Experimental Design for Policy Choice

Samuel D. Higbee

Department of Economics, University of Chicago

Click here for the latest version

samuelhigbee@uchicago.edu

December 4, 2024

We study how to design experiments for the objective of choosing optimal polices. An experimenter wants to choose a policy to maximize welfare subject to budget or other policy constraints. The effects of counterfactual policies are described by a structural econometric model governed by an unknown parameter. The experimenter has access to some pilot data, and has the opportunity to collect additional data through an experiment. The joint experimental design and policy choice problem is a dynamic optimization problem with a very high-dimensional state space, since the chosen policy depends on the realized data. We propose a low-dimensional approximation to the solution and show it is asymptotically optimal under Bayes expected welfare. The method applies to policies allocating discrete as well as continuous treatments, such as cash transfers, prices, or tax credits, which may be targeted on the basis of covariates. We demonstrate the method using the conditional cash transfer program Progresa, showing how to design an experiment to help choose a policy aimed at increasing graduation rates and reducing gender disparities in education. Compared to the original Progresa experiment, the optimal experiment requires 60% fewer observations to obtain equally effective policies.

I thank Stéphane Bonhomme, Max Tabord-Meehan, Guillaume Pouliot, Arun Chandrasekhar, Alex Torgovitsky, Lars Peter Hansen, Azeem Shaikh, Giovanni Compiani, Ali Hortaçsu, and seminar participants at the University of Chicago, The University of Notre Dame, and Brigham Young University for helpful comments.

1 Introduction

Economists often use experiments to evaluate, compare, and choose policies. Given data from an experiment and an econometric model, decision makers can evaluate counterfactual policies and choose the one that maximizes some measure of welfare. Experimental design can play an important role in this process, as better estimates of the effects of counterfactual policies can minimize the chances of choosing suboptimal policies. For a finite experimental budget, some experimental designs will be more informative about the optimal policy than others. Such experiments should be preferred when the ultimate goal of the experiment is to inform policy choice.

Despite the fact that choosing policies is a ubiquitous objective for decision makers, very little is known about how to optimally design experiments for this purpose. Instead, much of the guidance on experimental design assumes the objective of maximizing the precision of parameter or treatment effect estimates (see Athey and Imbens (2017), as well as Section 1.1 below for a review). However, not all parameters are equally important for choosing policies. For example, estimating the average effect of a welfare benefits program may not be helpful for choosing how to target benefits based on income or the number of dependents. Likewise, for the purpose of policy choice it is not helpful to estimate heterogeneous effects across dimensions that the policy cannot influence. Instead, the experiment should focus on learning about the effects of counterfactual policies on the policymaker's objective, prioritizing policies that are most likely to be welfare-enhancing. Because the objective and constraints of a policy choice problem differ from those of a parameter estimation problem, an experiment designed around one objective may not be well-suited for the other.

This paper provides a method for designing experiments specifically for the objective of policy choice. We consider an experimenter who wants to choose a policy to maximize expected welfare subject to budget or other policy constraints. The effects of counterfactual policies are described by a structural econometric model governed by an unknown finite-dimensional parameter. The experimenter has access to some pilot data, and has the opportunity to collect another wave of experimental data to learn about the parameter. Our goal is to design this wave to be as helpful as possible for choosing the best policy. Our proposed method uses pilot data to characterize the parts of the policy space that are likely to be effective, and then designs the main wave of the experiment to learn which of these policies is best. By taking into account both the objective and constraints

of the policy choice problem, the method identifies the most valuable margins for experimentation. The resulting experiment delivers data which is maximally informative for choosing the best policy.

Experimental design for the purpose of policy choice is a dynamic decision problem which is difficult to solve due to the high-dimensionality of the state space. In the first period, the experimenter designs the main wave of the experiment based on pilot data. In the second period, the experimenter chooses a policy based on the experimental data, subject to constraints. Since the policy depends on the realized data, a naive approach requires the experimenter to specify which policy they would choose for every possible dataset that the experiment could generate. Even in simple cases, the state can consist of thousands of continuous variables. As a practical matter, solving this problem directly is not possible.

We propose a tractable method for finding the optimal experiment based on two approximations justified by asymptotic theory. First, we approximate the finite-sample experiment with a limit experiment in which the experimenter observes only a Gaussian estimate of model parameters. This reduces the question of selecting an experimental design to that of selecting a variance-covariance matrix for the Gaussian estimate. Second, we use a quadratic approximation to the welfare function. Under this quadratic approximation, the optimal policy depends only on a low-dimensional linear function of parameter estimates— namely, the marginal effect of adjusting the policy. Together, these two approximations suggest that the intractable finite-sample experimental design problem can be replaced with a tractable, low-dimensional problem in which only estimates of the marginal effect of the policy are observed. This enables us to develop an efficient algorithm to solve for the optimal experiment in practice.

We justify our proposed method by showing that it leads to the best possible welfare asymptotically, as the size of the experiment grows. Specifically, an experimenter who (i) uses our method to select the experimental design, and (ii) uses the resulting data to choose a policy, will have the highest possible limiting welfare across all possible experimental designs. We achieve this by validating the two approximations underlying our method as the size of the experiment grows large. We validate the first approximation by characterizing the asymptotic behavior of policies as the experimental design varies. We provide a limit-of-experiments result (Le Cam 1972) that shows that any policy is asymptotically equivalent to a policy in a Gaussian environment, uniformly over all designs. We validate the second approximation by showing that the policy choice problem in the

limit experiment is equivalent to solving a quadratic program. Together, these results imply that our method is asymptotically equivalent to directly solving the intractable, finite-sample decision problem.

We demonstrate the method in an application to the Progresa conditional cash transfer experiment (Schultz 2004, Gertler 2004, Gertler, Martinez, and RubioCodina 2012, Parker and Todd 2017). As an example of a possible objective, consider an experimenter who wants to maximize school completion and minimize gender disparities in graduation rates. The policy choice problem is to choose the amount of the subsidy, which can vary by grade and gender, to maximize this objective subject to a budget constraint. The experimenter has a limited budget with which to learn about the optimal policy. Our goal is to design an experiment which most effectively helps the experimenter select the optimal policy.

In this application, we estimate that our method requires 60% fewer observations than the original experimental design to obtain equally effective policies. This dramatic improvement in the cost-effectiveness of the experiment comes from (i) focusing the experiment on subpopulations most responsive to the transfer, as indicated by pilot data, and (ii) delivering more precise estimates of the effect of the cash transfer on welfare on these subpopulations. We find that the optimal experiment only experiments on secondary school children, since this is where the marginal effect of the subsidy is highest (Todd and Wolpin 2006, Attanasio, Meghir, and Santiago 2012). The optimal experiment offers a large subsidy to children in secondary school, which delivers more precise estimates of the marginal effect of the subsidy for this group.

Our approach applies to a wide range of possible objectives, policies, and constraints. The method is applicable to both discrete and continuous treatments, so long as the set of policies under consideration exhibits some decreasing returns. Since the subsidy is continuous in the Progresa application, this means we require that educational attainment is concave in the size of the subsidy. For a binary treatment, the policy may exhibit decreasing returns when the treatment is targeted based on a continuous variable. For example, if the treatment effect is a decreasing function of a continuous covariate like income, then increasing the threshold for treatment eligibility exhibits decreasing returns.

The objective can be any function of the policy and the parameters of the model. This allows the experimenter to target objectives that may not be simple reduced-form functions of the data.

For example, anti-poverty programs in developing countries may aim to maximize the long-run subjective utility of the recipients. Firms may seek to maximize long-run profits, mediated through the choices of forward-looking agents. Finally, a government may not be able to directly experiment with a minimum wage, but may be able to choose the best minimum wage policy by learning about labor supply elasticities through a cash transfer experiment. The methods of this paper therefore enable experimenters to take advantage of the "best of both worlds" described by Todd and Wolpin (2023) to leverage experimental variation and economic structure to efficiently learn about policies and quantities of interest.

1.1 Related Literature

This paper links two complementary literatures in econometrics: experimental design and policy choice. The questions of how to design experiments, and of how to choose policies based on the results of experiments, are often treated separately.

The optimal design of experiments constitutes a vast literature in many fields. The classic approach aims to estimate the parameters of a parametric model under various optimality criteria (Silvey 2013, Pukelsheim 2006, Chaudhuri and Mykland 1993, Chaloner and Verdinelli 1995). More recent works in econometrics have focused on experimental design with the aim of efficiently estimating treatment effects with binary or discrete treatments in semiparametric settings. This literature is reviewed in Athey and Imbens (2017). Examples include Hahn, Hirano, and Karlan (2011), Bai (2022), Viviano (2022), Tabord-Meehan (2023), Cytrynbaum (2024), Bai et al. (2024). The methods proposed by these papers, like the method proposed here, rely on large pilot samples to inform the design of the main experiment. When this assumption fails, such methods can have poor finite-sample properties (Cai and Rafi 2024). Asymptotic efficiency bounds for ATE estimation across experimental designs are studied in Armstrong (2022). Our setting is parametric, but we emphasize that our contribution of an asymptotically optimal experiment for policy choice in parametric models is both novel and a necessary first step towards designing experiments for policy choice in more flexible settings.

Another large literature in econometrics focuses on policy choice, given experimental or observational data. This is often done in a nonparametric or semiparametric setting (Manski 2004, Bhattacharya and Dupas 2012, Kitagawa and Tetenov 2018, Athey and Wager 2021, Mbakop and

Tabord-Meehan 2021, Sakaguchi 2024, among others) or a Gaussian environment motivated by asymptotic theory (Stoye 2009, Stoye 2012). Counterfactual evaluation and policy choice is also a central goal of structural econometric models. Some examples in development economics which use structural models in conjunction with experiments to evaluate counterfactual policies include Todd and Wolpin 2006, Attanasio, Meghir, and Santiago 2012, and Duflo, Hanna, and Ryan 2012). This approach is surveyed in Todd and Wolpin (2023).

The econometric policy choice literature takes the data-generating process to be outside the control of the decision maker. We complement this literature by providing guidance on how to design an experiment that will be used to choose a policy. That is, we show how to choose the most favorable data-generating process for the policy choice problem from a given class of experiments. In particular, we extend the asymptotic analysis of treatment choice rules as in Hirano and Porter (2009), Hirano and Porter (2020), and Xu (2024) to general nonlinear constrained decision problems and optimize the value of this problem across experimental designs.

Adaptive experiments are commonly used to choose welfare-maximizing policies, with the multi-armed bandit literature providing many algorithms for both in-sample and out-of-sample welfare maximization (see Lattimore and Szepesvári (2020) and Russo et al. (2018) for recent surveys). In econometrics, statistical properties of multi-armed bandits have been studied in Hirano and Porter (2023) and Chen and Andrews (2023). The literature on optimal policies in bandits is vast, with examples including Adusumilli (2024), Kasy and Sautmann (2021), Krishnamurthy et al. (2023), Cesa-Bianchi, Colomboni, and Kasy (2024), and Viviano and Rudder (2024). The results of this paper are most closely related to Hirano and Porter (2023) and Adusumilli (2024). Both papers, like ours, work in a limit experiment framework to characterize the asymptotic behavior of adaptive experiments. As a result, we obtain policies which are optimal in the limit experiment, in contrast to the rate-optimality results in much of the multi-armed bandit literature. Unlike our paper, which considers a nonlinear welfare function, nonlinear constraints, and policies which may assign continuous treatments on the basis of covariates, the aforementioned papers work in a multi-armed bandit setting where the decision is how many units to assign to each arm.

1.2 Outline

In Section 2, we describe the general framework and the decision problem faced by the experimenter, and find that an exact solution is infeasible. In Section 3, we describe a tractable solution method where the difficult decision problem is replaced by a simpler Gaussian experiment with quadratic loss and linear constraints. In Section 4, we show the proposed method is asymptotically optimal, using a limit experiment framework. In Section 5, we study the application to Progresa in more detail. In Section 6, we discuss extensions to robust Bayes preferences and multi-wave experiments.

2 Environment and decision problem

In this section we present the decision problem faced by the experimenter. There are two decisions the experimenter has to make: how to design the main wave of the experiment, and what policy to choose after observing the results of the experiment. This constitutes a two-period dynamic decision problem which in principle can be solved by backwards induction, but is far too high-dimensional to solve in practice.

We describe each component of the decision problem in turn. They are (1) the set of experimental designs available, (2) the data-generating process as described by an econometric model, (3) the set of possible policies, and (4) the welfare or objective of the experiment. We then combine these components to formally state the decision problem. Finally, we discuss some specific settings in which the framework can be applied.

2.1 Experimental designs

The experimenter designs the main wave of the experiment by deciding how to randomly assign treatment to the sample on the basis of covariates.

Specifically, the experimenter has access to n experimental units indexed by $i \in \{1, \dots, n\}$. The first n_0 units constitute the pilot sample, and the remaining $n_1 = n - n_0$ units constitute the main sample. Associated with each observation is a vector of covariates x_i, a treatment assignment z_i, and an outcome y_i. The experimenter has no control over how the pilot sample is chosen, but can choose the design of the experiment for the main sample.

The assignment of treatment z_i can depend on covariates x_i. If the treatment is discrete, the

experimenter can choose the probability of treatment as a function of covariates. If the treatment is continuous, the experimenter may choose both the probability of treatment and the magnitude of treatment as a function of covariates. In general, the experimental designs are represented by conditional distributions of the form

$$p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})$$

where $\boldsymbol{\delta}$ is a finite-dimensional design vector parametrizing this distribution. For the n_0 units in the pilot sample, $\boldsymbol{\delta}$ is fixed to some value $\boldsymbol{\delta}_0$ outside the control of the experimenter. In contrast, the experimenter can choose $\boldsymbol{\delta}$ governing the treatment assignment of the remaining n_1 units in the main sample.

The choice of design is subject to some constraints. For example, the experimenter may have a budget constraint which limits the total cost of the experiment, or there may be constraints that a continuous treatment such as a price or subsidy must be nonnegative or bounded. We represent these constraints by the inequalities

$$f(\boldsymbol{\delta}) \leq 0.$$

Remark 2.1 (Endogenous sample size)**:** The experimenter need not observe all n_1 units in the main sample. In particular, if sampling units is costly a smaller experiment may be preferred. We can accommodate this by allowing $\boldsymbol{\delta}$ to include a probability of sampling each unit. In this case it is required that this probability be bounded away from zero. ◊

Remark 2.2 (I.i.d sampling)**:** We restrict attention to i.i.d. sampling, where the treatment of unit i is independent of the treatment of any other unit conditional on x_i. This simplifies the asymptotic analysis of Section 4. We defer the analysis of designs with dependence, such as complete randomization, biased-coin designs, and matched-pair designs, to future work. ◊

By changing $\boldsymbol{\delta}$, the experimenter can change the design of the experiment along many dimensions, including probability of treatment and the size or dosage of continuous treatments. We give a few examples of policies that may be used in practice.

Example 2.3 (Binary treatment)**:** Many interventions, such as a job training program, are implemented as binary treatments. In this case, the experimenter can choose the probability of treatment as a function of covariates. This may be desirable if the cost of treatment or variance of outcomes depends on covariates. One possible design is to assign treatment with probit probabilities, where treatment is assigned according to

$$z_i = \mathbb{1}[x_i'\boldsymbol{\delta} + \nu_i > 0]$$

where x_i is a vector of covariates and ν_i is a standard normal random variable independent of x_i. $\Diamond$

Example 2.4 (Continuous-valued treatments)**:** Prices, subsidies, unemployment benefits, and many other treatments can take on a continuum of values. Such treatments can be assigned according to a continuous distribution, or may be discretely distributed with support points chosen by the experimenter. If a continuous distribution is desired, one can assign treatment by

$$z_i = x_i'\delta_1 + \nu_i \times \delta_2$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2)$ and ν_i is a randomization device with a continuous distribution, such as a standard normal random variable.

On the other hand, it may be more practical to administer an experiment which implements only finitely many treatment values. For example, the experimenter may divide the population into treated and control groups and assign a treatment value which is a linear function of (potentially discretized) covariates within the treatment group. This takes the form

$$z_i = x_i'\delta_1 \times \mathbb{1}[\nu_i \leq \delta_2]$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2)$ and ν_i is a uniformly distributed randomization device which determines whether the unit is in the treatment group. This class of designs is used in the Progresa example of Section 5. $\Diamond$

Example 2.5 (Rankings and combinatorial treatments)**:** Our approach can also handle high-

dimensional discrete treatments like rankings in online search results (Ursu 2018, Compiani et al. 2023). Suppose $x_i = (x_{i1}, \ldots, x_{iJ})$ is a vector of covariates for J alternative products. The experimenter wants to assign z_i which is a permutation of $\{1, \ldots, J\}$ designating the position of each product in a list. This is a discrete treatment with $J!$ possible values. We can parametrize the distribution of z_i by the ranked logit probabilities, given by

$$p_{z|x}(z_i = z \mid x_i; \boldsymbol{\delta}) = \prod_{j=1}^{J-1} \frac{\exp(x'_{iz_j}\boldsymbol{\delta})}{\sum_{j'=j}^{J} \exp(x'_{iz_{j'}}\boldsymbol{\delta})}.$$

where $z_j \in \{1, \ldots, J\}$ is the jth element of the permutation z. $\lozenge$

2.2 Model

While the experimenter chooses the design governing the distribution of treatment, the experimenter must learn about the data-generating process for the outcome through the experiment.

The data-generating process is described by a model describing the distribution of the outcome y_i conditional on treatment z_i and covariates x_i. The model for the outcome is given by a conditional distribution

$$p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta})$$

parametrized by $\boldsymbol{\theta} \in \mathbb{R}^\ell$. As discussed in Section 1 and as will be discussed in examples throughout the paper, the model may encode key identifying assumptions as well as economic primitives used for welfare calculations.

Together, the model and the choice of design determine the data-generating process conditional on covariates, denoted

$$p(y_i, z_i \mid x_i; \boldsymbol{\theta}, \boldsymbol{\delta}) = p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})$$

for each unit i. We do not model the data-generating process for covariates, and perform the analysis conditional on covariates since the experimenter observes covariates before assigning treatment. Below, when writing expectations, we will leave the conditioning on covariates implicit.

Example 2.6 (Discrete choice)**:** Suppose y_i is a discrete choice generated by a multinomial logit model conditional on price z_i and covariates x_i. The data-generating process for y_i is

$$y_i = \text{argmax}_j \quad u_{ij}$$
$$u_{ij} = x_{ij}'\theta_1 + z_{ij}'\theta_2 + \epsilon_{ij}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and ϵ_{ij} is a T1EV random variable. This model specifies the distribution of y_i conditional on z_i and x_i. The experimenter can choose a design to vary the price z_i along the lines of Example 2.4. ◊

Example 2.7 (Welfare participation)**:** As a special case of the discrete choice model, consider a simple model of labor supply in a welfare program. Utility in the non-working and working states, respectively, are

$$u_{i0} = v_i + (1 + \theta_1)z_i$$
$$u_{i1} = v_i + w_i$$

where z_i is the benefit level and $x_i = (v_i, w_i)$ are the non-labor income and potential wage of the individual (assumed to be observed for this example). The term $\theta_1 z_i$ represents a potential stigma effect of receiving benefits, whereby benefits may be valued differently than equivalent sources of income. The decision of whether to work, indicated by y_i, is given by

$$y_i = \mathbb{1}[\theta_2 + \theta_3(u_{i1} - u_{i0}) > \epsilon_i]$$

where ϵ_i is a random utility shock. Variation in the treatment z_i is needed to identify the stigma effect θ_1. Under a distributional assumption on ϵ_i, the welfare and employment effects of counterfactual benefit levels can be estimated. While this simple model is static, one could alternatively use a dynamic model of labor supply (e.g. Card and Hyslop (2005)) to estimate the effects of benefits programs and maximize social returns over longer horizons. ◊

2.3 Policies

After the experiment is conducted, the experimenter chooses a policy which will be applied to the population. We assume the experimental sample is drawn from the same population to which the policy will be applied.

In contrast to the experimental design, the policy assigns the treatment as a deterministic function of covariates. That is, policies govern the value of z_i assigned to the population as a deterministic function of covariates through the function

$$z_i = z(x_i; \boldsymbol{\pi})$$

for a known function z parameterized by $\boldsymbol{\pi} \in \mathbb{R}^k$. This policy will assign treatment to units outside the experiment.

Just as the experimental design is subject to constraints, the policy may also be subject to constraints, denoted by

$$g(\boldsymbol{\pi}) \leq 0.$$

While it is natural that some constraints such as nonnegativity of the treatment are shared between the design and the policy, the policy may also be subject to constraints that are not present in the experimental design, or vice versa. Also, the budget constraint may differ between the experiment and the policy.

Example 2.8 (Targeted treatment)**:** Many interventions are tested with the goal of choosing how to allocate the invervention on the basis of covariates. In this case, the policy will be a deterministic assignment of the same treatment that was tested in the experiment. Following the examples of Section 2.1, the policy may take the form

$$z_i = \mathbb{1}[x_i'\boldsymbol{\pi} > 0]$$

for binary treatments and

$$z_i = x_i'\boldsymbol{\pi}.$$

for continuous treatments. ◊

Example 2.9 (Incentive schemes)**:** Some policies allocate a treatment based on outcomes rather than (or in addition to) covariates. For example, Duflo, Hanna, and Ryan (2012) experimentally evaluate the effect of financial incentives and monitoring on teacher attendance. Using a dynamic discrete choice model, they estimate the effects of counterfactual incentive schemes. They consider policies which give a bonus of π_1 rupees per day to teachers who attend school, for every day above a threshold π_2. These policies take the form

$$z_i = \pi_1 \times \max(y_i - \pi_2, 0)$$

where y_i is the number of days attended in a given month. ◊

Remark 2.10 (Mapping experimental variation to policy counterfactuals)**:** We can accommodate settings where the policy manipulates a different variable than the experiment. Specifically, suppose $z_i = (z_{i1}, z_{i2})$ and the experiment only manipulates z_{i1} while the policy only manipulates z_{i2}. So long as variation in z_{i1} identifies the effect of z_{i2}, the experimenter can use the experiment to choose the best policy. For example, Todd and Wolpin (2006) uses variation in wages to estimate the effect of a conditional cash transfer. Our method allows the experimenter to determine the best way to experiment with z_{i1} to learn about and choose the best policy for z_{i2}. ◊

2.4 Welfare

The experimenter's ultimate objective is to maximize the expected welfare resulting from the policy chosen at the end of the experiment. The experimenter has some welfare function

$$W(\boldsymbol{\theta}, \boldsymbol{\pi})$$

which gives the welfare of choosing the policy $\boldsymbol{\pi}$ when the true parameter is $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is unknown, the experimenter aggregates over possible values of $\boldsymbol{\theta}$ using a prior distribution $q(\boldsymbol{\theta})$ to obtain an objective function. This means the experimenter seeks to maximize

$$\mathbb{E}_{\boldsymbol{\delta}}[W(\boldsymbol{\theta}, \boldsymbol{\pi})]$$

where the expectation is taken with respect to the prior distribution $q(\boldsymbol{\theta})$ as well as the data-generating process for the pilot data and the main wave of the experiment. The decision maker's actions consist of the design $\boldsymbol{\delta}$, which can depend on the pilot data, and the policy $\boldsymbol{\pi}$, which can depend on both the pilot and main data.

