) simulations to approximate Step (2) from the previous paragraph. In each simulation b we draw a vector of true and estimated treatment effects $(\tau, \hat{\tau})$ for all sites from the joint distribution described in Section 2.3.4:

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

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 $\hat{\tau}[\mathscr{S}]$ (see Section 2.3.4 for the general formula and Appendix A for an example). We then compute the optimal site-level treatment vector T^*, a set of indicators for whether each site's posterior mean treatment effect exceeds 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

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

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

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

which gives us an entry in e.g. Table 5 for experimental site combination $\mathscr{S}$.

To select sites, we choose the $\mathscr{S}$ with the highest $\overline{W}[\mathscr{S}]$.

G Site Selection Rule Performance Under Alternative Priors

We now examine the welfare predicted under specific mixed priors when choosing sites using different priors. Figure 13 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 a different prior. In other words, sites are selected using the prior indicated by the line type 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 structural prior according to different mixed priors.

It is never optimal to use the pure structural prior to select sites since Figure 13 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 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 0.75, 0.8, and 0.55 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.

H Choosing Different Numbers of Experimental Sites

H.1 Best/Worst Site Combinations with Different Numbers of Sites

Table 9 shows the five best and worst choices when a single corridor is chosen to be experimented on in Bangladesh. Table 10 shows corresponding results for Pakistan, with the welfare values are based on the full set of admissible corridors instead of just marginal corridors.

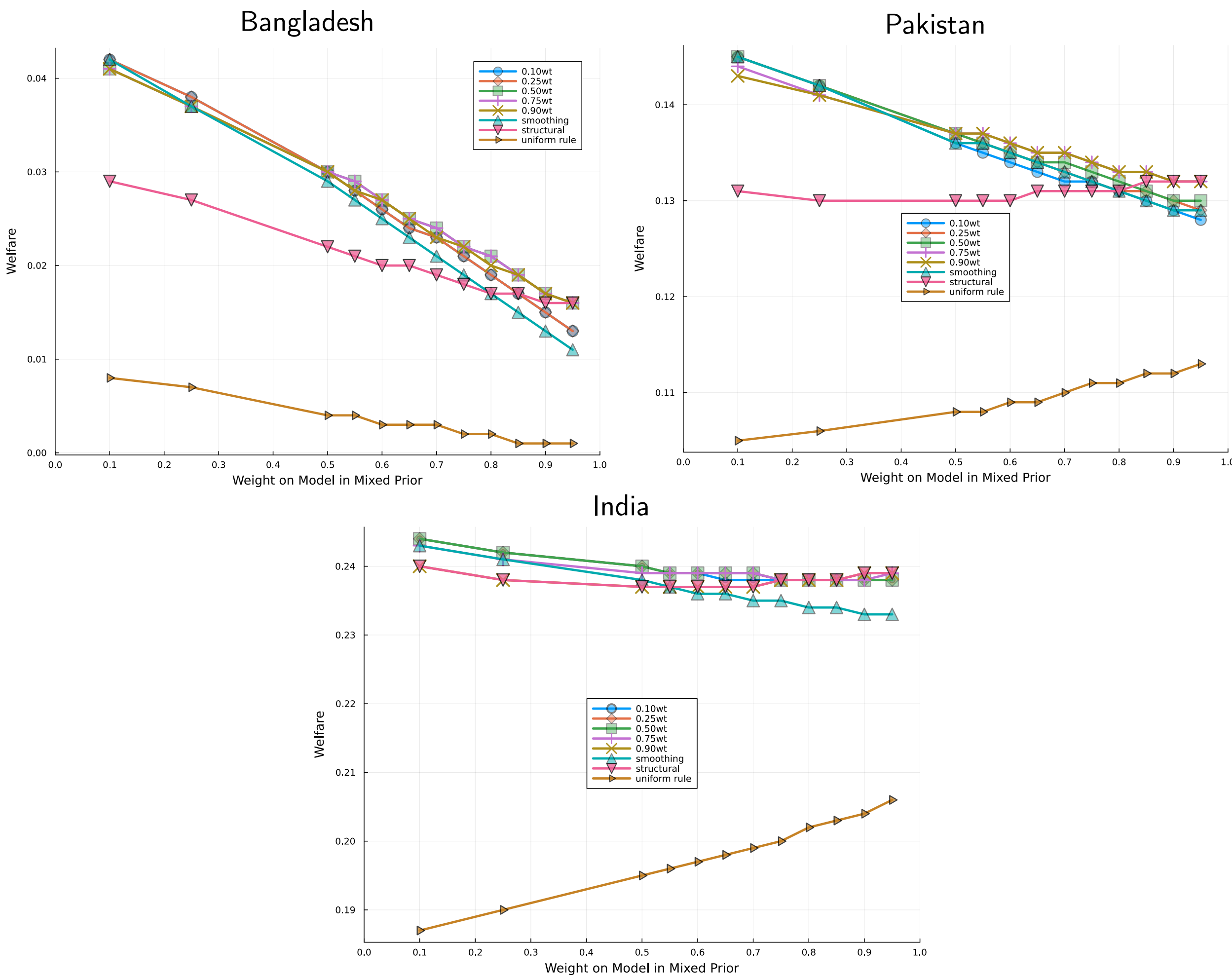
Table 9: Welfare when Choosing a Single Corridor in Bangladesh