Remark 2.11 (Identification and the role of the prior)**:** The experimenter observes the pilot data before choosing either the design or the policy. If $\boldsymbol{\theta}$ is identified from the pilot data, the influence of the prior on the posterior conditional on the pilot data vanishes as the pilot sample grows. Motivated by this asymptotic result, our proposed method ignores the prior when designing the experiment and choosing the policy. Instead, the design of the experiment and the chosen policy depend only on data observed in either the pilot or the main wave of the experiment. Our focus on Bayes welfare is motivated by analytical tractability rather than a reliance on prior information. When pilot data does not identify $\boldsymbol{\theta}$ or the pilot sample is small, the role of the prior does not vanish. In Section 6, we discuss a possible extension of our method to these settings. In these extensions, an informative prior will play a role. ♢

Remark 2.12 (Maximin welfare and minimax regret)**:** Other commonly used decision criteria include maximin welfare and minimax regret, where regret is the difference between the welfare of the first-best policy when $\boldsymbol{\theta}$ is known and the welfare of the chosen policy. See Manski 2021 for an overview of these approaches. ♢

In many treatment choice problems, the welfare function is the expectation of the outcome y_i when the true parameter is $\boldsymbol{\theta}$ and the policy $\boldsymbol{\pi}$ is chosen, although this need not be the case. By considering welfare functions that depend on $\boldsymbol{\theta}$ in more general ways, we allow the experimenter to target objectives such as agent welfare, consumer surplus, or counterfactual outcomes. As a special case, we can use our framework when the objective is to precisely estimate a particular parameter

or counterfactual.

Example 2.13 (Average outcome)**:** In the case that a direct measure of welfare is observed, such as mortality, revenue, or criminal recidivism, welfare is given by

$$W(\boldsymbol{\theta}, \boldsymbol{\pi}) = \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\pi}}[y_i].$$

where $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\pi}}$ denotes the expectation when the parameter is $\boldsymbol{\theta}$ and the policy is governed by $\boldsymbol{\pi}$. $\lozenge$

Example 2.14 (Consumer surplus)**:** Consider the discrete choice setting of Example 2.6. If the experimenter is interested in maximizing consumer surplus, the welfare function is

$$W(\boldsymbol{\theta}, \boldsymbol{\pi}) = \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\pi}} \log \sum_j \exp\left(x_{ij}'\theta_1 + z_{ij}'\theta_2\right)$$

up to a constant. $\lozenge$

Example 2.15 (Parameter or counterfactual estimation)**:** Suppose the experimenter wants to predict the effect of a counterfactual policy but cannot directly experiment with that policy. However, combining the experimental variation $p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})$ with a model of the outcome $p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta})$ allows the experimenter to identify the effect of the counterfactual policy. This counterfactual can be expressed as a function of the underlying parameters, so that some $\gamma(\boldsymbol{\theta})$ is the object of interest. The goal of the experimenter is to form an estimate of this effect, leading to the welfare function

$$W(\boldsymbol{\theta}, \boldsymbol{\pi}) = (\boldsymbol{\pi} - \gamma(\boldsymbol{\theta}))^2.$$

Using this welfare function is equivalent to minimizing the posterior variance on $\gamma(\boldsymbol{\theta})$. $\lozenge$

2.5 Decision problem

We now combine the previous components to formally state the decision problem the experimenter wants to solve. This is a two-period dynamic decision problem and in principle can be solved by backwards induction; however, this is practically impossible.

The experimenter seeks to solve

$$\max_{\boldsymbol{\delta},\boldsymbol{\pi}} \quad \mathbb{E}_{\boldsymbol{\delta}}[W(\boldsymbol{\theta},\boldsymbol{\pi})] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0, \quad g(\boldsymbol{\pi}) \leq 0 \tag{1a}$$

where the expectation is taken over the posterior distribution of $\boldsymbol{\theta}$ as well as over the data generated by both the pilot and the main wave of the experiment. Specifically, the data-generating process is described by

$$\begin{aligned} &(y_i, z_i) \sim p(y_i, z_i \mid x_i; \boldsymbol{\theta}, \boldsymbol{\delta}_0) \quad i = 1, \ldots, n_0 \\ &(y_i, z_i) \sim p(y_i, z_i \mid x_i; \boldsymbol{\theta}, \boldsymbol{\delta}) \quad i = n_0 + 1, \ldots, n \\ &\boldsymbol{\theta} \sim q(\boldsymbol{\theta}) \end{aligned} \tag{1b}$$

where $\boldsymbol{\delta}_0$ describes the (fixed) design of the pilot sample, $\boldsymbol{\delta}$ is a function of the pilot data, and $\boldsymbol{\pi}$ is a function of both the pilot and main data.

The problem (1a)-(1b) is a two-period dynamic decision problem, which in principle can be solved by backwards induction. Since $\boldsymbol{\pi}$ can depend on the data, the optimal policy maximizes the posterior expected welfare conditional on the full dataset. The value function characterizing the optimal policy as a function of the data is

$$V_n(\{y_i, z_i\}_{i=1}^n) = \max_{\boldsymbol{\pi}} \quad \mathbb{E}[W(\boldsymbol{\theta},\boldsymbol{\pi}) \mid \{y_i, z_i\}_{i=1}^n] \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0 \tag{2a}$$

Likewise, the optimal design maximizes the posterior expected value conditional on the pilot data only. This means the optimal $\boldsymbol{\delta}$ solves

$$\max_{\boldsymbol{\delta}} \quad \mathbb{E}[V_n(\{y_i, z_i\}_{i=1}^n) \mid \{y_i, z_i\}_{i=1}^{n_0}] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0 \tag{2b}$$

and the law of motion for the state is the data-generating process in (1b).

One could imagine trying to solve (2a)-(2b) by standard dynamic programming methods. However, since the dependence of $\boldsymbol{\pi}$ on previously observed data is unrestricted, the state space of V_n is the set of possible datasets that the main wave of the experiment could generate. This is

extremely high-dimensional even for moderate-sized experiments and low-dimensional variables[1]. For example, if $n_1 = 1000$ and y and z are one-dimensional then the state space of V_n is $\mathbb{R}^{2000}$.

This problem is a general feature of dynamic experiments, and features prominently in multi-armed bandit literature. More broadly, it is a feature of statistical decision problems where a decision maker must specify a decision rule which may depend on realizations of the data (Manski 2021). For this reason, much of the literature on adaptive experimentation has proposed algorithms which are approximate solutions to (2a)-(2b) (Lattimore and Szepesvári 2020).

We propose a general-purpose solution method for this problem motivated by asymptotic approximations. In contrast to much of the existing literature on adaptive experimentation, our method is not specific to a particular objective or model, such as the multi-armed bandit setting, but accomodates general constrained nonlinear decision problems of the form (1a)-(1b). Moreover, our method provides guarantees not just on the rate of convergence but on the asymptotic optimality of the method among rate-optimal experimental designs and policies. That is, no experimental design can achieve higher expected welfare asymptotically than the one we propose. This method is the subject of the next section.

Remark 2.16 (Parameter estimation)**:** When the experimenter's objective is to estimate a parameter, the value function simplifies. This is a commonly studied problem in the literature, especially with respect to estimating the average treatment effect in semiparametric models (Hahn, Hirano, and Karlan 2011, Bai 2022, Tabord-Meehan 2023, Cytrynbaum 2024, Bai et al. 2024). Suppose the experimenter's objective is to estimate the linear function of parameters $a'\boldsymbol{\theta}$. Then welfare function is

$$W(\boldsymbol{\theta}, \boldsymbol{\pi}) = -(a'\boldsymbol{\theta} - \boldsymbol{\pi})^2$$

and the optimal "policy" is the estimator

$$\boldsymbol{\pi} = \mathbb{E}[a'\boldsymbol{\theta} \mid \{y_i, z_i\}_{i=1}^n].$$

[1] Alternatively, one could use the posterior on $\boldsymbol{\theta}$ as the state, but if the likelihood and prior are not conjugate the state space will be the set of probability distributions on $\mathbb{R}^\ell$, which is infinite-dimensional. For ℓ small enough, it may be feasible to use approximation methods to represent the posterior.

As a result, in this case the value function (2a)-(2b) has the closed-form expression

$$V_n(\{y_i, z_i, x_i\}_{i=1}^n) = -\text{Var}(a'\boldsymbol{\theta} \mid \{y_i, z_i\}_{i=1}^n).$$

Hence, the optimal experimental design for parameter estimation problems focuses on minimizing the posterior variance, or minimizing the semiparametric efficiency bound for the parameter of interest in semiparametric models under a local asymptotic minimaxity criterion.

A similar situation arises when the welfare function is approximately quadratic and there are no constraints on the policy. Then the optimal policy is the solution to a first-order condition and is approximately equal to a linear function of parameters. Asymptotically optimal policies in such settings are overviewed in Hirano and Porter (2023). The present paper extends such results to optimize over experimental designs as well as policies.

For constrained decision problems, which are the focus of the present paper, the optimal policy is not a smooth function of the parameters and there is no such closed-form solution to the value function. This means that the value function must be approached by either brute-force solution, which is infeasible, or an approximate solution method. The multi-armed bandit literature provides a number of approximate solution methods for various objectives and models. Adusumilli (2024) provides an approximation for the multi-armed bandit problem with many waves which is motivated by asymptotic theory and hence is asymptotically optimal. The present paper is similar in spirit but instead focuses on constrained, nonlinear policy choice problems with only one adaptive wave. Like Adusumilli (2024), our approximation is motivated by asymptotic theory and is likewise asymptotically optimal for the class of problems we consider. ◊

2.6 Applications

The framework described above is general and can be used in a variety of situations. Here we present some examples to illustrate the breadth of applications.

Progresa Application: In the Progresa experiment, children in rural Mexico were randomly assigned to be eligible to receive a cash transfer conditional on attending school. The size of the transfer depended on the child's current grade level and gender. Here z_i is a vector describing the amount of the transfer offered to child i at every grade level, x_i is a vector of covariates including

the child's current grade level and gender, and y_i is an indicator for whether the child attended school.

We consider experimental designs and policies similar to Example 2.4 of the form

$$z_i = x_i'\delta_1 \times \mathbb{1}[\nu_i \leq \delta_2]$$

where ν_i is a uniformly distributed randomization device, and policies of the form

$$z_i = x_i'\boldsymbol{\pi}.$$

We use a dynamic school choice model for children's schooling decisions. Similar models have been used to the analyze the Progresa experiment by Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2012). The model relates observed schooling decisions y_i to long-run educational attainment, which is only observed for older children in the sample. by capturing the dynamic effect of offering future subsidies in later school grades on children's schooling decisions in earlier grades. This dynamic effect is found to be a key mechanism through which the effectiveness of the subsidy operates by both Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2012). One possible objective, considered in Section 5, is the goal of designing a subsidy schedule to maximize the eventual educational attainment of children and reduce gender disparities in graduation rates. Other objectives, like the subjective welfare of households, could also be considered.

The experimenter faces a budget constraint for both the experiment and the policy, although the budgets for each stage of the decision problem may differ. For a given experimental budget, the experimenter can offer large subsidies to a small treatment group or small subsidies to a large treatment group. Within the treatment group, giving higher subsidies to girls results in fewer resources available to give to boys, and likewise for earlier/later grade levels. Additionally, the subsidy must be nonnegative. ◊

Consumer Search Application: Expedia, an online platform for booking hotels, ran an experiment in 2012-2013 in which they varied the order in which hotels were displayed to consumers. The treatment z_i is an ordered list of J hotels and the observed covariates x_i contains attributes of the consumer and the hotels associated with a particular query. The outcome y_i is multidimensional

and for each hotel in the query indicates whether the consumer clicked on the hotel and whether they booked the hotel.

Although the treatment z_i is extremely high dimensional, with possible values being the set of permutations of subsets of hotels satisfying the query, we can parametrize distributions over z_i using the ranked logit policy class from Example 2.5. The experimenter chooses the vector $\boldsymbol{\delta}$ which determines the importance of product characteristics, such as ratings or price, in determining the order of the search results.

We can use a model of consumer search with recall as in Weitzman (1979), Ursu (2018), and Compiani et al. (2023) to model $p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta})$. These models impose structure on the relationship between rankings and consumer behavior, allowing the experimenter to evaluate counterfactual policies without having to estimate the effect of every possible ranking, of which there are $J!$. The model also allows the experimenter to evaluate consumer surplus.

A natural objective in this setting is revenue maximization. However, as discussed in Compiani et al. (2023), the experimenter may also be interested in maximizing consumer surplus in order to maximize the user base of the platform. Constraints on the experiment and policy may include a budget constraint, an incentive compatibility or participation constraint for hotels (i.e. a guarantee of a minimum level of visibility), or a restriction on how attributes of the consumer can affect the ranking. ◊

3 Solution method

Since the finite-sample problem is intractable, we propose an approximation to this problem that is low-dimensional and straightforward to solve. This approximation has two components: first, replacing the high-dimensional data with a low-dimensional Gaussian estimate, and second, replacing the nonlinear welfare with a quadratic approximation.

This section presents the method as it is implemented in practice. Formal justification of the approximations and the asymptotic optimality of the method are given in Section 4.

3.1 Gaussian approximation

The first step of the approximation is to assume that rather than observing the full pilot data and main data, the experimenter only observes a Gaussian estimate from each wave. The question of how to design the experiment is then reduced to the question of choosing the variance-covariance matrix of this estimate. Whereas the state of the finite-sample problem is the full data, the state of the Gaussian approximation to the problem is the posterior mean and variance on the model parameters.

After the pilot experiment, suppose the experimenter only observes

$$\hat{\boldsymbol{\theta}}_0 \sim N\left(\boldsymbol{\theta}, \frac{1}{n_0} J_0^{-1}\right)$$

where J_0 is the Fisher information matrix for the pilot experiment. Note that this is the limiting distribution of the maximum likelihood estimator, so we can interpret $\hat{\boldsymbol{\theta}}_0$ as the maximum likelihood estimator for the pilot data. Then, after the main wave of the experiment, the experimenter observes

$$\hat{\boldsymbol{\theta}}_1 \sim N\left(\boldsymbol{\theta}, \frac{1}{n_1} J(\boldsymbol{\delta})^{-1}\right)$$

where $J(\boldsymbol{\delta})$ is the Fisher information matrix when the design $\boldsymbol{\delta}$ is chosen.

By varying the experimental design $\boldsymbol{\delta}$, the experimenter varies the variance-covariance matrix of the resulting estimate $\hat{\boldsymbol{\theta}}_1$. The following example illustrates how the different designs affect the variance-covariance matrix of estimates of group means.

Example 3.1 (Group means)**:** Suppose y_i, z_i and x_i are all binary. Let $\theta_{z,x}$ be the probability that $y_i = 1$ for units with $z_i = z$ and $x_i = x$ so that

$$p_{y|z,x}(y_i = 1 \mid z_i, x_i; \boldsymbol{\theta}) = \sum_{z,x \in \{0,1\}} \theta_{z,x} \mathbb{1}[z_i = z, x_i = x].$$

Suppose treatment is assigned by

$$p_{z|x}(z_i = 1 \mid x_i; \boldsymbol{\delta}) = \delta_0(1 - x_i) + \delta_1 x_i,$$

where δ_0 and δ_1 are the treatment assignment probabilities for units with $x_i = 0$ and $x_i = 1$. Also suppose that the probability that $x_i = 1$ is $\frac{1}{2}$. The Fisher information matrix for this model is given by

$$J(\boldsymbol{\delta}) = \frac{1}{2}\begin{pmatrix} \frac{1-\delta_0}{\theta_{0,0}(1-\theta_{0,0})} & 0 & 0 & 0 \\ 0 & \frac{\delta_0}{\theta_{1,0}(1-\theta_{1,0})} & 0 & 0 \\ 0 & 0 & \frac{1-\delta_1}{\theta_{0,1}(1-\theta_{0,1})} & 0 \\ 0 & 0 & 0 & \frac{\delta_1}{\theta_{1,1}(1-\theta_{1,1})} \end{pmatrix}$$

This will be singular whenever $\delta_0 \in \{0, 1\}$ or $\delta_1 \in \{0, 1\}$, corresponding to a treatment assignment that does not randomize for the subpopulations characterized by $x_i = 0$ or $x_i = 1$. $\Diamond$

The variance-covariance matrix $J(\boldsymbol{\delta})$ affects the welfare of the policy chosen at the end of the experiment by determining the information available to the experimenter when choosing the policy. This is captured by the posterior distribution of the parameter of interest conditional on the Gaussian estimates $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$. So long as J_0 is nonsingular, the prior is dominated by the pilot data as n_0 grows large and the posterior after observing $\hat{\boldsymbol{\theta}}_0$ is approximately Gaussian with mean and variance given by

$$\mu_0 = \hat{\boldsymbol{\theta}}_0 \qquad\qquad \Sigma_0 = \frac{1}{n_0} J_0^{-1}.$$

We take this as the starting point for the main wave of the experiment, and in doing so ensure our method is insensitive to the prior q. Then, after picking the design $\boldsymbol{\delta}$ and observing $\hat{\boldsymbol{\theta}}_1$, the posterior is also Gaussian and is characterized by the Bayesian updating formula

$$\begin{aligned} \mu_1 &= \left(\Sigma_0^{-1} + n_1 J(\boldsymbol{\delta})\right)^{-1} \left(\Sigma_0^{-1}\mu_0 + n_1 J(\boldsymbol{\delta})\hat{\boldsymbol{\theta}}_1\right) \\ \Sigma_1 &= \left(\Sigma_0^{-1} + n_1 J(\boldsymbol{\delta})\right)^{-1}. \end{aligned} \tag{3}$$

Designs which are more informative about particular parameters will lead to posteriors which are more informative about those parameters. Whether a particular $\boldsymbol{\delta}$ is optimal in terms of its information depends on the objective and constraints of the experimenter. For a given policy choice problem, some parameters may be more important than others and it may be desirable to estimate

them more precisely.

We now present the value function under the Gaussian approximation. In the second period of the decision problem, the experimenter chooses a policy $\boldsymbol{\pi}$ to maximize posterior expected welfare, leading to the value function

$$V_n^G(\mu_1, \Sigma_1) = \max_{\boldsymbol{\pi}} \quad \mathbb{E}\left[W(\boldsymbol{\theta}, \boldsymbol{\pi}) \mid \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_1\right] \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0 \tag{4a}$$

where $\boldsymbol{\theta} \sim N(\mu_1, \Sigma_1)$ given $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$. To solve for the optimal design, the experimenter must first compute the value function V_n^G for each possible state (μ_1, Σ_1). Then, the experimenter picks the design which maximizes the expected value of V_n^G. That is, in the first period the optimal design solves

$$\max_{\boldsymbol{\delta}} \quad \mathbb{E}\left[V_n^G(\mu_1, \Sigma_1) \mid \hat{\boldsymbol{\theta}}_0\right] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0 \tag{4b}$$

where the law of motion for (μ_1, Σ_1) is given by the Bayesian updating formula (3). The expectation is taken over the distribution of $\hat{\boldsymbol{\theta}}_1$, which is governed by the choice of $\boldsymbol{\delta}$ through $J(\boldsymbol{\delta})$, as well as over $\boldsymbol{\theta} \sim N(\mu_0, \Sigma_0)$ given $\hat{\boldsymbol{\theta}}_0$. We use the G superscript to reflect the Gaussian approximation, and the n subscript because the variances of $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$ depend on the sample size.

Restricting attention to this Gaussian estimate reduces the state space of the dynamic program significantly. In fact, the state space no longer depends on the sample size. Rather, it depends on the dimension of the parameter $\boldsymbol{\theta}$. To solve V_n^G, it is necessary to keep track of the posterior mean μ_1 and the lower triangular part of the posterior covariance matrix Σ_1. For $\boldsymbol{\theta} \in \mathbb{R}^\ell$, this state is of dimension $\ell + \ell(\ell+1)/2$. This can still be large. For example, in the Progresa application discussed in Section 5, $\ell = 15$ and the state space is $\mathbb{R}^{135}$. However, a further simplification is achieved by using a quadratic approximation to the welfare function. We discuss this next.

Remark 3.2 (Singular Fisher information)**:** It is possible that the Fisher information matrix $J(\boldsymbol{\delta})$ is singular for some choices of $\boldsymbol{\delta}$. In fact, this may be the case for the optimal design. To see this, consider the setting of Example 3.1. Suppose that the pilot data gives a precise estimate of the control outcomes $\theta_{0,0}$ and $\theta_{0,1}$, but very imprecise estimates of the treatment outcomes $\theta_{1,0}$ and $\theta_{1,1}$. Then the optimal design may set $\delta_1 = \delta_2 = 1$, so that the entire population is treated in the

main wave. This leads to a singular $J(\boldsymbol{\delta})$ in the main wave, but by combining the pilot data and the main data, the experimenter obtains a good estimate of both treatment and control outcomes to inform the policy choice.

When $J(\boldsymbol{\delta})$ is singular, the MLE is not well-defined. To account for this possibility, our results in Section 4 will rely on the sample average of the scores. When $J(\boldsymbol{\delta})$ is nonsingular, observing the sample average of the scores is equivalent to observing the MLE. Otherwise, using the sample average of the scores allows us to generalize V_n^Q to remain valid in the case of singular $J(\boldsymbol{\delta})$. The law of motion for (μ_1, Σ_1) when $J(\boldsymbol{\delta})$ is singular is given by

$$\mu_1 = \left(\Sigma_0^{-1} + n_1 J(\boldsymbol{\delta})\right)^{-1} \left(\Sigma_0^{-1}\mu_0 + A_1\right)$$
$$\Sigma_1 = \left(\Sigma_0^{-1} + n_1 J(\boldsymbol{\delta})\right)^{-1}$$

where $A_1 \sim N\left(J(\boldsymbol{\delta})\boldsymbol{\theta}, \frac{1}{n_1} J(\boldsymbol{\delta})\right)$. When $J(\boldsymbol{\delta})$ is nonsingular, $A_1 = J(\boldsymbol{\delta})\hat{\boldsymbol{\theta}}_1$ and this law of motion is exactly the same as the standard Bayesian updating formula (3). ◊

3.2 Quadratic approximation

The second simplification is to replace the nonlinear welfare function with a quadratic approximation. Under this approximation, the optimal policy depends on the unknown parameter only through a low-dimensional linear function of parameters interpreted as the marginal effect of changing the policy. This reduces the dimension of the state from the posterior mean and variance on the entire parameter to only the posterior mean on this marginal effect.

Suppose that the true parameter $\boldsymbol{\theta}$ is in a neighborhood of some $\boldsymbol{\theta}_0$. This neighborhood will be formalized in Section 4. Let the optimal policy under $\boldsymbol{\theta}_0$ be

$$\boldsymbol{\pi}_0 = \arg\max_{\boldsymbol{\pi}} \quad W(\boldsymbol{\theta}_0, \boldsymbol{\pi}) \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0. \tag{5}$$

We construct a quadratic approximation to the welfare function around $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$. Let

$$C = \nabla_{\boldsymbol{\pi}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$$
$$D = \nabla^2_{\boldsymbol{\pi}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$$

$$H = \nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$$

be the marginal effect of the policy on welfare, the cross-partial derivative of welfare, and the second derivative of welfare with respect to $\boldsymbol{\pi}$, evaluated at $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$. Since the choice of policy is subject to constraints, it will not generally be true that $C = 0$. Using these quantities, we define the quadratic welfare function

$$W^Q(\boldsymbol{\theta}, \boldsymbol{\pi}) = (\boldsymbol{\pi} - \boldsymbol{\pi}_0)'[C + D(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] + \frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\pi}_0)' H (\boldsymbol{\pi} - \boldsymbol{\pi}_0).$$

This approximation is similar to a second-order Taylor expansion of welfare around $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, but ignores terms which do not involve $\boldsymbol{\pi}$, since these terms will not affect the choice of policy $\boldsymbol{\pi}$.

The key property of W^Q is that $\boldsymbol{\theta}$ enters the decision problem only through the linear function $D\boldsymbol{\theta}$. This quantity has the intuitive interpretation of capturing the marginal effect on welfare of changing the policy from $\boldsymbol{\pi}_0$ to some other policy $\boldsymbol{\pi}$. To see this, recall that in the main wave of the experiment the experimenter will observe some Gaussian estimate

$$\hat{\boldsymbol{\theta}}_1 \sim N\left(\boldsymbol{\theta}, \frac{1}{n_1} J(\boldsymbol{\delta})^{-1}\right).$$

By the Delta method, when $\boldsymbol{\theta}$ is close to $\boldsymbol{\theta}_0$ and n_1 is large, the maximum likelihood estimate of the marginal effect of the policy $\nabla_{\boldsymbol{\pi}} W(\boldsymbol{\theta}, \boldsymbol{\pi}_0)$ is approximately distributed as

$$\sqrt{n_1}(\nabla_{\boldsymbol{\pi}} W(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\pi}_0) - \nabla_{\boldsymbol{\pi}} W(\boldsymbol{\theta}, \boldsymbol{\pi}_0)) \sim N(0, DJ(\boldsymbol{\delta})^{-1} D').$$

which is the same distribution as $\sqrt{n_1}(D\hat{\boldsymbol{\theta}}_1 - D\boldsymbol{\theta})$. Thus, estimates of $\boldsymbol{\theta}$ affect the choice of policy only by providing estimates of (a linear approximation to) the marginal effect of the policy on welfare.

The intuition behind this quadratic approximation is that when choosing a policy, estimates of the parameters of the model are only relevant insofar as they are informative about the effects of counterfactual policies. For policies close to $\boldsymbol{\pi}_0$, the marginal effect of changing the policy characterizes these counterfactuals. Thus, when choosing a policy, the experimenter need only consider the posterior on this marginal effect. In the Progresa example, this marginal effect is the

increase in graduation rates resulting from giving an extra peso to primary school children versus giving an extra peso to secondary school children, or to girls versus boys.

This intuitive interpretation has important practical implications as well. Since W^Q is linear in $D\boldsymbol{\theta}$ we can write expected welfare under the quadratic approximation as a function only of the posterior mean on $D\boldsymbol{\theta}$:

$$\mathbb{E}\left[W^Q(D\boldsymbol{\theta}, \boldsymbol{\pi}) \mid \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_1\right] = W^Q(D\mu_1, \boldsymbol{\pi})$$

and therefore $D\mu_1$ is a sufficient state variable for the policy choice problem in the second period. That is, if the experimenter knows $D\mu_1$, then no other data is needed to choose the policy which maximizes posterior expected welfare. Since $D\boldsymbol{\theta}$ characterizes the marginal effect of the policy $\boldsymbol{\pi}$, $D\mu_1$ is of the same dimension as $\boldsymbol{\pi}$. For many policy classes used in the literature (see Kitagawa and Tetenov (2018) or Athey and Wager (2021) for examples), $\boldsymbol{\pi}$ is low-dimensional, perhaps much lower-dimensional than $\boldsymbol{\theta}$.

We now present the value function under the additional quadratic approximation. In the second period, the value function is defined by choosing the policy to maximize posterior expected welfare, leading to the value function

$$V_n^Q(D\mu_1) = \max_{\boldsymbol{\pi}} \quad W^Q(D\mu_1, \boldsymbol{\pi}) \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0. \tag{6a}$$

To solve for the optimal design, the experimenter must first compute the value function V_n^Q for each possible state $D\mu_1$. Then, the experimenter picks the design which maximizes the expected value, integrating over possible values of $D\mu_1$. That is, in the first period the optimal design solves

$$\max_{\boldsymbol{\delta}} \quad \mathbb{E}\left[V_n^Q(D\mu_1) \mid \hat{\boldsymbol{\theta}}_0\right] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0 \tag{6b}$$

where the law of motion for (μ_1, Σ_1) is given by the Bayesian updating formula (3). As before, the expectation is taken over both the distribution of $\hat{\boldsymbol{\theta}}_1$, which is governed by the choice of $\boldsymbol{\delta}$ through $J(\boldsymbol{\delta})$, as well as $\boldsymbol{\theta} \sim N(\mu_0, \Sigma_0)$ given $\hat{\boldsymbol{\theta}}_0$.

This value function is extremely tractable. Recall that the finite-sample value function V_n has a state space with potentially thousands of dimensions, increasing with the sample size. Using only

the Gaussian approximation, we obtained a value function V_n^G with a state space of dimension $\ell+\ell(\ell+1)/2$. Using both the quadratic approximation and the Gaussian approximation, we obtain a value function V_n^Q with a state space of dimension k, where k is the dimension of the policy $\boldsymbol{\pi}$. In the Progresa application discussed in Section 5, the policy $\boldsymbol{\pi}$ is six-dimensional. This means that the quadratic approximation reduces the state space of the decision problem from $\mathbb{R}^{135}$ to $\mathbb{R}^6$. The difficulty of solving the experimental design-policy choice problem is therefore dependent on the complexity of the policy $\boldsymbol{\pi}$, rather than the complexity of the model described by $\boldsymbol{\theta}$.

In practice, $\boldsymbol{\theta}_0$ and $\boldsymbol{\pi}_0$ are unknown, making solving V_n^Q infeasible. Instead, we use a consistent estimate of these quantities from the pilot data. Following the steps above, we solve for the policy $\hat{\boldsymbol{\pi}}_0$ which is optimal for $\hat{\boldsymbol{\theta}}_0$ and construct $\hat{C}, \hat{D}, \hat{H}$ by using $(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\pi}}_0)$ in place of $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$. We can then define $\hat{W}^Q$ and $\hat{V}_n^Q$ analogously to W^Q and V_n^Q, and solve for an estimate of the optimal design by maximizing the expected value of $\hat{V}_n^Q$.

In the next section, we justify both the Gaussian and quadratic approximations, as well as the use of pilot data to estimate $\hat{V}_n^Q$. We show that the error of these approximations is asymptotically negligible as n_0 and n_1 grow, and solving $\hat{V}_n^Q$ yields an asymptotically optimal experiment for policy choice.

Remark 3.3 (Computation with linear constraints)**:** Solving $\hat{V}_n^Q$ is especially tractable when the constraints $g(\boldsymbol{\pi}) \leq 0$ are linear, as is the case in the Progresa application in Section 5. When constraints are linear, $\hat{V}_n^Q(D\mu_1)$ is a quadratic program for any value of $D\mu_1$. To solve for $\hat{V}_n^Q$, we solve this quadratic program for many possible values of $D\mu_1$ and interpolate the corresponding values with a neural network. This is very fast when $D\mu_1$ is low-dimensional, such as the six-dimensional policy in the Progresa application. We discuss the specifics of our algorithm in Appendix F. $\lozenge$

Remark 3.4 (Approximating nonlinear constraints)**:** When the constraints are nonlinear, constructing $\hat{V}_n^Q$ can be performed by solving a nonlinear program for each value of $D\mu_1$, where the objective is quadratic and the constraints are nonlinear. Since this is more computationally intensive than the quadratic program we obtain with linear constraints, we provide a method for approximating the nonlinear constraints with linear constraints in Appendix C. Since V_n^Q may be only twice directionally Hadamard differentiable at $\boldsymbol{\theta}_0$ (Shapiro 1985, Shapiro 1991), "naive" estimates of V_n^Q may not be consistent (Fang and Santos 2018). To obtain a consistent estimate of

V_n^Q, we need a consistent estimate of the active set of constraints at $\boldsymbol{\theta}_0$. The method we propose in Appendix C achieves this at the expense of introducing an additional tuning parameter. $\lozenge$

4 Limit experiment and asymptotic optimality

The solution method presented in the previous method is asymptotically optimal. We establish this by showing that both the Gaussian and quadratic approximations have negligible error as the sizes of the pilot the main experiment grow. Together, these results imply that solving the approximated value function leads to experiments and policies which are asymptotically optimal.

4.1 Local asymptotic framework

The Gaussian and quadratic approximations are justified in an asympotic decision environment called the limit experiment. The limit experiment describes the possible asymptotic behaviors of any design and policy in the finite-sample environment. The benefit of working in the limit experiment is that it is much simpler than the finite-sample problem, allowing us to characterize the best possible design and policy in the limit.

The limit experiment is constructed in a local asymptotic framework. Local asymptotics are commonly used to study optimality of estimators, tests, and the consequences of model misspecification, among other econometric and statistical problems (see Le Cam (1972) and Van der Vaart (2000) for classic results on estimation and testing and Staiger and Stock (1994) for a seminal analysis of weak instruments in a local asymptotic framework). In local asymptotics, we analyze the performance of our method under parameter values that are difficult to distingish from each other even in large samples. In particular, we model

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h/\sqrt{n}$$

for some "local parameter" h and reference value of the parameter $\boldsymbol{\theta}_0$. The $1/\sqrt{n}$ scaling means that the difference between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ is of the same order of magnitude as the standard error of typical estimators of $\boldsymbol{\theta}$. Likewise, we will consider local alternatives to the policy $\boldsymbol{\pi}$, where

$$\boldsymbol{\pi} = \boldsymbol{\pi}_0 + c/\sqrt{n}$$

so that the experimenter considers policies in a neighborhood of the reference policy $\boldsymbol{\pi}_0$ of the same order of magnitude as sampling uncertainty.

Like all asymptotics, local asymptotics are not meant to reflect a data-generating process that is literally changing as the sample size grows, but rather is intended to capture the finite-sample property that $\boldsymbol{\theta}$ is difficult to distinguish from $\boldsymbol{\theta}_0$ at the current sample size in two senses which are material for our approximation. First, $\boldsymbol{\theta}_0$ is the value at which the Fisher information matrix is evaluated. Therefore, for local asymptotics to be accurate we need the covariance matrix of $\hat{\boldsymbol{\theta}}_1$ to be well-approximated by $J(\boldsymbol{\delta})$, the Fisher information at $\boldsymbol{\theta}_0$. Second, $\boldsymbol{\theta}_0$ is used to define the quadratic approximation to welfare. Therefore we need $\boldsymbol{\theta}$ to be close enough to $\boldsymbol{\theta}_0$ that the quadratic approximation is valid. Since we use the pilot estimate $\hat{\boldsymbol{\theta}}_0$ to construct $J(\boldsymbol{\delta})$ and the quadratic approximation in practice, we may gauge the strength of this assumption on the basis of whether these approximations seem reasonable for values of $\boldsymbol{\theta}$ that are difficult to distinguish from $\hat{\boldsymbol{\theta}}_0$ in the pilot data.

4.2 Justifying the Gaussian approximation

To justify the Gaussian approximation of V_n^G in (4a)-(4b), we show that any sequence of policies $\boldsymbol{\pi}_n$ in the finite-sample problems indexed by n converges in distribution to a policy that depends only on Gaussian estimates of the parameter.

Our first result characterizes the asymptotic behavior of any sequence of policies in the finite-sample experiment. A key (and standard) assumption for local asymptotic optimality results is that the model is smooth in $\boldsymbol{\theta}$ in the sense of differentiability in quadratic mean.

Assumption 4.1 (Differentiability in Quadratic Mean)**:** *There exists a function $\psi(y, z, x)$, called the score of the outcome model $p_{y|z,x}$ at $\boldsymbol{\theta}_0$, such that*

$$\begin{aligned}\sup_{z,x} \int \Bigg[& \sqrt{p_{y|z,x}(y \mid z, x; \boldsymbol{\theta} + h)} \\ & - \sqrt{p_{y|z,x}(y \mid z, x; \boldsymbol{\theta}_0)} - \frac{1}{2} h' \psi(y, z, x) \sqrt{p_{y|z,x}(y, z \mid x; \boldsymbol{\theta}_0)} \Bigg]^2 d\lambda(y) \\ & = o(\|h\|^2)\end{aligned}$$

as $h \to 0$, for some dominating measure λ.

At a high level, this means the square root of the density is differentiable in $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. This condition allows us to approximate the model $p_{y|z,x}(y \mid z, x; \boldsymbol{\theta})$ by a Gaussian model for values of $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. Typically, the score $\psi(y, z, x)$ is the gradient of the log-likelihood of the model:

$$\psi(y, z, x) = \nabla_{\boldsymbol{\theta}} \log p(y \mid z, x; \boldsymbol{\theta}_0).$$

See Van der Vaart (2000) Lemma 7.6 for sufficient conditions for this formulation. The corresponding Fisher information matrix of the model at $\boldsymbol{\theta}_0$ depends also on the design $\boldsymbol{\delta}$ and is given by

$$J(\boldsymbol{\delta}) = \mathbb{E}_{\boldsymbol{\theta}_0, \boldsymbol{\delta}}\Big[\psi(y_i, z_i, x_i)\psi(y_i, z_i, x_i)'\Big].$$

Because an experimenter may find it optimal to choose a policy $\boldsymbol{\delta}$ that induces a singular Fisher information matrix (for example, if there is no need to experiment on certain sub-populations in Example 3.1), our results use the sample average of the scores as a sufficient statistic for each wave which remains well-defined even when the maximum likelihood estimator is not. We assume that y_i and z_i are generated by a potential outcome model where $y_i = y(z_i, x_i, \epsilon_i; \boldsymbol{\theta})$ where ϵ_i is a structural error term and $z_i = z(x_i, \nu_i; \boldsymbol{\delta})$ where ν_i is a randomization device. Define the stochastic process $A_{1,n}(\cdot)$ by

$$A_{1,n}(\boldsymbol{\delta}) = \frac{1}{\sqrt{n_1}} \sum_{i=n_0+1}^{n_1} \psi\Big(y\Big(z(x_i, \nu_i; \boldsymbol{\delta}), x_i, \epsilon_i; \boldsymbol{\theta}_0\Big), z(x_i, \nu_i; \boldsymbol{\delta}), x_i\Big).$$

We will restrict our attention to policies and models for which the mapping between $\boldsymbol{\delta}$ and the sample average of the scores is well-behaved in the following sense:

Assumption 4.2 (Equicontinuity of scores)**:** *The sample average of the scores $A_{1,n}(\boldsymbol{\delta})$ is stochastically equicontinuous.*

Stochastic equicontinuity is a high-level condition that ensures that the score $A_{1,n}$ has relatively smooth sample paths for large enough n. Many lower-level sufficient conditions are available in the literature (e.g. Andrews (1994), Van der Vaart and Wellner (2013)). Assumption 4.2 is a joint

restriction on both the treatment assignment mechanism $p_z(z_i \mid x_i; \boldsymbol{\delta})$ and the outcome model $p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta})$. Both the treatment assignment and the model must be smooth enough that small changes in $\boldsymbol{\delta}$ lead to small changes in the score $A_{1,n}$. In Appendix B we give a precise definition of stochastic equicontinuity, give lower-level sufficient conditions which are natural for our setting, and verify that they are satisfied for the class of polices and model which we use in the Progresa application in Section 5. Specifically, we show that for a class of designs nesting Example 2.4, the treatment assignment mechanism has finite bracketing entropy. Combined with a smoothness condition on the score of the model which we verify, this implies stochastic equicontinuity of A_n.

Finally, we require that both the size of the pilot experiment and the size of the main experiment grow at the same rate, allowing us to apply asymptotic analysis to both waves of the experiment.

Assumption 4.3 (Large waves)**:** *The pilot sample size n_0 satisfies $0 < \lim_{n\to\infty} n_0/n < 1$.*

Our first result states that under these assumptions, any convergent design and policy in the finite-sample experiment converges to some design and policy in the limit experiment. The limit experiment is the Gaussian environment discussed in Section 3. When we say $(\boldsymbol{\delta}, c)$ is a design and policy in the limit experiment, we mean $\boldsymbol{\delta} = \boldsymbol{\delta}(A_0, U_0)$ and $c = c(A_0, A_1(\boldsymbol{\delta}), U_0, U_1)$ where $A_0 \sim N(J_0 h, J_0)$, $A_1(\cdot)$ is a Gaussian process which is the limit of $A_{1,n}(\cdot)$, and U_0 and U_1 are uniformly distributed randomization devices.

Lemma 4.4: *Suppose Assumptions 4.1, 4.2, and 4.3 hold. Let $(\boldsymbol{\delta}_n, \boldsymbol{\pi}_n)$ be a sequence of designs and policies in the finite-sample experiment, and define $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)$. Suppose $\boldsymbol{\delta}_n$ and c_n jointly converge in distribution under $\boldsymbol{\theta}_0$. Then there exists an experimental design and policy $(\boldsymbol{\delta}, c)$ in the limit experiment such that*

$$(\boldsymbol{\delta}_n, c_n) \overset{h}{\rightsquigarrow} (\boldsymbol{\delta}, c)$$

where $\overset{h}{\rightsquigarrow}$ denotes convergence in distribution along the sequence $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h/\sqrt{n}$.

Proof. See Appendix A.1. □

This result shows that any weakly convergent sequence of decisions in the finite-sample problem is asymptotically equivalent in distribution to a design and policy in the limit experiment, where

only the Gaussian random variables A_0 and A_1 are observed. This suggests that to find an asymptotically optimal policy, we find the optimal policy in the limit experiment. This is the reason that V_n^G depends only on Gaussian estimates.

Remark 4.5 (Relation to previous work)**:** Lemma 4.4 is related to previous work on asymptotic representation theorems for statistical models. For the standard case, see Le Cam (1972) and Van der Vaart (2000). For static, binary treatment choice problems with fixed data distribution, Hirano and Porter (2009) derive optimal treatment rules in a limit experiment framework.

Hirano and Porter (2023) establish an asymptotic representation theorem for batched multi-armed bandits. In Section 6 we will show that our result extends to dynamic experiments as well. The main distinction between our result and Hirano and Porter (2023) is the latter's focus on the multi-armed bandit setting. Our result builds on this work by extending their argument to more general decision problems, including settings with a continuum of treatments or targeting based on covariates. This generality comes at the expense of the additional Assumption 4.2, which is not required in the multi-armed bandit setting. Part of our contribution is verifying this high-level condition for a class of experimental designs with continuous treatments in Appendix B. Finally, Hirano and Porter (2023) do not solve for optimal policies, which motivates the quadratic approximations of the next section. ◊

For our purposes, Lemma 4.4 is helpful because it implies that the welfare of any sequence $(\boldsymbol{\delta}_n, c_n)$ in the finite-sample environment converges to the welfare of some policy in the Gaussian environment of V_n^G. We now state regularity conditions on the decision problem that ensure the Gaussian value function V_n^G is a good approximation to the finite-sample value function V_n. Before doing so, we define the regret of a policy $\boldsymbol{\pi}$, since some of our regularity conditions will be stated in terms of regret. The regret of a policy $\boldsymbol{\pi}$ is

$$R(\boldsymbol{\theta}, \boldsymbol{\pi}) = \left[\max_{\tilde{\boldsymbol{\pi}}} W(\boldsymbol{\theta}, \tilde{\boldsymbol{\pi}}) \text{ s.t. } g(\tilde{\boldsymbol{\pi}}) \leq 0\right] - W(\boldsymbol{\theta}, \boldsymbol{\pi}).$$

This is the difference between the welfare of the optimal policy and the welfare of the chosen policy $\boldsymbol{\pi}$. Because regret recenters welfare, it is does not diverge as $n \to \infty$, which is convenient for asymptotic analysis. Since the recentering term in brackets does not involve $\boldsymbol{\pi}$, policies which maximize welfare also minimize regret. We also define $(\boldsymbol{\delta}_n^*, \boldsymbol{\pi}_n^*)$ as the optimal design and policy

in the finite-sample problem V_n, and $(\boldsymbol{\delta}_n^G, \boldsymbol{\pi}_n^G)$ as the optimal design and policy in the Gaussian problem V_n^G.

Assumption 4.6: *1. Welfare is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ at $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$.*

2. The prior density $q(\boldsymbol{\theta})$ is continuous and positive at $\boldsymbol{\theta}_0$.

3. Let $\tilde{\boldsymbol{\pi}}_n^G$ be the finite-sample analog of $\boldsymbol{\pi}_n^G$, where A_0 and A_1 are replaced by $A_{0,n}$ and $A_{1,n}$. There exists a function $\bar{W}(\boldsymbol{\theta})$ such that $R(\boldsymbol{\theta}, \tilde{\boldsymbol{\pi}}_n^G) \leq n^{-1}\bar{W}(\boldsymbol{\theta})$ and $\int |\bar{W}(\boldsymbol{\theta})|^{1+\iota} dQ(\boldsymbol{\theta}) < \infty$ for some $\iota > 0$.

4. $\sqrt{n}(\boldsymbol{\pi}_n^* - \boldsymbol{\pi}_0)$ is bounded in probability for any $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h/\sqrt{n}$.