Best Choice		Worst Choice	
Corridor	Avg. Welfare	Corridor	Avg. Welfare
Dhaka-Noakhali	0.0218	Dhaka-Panchagarh	0.004
Dhaka-Feni	0.0216	Gazipur-Panchagarh	0.003
Dhaka-Kishoregonj	0.0216	Dhaka-Gopalganj	0.001
Chittagong-Comilla	0.0211	Dhaka-Madaripur	0.001
Dhaka-Pirojpur	0.0210	Dhaka-Bhola	0.001

Table 10: Welfare when Choosing a Single Corridor in Pakistan

Best Choice		Worst Choice	
Corridor	Avg. Welfare	Corridor	Avg. Welfare
Multan-D.G. Khan	0.129	Karachi-Swat	0.101
R. Y. Khan-Ghotki	0.129	Karachi-Lodhran	0.099
Rawalpindi-Mardan	0.129	Karachi-Sukkur	0.099
Swat-Shangla	0.128	Rawalpindi-Islamabad	0.098
Lahore-Narowal	0.128	Chitral-Islamabad	0.098

Figure 13: Welfare of Different Rules Under Different Mixed Priors For Different Countries



H.2 Calculating the Total 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 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 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 11 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 11: 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

I 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 12 shows the best choices for four migration corridors with two each from Bangladesh and Pakistan, respectively, and Table 13 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 12: Best Site Choices when Choosing Jointly across Bangladesh and Pakistan

Bangladesh		Pakistan		
Corridor 1	Corridor 2	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	Lahore-Faisalabad	Peshawar-Mardan	0.0978
Dhaka-Magura	Dhaka-Noakhali	Karachi-Hyderabad	Peshawar-Mardan	0.0977
Dhaka-Magura	Dhaka-Noakhali	Karachi-Hyderabad	Nowshera-Mardan	0.0977
Dhaka-Kishoregonj	Dhaka-Noakhali	Lahore-Faisalabad	Peshawar-Mardan	0.0977
Dhaka-Magura	Dhaka-Noakhali	Peshawar-Mardan	Okara-Pak Pattan	0.0977

Table 13: Worst Site Choices when Choosing Jointly across Bangladesh and Pakistan

Bangladesh		Pakistan		
Corridor 1	Corridor 2	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Madaripur	Gazipur-Panchagarh	Rawalpindi-Islamabad	Karachi-Sukkur	0.0652
Dhaka-Gopalganj	Dhaka-Lalmonirhat	Chitral-Islamabad	Karachi-Sukkur	0.0649
Dhaka-Gopalganj	Gazipur-Panchagarh	Chitral-Islamabad	Karachi-Sukkur	0.0647
Dhaka-Lalmonirhat	Dhaka-Madaripur	Chitral-Islamabad	Karachi-Sukkur	0.0646
Dhaka-Madaripur	Gazipur-Panchagarh	Chitral-Islamabad	Karachi-Sukkur	0.0645

J Aggregate Variance Reduction vs. Full Decision-Theoretic Analysis

In this section we consider the possibility of choosing experimental sites based on a heuristic criterion: achieving the greatest reduction in the variance of beliefs about individual site effects. Suppose we select some subset of sites $\mathscr{S}$ for experimentation. If the prior distribution on the entire vector of site-level effects τ is multivariate normal with mean vector μ_τ and variance matrix Σ_τ, then by the results in Section 2.3.4, the posterior variance after observing the experimental estimates will be

$$\Sigma_\tau - \Sigma_\tau[:,\mathscr{S}]\,\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S},\mathscr{S}]\}^{-1}\,\Sigma_\tau[:,\mathscr{S}]'.$$