5. The set of $\boldsymbol{\delta}$ satisfying $f(\boldsymbol{\delta}) \leq 0$ is compact.

Continuity and nonnegativity of the prior ensures that the post-pilot posterior is well-approximated by the flat-prior posterior used in the Gaussian problem V_n^G. The dominance condition ensures that the regret of the feasible policy $\tilde{\boldsymbol{\pi}}_n^G$, which is greater than the regret of $\boldsymbol{\pi}_n^*$, is uniformly integrable. The n^{-1} rate is the natural rate for regret when welfare is approximately quadratic and estimates are $\sqrt{n}$-consistent. The assumption that $\sqrt{n}(\boldsymbol{\pi}_n^* - \boldsymbol{\pi}_0)$ is bounded in probability likewise ensures that the $\sqrt{n}$ scaling of the limit experiment is appropriate. Finally, the compactness of the feasible set for $\boldsymbol{\delta}$ could be replaced by an assumption that $\boldsymbol{\delta}_n^*$ is bounded in probability.

Theorem 4.7: *Suppose Assumptions 4.1, 4.2, 4.3, and 4.6 hold. Additionally assume that $\boldsymbol{\pi}_n^G$ is continuous in A_0 and A_1 and that J_0 is nonsingular. Then*

$$\mathbb{E}_{\boldsymbol{\delta}_n^*}\left[V_n(\{y_i, z_i\}_{i=1}^n)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^G}\left[V_n^G(\mu_1, \Sigma_1)\right] + o(n^{-1})$$

Proof. See Appendix A.1. □

This result justifies the Gaussian approximation of the value function V_n^G in (4a)-(4b) by showing the solution to the finite-sample problem is asymptotically equivalent to the solution to the Gaussian problem in terms of welfare. The additional assumption that $\boldsymbol{\pi}_n^G$ are continuous in A_0 and A_1 is implied by the regularity conditions we assume for the quadratic approximation below.

Remark 4.8 (Estimating policies at the parametric rate)**:** While the assumption that the optimal policy be $\sqrt{n}$-rate estimable often holds in parametric models, it can fail in nonparametric empirical welfare maximization problems as in Kitagawa and Tetenov (2018) and Athey and Wager (2021). In a semiparametric setting, Bhattacharya and Dupas (2012) discusses $\sqrt{n}$-rate estimation of the optimal policy under parametrized treatment effects and kernel smoothing of the covariate distribution, as well as the slower convergence rate in the nonparametric case. In fully parametric structural models, $\sqrt{n}$-rate estimation of optimal policies is less demanding, so long as the optimal policy is a continuous (though not necessarily differentiable) function of $\boldsymbol{\theta}$. $\lozenge$

4.3 Justifying the quadratic approximation

To justify the quadratic approximation of the value function V_n^Q in (6a)-(6b), we show that solving the policy choice problem is equivalent to solving a quadratic program in the limit experiment. This further requires that the welfare and constraints are smooth, so that the quadratic approximation exists and is accurate in a neighborhood of $\boldsymbol{\theta}_0$.

Assumption 4.9: *$W(\boldsymbol{\theta}, \boldsymbol{\pi})$ is twice continuously differentiable at $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, and $g(\boldsymbol{\pi})$ is twice continuously differentiable at $\boldsymbol{\pi}_0$.*

In addition to smoothness of the welfare and constraints, the decision problem must satisfy a number of regularity conditions in a neighborhood of $\boldsymbol{\theta}_0$. Let $\boldsymbol{\lambda}_0$ be the Lagrange multipliers corresponding to $\boldsymbol{\pi}_0$.

Assumption 4.10: *The decision problem (5) under $\boldsymbol{\theta}_0$ satisfies the following:*

1. *There exists a number α and a compact set $S \subset \mathbb{R}^k$ such that $\alpha > -W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$ and $\{\boldsymbol{\pi} : g(\boldsymbol{\pi}) \leq 0,\ -W(\boldsymbol{\theta}, \boldsymbol{\pi}) \leq \alpha\} \subseteq S$ for all $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.*

2. *The optimal $\boldsymbol{\pi}_0$ in (5) is unique.*

3. *The rows of $\nabla g(\boldsymbol{\pi}_0)$ are linearly independent.*

4. *Let $\mathcal{J}_1$ be the set of indices j such that $\lambda_{0j} > 0$. Let $\mathcal{J}_2$ be the set of indices j such that $g_j(\boldsymbol{\pi}_0) = 0$ and $\lambda_{0j} = 0$. For every nonzero vector c in $\mathcal{C} = \{c : \nabla g_j(\boldsymbol{\pi}_0)'c = 0, j \in \mathcal{J}_1;\ \nabla g_j(\boldsymbol{\pi}_0)'c \leq 0, j \in \mathcal{J}_2\}$ we have that $c'Hc < 0$.*

The first assertion of Assumption 4.10 requires the decision problem to be feasible and bounded in a neighborhood of $\boldsymbol{\theta}_0$. The third assertion is equivalent to uniqueness of the Lagrange multipliers corresponding to $\boldsymbol{\pi}_0$ and ensures the feasible set under the linear constraints is nonempty (Shapiro 1985). The fourth is a strict second-order sufficient condition which ensures that the Lagrangian is locally convex in $\boldsymbol{\pi}$ at $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0)$.

Another common regularity condition for sensitivity analysis of nonlinear optimization problems is the strict complementary slackness condition, which ensures that all constraints may be treated as equality constraints in a neighborhood of $\boldsymbol{\theta}_0$. This condition is not necessary for the quadratic approximation to yield an accurate approximation of the decision problem, which means our asymptotic analysis reflects finite-sample uncertainty about which constraints are binding. However, when strict complementary slackness does not hold, the value function may be only twice Hadamard directionally differentiable (Shapiro 1985).

Lemma 4.11: *Suppose Assumptions 4.3, 4.9 and 4.10 hold, and that J_0 is nonsingular. Then*

$$\begin{aligned} V_n^G(\mu_1, \Sigma_1) - V_n^Q(D\mu_1) = W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) + (\mu_1 - \boldsymbol{\theta}_0)' \nabla_\theta W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \\ + \frac{1}{2}\Big(trace(\Sigma_1 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)) + (\mu_1 - \boldsymbol{\theta}_0)' \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)(\mu_1 - \boldsymbol{\theta}_0) \Big) \\ + o_p(n^{-1}) \end{aligned}$$

Proof. See Appendix A.2. □

Asymptotic analysis of statistical decision rules often employ local quadratic approximations to loss functions (Hirano and Porter 2020). Lemma 4.11 is a generalization of this idea to constrained optimization problems. The particular form of the quadratic approximation is a consequence of results in sensitivity analysis of nonlinear optimization (Shapiro (1985), Shapiro (1988), Bonnans and Shapiro (2013)). This result shows that V_n^Q is a second-order approximation to V_n^G in a neighborhood of $\boldsymbol{\theta}_0$, up to terms which do not depend on $\boldsymbol{\pi}$ and therefore do not affect the chosen design and policy.

Our next result establishes that solving $\hat{V}_n^Q$ is asymptotically equivalent to solving V_n^G. We define $(\boldsymbol{\delta}_n^Q, \boldsymbol{\pi}_n^Q)$ as the optimal design and policy in the quadratic approximation V_n^Q, and $(\hat{\boldsymbol{\delta}}_n^Q, \hat{\boldsymbol{\pi}}_n^Q)$ as the estimated counterparts.

Theorem 4.12: *Suppose Assumptions 4.3, 4.6, 4.9, and 4.10 hold. Assume there exists a function* $\bar{W}(\boldsymbol{\theta})$ *such that* $R(\boldsymbol{\theta}, \tilde{\boldsymbol{\pi}}_n^Q) \leq n^{-1}\bar{W}(\boldsymbol{\theta})$ *and* $\int |\bar{W}(\boldsymbol{\theta})|^{1+\iota} d\boldsymbol{\theta} < \infty$ *for some* $\iota > 0$*. Then*

$$\mathbb{E}_{\boldsymbol{\delta}_n^G}\left[V_n^G(\mu_1, \Sigma_1)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

where $M(\mu_0, \Sigma_0) = W(\mu_0, \boldsymbol{\pi}_0) + \frac{1}{2} trace(\Sigma_0 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\mu_0, \boldsymbol{\pi}_0))$*. Further,*

$$\mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1)\right] = \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[\hat{V}_n^Q(D\mu_1)\right] + o(n^{-1})$$

Proof. See Appendix A.2. □

This result shows that the estimated quadratic approximation is asymptotically equivalent to the Gaussian value function, up to a term $M(\mu_0, \Sigma_0)$ which does not depend on the chosen design $\boldsymbol{\delta}$ or policy $\boldsymbol{\pi}$. This means that solving $\hat{V}_n^Q$ and implementing the resulting design and policy is asymptotically equivalent to solving V_n^G. The additional dominance assumption ensures that regret is uniformly integrable in the Gaussian problem where $dQ(\boldsymbol{\theta})$ is replaced by Lebesgue measure. Assumption 4.6 ensures this only for the finite-sample problem with the original prior $dQ(\boldsymbol{\theta})$.

4.4 Asymptotic optimality

We now combine Theorems 4.7 and 4.12 to show that the solution to V_n^Q provides an upper bound for the asymptotic performance of any sequence of designs and policies in the finite-sample experiment. We also show our method attains this bound– solving $\hat{V}_n^Q$ and using the resulting design in the main wave is asymptotically optimal. This establishes that no other feasible design and policy in the finite-sample can asymptotically outperform the solution to $\hat{V}_n^Q$.

Theorem 4.13: *Maintain the assumptions of Theorems 4.7 and 4.12. Then* V_n^Q *provides an asymptotic upper bound on the welfare of any sequence of designs and policies in the finite-sample experiment. That is, if* $(\boldsymbol{\delta}_n, \boldsymbol{\pi}_n)$ *is a sequence of feasible designs and policies in the finite-sample experiment with* $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)$*, then*

$$\limsup_{n\to\infty} \mathbb{E}_{\boldsymbol{\delta}_n}\left[W\left(\boldsymbol{\theta}, \boldsymbol{\pi}_n\right)\right] \leq \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[W_n^Q\left(D\boldsymbol{\theta}, \boldsymbol{\pi}_n^Q\right) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

where $M(\mu_0, \Sigma_0)$ is as in Theorem 4.12. Moreover, this upper bound is attained by solving $\hat{V}_n^Q$, using $\hat{\boldsymbol{\delta}}_n^Q$ in the main wave and then solving resulting finite-sample policy choice problem:

$$\lim_{n\to\infty} \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[W_n^Q\left(D\boldsymbol{\theta}, c_\infty\right) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

i.e. the design $\hat{\boldsymbol{\delta}}_n^Q$ is asymptotically optimal.

Proof. See Appendix A.3. □

This is the main result of the paper, which justifies the use of the method proposed in Section 3. While stated as a theorem, it is a straightforward consequence of Theorems 4.7 and 4.12. The first assertion of Theorem 4.13 states that no feasible design and policy in the finite-sample experiment can achieve higher welfare than V_n^Q with the optimal design $\boldsymbol{\delta}_n^Q$. However, V_n^Q describes a decision problem in the fictional limit experiment in which only the Gaussian random variables A_0 and A_1 are observed, and in which the quadratic approximation to welfare is constructed around the unknown parameter $\boldsymbol{\theta}_0$. This bound is useful only insofar as the design and policy we use in practice can actually attain this bound.

To this end, the second assertion of Theorem 4.13 establishes that the design $\hat{\boldsymbol{\delta}}_n^Q$ obtained from solving $\hat{V}_n^Q$ achieves this bound. An experimenter who (i) implements $\hat{\boldsymbol{\delta}}_n^Q$ in the main wave of the experiment and then (ii) solves a policy choice problem conditional on the data from this experiment will asymptotically achieve the welfare of the limit experiment. We may restate the second result of this theorem by saying that for any sequence of designs $\boldsymbol{\delta}_n$ in the finite-sample experiment,

$$\limsup_{n\to\infty} \mathbb{E}_{\boldsymbol{\delta}_n}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right] \leq \lim_{n\to\infty} \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right].$$

Therefore, this theorem shows that using $\hat{\boldsymbol{\delta}}_n^Q$ is asymptotically equivalent to solving the finite-sample problem by backwards induction exactly.

Remark 4.14 (Binary treatment choice)**:** Theorems 4.7 and 4.12 rely on the continuity of the policy c_n^Q in the design $\boldsymbol{\delta}$. This is established in the proof of Theorem 4.12. This highlights a distinction between the types of decision problems considered here versus the problem of whether

to administer a binary treatment to a population (e.g. Hirano and Porter (2009)). In such settings, the payoff is the average treatment effect which implies the welfare function is linear in the policy. This leads to a discontinuous cutoff rule. With a linear payoff the strict second-order sufficient condition in Assumption 4.10 is not satisfied. To apply the results of this section, it is necessary that the welfare function be strictly concave in the policy. This may still be satisfied in settings with a binary treatment if the policy assigns treatment on the basis of continuous variables, as in Bhattacharya and Dupas (2012). ◊

5 Application to Progresa

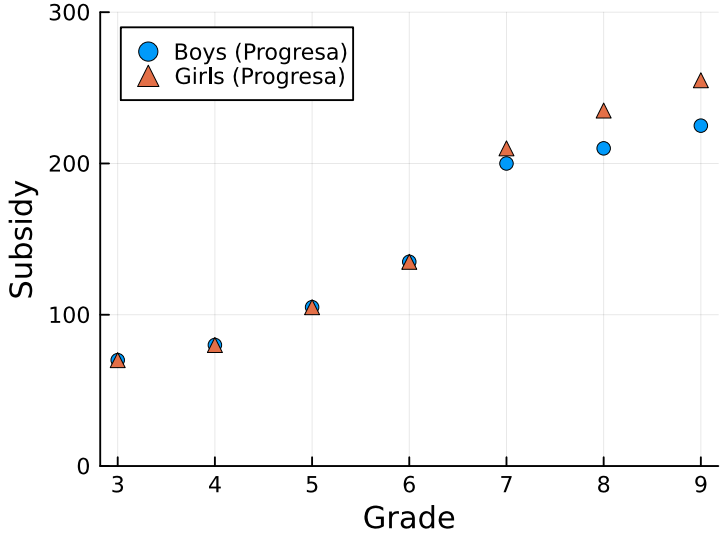
We demonstrate how to apply our method in practice in an application to the Progresa cash transfer program in Mexico. We consider the policy choice problem of choosing the size and targeting of the cash transfer to maximize the secondary school completion rate and reduce gender disparities in school completion. We consider counterfactual experiments that an experimenter could have conducted to learn about the optimal policy. We estimate that designing the experiment optimally for this policy choice problem can deliver equally effective policies as the original Progresa experiment with 60% fewer observations.

5.1 Setting

Progresa was a conditional cash transfer program administered by the Mexican government beginning in 1998. The program gave cash to households for every child enrolled in school. The size of the transfer depended on children's gender and grade of enrollment, and the magnitude was economically significant— the ninth grade subsidy constituted about 40% of an adult male's wage and about 66% of a child's wage (Schultz 2004). The government conducted an experimental evaluation of the program before a larger rollout in 2000.

Figure 1 shows the original subsidy schedule used in the Progresa experiment. 62% of the experimental sample was treated. After observing the results of the experiment, the government enacted a policy that gave the exact same subsidy that was used in the experiment to a wider set of villages.

Figure 1: Progresa subsidy schedule



Note: The Progresa subsidy schedule that was used in both the experimental evaluation and the subsequent policy rollout.

5.2 Model

We analyze the effect of the subsidy on graduation rates using a dynamic school choice model similar to those of Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2012). There are two key considerations captured in the dynamic model. First, households may be forward-looking in their schooling decisions. Subsidies offered in secondary school may affect the decision to enroll in primary school. Also, the decision to enroll in school one period can have affect future schooling decisions through dynamic complementarities in human capital accumulation (Cunha and Heckman 2007). Hence, the dynamic structure of the subsidy is a key mechansim through which the subsidy affects graduation rates. The second consideration captured by the model is that graduation rates are a long-term outcome that is not observed in the short-run experiment. A dynamic model allows us to forecast the long-term effect of the subsidy on graduation rates using short-run experimental data (see Athey et al. (2019) for a related but less structural approach to this problem).

We now describe the model, supressing the unit index i for simplicity. At each age τ before age 18, households may choose whether or not a child in grade s_τ will attend school, denoted $y_\tau = 1$ or $y_\tau = 0$. If the child attends school, they recieve a subsidy z_{s_τ} which depends on the grade of enrollment. Otherwise, the child works and earns a wage w_{τ,s_τ} which can depend on age as well as education level. Covariates $x_\tau = (w_{\tau,s_\tau}, b_\tau)$ include the wage as well as variables b_τ consisting

of age τ, current grade of enrollment s_τ, gender, and interactions. We model the choice-specific utilities as

$$u_{1\tau} = \theta_0 + \theta_1' b_\tau + \theta_2 z_{s_\tau} + \theta_3' b_\tau z_{s_\tau} + \epsilon_{1\tau}$$
$$u_{0\tau} = \theta_4 w_{\tau,s_\tau} + \theta_5' b_\tau w_{\tau,s_\tau} + \epsilon_{0\tau}.$$

where both the wage and the subsidy are interacted with covariates, allowing the subsidy to have different effects on different subpopulations.

Housholds choose a sequence of school attendance y_τ to maximize long-term utility. The level of schooling completed by age τ is s_τ. Each period, if the child was enrolled in school they may advance a grade with probability $r(\tau, s_\tau)$ depending on age and grade, or they may fail and remain at the same grade. Failure is assumed to be exogenous. Finally, at age $\tau = 18$ households obtain a terminal value

$$v_{18} = \theta_6 \mathbb{1}[s_{18} \geq 6] + \theta_7 \mathbb{1}[s_{18} = 9]$$

which depends on whether the child completed primary school and whether the child completed secondary school. This results in the following optimization problem that households solve:

$$\max_{\{y_\tau\}_{\tau=6}^{17}} \sum_{\tau=0}^{17} \beta^\tau \mathbb{E}\left[y_\tau u_{1\tau} + (1 - y_\tau) u_{0\tau}\right] + \beta^{18} \mathbb{E}\left[v_{18}\right]$$
$$s_{\tau+1} = s_\tau + y_\tau \text{Bernoulli}(r(\tau, s_\tau))$$

We assume that $\epsilon_{y\tau}$ have a type 1 extreme value distribution, which enables us to compute the value function for the households by backwards induction. We can then estimate the model by maximum likelihood from a cross-section of children so long as there is sufficient variation in the subsidy for the model to be identified.

Remark 5.1 (Stigma effect and lack of identification in pilot data)**:** If money is fungible to the household, we may expect that the coefficients on the wage and the subsidy are equal in magnitude but opposite in sign. The model considered here allows for the possibility that the subsidy has a different effect on utility than the wage. This "stigma effect" means that the model is not identified

without variation in the subsidy (Todd and Wolpin 2006, Attanasio, Meghir, and Santiago 2012). In this section, we assume the pilot data follows the original Progresa experimental design and therefore the model is identified from pilot data. If the stigma effect is not identified (for example, if the pilot data contains no variation in the subsidy) then we cannot construct the pilot estimate $\hat{\boldsymbol{\theta}}_0$ and the method of Section 3 cannot be applied. One possible approach is to make an additional identifying assumption (for example, that money is fungible to the household). The experiment may be used to learn about possible local failures of this assumption and subsequently choose a policy that is robust to these failures. We discuss this more in Section 6. ◊

5.3 Decision problem

The experimenter's decision problem is to choose (i) an experimental design and (ii) a policy that maximizes the fraction of children graduating from secondary school while minimizing the discrepancy in graduation rates between boys and girls. The welfare function we use is

$$W(\boldsymbol{\theta}, \boldsymbol{\pi}) = (1-\kappa)\mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\pi}}[s_{18} \geq 9] - \kappa \left(\mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\pi}}[s_{18} \geq 9 \mid \text{boy}] - \mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\pi}}[s_{18} \geq 9 \mid \text{girl}]\right)^2$$

where κ is a preference parameter that trades off how much the experimenter cares about the overall graduation rate versus gender inequality in graduation rates. We will present results for a variety of values of $\kappa \in [0, 1]$.

We suppose the experimenter has access to some pilot data and wants to run a bigger experiment to learn about the optimal policy. This pilot experiment consists of a sample of $n_0 = 500$ children from the original Progresa experiment as described above. However, the design of the main experiment can differ from the original Progresa experiment in two regards. The first way in which the design can differ is that the experimenter can change the peso amount of the subsidy offered to children in each grade. We consider subsidies which are piecewise linear in grade, and which differ by gender. The subsidy is 7-dimensional, consisting of the subsidy amount in grades $s = 3$ through $s = 9$, and is given by

$$z_s = \begin{cases} \pi_1 + \pi_2 s + \pi_3 (s-6)\mathbb{1}[s > 6] & \text{if boy} \\ \pi_4 + \pi_5 s + \pi_6 (s-6)\mathbb{1}[s > 6] & \text{if girl} \end{cases}$$

The entire 7-dimensional vector of subsidies by grade is relevant to the household's decision because households are forward-looking in the model and subsidies in later grades may affect the decision to enroll in school in the current period. The second way in which the design can differ from the original Progresa experiment is that the experimenter can let the probability of treatment vary by gender and by whether the child is in primary or secondary school. We will consider several possible sample sizes for the main wave of the experiment, from $n_1 = 500$ (equal to the size of the pilot wave) to $n_1 = 4000$.

Having run the main experiment, the experimenter will choose how to structure the subsidy policy across grade levels to maximize expected welfare. As in the experiment, the peso amount of the subsidy is piecewise linear in grade and varies by gender.

Both the experimental subsidy and the chosen policy subsidy are subject to constraints. First, the total amount of the chosen subsidy cannot exceed the total amount of the original Progresa subsidy, although the experimenter can allocate the subsidy differently across gender. Second, the subsidy must be nonnegative, preventing the experimenter from taxing children in some grades to subsidize children in other grades. Both of these constraints bind in practice, although the experimenter may be unsure about for which grades the nonnegativity constraint will bind and for which grades there will be a strictly positive subsidy.

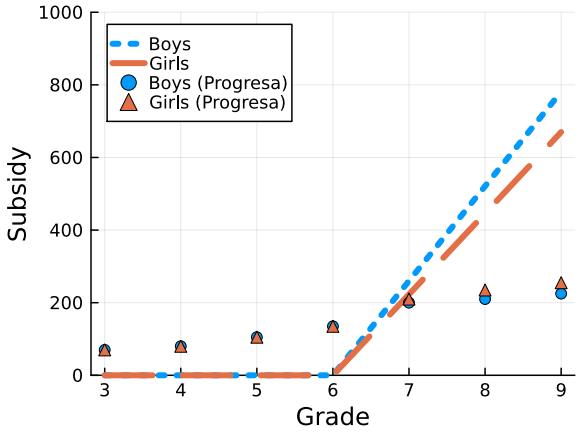
5.4 Optimal experiment

We now present the result of applying our solution method to this decision problem. Implementation details are provided in Appendix F.