Thus the variance is reduced by

$$\mathrm{VR}(\mathscr{S}) = \Sigma_\tau[:,\mathscr{S}]\,\{(\Sigma_\tau + \Sigma_\epsilon)[\mathscr{S},\mathscr{S}]\}^{-1}\,\Sigma_\tau[:,\mathscr{S}]',$$

which accounts for the estimation error in the estimates from the experimental sites (through Σ_ϵ) and the correlation across sites in Σ_τ that enables estimates from the experimental sites to be informative about nonexperimental sites.

The (s,s) diagonal element of $\mathrm{VR}(\mathscr{S})$ measures how much the uncertainty about a single site-level effect τ_s is reduced by experimenting on $\mathscr{S}$. We consider a simple sum of these individual variance improvements as a heuristic measure of informativeness of a design:

$$\sum_{s=1}^{S} \mathrm{VR}(\mathscr{S})_{s,s} = \mathrm{trace}\,(\mathrm{VR}(\mathscr{S}))\,.$$

We compare the site combinations $\mathscr{S}$ that lead to high values for this sum with the site combinations that lead to highest welfare under the structural prior using our full decision-theoretic analysis. Figure 14 displays any migration corridor that appears in the top ten combinations of two sites under the structural prior. The figure also superimposes corridors that appear in the top ten combinations which lead to the greatest reduction in variance as described above. In Bangladesh and Pakistan, there is a high degree of overlap in the top corridors as defined according to these two criteria.

Figure 14: Sites Leading to Greatest Reduction in Variance and Optimal Sites under the Structural Prior

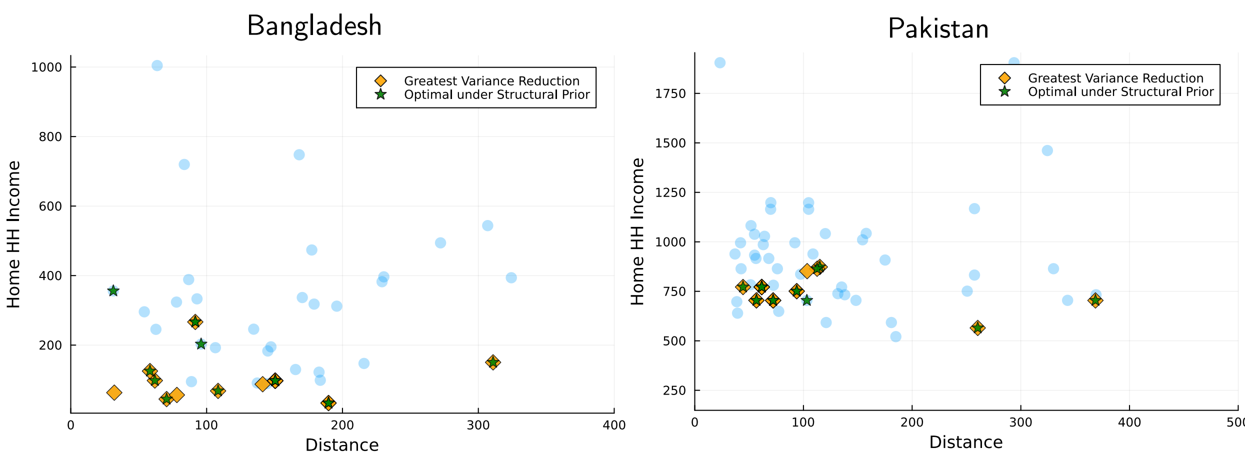


Figure 15 shows a similar comparison for India, but there is hardly any overlap in top sites. Figure 16 examines this in more detail. The left plot of Figure 16 shows the reduction in the aggregate variance generated by different site combinations, in increasing order. The site combinations that are selected by the welfare-maximization problem are superimposed as diamonds, and indicate that the welfare-optimal site combinations do lead to a large reduction in the overall variance measure. The right plot of Figure 16 shows the welfare under the structural prior from all possible site combinations in India, in ascending order. Combinations that lead to the greatest reduction in variance are again superimposed as diamonds, and indicate that the site combinations that lead to the highest variance reduction may not lead to high social welfare. This could be due, in part, to the fact that in India most sites are not "marginal" in the sense of Section 5.6, so that reducing the uncertainty about the treatment effects at those sites does not improve welfare very much.