We begin by presenting the optimal experiment for a small main wave of $n_1 = 500$ children and an inequality weight of $\kappa = 0.5$. We show the optimal experimental subsidy in Figure 2. The optimal experimental subsidy is zero in primary school, and increases steeply in secondary school. The total amount of the experimental subsidy is also larger than Progresa, which is feasible because the experimenter is not offering the subsidy (including the promise of subsidies in future grades) to primary school children. The optimal treatment probabilities are 67% for secondary school boys, 77% for secondary school girls, and 0% for primary school children. Results for other values of n_1 and κ are presented in Appendix F and are broadly consistent with these results.

The expected welfare resulting from the optimal experiment for $\kappa = 0.5$ across a variety of

Figure 2: Optimal experimental design



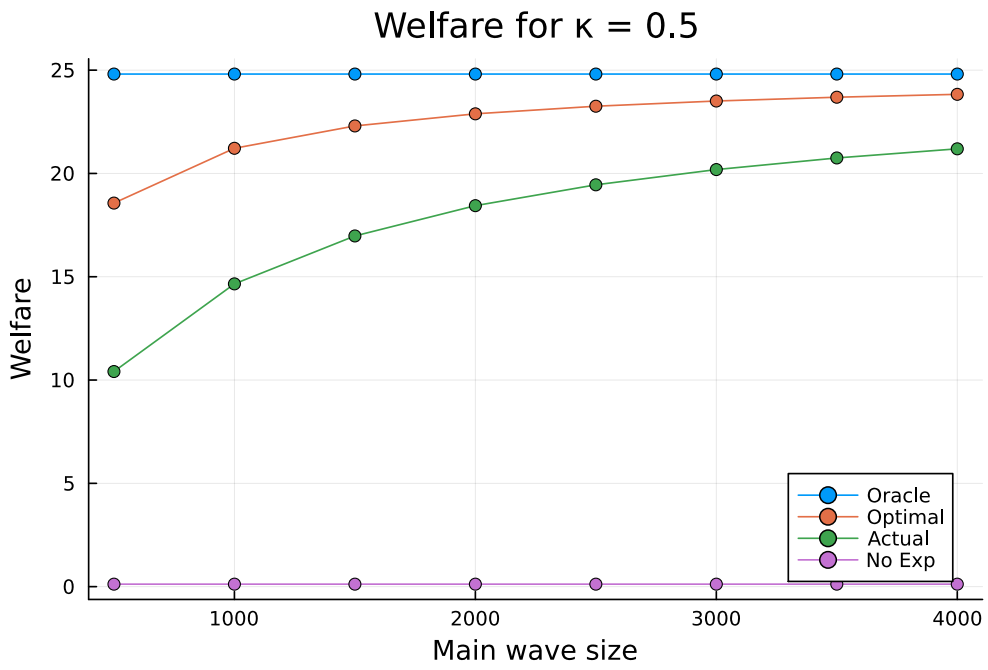
Note: the lines show the optimal experimental subsidy and the points show the actual Progresa subsidy. The corresponding treated fraction is 0% for primary school boys and girls, 67% for secondary school boys, and 77% for secondary school girls.

values of n_1 is shown in Figure 3. The gains from the optimal experiment are substantial. Using the original Progresa experimental design requires $n_1 = 2000$ observations to obtain the welfare attained by the optimal experiment with only $n_1 = 500$. By focusing the experiment on the policies and subpopulations where the subsidy is most likely to be effective, the optimal experiment makes much more effective use of the experimental budget. In Appendix F, we show results for other values of κ and decompose these results into the increase in graduation rates and the reduction in gender disparities. Most of the gains from the optimal experiment come from reducing gender disparities, which requires precise estimates of the response of boys and girls in secondary school to the subsidy.

We compute expected regret of both the optimal experiment and the original Progresa experiment in Table 1. Expected regret is the difference between the welfare of the optimal policy (i.e. the policy that we would choose if $\boldsymbol{\theta}$ were known exactly) and the welfare of the policy chosen with the experimental estimates. Across a wide range of values of n_1 and κ, regret is often less than half as large for the optimal experiment as for the Progresa experiment.

To interpret these results, we discuss how the pilot estimates inform the optimal experiment. When considering how to structure the subsidy across grades, there are two competing effects to consider. First, there is the option value of subsidies in later grades. Even if a child is not currently offered a subsidy, offering a subsidy in later grades may induce the child to enroll in school in the

Figure 3: Expected welfare from optimal experiment



Note: The expected welfare from the optimal experiment when $\kappa = 0.5$ is shown for a variety of values of n_1. "Oracle" refers to the expected welfare from the infeasible, optimal policy given the true parameter values. "Optimal" refers to the expected welfare from running the optimal experiment. "Actual" refers to the expected welfare from running the original Progresa experiment. "No Exp" refers to the expected welfare from using the pilot data only.

Table 1: Expected regret from optimal experiment

	$\kappa = 0.0$		$\kappa = 0.1$		$\kappa = 0.5$		$\kappa = 0.9$	
n_1	Optimal	Actual	Optimal	Actual	Optimal	Actual	Optimal	Actual
500	0.09	0.14	1.12	2.54	6.25	14.4	11.45	26.03
1000	0.05	0.1	0.65	1.79	3.59	10.16	6.74	18.54
1500	0.04	0.08	0.46	1.39	2.51	7.84	4.78	14.43
2000	0.03	0.06	0.35	1.13	1.92	6.37	3.71	11.82
2500	0.03	0.05	0.28	0.95	1.56	5.36	3.03	10.02
3000	0.02	0.05	0.24	0.82	1.3	4.62	2.56	8.69
3500	0.02	0.04	0.2	0.72	1.12	4.06	2.22	7.67
4000	0.02	0.04	0.18	0.64	0.98	3.62	1.96	6.87

Note: Expected regret is computed via Monte Carlo simulation over the post-pilot posterior and is therefore conditional on the pilot data.

current period so they have the option of receiving the subsidy in later grades. This potentially makes subsidies in later grades more effective. Second, there is the dynamic effect of human capital accumulation. If a child is offered a subsidy in the current period and enrolls in school, they may be more likely to enroll in school in future periods. This potentially makes subsidies in early grades more effective. Which effect dominates in practice is an empirical question.

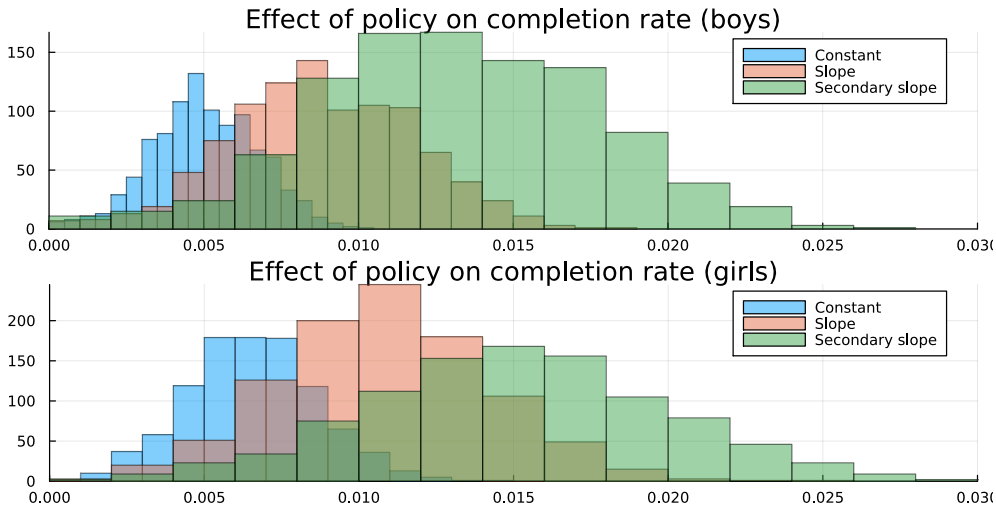
Our pilot estimates of the dynamic model indicate that the first effect dominates the second. This is in line with the results of Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2012). Our estimates indicate that increasing the subsidy in later grades is more effective at increasing graduation rates than increasing the subsidy in early grades. As a result, after observing the pilot data the experimenter is confident that the optimal policy will involve offering a subsidy in later grades, and does not need to experiment with sub-optimal policies that offer subsidies in early grades. Full pilot estimates of parameters are presented in Appendix F.

To see this, we plot the post-pilot posterior on the effect of adjusting our policy on graduation rates in Figure 4. Our policy is a piecewise linear subsidy in grade, where $\boldsymbol{\pi}$ consists of a constant, a slope, and an increase in slope starting in secondary school, separately for boys and girls. We consider increasing each of these parameters by 1 peso, and plot draws from the post-pilot posterior on the resulting increase in graduation rates in Figure 4. For both boys and girls, it is clear that increasing the steepness of the subsidy in secondary school is more effective than adjusting the policy in other dimensions. The principle question that remains for the experimenter is whether this effect is larger for boys or girls.

Given that subsidies in later grades are likely to be more effective, the optimal experiment focuses on learning about the effects of making marginal adjustments to this pilot estimate. This is where the value of information is highest, and therefore the optimal experiment focuses on discerning the differential effect of secondary school subsidies on boys versus girls. Compared to the original Progresa experiment, the posterior variance on these marginal effects is 57% smaller. Hence, the optimal experiment is more effectively able to learn about the decision-relevant marginal effect of adjusting the subsidy policy.

To summarize, the optimal experiment is more effective for two reasons. First, the optimal experiment focuses on learning about the effects of counterfactual policies through the marginal effect of policy changes. To choose optimal policies, the experimenter only needs to know the effect

Figure 4: Pilot estimates of marginal effects



Note: draws from the post-pilot posterior on the effect on graduation rates of increasing the subsidy in each dimension of the policy. These effects are highly correlated (> 0.9) within gender, but essentially uncorrelated across gender.

of increasing the subsidy along each dimension of the policy. In this setting, this is the effect of increasing the slope of the subsidy with respect to grade. Second, the optimal experiment takes into account policy constraints to determine which of these marginal effects are most important to learn about. The pilot data indicates that policies which give money to primary school children are unlikely to be implemented, so this marginal effect is less important to learn about than the effect of changing the secondary school subsidy. By focusing the experiment on learning about the decision-relevant marginal effects of policy changes, the optimal experiment is able to more effectively learn about the optimal policy.

6 Extensions

In this section we provide some extensions of our method to three practically relevant settings. First, we consider a multi-wave experiment in which the experimenter can design several waves of an experiment after observing pilot data to sequentially refine estimates of the optimal policy. Second, we consider a situation in which pilot data only weakly or does not identify all parameters of the model. We approach this by supposing the experimenter has a family of priors even after observing pilot data, and wants to be robust to this family. Third, we consider situations where

formulating a prior about the marginal effect of the policy is feasible, but the experimenter does not want to commit to a full prior on the model parameters.

6.1 Multi-wave experiments

If the experimenter has the opportunity to run several waves of experimentation, it may be beneficial to sequentially adjust the design of the experiment as the experimenter learns more about the optimal policy.

Suppose the experiment lasts for T waves, each consisting of n_t experimental units. We will assume $\frac{n_t}{n}$ has a strictly positive limit as $n \to \infty$ for every $t \in \{1, \dots, T\}$, allowing asymptotic analysis to apply to each wave. For each wave $t \in \{1, \dots, T\}$, the experimenter chooses a design $\boldsymbol{\delta}_t$. In the last period $T+1$, the experimenter chooses a policy $\boldsymbol{\pi}$. The benefit of revising the design $\boldsymbol{\delta}_t$ throughout the experiment is that the experimenter can refine the estimate of the optimal policy and the marginal effect of the policy. For example, the experimenter can stop treating certain subpopulations if the treatment is not effective for them, and reallocate the budget to other subpopulations which are more relevant for the policy choice problem.

The multi-wave generalization of (6a)-(6b) is

$$V^Q_{T+1,n}\left(D\mu_{T+1}\right) = \max_{\boldsymbol{\pi}} \quad W^Q(D\mu_{T+1}, \boldsymbol{\pi}) \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0. \tag{7a}$$

The optimal design in each period $t \in \{1, \dots, T\}$ solves

$$V^Q_{t,n}\left(\mu_t, \Sigma_t\right) = \max_{\boldsymbol{\delta}_t} \quad \mathbb{E}\left[V^Q_{t+1,n}\left(\mu_{t+1}, \Sigma_{t+1}\right) \mid \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{t-1}\right] \quad \text{s.t.} \quad f(\boldsymbol{\delta}_t) \leq 0 \tag{7b}$$

where $\boldsymbol{\theta} \sim N(\mu_t, \Sigma_t)$ given $\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{t-1}$ and the standard Bayesian updating formula (3) is used to update the posterior every period.

In Appendix E, we show that the solution to (7a)-(7b) delivers an asymptotically optimal experiment and policy. However, compared to the single-period case, the optimal dynamic experiment and policy in (7a)-(7b) can be difficult to solve in practice. While the terminal value function $V^Q_{T+1,n}$ is just as low-dimensional as in the single-period case, the intermediate value functions $V^Q_{t,n}$ depend on the full posterior distribution of $\boldsymbol{\theta}$ characterized by (μ_t, Σ_t) rather than the lower-dimensional

$(D\mu_t, D\Sigma_t D')$. As was seen in Section 3, the full posterior distribution can be high-dimensional enough to make this difficult.

Therefore, these results provide a benchmark against which many heuristic methods for dynamic experiments can be compared in a simpler Gaussian environment than the finite-sample experiment. These results may also open the door to new heuristic methods based on the Gaussian and quadratic approximations we propose.

6.2 Lack of identification and multiple priors

Thus far we have considered the problem of designing an optimal experiment when the posterior after observing pilot data is well approximated by a unique Gaussian distribution. This requires that J_0 be nonsingular so that $\sqrt{n}$-consistent estimates of $\boldsymbol{\theta}$ are available. As a result, the solution method described in Section 3 does not depend on the pre-pilot prior q. This may not be the case, for example, if pilot data only consists of a pre-treatment survey, with no variation in treatment status.

We propose placing a family of priors on the unidentified parameters. It is required that these priors are asymptotically non-vanishing, i.e. that the prior is on the local parameter h of Section 4. For this section, we will take as given the Gaussian-quadratic limiting environment described in Section 4 and consider modifications of the value function V_n^Q to account for these concerns.

We specify a family of priors which are close to a reference prior q_0 in the sense of Kullback-Leibler divergence. Kullback-Leibler balls have been fruitfully employed in dynamic decision environments under ambiguity in macroeconomics; see Hansen and Sargent (2001), Hansen and Sargent (2008), and Hansen and Sargent (2010). For a given preference parameter $\kappa \geq 0$ governing the degree of concern for misspecification, the value function is given by

$$V_{\kappa,n}^Q\left(\hat{\boldsymbol{\theta}}_1, J(\boldsymbol{\delta})\right) = \max_{\boldsymbol{\pi}} \min_{q} \quad \mathbb{E}_q[W_n^Q(D\boldsymbol{\theta}, \boldsymbol{\pi})] - \kappa \int \log \frac{q(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_1, J(\boldsymbol{\delta}))} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{8a}$$
$$\text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0$$

where $\boldsymbol{\theta} \sim q$ and q is penalized for being far from the posterior under the reference prior. Having

computed the value function $V^Q_{\kappa,n}$, the optimal design solves

$$\max_{\boldsymbol{\delta}} \min_{q} \quad \mathbb{E}_q\left[V^Q_{\kappa,n}\left(\hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta})\right)\right] - \kappa \int \log \frac{q(\boldsymbol{\theta})}{q_0(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{8b}$$
$$\text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0$$

where $\boldsymbol{\theta} \sim q$ and q is penalized for being far from the reference prior. In both periods, q ranges over all possible distributions on $\boldsymbol{\theta}$.

The value function (8a)-(8b) appears difficult to solve due to the max-min structure and the lack of a single prior on $D\boldsymbol{\theta}$ characterizing the state. However, we show in Appendix D that the solution to (8a)-(8b) is given by a simpler value function. A standard calculation for entropy-regularized optimization problems (e.g. Dupuis and Ellis (2011)) characterizes the least favorable prior in each period and the corresponding value function.

Remark 6.1 (Dynamic consistency)**:** In dynamic decision problems with multiple priors, dynamic consistency is a concern. That is, having decided on a contingency plan by solving the ex-ante problem under commitment, the experimenter may find it optimal to deviate from this plan in the second period. This means the decision problem cannot be solved by backwards induction. We follow Hansen and Sargent (2001) in solving a "multiplier problem" each period where the worst-case prior is implicitly chosen by penalizing the distance between the adversarially chosen prior q and the reference prior q_0. This amounts to allowing the decision maker to consider a different worst-case prior at each stage of the decision tree. ◊

Remark 6.2 (Choice of reference prior)**:** The choice of reference prior is important. One way of constructing a reference prior involves specifying a restricted model which is identified with the pilot data, and for which preliminary estimates can be obtained. The experimenter then considers relaxing the model to allow for misspecification of the restricted model. Details are given in Appendix D. ◊

6.3 Robustness to nuisance parameters

An experimenter may not wish to commit to a prior on the full parameter vector $\boldsymbol{\theta}$. Because of this, it is often desireable to be robust to a wide class of priors. While a fully minimax analysis is outside the scope of this paper, it is possible to be minimax over all parameters which do not

directly enter the policy choice problem. As with the previous section, we will take as given the Gaussian-quadratic limiting environment described in Section 4 and consider modifications of the value function V_n^Q to account for these concerns.

The method presented in Section 3 and the results of Section 4 demonstrate that $\boldsymbol{\theta}$ enters the policy choice problem only through $D\boldsymbol{\theta}$. We can therefore consider $D\boldsymbol{\theta}$ as the parameter of interest and $D^\perp\boldsymbol{\theta}$ as the nuisance parameter, where the rows of $D^\perp$ constitute an orthonormal basis for the null space of the rows of D.

Suppose the experimenter can specify a prior on the marginal effect of the policy, but struggles to specify a prior on the full model. We seek an experimental design which is robust to the set of all priors on $\boldsymbol{\theta}$ which imply the same prior on $D\boldsymbol{\theta}$. That is, we consider the set of priors

$$\mathcal{Q} = \left\{ q : \int \mathbb{1}[D\boldsymbol{\theta} \in B] q(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int \mathbb{1}[D\boldsymbol{\theta} \in B] q_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad \forall \text{ Borel } B \subseteq \mathbb{R}^k \right\}$$

where $q_0(\boldsymbol{\theta}) = dN(\boldsymbol{\theta}; \mu_0, \Sigma_0)$. This is the set of priors q which have the same marginal distribution on $D\boldsymbol{\theta}$ as the reference prior q_0.

Example 6.3 (Progresa)**:** It is possible that an experimenter or policymaker is be able to express beliefs about the effect of giving an extra peso to boys versus girls, but struggle to express beliefs about model primitives such as household preferences. Recall that $\boldsymbol{\theta}$ is 15-dimensional and includes utility parameters such as the stigma effect of the subsidy and subjective costs of primary versus secondary school attendance. In contrast, a prior on $D\boldsymbol{\theta}$ can be constructed by eliciting beliefs about the effects of different subsidy levels on overall graduation rates. ♢

With this set of priors, the value function the experimenter solves is

$$V_{\perp,n}^Q \left(\hat{\boldsymbol{\theta}}_1, J(\boldsymbol{\delta}) \right) = \max_{\boldsymbol{\pi}} \inf_{q \in \mathcal{Q}} \quad \mathbb{E}_q[W_n^Q(D\boldsymbol{\theta}, \boldsymbol{\pi}) \mid \hat{\boldsymbol{\theta}}_1] \quad \text{s.t.} \quad g_\infty(\boldsymbol{\pi}) \leq 0 \tag{9a}$$

where $\boldsymbol{\theta} \sim q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_1)$. The optimal design solves

$$\max_{\boldsymbol{\delta}} \inf_{q \in \mathcal{Q}} \quad \mathbb{E}_q \left[V_{\perp,n}^Q \left(\hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta}) \right) \right] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0 \tag{9b}$$

where $\boldsymbol{\theta} \sim q(\boldsymbol{\theta})$.

This value function, like (8a)-(8b), appears difficult to solve due to the max-min structure and the lack of a single prior on $D\boldsymbol{\theta}$ characterizing the state. However, we show in Appendix D that the solution to (9a)-(9b) is given by a simpler value function. The optimal design ignores estimates of the nuisance parameter and updates the prior as if only estimates of $D\boldsymbol{\theta}$ were available. For any particular prior q on the full vector $\boldsymbol{\theta}$, this update will not generally coincide with the correct Bayesian updating formula for the full posterior. However, Theorem D.2 shows that this updating formula coincides with that of the least favorable prior in $\mathcal{Q}$.

The intuition behind this result is that the least favorable prior in $\mathcal{Q}$ is one which is uncorrelated with the parameters of interest. This is because prior correlation between parameters of interest and nuisance parameters can only help the experimenter in formulating a more precise posterior. Therefore, the least favorable prior places no correlation between the parameters of interest and the nuisance parameters and estimates of the nuisance parameters will not affect the posterior on the parameters of interest. In response, the experimenter will ignore the nuisance parameter when forming the posterior. The resulting design is minimax over all priors in $\mathcal{Q}$.

References

Adusumilli, Karun (Jan. 31, 2024). *Risk and Optimal Policies in Bandit Experiments*. arXiv: 2112.06363 [cs, econ]. URL: http://arxiv.org/abs/2112.06363 (visited on 02/26/2024). Pre-published.

Andrews, Donald W.K. (1994). "Chapter 37 Empirical Process Methods in Econometrics". In: *Handbook of Econometrics*. Vol. 4. Elsevier, pp. 2247–2294. ISBN: 978-0-444-88766-5. DOI: 10.1016/S1573-4412(05)80006-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S1573441205800066 (visited on 03/30/2024).

Armstrong, Timothy B. (May 5, 2022). *Asymptotic Efficiency Bounds for a Class of Experimental Designs*. arXiv: 2205.02726 [stat]. URL: http://arxiv.org/abs/2205.02726 (visited on 05/24/2024). Pre-published.

Athey, S. and G.W. Imbens (2017). "The Econometrics of Randomized Experiments". In: *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 73–140. ISBN: 978-0-444-63324-8. DOI: 10.1016/bs.hefe.2016.10.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214658X16300174 (visited on 02/26/2024).

Athey, Susan and Stefan Wager (2021). "Policy Learning With Observational Data". In: *Econometrica* 89.1, pp. 133–161. ISSN: 0012-9682. DOI: 10.3982/ECTA15732. URL: https://www.econometricsociety.org/doi/10.3982/ECTA15732 (visited on 07/24/2024).

Athey, Susan et al. (Nov. 2019). *The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely*. w26463. Cambridge, MA: National Bureau of Economic Research, w26463. DOI: 10.3386/w26463. URL: http://www.nber.org/papers/w26463.pdf (visited on 10/13/2024).

Attanasio, Orazio P., Costas Meghir, and Ana Santiago (Jan. 1, 2012). "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA". In: *The Review of Economic Studies* 79.1, pp. 37–66. ISSN: 1467-937X, 0034-6527. DOI: 10.1093/restud/rdr015. URL: https://academic.oup.com/restud/article/79/1/37/1562110 (visited on 12/16/2023).

Bai, Yuehao (Dec. 1, 2022). "Optimality of Matched-Pair Designs in Randomized Controlled Trials". In: *American Economic Review* 112.12, pp. 3911–3940. ISSN: 0002-8282. DOI: 10.1257/

aer.20201856. URL: https://pubs.aeaweb.org/doi/10.1257/aer.20201856 (visited on 07/24/2024).

Bai, Yuehao et al. (June 13, 2024). *On the Efficiency of Finely Stratified Experiments*. arXiv: 2307.15181 [econ, math, stat]. URL: http://arxiv.org/abs/2307.15181 (visited on 07/24/2024). Pre-published.

Bhattacharya, Debopam and Pascaline Dupas (Mar. 2012). "Inferring Welfare Maximizing Treatment Assignment under Budget Constraints". In: *Journal of Econometrics* 167.1, pp. 168–196. ISSN: 03044076. DOI: 10.1016/j.jeconom.2011.11.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407611002697 (visited on 10/17/2024).

Bonnans, J Frédéric and Alexander Shapiro (2013). *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media. ISBN: 1-4612-1394-0.

Cai, Yong and Ahnaf Rafi (May 2024). "On the Performance of the Neyman Allocation with Small Pilots". In: *Journal of Econometrics* 242.1, p. 105793. ISSN: 03044076. DOI: 10.1016/j.jeconom.2024.105793. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407624001398 (visited on 10/26/2024).

Card, David and Dean R. Hyslop (Nov. 2005). "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers". In: *Econometrica* 73.6, pp. 1723–1770. ISSN: 0012-9682, 1468-0262. DOI: 10.1111/j.1468-0262.2005.00637.x. URL: http://doi.wiley.com/10.1111/j.1468-0262.2005.00637.x (visited on 10/17/2024).

Carlier, Guillaume, Victor Chernozhukov, and Alfred Galichon (June 1, 2016). "Vector Quantile Regression: An Optimal Transport Approach". In: *The Annals of Statistics* 44.3. ISSN: 0090-5364. DOI: 10.1214/15-AOS1401. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-44/issue-3/Vector-quantile-regression-An-optimal-transport-approach/10.1214/15-AOS1401.full (visited on 10/05/2024).

Cesa-Bianchi, Nicolo, Roberto Colomboni, and Maximilian Kasy (July 29, 2024). *Adaptive Maximization of Social Welfare*. arXiv: 2310.09597 [cs, econ, stat]. URL: http://arxiv.org/abs/2310.09597 (visited on 08/08/2024). Pre-published.

Chaloner, Kathryn and Isabella Verdinelli (Aug. 1, 1995). "Bayesian Experimental Design: A Review". In: *Statistical Science* 10.3. ISSN: 0883-4237. DOI: 10.1214/ss/1177009939. URL: https:

//projecteuclid.org/journals/statistical-science/volume-10/issue-3/Bayesian-Experimental-Design-A-Review/10.1214/ss/1177009939.full (visited on 03/11/2024).

Chaudhuri, Probal and Per A. Mykland (1993). "Nonlinear Experiments: Optimal Design and Inference Based on Likelihood". In: *Journal of the American Statistical Association* 88.422, pp. 538–546. ISSN: 0162-1459. DOI: 10.2307/2290334. JSTOR: 2290334. URL: https://www.jstor.org/stable/2290334 (visited on 01/16/2024).

Chen, Jiafeng and Isaiah Andrews (Sept. 21, 2023). *Optimal Conditional Inference in Adaptive Experiments*. arXiv: 2309.12162 [cs, econ, math, stat]. URL: http://arxiv.org/abs/2309.12162 (visited on 02/26/2024). Pre-published.

Compiani, Giovanni et al. (Dec. 21, 2023). "Online Search and Optimal Product Rankings: An Empirical Framework". In: *Marketing Science*, mksc.2022.0071. ISSN: 0732-2399, 1526-548X. DOI: 10.1287/mksc.2022.0071. URL: https://pubsonline.informs.org/doi/10.1287/mksc.2022.0071 (visited on 04/12/2024).

Cunha, Flavio and James Heckman (2007). "The Technology of Skill Formation". In: 97.2.

Cytrynbaum, Max (2024). "Optimal Stratification of Survey Experiments". In:

Duflo, Esther, Rema Hanna, and Stephen P Ryan (June 1, 2012). "Incentives Work: Getting Teachers to Come to School". In: *American Economic Review* 102.4, pp. 1241–1278. ISSN: 0002-8282. DOI: 10.1257/aer.102.4.1241. URL: https://pubs.aeaweb.org/doi/10.1257/aer.102.4.1241 (visited on 07/23/2024).

Dupuis, Paul and Richard S Ellis (2011). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons. ISBN: 1-118-16589-6.

Fang, Zheng and Andres Santos (Sept. 11, 2018). "Inference on Directionally Differentiable Functions". In: *The Review of Economic Studies*. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdy049. URL: https://academic.oup.com/restud/advance-article/doi/10.1093/restud/rdy049/5094886 (visited on 10/11/2024).

Gertler, Paul (Apr. 1, 2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment". In: *American Economic Review* 94.2, pp. 336–341. ISSN: 0002-8282. DOI: 10.1257/0002828041302109. URL: https://pubs.aeaweb.org/doi/10.1257/0002828041302109 (visited on 12/16/2023).

Gertler, Paul J, Sebastian W Martinez, and Marta RubioCodina (2012). "Investing Cash Transfers to Raise Long-Term Living Standards". In:

Hahn, Jinyong, Keisuke Hirano, and Dean Karlan (Jan. 2011). "Adaptive Experimental Design Using the Propensity Score". In: *Journal of Business & Economic Statistics* 29.1, pp. 96–108. ISSN: 0735-0015, 1537-2707. DOI: 10.1198/jbes.2009.08161. URL: http://www.tandfonline.com/doi/abs/10.1198/jbes.2009.08161 (visited on 09/04/2024).

Hansen, Lars Peter and Thomas J Sargent (2001). "Robust Control and Model Uncertainty". In:

— (2008). *Robustness*. Princeton university press. ISBN: 1-4008-2938-0.

Hansen, Lars Peter and Thomas J. Sargent (2010). "Wanting Robustness in Macroeconomics". In: *Handbook of Monetary Economics*. Vol. 3. Elsevier, pp. 1097–1157. ISBN: 978-0-444-53470-5. DOI: 10.1016/B978-0-444-53454-5.00008-6. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444534545000086 (visited on 05/14/2024).

Hirano, Keisuke and Jack R. Porter (2009). "Asymptotics for Statistical Treatment Rules". In: *Econometrica* 77.5, pp. 1683–1701. ISSN: 1468-0262. DOI: 10.3982/ECTA6630. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6630 (visited on 02/26/2024).

— (2020). "Asymptotic Analysis of Statistical Decision Rules in Econometrics". In: *Handbook of Econometrics*. Vol. 7. Elsevier, pp. 283–354. ISBN: 978-0-444-63649-2. DOI: 10.1016/bs.hoe.2020.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1573441220300040 (visited on 01/29/2024).

— (Feb. 6, 2023). *Asymptotic Representations for Sequential Decisions, Adaptive Experiments, and Batched Bandits*. arXiv: 2302.03117 [econ]. URL: http://arxiv.org/abs/2302.03117 (visited on 12/16/2023). Pre-published.

Kasy, Maximilian and Anja Sautmann (2021). "Adaptive Treatment Assignment in Experiments for Policy Choice". In: *Econometrica* 89.1, pp. 113–132. ISSN: 0012-9682. DOI: 10.3982/ECTA17527. URL: https://www.econometricsociety.org/doi/10.3982/ECTA17527 (visited on 05/20/2024).

Kitagawa, Toru and Aleksey Tetenov (2018). "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice". In: *Econometrica* 86.2, pp. 591–616. ISSN: 1468-0262. DOI: 10.3982/ECTA13288. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13288 (visited on 02/26/2024).

Krishnamurthy, Sanath Kumar et al. (2023). "Proportional Response: Contextual Bandits for Simple and Cumulative Regret Minimization". In:

Lattimore, Tor and Csaba Szepesvári (July 31, 2020). *Bandit Algorithms*. 1st ed. Cambridge University Press. ISBN: 978-1-108-57140-1 978-1-108-48682-8. DOI: 10.1017/9781108571401. URL: https://www.cambridge.org/core/product/identifier/9781108571401/type/book (visited on 02/26/2024).

Le Cam, Lucien (1972). "Limits of Experiments". In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. University of California Press Berkeley-Los Angeles, pp. 245–261.

Manski, Charles F. (2004). "Statistical Treatment Rules for Heterogeneous Populations". In: *Econometrica* 72.4, pp. 1221–1246. ISSN: 1468-0262. DOI: 10.1111/j.1468-0262.2004.00530.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2004.00530.x (visited on 02/26/2024).

— (2021). "Econometrics for Decision Making: Building Foundations Sketched by Haavelmo and Wald". In: *Econometrica* 89.6, pp. 2827–2853. ISSN: 0012-9682. DOI: 10.3982/ECTA17985. URL: https://www.econometricsociety.org/doi/10.3982/ECTA17985 (visited on 07/29/2024).

Mbakop, Eric and Max Tabord-Meehan (2021). "Model Selection for Treatment Choice: Penalized Welfare Maximization". In: *Econometrica* 89.2, pp. 825–848. ISSN: 0012-9682. DOI: 10.3982/ECTA16437. URL: https://www.econometricsociety.org/doi/10.3982/ECTA16437 (visited on 08/09/2024).

Parker, Susan W. and Petra E. Todd (2017). "Conditional Cash Transfers: The Case of "Progresa/Oportunidades"". In: *Journal of Economic Literature* 55.3, pp. 866–915. ISSN: 0022-0515. JSTOR: 26303307. URL: https://www.jstor.org/stable/26303307 (visited on 12/16/2023).

Pukelsheim, Friedrich (2006). *Optimal Design of Experiments*. SIAM. ISBN: 0-89871-604-7.

Russo, Daniel J. et al. (2018). "A Tutorial on Thompson Sampling". In: *Foundations and Trends® in Machine Learning* 11.1, pp. 1–96. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000070. URL: http://www.nowpublishers.com/article/Details/MAL-070 (visited on 02/26/2024).

Sakaguchi, Shosei (Apr. 10, 2024). *Estimation of Optimal Dynamic Treatment Assignment Rules under Policy Constraints*. arXiv: 2106.05031 [econ, stat]. URL: http://arxiv.org/abs/2106.05031 (visited on 05/22/2024). Pre-published.

Schultz, Paul (June 2004). "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program". In: *Journal of Development Economics* 74.1, pp. 199–250. ISSN: 03043878. DOI: 10.1016/j.jdeveco.2003.12.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304387803001858 (visited on 12/16/2023).

Shapiro, Alexander (Dec. 1985). "Second Order Sensitivity Analysis and Asymptotic Theory of Parametrized Nonlinear Programs". In: *Mathematical Programming* 33.3, pp. 280–299. ISSN: 0025-5610, 1436-4646. DOI: 10.1007/BF01584378. URL: http://link.springer.com/10.1007/BF01584378 (visited on 08/02/2024).

— (May 1988). "Sensitivity Analysis of Nonlinear Programs and Differentiability Properties of Metric Projections". In: *SIAM Journal on Control and Optimization* 26.3, pp. 628–645. ISSN: 0363-0129, 1095-7138. DOI: 10.1137/0326037. URL: http://epubs.siam.org/doi/10.1137/0326037 (visited on 08/02/2024).

— (Dec. 1991). "Asymptotic Analysis of Stochastic Programs". In: *Annals of Operations Research* 30.1, pp. 169–186. ISSN: 0254-5330, 1572-9338. DOI: 10.1007/BF02204815. URL: http://link.springer.com/10.1007/BF02204815 (visited on 10/11/2024).

Silvey, Samuel (2013). *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Vol. 1. Springer Science & Business Media. ISBN: 94-009-5912-5.

Staiger, Douglas O and James H Stock (1994). "Instrumental Variables Regression with Weak Instruments". In:

Stoye, Jörg (July 2009). "Minimax Regret Treatment Choice with Finite Samples". In: *Journal of Econometrics* 151.1, pp. 70–81. ISSN: 03044076. DOI: 10.1016/j.jeconom.2009.02.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407609000724 (visited on 08/09/2024).

— (Jan. 2012). "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments". In: *Journal of Econometrics* 166.1, pp. 138–156. ISSN: 03044076. DOI: 10.1016/j.jeconom.2011.06.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407611001254 (visited on 08/09/2024).

Tabord-Meehan, Max (Sept. 5, 2023). "Stratification Trees for Adaptive Randomisation in Randomised Controlled Trials". In: *Review of Economic Studies* 90.5, pp. 2646–2673. ISSN: 0034-

6527, 1467-937X. DOI: 10.1093/restud/rdac083. URL: https://academic.oup.com/restud/article/90/5/2646/6931814 (visited on 07/24/2024).

Todd, Petra E and Kenneth I Wolpin (Nov. 1, 2006). "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility". In: *American Economic Review* 96.5, pp. 1384–1417. ISSN: 0002-8282. DOI: 10.1257/aer.96.5.1384. URL: https://pubs.aeaweb.org/doi/10.1257/aer.96.5.1384 (visited on 12/16/2023).

Todd, Petra E. and Kenneth I. Wolpin (Mar. 1, 2023). "The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling". In: *Journal of Economic Literature* 61.1, pp. 41–85. ISSN: 0022-0515. DOI: 10.1257/jel.20211652. URL: https://pubs.aeaweb.org/doi/10.1257/jel.20211652 (visited on 04/12/2024).

Ursu, Raluca M. (Aug. 2018). "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions". In: *Marketing Science* 37.4, pp. 530–552. ISSN: 0732-2399, 1526-548X. DOI: 10.1287/mksc.2017.1072. URL: https://pubsonline.informs.org/doi/10.1287/mksc.2017.1072 (visited on 04/12/2024).

Van der Vaart, Aad W (2000). *Asymptotic Statistics*. Vol. 3. Cambridge university press. ISBN: 0-521-78450-6.

Van der Vaart, Aad W and Jon Wellner (2013). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media. ISBN: 1-4757-2545-0.

Viviano, Davide (July 20, 2022). *Experimental Design under Network Interference*. arXiv: 2003.08421 [econ, stat]. URL: http://arxiv.org/abs/2003.08421 (visited on 07/24/2024). Pre-published.

Viviano, Davide and Jess Rudder (2024). "Policy Design in Experiments with Unknown Interference". In:

Weitzman, Martin (1979). "Optimal Search for the Best Alternative". In: *Econometrica*.

Xu, Han (2024). "Asymptotic Analysis of Point Decisions with General Loss Functions". In: URL: https://sites.google.com/view/han-xu/research?authuser=0.

A Proofs of main results

A.1 Proof of Theorem 4.7

We first set some notation for the proofs. Let

$$p_{1,n,\boldsymbol{\theta}}(\boldsymbol{\delta}) = \prod_{i=n_0+1}^{n} p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})$$

be the density of the data generated by the main wave of the experiment, as a function of the design. Likewise, let

$$p_{0,n,\boldsymbol{\theta}} = \prod_{i=1}^{n_0} p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta}_0)$$

be the density of the pilot wave at the fixed pilot design $\boldsymbol{\delta}_0$. We also use the shorthand

$$\psi_i(\boldsymbol{\delta}) = \psi(y_i(z_i(\boldsymbol{\delta})), z_i(\boldsymbol{\delta}), x_i)$$

where $y_i(z_i) = y(z_i, x_i, \epsilon_i; \boldsymbol{\theta})$ and $z_i(\boldsymbol{\delta}) = z(x_i, \nu_i; \boldsymbol{\delta})$.

We now present the proof of the asymptotic representation result.

Lemma 4.4: *Suppose Assumptions 4.1, 4.2, and 4.3 hold. Let $(\boldsymbol{\delta}_n, \boldsymbol{\pi}_n)$ be a sequence of designs and policies in the finite-sample experiment, and define $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)$. Suppose $\boldsymbol{\delta}_n$ and c_n jointly converge in distribution under $\boldsymbol{\theta}_0$. Then there exists an experimental design and policy $(\boldsymbol{\delta}, c)$ in the limit experiment such that*

$$(\boldsymbol{\delta}_n, c_n) \overset{h}{\rightsquigarrow} (\boldsymbol{\delta}, c)$$

where $\overset{h}{\rightsquigarrow}$ denotes convergence in distribution along the sequence $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h/\sqrt{n}$.

Proof. We begin by deriving the weak limit of $A_{1,n}(\cdot)$. First, by the multivariate central limit theorem, for every finite set of points $b, \ldots, b_K \in \Delta$ the random vector $(A_{1,n}(b_1), \ldots, A_{1,n}(b_K))$ converges in distribution in $\mathbb{R}^K$ to a Gaussian vector with mean zero and covariance

$$\mathrm{Cov}\Big(A_{1,n}(b_j), A_{1,n}(b_{j'})\Big) = \mathrm{Cov}\Big(\psi_i(b_j), \psi_i(b_{j'})\Big)$$

for $i \in \{n_0+1, \ldots, n\}$. Second, $A_{1,n}(\cdot)$ is stochastically equicontinuous by assumption. From these two claims it follows by Van der Vaart (2000) Theorem 18.14 that

$$A_{1,n}(\cdot) \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} A_1(\cdot)$$

in $L_\infty(\Delta)$ for some tight random element $A_1(\cdot)$. Specifically, $A_1(\cdot)$ is a Gaussian process with mean zero and covariance

$$\mathrm{Cov}\Big(A_1(b), A_1(b')\Big) = \mathrm{Cov}\Big(\psi_i(b), \psi_i(b')\Big).$$

for b and b' in Δ, and where $i \in \{n_0+1, \ldots, n\}$. In particular, for any $h \in \mathbb{R}^\ell$,

$$h'A_{1,n}(\cdot) - \frac{1}{2}h'J(\cdot)h \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} h'A_1(\cdot) - \frac{1}{2}h'J(\cdot)h$$

We use this to derive the limiting distribution of the log-likelihood ratio for the main wave. Specifically,

$$\begin{aligned}\log \frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\cdot)}{p_{1,n,\boldsymbol{\theta}_0}(\cdot)} &= \left(\log \frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\cdot)}{p_{1,n,\boldsymbol{\theta}_0}(\cdot)} - h'A_{1,n}(\cdot) + \frac{1}{2}h'J(\cdot)h\right) + h'A_{1,n}(\cdot) - \frac{1}{2}h'J(\cdot)h \\ &\overset{\boldsymbol{\theta}_0}{\rightsquigarrow} h'A_1(\cdot) - \frac{1}{2}h'J(\cdot)h\end{aligned}$$

because the first two terms converge to zero in probability in $L_\infty(\Delta)$ by Lemma A.1 below. We establish that

$$\log \frac{p_{0,n,\boldsymbol{\theta}_0+h/\sqrt{n}}}{p_{0,n,\boldsymbol{\theta}_0}} \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} h'A_0 - \frac{1}{2}h'J_0h$$

by a similar argument, the only difference being that $\boldsymbol{\delta}_0$ is fixed so that A_0 is a random variable rather than a process.

We next establish the limiting distribution of $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)$ under $\boldsymbol{\theta}_0$. Since $(\boldsymbol{\delta}_n, c_n)$ converges in distribution and the log-likelihood ratios of both waves converge marginally in distribution under $\boldsymbol{\theta}_0$, these four random elements are jointly uniformly tight. By Prohorov's theorem (Van der

Vaart (2000) Theorem 18.12), there exists a subsequence along which

$$\left(\boldsymbol{\delta}_n, c_n, \log \frac{p_{0,n,\boldsymbol{\theta}_0+h/\sqrt{n}}}{p_{0,n,\boldsymbol{\theta}_0}}, \log \frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\cdot)}{p_{1,n,\boldsymbol{\theta}_0}(\cdot)}\right) \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} \left(\boldsymbol{\delta}, c, h'A_0 - \frac{1}{2}h'J_0h, h'A_1(\cdot) - \frac{1}{2}h'J(\cdot)h\right).$$

Along this subsequence, continuity of the sample paths of $A_1(\cdot)$ and uniform convergence of $A_{1,n}(\cdot)$ to $A_1(\cdot)$ implies that

$$\left(\boldsymbol{\delta}_n, c_n, \log \frac{p_{0,n,\boldsymbol{\theta}_0+h/\sqrt{n}}}{p_{0,n,\boldsymbol{\theta}_0}}, \log \frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta}_n)}{p_{1,n,\boldsymbol{\theta}_0}(\boldsymbol{\delta}_n)}\right) \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} \left(\boldsymbol{\delta}, c, h'A_0 - \frac{1}{2}h'J_0h, h'A_1(\boldsymbol{\delta}) - \frac{1}{2}h'J(\boldsymbol{\delta})h\right)$$

Since the log-likelihood ratio of the entire experiment is the sum of the log-likelihood ratios of the two waves, we have established joint convergence of $(\boldsymbol{\delta}_n, c_n)$ and the log-likelihood ratio along this subsequence.

We can now apply Le Cam's third lemma (Van der Vaart (2000) Theorem 6.6) to derive the limiting distribution of $(\boldsymbol{\delta}_n, c_n)$ along this subsequence under local alternatives. For any Borel set $B \subseteq (\Delta \times \mathbb{R}^k)$, the limiting distribution is given by

$$\mathbb{P}_h[(\boldsymbol{\delta}, c) \in B] = \mathbb{E}_{\boldsymbol{\theta}_0}\mathbb{1}[(\boldsymbol{\delta}, c) \in B] \exp\left(h'A_0 - \frac{1}{2}h'J_0h\right) \exp\left(h'A_1(\boldsymbol{\delta}) - \frac{1}{2}h'J(\boldsymbol{\delta})h\right).$$

Since the full sequence $(\boldsymbol{\delta}_n, c_n)$ also converges under local alternatives by assumption, the expression above is also the limiting distribution of the full sequence $(\boldsymbol{\delta}_n, c_n)$.