K 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:

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

Figure 15: Sites Leading to Greatest Reduction in Variance and Optimal Sites under the Structural Prior

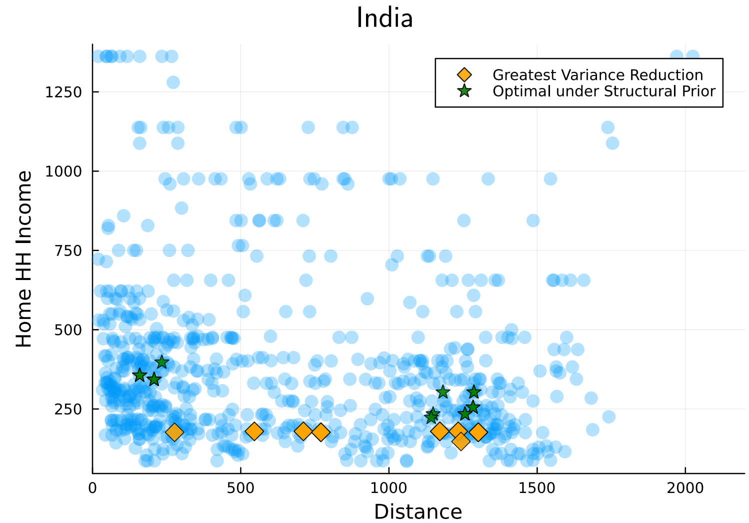
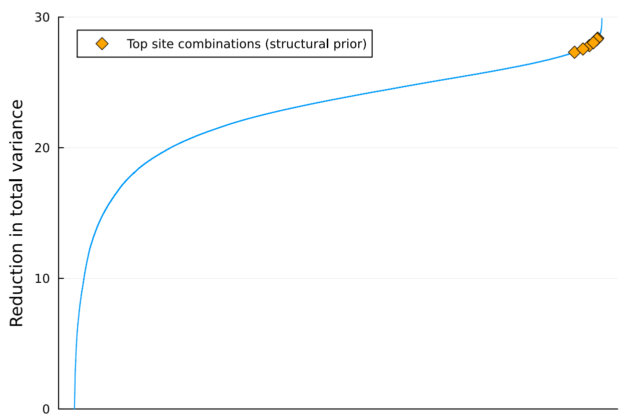
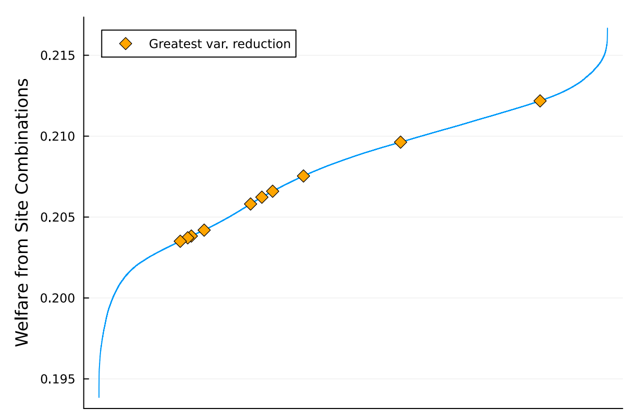


Figure 16: Sites Leading to Greatest Reduction in Variance and Optimal Sites under the Structural Prior





Notes: Sites are ordered from smallest to largest in terms of the variable on the Y-axis.

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 14–16. In all the estimated regressions, the constant term is large. This implies that the structural prior imposes very high correlations in outcomes 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 14: Regressing Structural Correlations on Differences in Site Characteristics in Bangladesh

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9926	0.00019	0.9922	0.993
Distance	-0.0006	0.0002	-0.0009	-0.0002
Home income	-0.0083	0.00034	-0.009	-0.0077
Migrant Wage	-0.0015	0.0002	-0.0019	-0.0011
Home HH size	0.0003	0.00023	-0.0002	0.0007
Operator Density	-0.0013	0.00025	-0.0018	-0.0009
Home income std. dev.	0.0053	0.00031	0.0047	0.0059
Migrant wage std. dev.	-0.0002	0.0002	-0.0006	0.0002
Home HH size std. dev.	-0.0003	0.00025	-0.0008	0.0002
Number of Corrs.	820		F	130.4816
SSR	0.0247		(Prob. >F)	0.0000
R^2	0.5628			

Table 15: Regressing Structural Correlations on Differences in Site Characteristics in Pakistan

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9991	0.00001	0.9991	0.9992
Distance	-	0.00001	-	-
Home income	-0.0001	0.00002	-0.0001	-
Migrant Wage	-0.0001	0.00002	-0.0001	-
Home HH size	-	0.00002	-	0.0001
Operator Density	-0.0010	0.00001	-0.0011	-0.0010
Home income std. dev.	-	0.00002	-	-
Migrant wage std. dev.	-	0.00002	-0.0001	-
Home HH size std. dev.	-0.0001	0.00002	-0.0001	-
Number of Corrs.	1770		F	1207.9341
SSR	0.0003		(Prob. >F)	0.0000
R^2	0.8459			

Table 16: Regressing Structural Correlations on Differences in Site Characteristics in India

Variable	Coef.	Std. Error	Lower 95%	Upper 95%
Intercept	0.9225	0.00029	0.922	0.9231
Distance	0.0017	0.0003	0.0011	0.0023
Home income	0.0016	0.00039	0.0008	0.0023
Migrant Wage	0.0004	0.00033	-0.0003	0.001
Home HH size	0.0012	0.00034	0.0005	0.0018
Operator Density	-0.0802	0.00029	-0.0808	-0.0796
Home income std. dev.	0.006	0.00038	0.0052	0.0067
Migrant wage std. dev.	-0.0024	0.00033	-0.0031	-0.0018
Home HH size std. dev.	0.0011	0.00034	0.0005	0.0018
Number of Corrs.	273430		F	9525.6627
SSR	6402.9309		(Prob. >F)	0.0000
R^2	0.2180			

L 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 likely to be 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 experiment data of Lee et al. (2022) 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 17–20 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. Compared to Table 5, combinations with the best or worst welfare values are largely the same as with our estimated value of ψ.

Table 17: Best and Worst Site Combinations in Bangladesh under Mixed Prior; ψ Magnified by 0.1

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.008	Dhaka-Gopalganj	Dhaka-Panchagarh	0.001
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.008	Dhaka-Bhola	Dhaka-Thakurgaon	0.001
Dhaka-Kishoregonj	Dhaka-Noakhali	0.008	Dhaka-Gopalganj	Dhaka-Thakurgaon	0.001
Dhaka-Noakhali	Dhaka-Shariatpur	0.008	Dhaka-Bhola	Dhaka-Madaripur	0.001
Dhaka-Magura	Dhaka-Pirojpur	0.008	Dhaka-Bhola	Dhaka-Gopalganj	0.001

Table 18: Best and Worst Site Combinations in Bangladesh under Mixed Prior; ψ Magnified by 0.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.015	Dhaka-Bhola	Dhaka-Panchagarh	0.002
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.015	Dhaka-Gopalganj	Dhaka-Thakurgaon	0.002
Dhaka-Barguna	Dhaka-Kishoregonj	0.015	Dhaka-Bhola	Dhaka-Thakurgaon	0.002
Dhaka-Kishoregonj	Dhaka-Noakhali	0.015	Dhaka-Bhola	Dhaka-Madaripur	0.001
Dhaka-Gaibandha	Dhaka-Kishoregonj	0.014	Dhaka-Bhola	Dhaka-Gopalganj	0.001

Table 19: Best and Worst Site Combinations in Bangladesh under Mixed Prior; ψ Magnified by 1.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.055	Dhaka-Gopalganj	Gazipur-Panchagarh	0.020
Dhaka-Barguna	Dhaka-Kishoregonj	0.054	Dhaka-Bhola	Gazipur-Panchagarh	0.020
Dhaka-Kishoregonj	Dhaka-Noakhali	0.054	Dhaka-Madaripur	Gazipur-Panchagarh	0.019
Dhaka-Faridpur	Dhaka-Noakhali	0.054	Dhaka-Bhola	Dhaka-Madaripur	0.017
Dhaka-Feni	Dhaka-Narsingdi	0.054	Dhaka-Bhola	Dhaka-Gopalganj	0.017

Table 20: Best and Worst Site Combinations in Bangladesh under Mixed Prior; ψ Magnified by 2.5