Next, we construct a statistic in the limit experiment with this distribution. The construction is similiar to the proof of Theorem 3 in Hirano and Porter (2023) and uses conditional vector quantile functions (Carlier, Chernozhukov, and Galichon 2016). Let A_0^h be a Gaussian random variable with mean J_0h and variance J_0. Let $A_1^h(\cdot)$ be a Gaussian process with mean $J(\cdot)h$ and the same covariance process as $A(\cdot)$. Let U_0, U_1 be uniform random variables independent of each other and of A_0^h, $A_1^h(\cdot)$. Note that U_0 and U_1 can be constructed from a single uniform random variable U.

Our goal is to construct $(\boldsymbol{\delta}^h, c^h)$ with distribution $\mathbb{P}_h$ under local alternatives, where $\boldsymbol{\delta}^h$ depends only on A_0^h and U_0 and where c^h depends only on A_0^h, $A_1^h(\cdot)$, and U_1. Let $q_{\boldsymbol{\delta}|A_0}(u \mid a_0)$ be the conditional vector quantile function of $\boldsymbol{\delta}$ given $A_0 = a_0$. Let $q_{c|A_0,\boldsymbol{\delta},A_1(\boldsymbol{\delta})}(u \mid a_0, \delta, a_1)$ be the

conditional vector quantile function of c given $A_0 = a_0$, $\boldsymbol{\delta} = \delta$, and $A_1(\delta) = a_1$. Define

$$\begin{aligned}\boldsymbol{\delta}^h &= q_{\boldsymbol{\delta}|A_0}(U_0 \mid A_0^h) \\ c^h &= q_{c|A_0,\boldsymbol{\delta},A_1(\boldsymbol{\delta})}(U_1 \mid A_0^h, \boldsymbol{\delta}^h, A_1^h(\boldsymbol{\delta}^h)).\end{aligned}$$

Then for $h = 0$, it follows from the definition of conditional vector quantile functions that

$$\left(\boldsymbol{\delta}^h, c^h\right) \sim (\boldsymbol{\delta}, c)$$

and therefore $(\boldsymbol{\delta}_n, c_n) \overset{\boldsymbol{\theta}_0}{\rightsquigarrow} (\boldsymbol{\delta}^0, c^0)$.

We now verify that the statistic $(\boldsymbol{\delta}^h, c^h)$ has the desired distribution under $h \neq 0$ as well. For some values of δ in the support of $\boldsymbol{\delta}^h$, $J(\delta)$ may not be invertible, and therefore $A_1^h(\delta)$ may not admit a density with respect to Lebesgue measure on $\mathbb{R}^\ell$. However, it admits a density with respect to Lebesgue measure on the support of $A_1^h(\delta)$, which is the affine subspace

$$\mathcal{A}(\delta) = \left\{ J(\delta)h + J(\delta)^{1/2}a : a \in \mathbb{R}^\ell \right\}.$$

This density is given by

$$\frac{1}{\sqrt{\det{}^-(2\pi J(\delta))}} \exp\left(-\frac{1}{2}(a - J(\delta)h)' J(\delta)^- (a - J(\delta)h)\right)$$

where $\det^-$ denotes the pseudo-determinant and $J(\delta)^-$ denotes the Moore-Penrose pseudo-inverse of $J(\delta)$.

Let B_1, B_2 be Borel sets in Δ and $\mathbb{R}^k$ respectively. With abuse of notation, let $J(a_0, U_0) = J(q_{\boldsymbol{\delta}|A_0}(U_0 \mid a_0))$ and $\mathcal{A}(a_0, U_0) = \mathcal{A}(q_{\boldsymbol{\delta}|A_0}(U_0 \mid a_0))$. Then

$$\begin{aligned}&\mathbb{P}_h[\boldsymbol{\delta}^h \in B_1, c^h \in B_2] \\ &= \mathbb{E}\left[\mathbb{1}[\boldsymbol{\delta}^h \in B_1, c^h \in B_2]\right] \\ &= \mathbb{E}\Bigg[\int_\Delta \int_{\mathcal{A}(a_0,U_0)} \mathbb{1}\left[q_{\boldsymbol{\delta}|A_0}(U_0 \mid a_0) \in B_1, q_{c|A_0,\boldsymbol{\delta},A_1(\boldsymbol{\delta})}(U_1 \mid a_0, U_0, a_1) \in B_2\right] \\ &\quad \times \frac{1}{\sqrt{\det(2\pi J_0)}} \exp\left(-\frac{1}{2}(a_0 - J_0 h)' J_0^{-1}(a_0 - J_0 h)\right)\end{aligned}$$

$$
\begin{aligned}
&\times \frac{1}{\sqrt{\det{}^{-}(2\pi J(\delta))}} \exp\left(-\frac{1}{2}(a_1 - J(a_0^h, U_0)h)' J(a_0^h, U_0)^{-}(a_1 - J(a_0^h, U_0)h)\right) da_1 da_0\Big] \\
&= \mathbb{E}\Big[\int_\Delta \int_{\mathcal{A}(a_0,U_0)} \mathbb{1}\left[q_{\boldsymbol{\delta}|A_0}(U_0 \mid a_0) \in B_1, q_{c|A_0,\boldsymbol{\delta},A_1(\boldsymbol{\delta})}(U_1 \mid a_0, U_0, a_1) \in B_2\right] \\
&\quad \times \exp\left(h'a_0 - \frac{1}{2}h'J_0 h\right) \exp\left(h'a_1 - \frac{1}{2}h'J(a_0,U_0)h\right) \\
&\quad \times \frac{1}{\sqrt{\det(2\pi J_0)}} \exp\left(-\frac{1}{2}a_0 J_0^{-1} a_0\right) \frac{1}{\sqrt{\det{}^{-}(2\pi J(a_0,U_0))}} \exp\left(-\frac{1}{2}a_1 J(a_0,U_0)^{-} a_1\right) da_1 da_0\Big] \\
&= \mathbb{E}\left[\mathbb{1}[(\boldsymbol{\delta}^0, c^0) \in B_1 \times B_2] \exp\left(h'A_0 - \frac{1}{2}h'J_0 h\right) \exp\left(h'A_1(\boldsymbol{\delta}^0) - \frac{1}{2}h'J(\boldsymbol{\delta}^0)h\right)\right]
\end{aligned}
$$

where the second equality uses the formula for the density of A_0^h and $A_1^h(\cdot)$, the third equality uses the fact that for any δ in the support of $\boldsymbol{\delta}^h$, $J(\delta)J(\delta)^{-}a = a$ for $a \in \mathcal{A}(\delta)$, and the fourth uses the formula for the density of A_0^0 and $A_1^0(\cdot)$. □

The following lemma ensures that the log-likelihood ratio of the experiment behaves asymptotically like a sample average. It generalizes Theorem 7.2 of Van der Vaart (2000) to ensure the approximation is uniform in $\boldsymbol{\delta}$.

Lemma A.1: *Suppose Θ is an open subset of $\mathbb{R}^\ell$ and that the model $\{p_{y|z,x}(y \mid z, x; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ satisfies Assumption 4.1. Also suppose Assumption 4.2 holds. Assume Δ is a compact subset of euclidean space. Then for any $\varepsilon > 0$,*

$$
\mathbb{P}\left(\sup_{\boldsymbol{\delta}\in\Delta}\left|\frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta})}{p_{1,n,\boldsymbol{\theta}_0}(\boldsymbol{\delta})} - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n h'\psi_i(\boldsymbol{\delta}) - \frac{1}{2}h'J(\boldsymbol{\delta})h\right)\right| > \varepsilon\right) \to 0
$$

Proof. We first note that the likelihood factors so that

$$
\begin{aligned}
\log \frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta})}{p_{1,n,\boldsymbol{\theta}_0}(\boldsymbol{\delta})} &= \log \prod_{i=n_0+1}^n \frac{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0 + h/\sqrt{n}) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})}{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta})} \\
&= \log \prod_{i=n_0+1}^n \frac{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0 + h/\sqrt{n})}{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0)} \\
&= \sum_{i=n_0+1}^n \log \frac{p_n(y_i \mid z_i, x_i)}{p(y_i \mid z_i, x_i)}
\end{aligned}
$$

where in the last line we have used the shorthand $p_n := p_{y|z,x}(\cdot \mid \cdot, \cdot; \boldsymbol{\theta}_0 + h/\sqrt{n})$ and $p := p_{y|z,x}(\cdot \mid \cdot, \cdot; \boldsymbol{\theta}_0)$.

Define the random variable

$$S_{ni}(\boldsymbol{\delta}) := 2\left(\sqrt{\frac{p_n}{p}}(y_i \mid z_i, x_i) - 1\right)$$

(recall y_i and z_i are functions of $\boldsymbol{\delta}$) which is well-defined $\mathbb{P}$-almost everywhere. We write a Taylor expansion of the log-likelihood ratio in terms of $S_{ni}(\boldsymbol{\delta})$:

$$\begin{aligned}
&\sum_{i=n_0+1}^{n} \log \frac{p_n(y_i \mid z_i, x_i)}{p(y_i \mid z_i, x_i)} \\
&= 2\sum_{i=n_0+1}^{n} \log\left(1 + \frac{1}{2}S_{ni}(\boldsymbol{\delta})\right) \\
&= \sum_{i=n_0+1}^{n}\left(S_{ni}(\boldsymbol{\delta}) - \frac{1}{4}\sum_i S_{ni}(\boldsymbol{\delta})^2 + \frac{1}{2}\sum_i S_{ni}(\boldsymbol{\delta})^2 r(S_{ni}(\boldsymbol{\delta}))\right) \qquad (10)
\end{aligned}$$

where $r(s) \to 0$ as $s \to 0$. We will evaluate each term of this expansion in turn to show that the expression is of the form claimed in the lemma.

We will start by evaluating the expectation of the first term. We have

$$\begin{aligned}
\mathbb{E}\left[\sum_i S_{ni}(\boldsymbol{\delta})\right] &= n\mathbb{E}\left[S_{ni}(\boldsymbol{\delta})\right] \\
&= n\mathbb{E}\left[\mathbb{E}\left[S_{ni}(\boldsymbol{\delta}) \mid z_i, x_i\right]\right] \\
&= 2n\mathbb{E}\left[\int \sqrt{p_n(y_i \mid z_i, x_i)}\sqrt{p(y_i \mid z_i, x_i)}d\lambda - 1\right] \\
&= -n\mathbb{E}\left[\int p_n(y_i \mid z_i, x_i)d\lambda\right. \\
&\qquad - 2\int \sqrt{p_n(y_i \mid z_i, x_i)}\sqrt{p(y_i \mid z_i, x_i)}d\lambda \\
&\qquad \left. + \int p(y_i \mid z_i, x_i)d\lambda\right] \\
&= -n\mathbb{E}\left[\int \left(\sqrt{p_n(y_i \mid z_i, x_i)} - \sqrt{p(y_i \mid z_i, x_i)}\right)^2 d\lambda\right] \\
&= -\frac{1}{4}\mathbb{E}\left[\mathbb{E}[(h'\psi_i(\boldsymbol{\delta}))^2 \mid z_i, x_i] + o(1)\right] \\
&= -\frac{1}{4}h'J(\boldsymbol{\delta})h + o(1)
\end{aligned}$$

where the second to last line follows from differentiability in quadratic mean (Assumption 4.1) since

$\|h/\sqrt{n}\|^2 = o(n^{-1})$. The last line follows from the fact that DQM is in fact uniform in z, x. We have that

$$\sup_{\boldsymbol{\delta}\in\Delta} \left| n\mathbb{E}\left[S_{ni}(\boldsymbol{\delta})\right] + \frac{1}{4}h'J(\boldsymbol{\delta})h \right| = o(1).$$

and, since the score is mean zero, we conclude that

$$\sup_{\boldsymbol{\delta}\in\Delta} \left| \mathbb{E}\left[\sum_i S_{ni}(\boldsymbol{\delta}) - \frac{1}{\sqrt{n}} \sum_i h'\psi_i(\boldsymbol{\delta}) + \frac{1}{4}h'J(\boldsymbol{\delta})h \right] \right| = o(1).$$

We now turn to the variance of the term inside this expectation. It is

$$\begin{aligned}
&\mathrm{Var}\left(\sum S_{ni}(\boldsymbol{\delta}) - 1/\sqrt{n} \sum h'\psi_i(\boldsymbol{\delta}) + \frac{1}{4}h'J(\boldsymbol{\delta})h \right) \\
&= \mathrm{Var}\left(\sum S_{ni}(\boldsymbol{\delta}) - 1/\sqrt{n} \sum h'\psi_i(\boldsymbol{\delta}) \right) \\
&= n\mathrm{Var}\left(S_{ni}(\boldsymbol{\delta}) - 1/\sqrt{n} h'\psi_i(\boldsymbol{\delta}) \right) \\
&\leq n\mathbb{E}\left((S_{ni}(\boldsymbol{\delta}) - 1/\sqrt{n} h'\psi_i(\boldsymbol{\delta}))^2 \right) \\
&= \mathbb{E}\left(\mathbb{E}\left[(\sqrt{n}S_{ni}(\boldsymbol{\delta}) - h'\psi_i(\boldsymbol{\delta}))^2 \mid z_i, x_i \right] \right) \\
&= \mathbb{E}\left(2\sqrt{n} \int \left(\sqrt{p_n} - \sqrt{p} - \frac{1}{2}h'\psi_i(\boldsymbol{\delta})\sqrt{p} \right)^2 d\lambda \right) \\
&= \mathbb{E}\left(2\sqrt{n} o(n^{-1}) \right)
\end{aligned} \tag{11}$$

where the last line is again by differentiability in quadratic mean. Again, since this is uniform in z, x and therefore in $\boldsymbol{\delta}$, we conclude that the variance of the sum is $o(1)$.

To summarize, we have shown that the mean and variance of

$$\sum S_{ni}(\boldsymbol{\delta}) - 1/\sqrt{n} \sum h'\psi_i(\boldsymbol{\delta}) + \frac{1}{4}h'J(\boldsymbol{\delta}_{nt})h$$

converge uniformly to zero in $\boldsymbol{\delta}$, and therefore we conclude that

$$\sup_{\boldsymbol{\delta}\in\Delta} \left| \sum_i S_{ni}(\boldsymbol{\delta}) - \left(\frac{1}{\sqrt{n}} \sum_i h'\psi_i(\boldsymbol{\delta}) - \frac{1}{4}h'J(\boldsymbol{\delta})h \right) \right| = o_p(1).$$

Now we turn to the second term in the Taylor expansion. Define

$$B_{ni} := \sqrt{n} S_{ni}(\boldsymbol{\delta}) - h'\psi_i(\boldsymbol{\delta})$$

and note that by (11), $\mathbb{E}B_{ni}^2 \to 0$. We have $nS_{ni}(\boldsymbol{\delta})^2 = (h'\psi_i(\boldsymbol{\delta}))^2 + 2h'\psi_i(\boldsymbol{\delta})B_{ni} + B_{ni}^2$ and therefore

$$\sum_i S_{ni}(\boldsymbol{\delta})^2 = \frac{1}{n}\sum_i (h'\psi_i(\boldsymbol{\delta}))^2 + \frac{1}{n}\sum_i h'\psi_i(\boldsymbol{\delta})B_{ni} + \frac{1}{n}\sum_i B_{ni}^2$$

where the second and third terms are $o_p(1)$ because $\mathbb{E}[B_{ni}^2] = o(1)$. Next,

$$\mathbb{E}\left(\frac{1}{n}\sum_i (h'\psi_i(\boldsymbol{\delta}))^2 - h'J(\boldsymbol{\delta})h\right) = 0$$
$$\mathrm{Var}\left(\frac{1}{n}\sum_i (h'\psi_i(\boldsymbol{\delta}))^2 - h'J(\boldsymbol{\delta})h\right) = o(1)$$

where the $o(1)$ term is uniform by the Glivenko-Cantelli theorem by Assumption 4.2. Hence

$$\sum_i S_{ni}(\boldsymbol{\delta})^2 = \frac{1}{4}h'J(\boldsymbol{\delta})h + o_p(1)$$

and so the second term in (10) is $-\frac{1}{4}h'J(\boldsymbol{\delta})h + o_p(1)$, where convergence in probability is in $L_\infty(\Delta)$.

For the last term, observe that

$$\sum_i S_{ni}(\boldsymbol{\delta})^2 r(S_{ni}(\boldsymbol{\delta})) \le \max_i |r(S_{ni}(\boldsymbol{\delta}))| \sum_i S_{ni}(\boldsymbol{\delta})^2$$

and

$$\begin{aligned}
\mathbb{P}(\max_i |S_{ni}(\boldsymbol{\delta})| > \epsilon) &\le n\mathbb{P}(|S_{ni}(\boldsymbol{\delta})| > \epsilon) \\
&\le n\mathbb{P}\left((h'\psi_i(\boldsymbol{\delta}))^2 > \frac{1}{2}n\epsilon^2\right) + n\mathbb{P}(B_{ni}^2 > \frac{1}{2}n\epsilon^2) \\
&= n\mathbb{P}\left((h'\psi_i(\boldsymbol{\delta}))^2 \mathbb{1}[(h'\psi_i(\boldsymbol{\delta}))^2 > \frac{1}{2}n\epsilon^2] > \frac{1}{2}n\epsilon^2\right) + n\mathbb{P}(B_{ni}^2 > \frac{1}{2}n\epsilon^2) \\
&\le 2\epsilon^{-2}\mathbb{E}\left[(h'\psi_i(\boldsymbol{\delta}))^2 \mathbb{1}[(h'\psi_i(\boldsymbol{\delta}))^2 > \frac{1}{2}n\epsilon^2]\right] + 2\epsilon^{-2}\mathbb{E}\left[B_{ni}^2\right] \\
&\le 2\epsilon^{-2}\mathbb{E}\left[(h'\psi_i(\boldsymbol{\delta}))^2\right] \mathbb{P}\left[(h'\psi_i(\boldsymbol{\delta}))^2 > \frac{1}{2}n\epsilon^2\right] + 2\epsilon^{-2}\mathbb{E}\left[B_{ni}^2\right]
\end{aligned}$$

$$\to 0$$

where, in the last line, the first term goes to zero because $\psi_i(\boldsymbol{\delta})$ is uniformly bounded in probability and the second term is again by (11). Hence, $\max_i |r(S_{nit})| \to 0$ in probability, and the last term in the Taylor expansion is $o_p(1)O_p(1) = o_p(1)$. □

We can now prove the main theorem of this section, which justifies the use of the Gaussian value function.

Theorem 4.7: *Suppose Assumptions 4.1, 4.2, 4.3, and 4.6 hold. Additionally assume that $\boldsymbol{\pi}_n^G$ is continuous in A_0 and A_1 and that J_0 is nonsingular. Then*

$$\mathbb{E}_{\boldsymbol{\delta}_n^*}\left[V_n(\{y_i, z_i\}_{i=1}^n)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^G}\left[V_n^G(\mu_1, \Sigma_1)\right] + o(n^{-1})$$

Proof. By Fatou's lemma, since regret is nonnegative,

$$\begin{aligned}
&\liminf_{n\to\infty} \int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] dQ\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}\right) \\
&\geq \int \liminf_{n\to\infty} \left(n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}\right)\right) dh.
\end{aligned}$$

Since q is positive and continuous in a neighborhood of $\boldsymbol{\theta}_0$, $q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}\right) \to q(\boldsymbol{\theta}_0)$ pointwise in h. Next, we derive the limit of the inner expectation.

There exists a subsequence n_j along which the lim inf is attained as a limit. Moreover, since by assumption $(\boldsymbol{\delta}_n^*, c_n^*)$ is bounded in probability, there exists a further subsequence n_{j_k} along which $(\boldsymbol{\delta}_n^*, c_n^*)$ converges in distribution. By Lemma 4.4, $(\boldsymbol{\delta}_n^*, c_n^*) \overset{h}{\rightsquigarrow} (\boldsymbol{\delta}^*, c^*)$ for some $(\boldsymbol{\delta}^*, c^*)$ in the limit experiment, where the limiting regret along n_{j_k} is the same as the limiting regret along n_j. For simplicity, convergence in what follows will be along this subsequence.

Since $c_n^* \overset{h}{\rightsquigarrow} c^*$ and welfare (and therefore regret) is continuous, the extended continuous mapping theorem implies

$$nR\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right) - nR\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c^*}{\sqrt{n}}\right) = o_p(1)$$

where we may represent c^* in the same probability space as c_n^* by the Skorohod representation

theorem.

The dominance condition on the regret of $\tilde{\boldsymbol{\pi}}_n^G$ ensures $R(\boldsymbol{\theta}, \tilde{\boldsymbol{\pi}}_n^G)$ is uniformly integrable. Since $\boldsymbol{\pi}_n^*$ is optimal in the finite-sample experiment, its regret is bounded by the regret of $\tilde{\boldsymbol{\pi}}_n^G$, and therefore its regret is also uniformly integrable. Therefore

$$n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] - n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c^*}{\sqrt{n}}\right)\right] = o(1)$$

by Theorem 2.20 of Van der Vaart (2000).