Best Combinations			Worst Combinations		
Corridor 1	Corridor 2	Avg. Welfare	Corridor 1	Corridor 2	Avg. Welfare
Dhaka-Magura	Dhaka-Noakhali	0.113	Dhaka-Gopalganj	Gazipur-Panchagarh	0.082
Dhaka-Faridpur	Dhaka-Noakhali	0.112	Dhaka-Bhola	Gazipur-Panchagarh	0.081
Dhaka-Kishoregonj	Dhaka-Noakhali	0.112	Dhaka-Madaripur	Gazipur-Panchagarh	0.080
Dhaka-Narsingdi	Dhaka-Noakhali	0.112	Dhaka-Bhola	Dhaka-Madaripur	0.080
Dhaka-Magura	Dhaka-Pirojpur	0.111	Dhaka-Bhola	Dhaka-Gopalganj	0.079

M Further Restricting Sites by Migrant Density

Field teams conducting initial scoping work in the three countries were concerned that they might not be able to locate and survey a sufficient number of home-household-migrant pairs in some of the top corridor combinations. We consider restricting the set of migration corridors to sites that satisfy a higher migrant density threshold than what is used to construct the initial set of viable sites, as described in Subsection 5.2.

Tables 21–23 show the top combinations as migrant density thresholds are increased. The welfare values are computed under the same prior as in Table 5 but we only report the site combinations whose constituent sites satisfy the more conservative migrant density threshold. Only four corridors in Pakistan satisfy the highest migrant density threshold, which is why one of the entries in Table 22 is blank.

Table 21: Top Corridor Combinations in Bangladesh as Migrant Density Threshold is Increased

Migrant Density: 0.02			Migrant Density: 0.05		
Corridor 1	Corridor 2	Welfare	Corridor 1	Corridor 2	Welfare
Dhaka-Bagerhat	Dhaka-Kishoregonj	0.030	Dhaka-Kishoregonj	Dhaka-Noakhali	0.030
Dhaka-Kishoregonj	Dhaka-Noakhali	0.030	Dhaka-Gaibandha	Dhaka-Kishoregonj	0.029
Dhaka-Gaibandha	Dhaka-Kishoregonj	0.029	Dhaka-Noakhali	Dhaka-Shariatpur	0.029
Dhaka-Pirojpur	Chittagong-Rangamati	0.029	Dhaka-Jhalokati	Dhaka-Shariatpur	0.027
Dhaka-Chandpur	Dhaka-Pirojpur	0.029	Dhaka-Gaibandha	Dhaka-Shariatpur	0.027

Table 22: Top Corridor Combinations in Pakistan as Migrant Density Threshold is Increased

Migrant Density: 0.01			Migrant Density: 0.015		
Corridor 1	Corridor 2	Welfare	Corridor 1	Corridor 2	Welfare
Peshawar-Kohat	Bahawalpur-Lodhran	0.137	Rawalpindi-Abbottabad	Lahore-Okara	0.133
Bahawalpur-Lodhran	Rawalpindi-Mardan	0.136	Karachi-Kohat	Lahore-Okara	0.117
Bahawalpur-Lodhran	Rawalpindi-Mansehra	0.136	Karachi-Mansehra	Lahore-Okara	0.116
Peshawar-Kohat	Multan-Lodhran	0.136	Lahore-Okara	Karachi-Swat	0.114
Multan-Lodhran	Rawalpindi-Mardan	0.136			

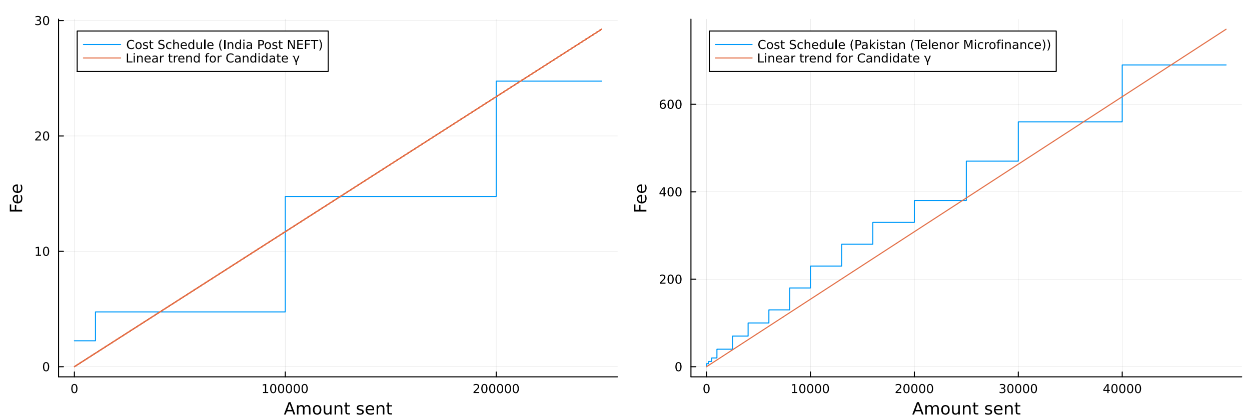
Table 23: Top Corridor Combinations in India as Migrant Density Threshold is Increased

Migrant Density: 0.01, Migrants per unit area: 0.75			Migrant Density: 0.03, Migrants per unit area: 1.0		
Corridor 1	Corridor 2	Welfare	Corridor 1	Corridor 2	Welfare
Thane-Balrampur	New Delhi-Dausa	0.217	Thane-Balrampur	New Delhi-Dausa	0.217
Thane-Balrampur	New Delhi-Hardwar	0.217	Thane-Basti	New Delhi-Dausa	0.217
Thane-Basti	New Delhi-Dausa	0.217	New Delhi-Dausa	Thane-Sant Kabir Nagar	0.217
New Delhi-Dausa	Thane-Sant Kabir Nagar	0.217	New Delhi-Dausa	Thane-Mirzapur	0.216
New Delhi-Hardwar	Thane-Sant Kabir Nagar	0.217	New Delhi-Dausa	Surat-Sant Kabir Nagar	0.216

N Defining 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. (2022). 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 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 17 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 17: Cost Schedules for Local Digital Money Transfer Services



O Additional Figures

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

Figures 18-19 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 18: Mean Welfare TE (Structural Prior)

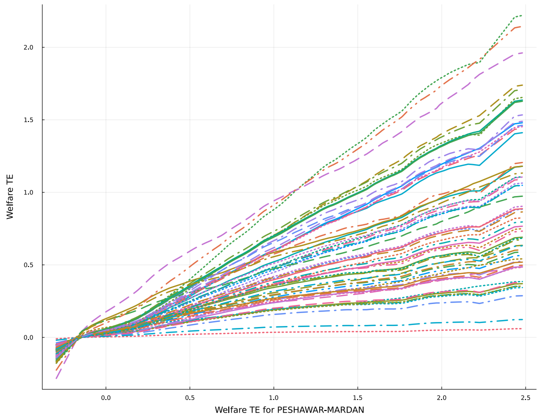
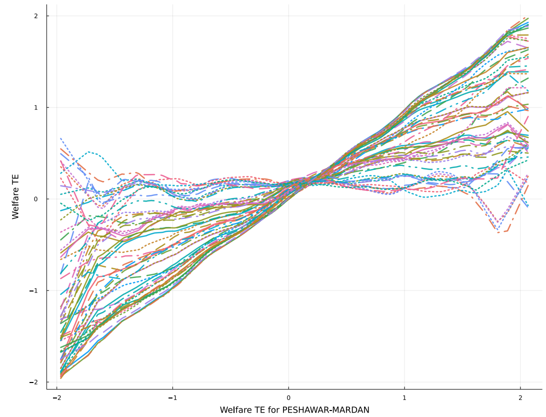


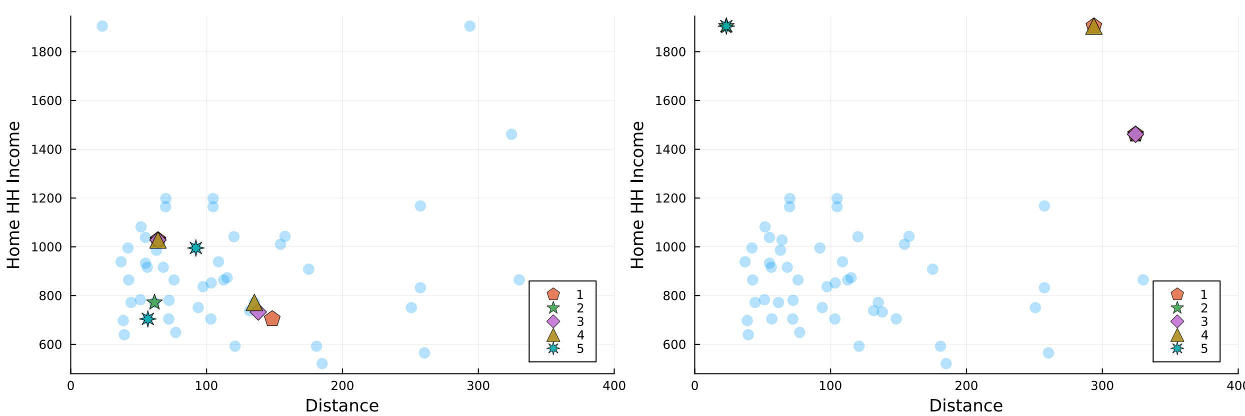
Figure 19: Mean Welfare TE (Smoothing Prior)



O.2 Best/Worst Corridor Combinations for Pakistan and India

Figures 20 and 21 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, analogously to Figure 6 in the main text. As in the main text, the best site combinations are located in the middle of the distribution of characteristics while the worst site combinations are outliers.

Figure 20: Best (left) and Worst (right) Migration Corridor Combinations in Pakistan

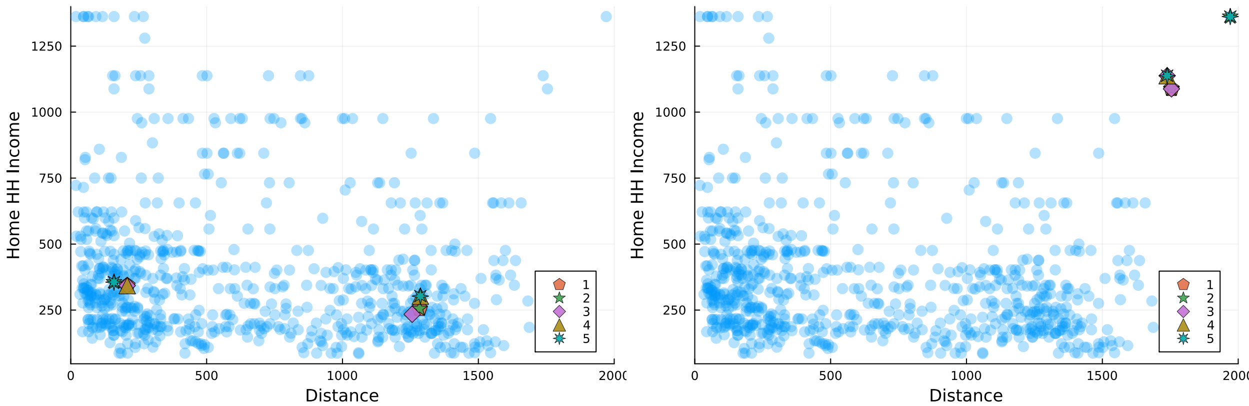


Notes: light dots represent candidate sites. Numbers in the legend refer to rank among the top or bottom sites. E.g. 1 refers to the best site pair in the left panel and the worst pair in the right panel.

P Price Index Computation

Fix a particular locality l and product j. Then, for this (l, j) pair, we have in the raw dataset two key pieces of information: quantity and expenditure for each household i that is surveyed. Say there are N

Figure 21: Best (left) and Worst (right) Migration Corridor Combinations in India



Notes: light dots represent candidate sites which are not among the best (worst) pairs. Numbers in the legend refer to rank among the best (worst), e.g. 1 refers to the best (worst) site pair in the left (right) panel.

households surveyed for this (l, j) pair. We first measure the sum of quantities consumed $\left(q \equiv \sum_i q_i\right)$ and the sum of expenditure $\left(\text{exp} = \sum_i \text{exp}_i\right)$.

We can then obtain the average expenditure by a household as simply $\frac{\text{exp}}{\text{q}}$. We can then obtain the average price $\text{avg\_p} = \frac{\text{p}}{\text{N}}$ and average consumption $\left(\text{avg\_q} = \frac{q}{N}\right)$ of each product in each locality.

We then calculate the index matrix, going by origin locality I get both origin prices (Sp) and origin quantities (Sq) and I compare it to destination Prices and Quantities. Finally I calculate the expenditure given the destination prices of the origin basket (Cexp) and the expenditure given both origin basket and origin prices (Sexp). The Laspeyres price index is then $(\frac{C_{\text{exp}}}{S_{\text{exp}}})$, dividing Expenditure over Quantity.