Furthermore, since $(\boldsymbol{\delta}_n^*, c_n^*)$ is a statistic in the limit experiment, it is feasible in the Gaussian problem, and therefore

$$n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c^*}{\sqrt{n}}\right)\right] \geq n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^G}{\sqrt{n}}\right)\right].$$

Combining the previous three displays, we have shown

$$\begin{aligned}
&\liminf_{n\to\infty} \int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] dQ(\boldsymbol{\theta}_0 + h/\sqrt{n}) \\
&\geq \int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^G}{\sqrt{n}}\right)\right] dh + o(1).
\end{aligned}$$

We now turn to the upper bound. Since regret is dominated by nonnegative, integrable $\bar{W}$, we can use Fatou's lemma in the other direction to get

$$\begin{aligned}
&\limsup_{n\to\infty} \int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] dQ\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}\right) \\
&\leq \int \limsup_{n\to\infty} n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] dh
\end{aligned}$$

Let $(\tilde{\boldsymbol{\delta}}_n^G, \tilde{c}_n^G)$ be the finite-sample analog of the optimal policy in the Gaussian problem, where we observe $A_{0,n}$ and $A_{1,n}$ instead of A_0 and A_1. Since $(\boldsymbol{\delta}_n^*, c_n^*)$ is optimal and $(\tilde{\boldsymbol{\delta}}_n^G, \tilde{c}_n^G)$ is feasible for the finite-sample problem,

$$n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^*}{\sqrt{n}}\right)\right] \leq n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\tilde{c}_n^G}{\sqrt{n}}\right)\right].$$

Since c_n^G is continuous in A_0 and A_1, and since welfare is continuous, by the continuous mapping

theorem we conclude

$$n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{\tilde{c}_n^G}{\sqrt{n}}\right)\right]-n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{c_n^G}{\sqrt{n}}\right)\right]\to 0$$

as before. As a result,

$$\begin{aligned}&\limsup_{n\to\infty}\int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{c_n^*}{\sqrt{n}}\right)\right]dQ\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}}\right)\\&\leq\int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{c_n^G}{\sqrt{n}}\right)\right]dh+o(1).\end{aligned}$$

Combining our upper and lower bounds and canceling the common recentering term in regret, we have

$$\begin{aligned}&\int n\mathbb{E}\left[W\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{c_n^*}{\sqrt{n}}\right)\right]dQ\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}}\right)\\&\quad-\int n\mathbb{E}\left[W\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{c_n^G}{\sqrt{n}}\right)\right]dh\\&=o(1)\end{aligned}$$

which is the desired result. □

A.2 Proof of Theorem 4.12

Lemma 4.11: *Suppose Assumptions 4.3, 4.9 and 4.10 hold, and that J_0 is nonsingular. Then*

$$\begin{aligned}V_n^G(\mu_1,\Sigma_1)-V_n^Q(D\mu_1)&=W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)+(\mu_1-\boldsymbol{\theta}_0)'\nabla_\theta W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)\\&\quad+\frac{1}{2}\Big(trace(\Sigma_1\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0))+(\mu_1-\boldsymbol{\theta}_0)'\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)(\mu_1-\boldsymbol{\theta}_0)\Big)\\&\quad+o_p(n^{-1})\end{aligned}$$

Proof. We first rewrite the value function V_n^G by introducing $\boldsymbol{\xi}\sim N(0,I)$ and letting $\vartheta=(\mu,\Sigma^{1/2})$ parameterize this problem. Define the objective function

$$r(\vartheta,\boldsymbol{\pi})=\mathbb{E}[W(\mu+\Sigma^{1/2}\boldsymbol{\xi},\boldsymbol{\pi})]$$

so that we can write the value function in terms of ϑ as

$$V_n^G(\vartheta) = \max_{\boldsymbol{\pi}} \quad r(\vartheta, \boldsymbol{\pi}) \quad \text{s.t.} \quad g(\boldsymbol{\pi}) \leq 0.$$

We will approximate this problem in a neighborhood of $\vartheta_0 = (\boldsymbol{\theta}_0, 0)$. By Theorem 3.1 of Shapiro (1985) (see Theorem A.6 below) we have that

$$V_n^G(\vartheta) - V_n^G(\vartheta_0) = (\vartheta - \vartheta_0)' \nabla_\vartheta r(\vartheta_0, \boldsymbol{\pi}_0) + \frac{1}{2} \zeta_n^G(\vartheta - \vartheta_0) + o(\|\vartheta - \vartheta_0\|^2)$$

where

$$\begin{aligned} \zeta_n^G(\vartheta - \vartheta_0) = \min_{\boldsymbol{\pi}} \quad & (\vartheta - \vartheta_0)' \nabla^2_{\vartheta\vartheta} r(\vartheta_0, \boldsymbol{\pi}_0)(\vartheta - \vartheta_0) \\ & + 2(\vartheta - \vartheta_0)' \nabla^2_{\vartheta\boldsymbol{\pi}} r(\vartheta_0, \boldsymbol{\pi}_0)(\boldsymbol{\pi} - \boldsymbol{\pi}_0) \\ & + (\boldsymbol{\pi} - \boldsymbol{\pi}_0)' [\nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}} r(\vartheta_0, \boldsymbol{\pi}_0) + \boldsymbol{\lambda}_0' \nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}} g(\boldsymbol{\pi}_0)](\boldsymbol{\pi} - \boldsymbol{\pi}_0) \\ & \text{s.t.} \\ & \boldsymbol{\pi}' \nabla g_j(\boldsymbol{\pi}_0) = 0, j \in \mathcal{J}_1 \\ & \boldsymbol{\pi}' \nabla g_j(\boldsymbol{\pi}_0) < 0, j \in \mathcal{J}_2 \end{aligned}$$

and derivatives with respect to $\Sigma^{1/2}$ are understood to be with respect to the vectorized elements of $\Sigma^{1/2}$.

We now calculate the terms appearing in ζ_n^G. To begin, the constant on the left-hand side is

$$V_n^G(\vartheta_0) = W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0).$$

Next, the first-order term in ϑ is

$$\begin{aligned} (\vartheta - \vartheta_0)' \nabla_\vartheta r(\vartheta_0, \boldsymbol{\pi}_0) &= \mathbb{E}[(\mu - \boldsymbol{\theta}_0 + \Sigma^{1/2} \boldsymbol{\xi})' \nabla_{\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)] \\ &= (\mu - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0). \end{aligned}$$

The cross-partial term is

$$\begin{aligned}(\vartheta-\vartheta_0)'\nabla^2_{\vartheta\boldsymbol{\pi}}r(\vartheta_0,\boldsymbol{\pi}_0)(\boldsymbol{\pi}-\boldsymbol{\pi}_0) &= \mathbb{E}[(\mu-\boldsymbol{\theta}_0+\Sigma^{1/2}\boldsymbol{\xi})'\nabla^2_{\boldsymbol{\theta}\boldsymbol{\pi}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)](\boldsymbol{\pi}-\boldsymbol{\pi}_0)\\ &= (\mu-\boldsymbol{\theta}_0)'D'(\boldsymbol{\pi}-\boldsymbol{\pi}_0)\end{aligned}$$

and the second-order term in ϑ is

$$\begin{aligned}(\vartheta-\vartheta_0)'\nabla^2_{\vartheta\vartheta}r(\vartheta_0,\boldsymbol{\pi}_0)(\vartheta-\vartheta_0) &= \mathbb{E}[(\mu-\boldsymbol{\theta}_0+\Sigma^{1/2}\boldsymbol{\xi})'\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)(\mu-\boldsymbol{\theta}_0+\Sigma^{1/2}\boldsymbol{\xi})]\\ &= \text{trace}(\Sigma\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0))+(\mu-\boldsymbol{\theta}_0)'\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)(\mu-\boldsymbol{\theta}_0)\end{aligned}$$

Finally, the second-order term in $\boldsymbol{\pi}$ is

$$(\boldsymbol{\pi}-\boldsymbol{\pi}_0)'[\nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}}r(\vartheta_0,\boldsymbol{\pi}_0)+\boldsymbol{\lambda}_0'\nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}}g(\boldsymbol{\pi}_0)](\boldsymbol{\pi}-\boldsymbol{\pi}_0) = (\boldsymbol{\pi}-\boldsymbol{\pi}_0)'H(\boldsymbol{\pi}-\boldsymbol{\pi}_0)$$

Putting these terms together, we have shown that

$$\begin{aligned}V_n^G(\vartheta)-\tilde{V}_n^Q(\vartheta) = {}& W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)+(\mu-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)\\ &+\frac{1}{2}\left(\text{trace}(\Sigma\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0))+(\mu-\boldsymbol{\theta}_0)'\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}}W(\boldsymbol{\theta}_0,\boldsymbol{\pi}_0)(\mu-\boldsymbol{\theta}_0)\right)\\ &+o(\|\vartheta-\vartheta_0\|^2)\end{aligned}$$

where

$$\begin{aligned}\tilde{V}_n^Q(\vartheta) = \max_{\boldsymbol{\pi}}\quad & (\boldsymbol{\pi}-\boldsymbol{\pi}_0)'D(\mu-\boldsymbol{\theta}_0)+\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\pi}_0)'[H+\boldsymbol{\lambda}_0'\nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}}g(\boldsymbol{\pi}_0)](\boldsymbol{\pi}-\boldsymbol{\pi}_0)\\ & \text{s.t.}\\ & \boldsymbol{\pi}'\nabla g_j(\boldsymbol{\pi}_0)=0, j\in\mathcal{J}_1\\ & \boldsymbol{\pi}'\nabla g_j(\boldsymbol{\pi}_0)<0, j\in\mathcal{J}_2.\end{aligned}$$

The program $\tilde{V}_n^Q$ is not quite the same as the program V_n^Q in the statement of the lemma. The differences are (i) that $\tilde{V}_n^Q$ uses a linear approximation to the nonlinear constraints $g(\boldsymbol{\pi})\leq 0$ and contains a nonlinear constraint term in the objective, (ii) that $\tilde{V}_n^Q$ imposes equality constraints for

indices j where strict complementarity holds in the original problem, and (iii) that $\tilde{V}_n^Q$ does not contain the term $(\boldsymbol{\pi} - \boldsymbol{\pi}_0)'C$ appearing in V_n^Q. However, another application of Theorem A.6 to V_n^Q shows that

$$V_n^Q(\vartheta) - \tilde{V}_n^Q(\vartheta) = o(\|\vartheta - \vartheta_0\|^2)$$

by the same arguments as above.

By Corollary 3.1 of Shapiro (1985), the sequence of functions

$$n^{-1}\left(V_n^G(\vartheta_0 + n^{-1/2}m_n) - V_n^Q(\vartheta_0 + n^{-1/2}m_n)\right)$$

converges to zero for every sequence $m_n \to m$. Note that

$$\begin{aligned}\mu_1 &= (n_0 J_0 + n_1 J(\boldsymbol{\delta}))^{-1}\left(n_0 J_0 \hat{\boldsymbol{\theta}}_0 + n_1 J(\boldsymbol{\delta})\hat{\boldsymbol{\theta}}_1\right)\\ \Sigma_1 &= (n_0 J_0 + n_1 J(\boldsymbol{\delta}))^{-1}\end{aligned}$$

so that $\mu_1 - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$ and $\Sigma_1^{1/2} = o(n^{-1/2})$. By the continuous mapping theorem we conclude that

$$V_n^G(\mu_1, \Sigma_1) - V_n^Q(D\mu_1) = o_p(n^{-1})$$

□

We now state the result from Shapiro (1985) used above. We first state the assumptions, which are weaker than those we actually made in Assumption 4.10. Consider the following nonlinear program

$$\begin{aligned} v(\vartheta) = \min_{\boldsymbol{\pi}} \quad & r(\vartheta, \boldsymbol{\pi}) \\ & \text{s.t.} \\ & g(\boldsymbol{\pi}) \leq 0 \end{aligned} \tag{12}$$

for some loss r parameterized by a Euclidean parameter ϑ. Let $\boldsymbol{\pi}$ be the optimal solution to (12)

under ϑ_0. Let $\boldsymbol{\lambda}_0 = (\lambda_{01}, \ldots, \lambda_{0m})$ be the optimal Lagrange multiplier corresponding to $\boldsymbol{\pi}_0$. Let $\mathcal{J}_1$ be the set of indices j such that $\lambda_{0j} > 0$. Let $\mathcal{J}_2$ be the set of indices j such that $g_j(\boldsymbol{\pi}_0) = 0$ and $\lambda_{0j} = 0$.

Assumption A.2: *There exists a number α and a compact set $S \subset \mathbb{R}^k$ such that $\alpha > v(\vartheta)$ and*

$$\{\boldsymbol{\pi} : g(\boldsymbol{\pi}) \leq 0,\ r(\vartheta, \boldsymbol{\pi}) \leq \alpha\} \subseteq S$$

for all ϑ in a neighborhood of ϑ_0.

Assumption A.3: *The optimal $\boldsymbol{\pi}$ under (ϑ_0), denoted $\boldsymbol{\pi}_0$, is unique.*

Assumption A.4: *The vectors*

$$\{\nabla g_j(\boldsymbol{\pi}_0) : j \in \mathcal{J}_1\}$$

are linearly independent, and there exists a vector $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi}' \nabla g_j(\boldsymbol{\pi}_0) = 0, j \in \mathcal{J}_1$$
$$\boldsymbol{\pi}' \nabla g_j(\boldsymbol{\pi}_0) < 0, j \in \mathcal{J}_2.$$

The previous assumption is implied by Assumption 4.10.3, that the rows of $\nabla g(\boldsymbol{\pi}_0)$ are linearly independent.

Assumption A.5: *Letting*

$$L(\vartheta, \boldsymbol{\pi}, \boldsymbol{\lambda}) = r(\vartheta, \boldsymbol{\pi}) + \boldsymbol{\lambda}' g(\boldsymbol{\pi})$$

be the Lagrangian, define

$$\begin{bmatrix} H_{\vartheta\vartheta} & H_{\vartheta\boldsymbol{\pi}} \\ H_{\boldsymbol{\pi}\vartheta} & H_{\boldsymbol{\pi}\boldsymbol{\pi}} \end{bmatrix} = \begin{bmatrix} \nabla^2_{\vartheta\vartheta} L(\vartheta_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0) & \nabla^2_{\vartheta\boldsymbol{\pi}} L(\vartheta_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0) \\ \nabla^2_{\boldsymbol{\pi}\vartheta} L(\vartheta_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0) & \nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}} L(\vartheta_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0) \end{bmatrix}$$

as the Hessian of the Lagrangian at the reference values. Then $c' H_{\boldsymbol{\pi}\boldsymbol{\pi}} c > 0$ for every nonzero vector

c such that

$$c'\nabla g_j(\boldsymbol{\pi}_0) \leq 0, j \in \mathcal{J}_1$$
$$c'\nabla g_j(\boldsymbol{\pi}_0) = 0, j \in \mathcal{J}_2.$$

Theorem A.6 (Shapiro (1985) Theorem 3.1)**:** *Suppose Assumptions A.2, A.3, A.4, and A.5 hold. Then*

$$v(\vartheta) - v(\vartheta_0) = (\vartheta - \vartheta_0)'\nabla_\vartheta r(\vartheta_0, \boldsymbol{\pi}_0) + \frac{1}{2}\zeta(\vartheta - \vartheta_0) + o(\|\vartheta - \vartheta_0\|^2)$$

where

$$\begin{aligned}\zeta(\vartheta - \vartheta_0) = \min_{\boldsymbol{\pi}} \quad & (\vartheta - \vartheta_0)'H_{\vartheta\vartheta}(\vartheta - \vartheta_0) + 2(\vartheta - \vartheta_0)'H_{\vartheta\boldsymbol{\pi}}(\boldsymbol{\pi} - \boldsymbol{\pi}_0) + (\boldsymbol{\pi} - \boldsymbol{\pi}_0)'H_{\boldsymbol{\pi}\boldsymbol{\pi}}(\boldsymbol{\pi} - \boldsymbol{\pi}_0) \\ & s.t. \\ & \boldsymbol{\pi}'\nabla g_j(\boldsymbol{\pi}_0) = 0, j \in \mathcal{J}_1 \\ & \boldsymbol{\pi}'\nabla g_j(\boldsymbol{\pi}_0) < 0, j \in \mathcal{J}_2.\end{aligned}$$

Moreover, let $\Phi(\vartheta)$ be the set of optimal solutions to $\zeta(\vartheta - \vartheta_0)$ and let $\boldsymbol{\pi}(\vartheta)$ be an optimal solution to (12). Then

$$\lim_{\vartheta \to \vartheta_0} \frac{dist(\boldsymbol{\pi}(\vartheta) - \boldsymbol{\pi}_0, \Phi(\vartheta - \vartheta_0))}{\|\vartheta - \vartheta_0\|} = 0.$$

Corollary A.7 (Shapiro (1985) Corollary 3.1)**:** *Under the assumptions of Theorem A.6,*

$$\lim_{t \downarrow 0, m \to m_0} t^{-2}\left(v(\vartheta_0 + tm) - v(\vartheta_0) - tm'\nabla_\vartheta r(\vartheta_0, \boldsymbol{\pi}_0)\right)$$

exists and is equal to $\frac{1}{2}\zeta(m_0)$ for every m_0.

We now prove the main result of this section, which justifies the quadratic approximation to the value function.

Theorem 4.12: *Suppose Assumptions 4.3, 4.6, 4.9, and 4.10 hold. Assume there exists a function*

$\bar{W}(\boldsymbol{\theta})$ *such that* $R(\boldsymbol{\theta}, \tilde{\boldsymbol{\pi}}_n^Q) \leq n^{-1}\bar{W}(\boldsymbol{\theta})$ *and* $\int |\bar{W}(\boldsymbol{\theta})|^{1+\iota} d\boldsymbol{\theta} < \infty$ *for some* $\iota > 0$. *Then*

$$\mathbb{E}_{\boldsymbol{\delta}_n^G}\left[V_n^G(\mu_1, \Sigma_1)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

where $M(\mu_0, \Sigma_0) = W(\mu_0, \boldsymbol{\pi}_0) + \frac{1}{2} trace(\Sigma_0 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\mu_0, \boldsymbol{\pi}_0))$. *Further,*

$$\mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1)\right] = \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[\hat{V}_n^Q(D\mu_1)\right] + o(n^{-1})$$

Proof. We start with the first claim. By Lemma 4.11, we have that

$$\begin{aligned} V_n^G(\mu_1, \Sigma_1) - V_n^Q(D\mu_1) = W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) + (\mu_1 - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \\ + \frac{1}{2}\left(\text{trace}(\Sigma_1 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)) + (\mu_1 - \boldsymbol{\theta}_0)' \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)(\mu_1 - \boldsymbol{\theta}_0)\right) \\ + o_p(n^{-1}) \end{aligned}$$

Note that regardless of the chosen design,

$$\begin{aligned} &\mathbb{E}\left[\text{trace}(\Sigma_1 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)) + (\mu_1 - \boldsymbol{\theta}_0)' \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)(\mu_1 - \boldsymbol{\theta}_0) \mid \hat{\boldsymbol{\theta}}_0\right] \\ &= \text{trace}(\Sigma_0 \nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)) \end{aligned}$$

and

$$\mathbb{E}\left[(\mu_1 - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} W(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \mid \hat{\boldsymbol{\theta}}_0\right] = 0.$$

Since the regret of $(\boldsymbol{\pi}_n^G)$ is uniformly integrable, for any fixed h we have that

$$\mathbb{E}\left[nR\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^G}{\sqrt{n}}\right)\right] - \mathbb{E}\left[nR^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right)\right] = o(1)$$

where R^Q is the regret under the welfare function W^Q.

Then we have

$$\lim_{n\to\infty} \int n\mathbb{E}\left[R\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^G}{\sqrt{n}}\right)\right] dh$$

$$
\begin{aligned}
&= \int \lim_{n\to\infty} \mathbb{E}\left[nR\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^G}{\sqrt{n}}\right)\right] dh \\
&= \int \lim_{n\to\infty} \mathbb{E}\left[nR^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right)\right] dh \\
&= \lim_{n\to\infty} \int n\mathbb{E}\left[R^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right)\right] dh
\end{aligned}
$$

where the first line is by the dominated convergence theorem, the second by Lemma 4.11, the third by optimality of $\boldsymbol{\delta}_n^Q$ for V_n^Q, and the fourth again by the dominated convergence theorem. Subtracting the first and last lines and canceling the centering term in regret, we have

$$
\lim_{n\to\infty}\left(\int \left(n\mathbb{E}_{\boldsymbol{\delta}_n^G}\left[V_n^G(\mu_1,\Sigma_1)\right] - n\mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1) + M(\mu_0,\Sigma_0)\right]\right) dh\right) = 0
$$

and the first claim is proved.

We now move on to the second claim. Conditional on h and scaled by n, the difference in welfare is

$$
\begin{aligned}
&nW^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) - n\hat{W}^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right) \\
&\geq nW^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right) - n\hat{W}^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right)
\end{aligned}
$$

where the inequality follows because c_n^Q is optimal for the loss W^Q over a feasible set which includes $\hat{c}_n^Q$. Since $\hat{W}^Q$ is constructed from the estimates $(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\pi}}_0)$ which are consistent, by the extended continuous mapping theorem this difference is $o_p(1)$. Since regret is uniformly integrable, the expectation is $o(1)$.

For the other direction, we proceed similarly. First, we can write the difference in welfare as

$$
\begin{aligned}
&nW^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) - n\hat{W}^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right) \\
&\leq nW^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) - n\hat{W}^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right)
\end{aligned}
$$

because $\hat{c}_n^Q$ is optimal for the loss $\hat{W}^Q$ over a feasible set which includes c_n^Q. Again, since $(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\pi}}_0) \xrightarrow{p} (\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, it follows that the difference in welfare is $o_p(1)$ and by uniform integrability, the expectation

is $o(1)$. □

A.3 Proof of Theorem 4.13

Theorem 4.13: *Maintain the assumptions of Theorems 4.7 and 4.12. Then V_n^Q provides an asymptotic upper bound on the welfare of any sequence of designs and policies in the finite-sample experiment. That is, if $(\boldsymbol{\delta}_n, \boldsymbol{\pi}_n)$ is a sequence of feasible designs and policies in the finite-sample experiment with $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)$, then*

$$\limsup_{n\to\infty} \mathbb{E}_{\boldsymbol{\delta}_n}\left[W\left(\boldsymbol{\theta}, \boldsymbol{\pi}_n\right)\right] \leq \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[W_n^Q\left(D\boldsymbol{\theta}, \boldsymbol{\pi}_n^Q\right) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

where $M(\mu_0, \Sigma_0)$ is as in Theorem 4.12. Moreover, this upper bound is attained by solving $\hat{V}_n^Q$, using $\hat{\boldsymbol{\delta}}_n^Q$ in the main wave and then solving resulting finite-sample policy choice problem:

$$\lim_{n\to\infty} \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right] = \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[W_n^Q\left(D\boldsymbol{\theta}, c_\infty\right) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})$$

i.e. the design $\hat{\boldsymbol{\delta}}_n^Q$ is asymptotically optimal.

Proof. For the first part, we have

$$\begin{aligned}\mathbb{E}\left[nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n}{\sqrt{n}}\right)\right] &\leq \mathbb{E}_{\boldsymbol{\delta}_n^*}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right] \\ &= \mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1) + M(\mu_0, \Sigma_0)\right] + o(n^{-1})\end{aligned}$$

where the first inequality follows from the optimality of $(\boldsymbol{\delta}_n^*, c_n^*)$ and the equality from Theorems 4.7 and 4.12. For the second part, let $(\tilde{\boldsymbol{\delta}}_n^Q, \tilde{c}_n^Q)$ be the finite-sample analogs of $(\hat{\boldsymbol{\delta}}_n^Q, \hat{c}_n^Q)$, where A_0 and A_1 are replaced by $A_{0,n}$ and $A_{1,n}$. We have

$$\mathbb{E}_{\tilde{\boldsymbol{\delta}}_n^Q}\left[V_n(\{y_i, z_i, x_i\}_{i=1}^n)\right] \geq \mathbb{E}\left[W\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\tilde{c}_n^Q}{\sqrt{n}}\right)\right].$$

Since V_n^G is continuous and regret is uniformly integrable,

$$\mathbb{E}\left[nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\tilde{c}_n^Q}{\sqrt{n}}\right)\right] - \mathbb{E}\left[nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right)\right] \to 0$$

and

$$\mathbb{E}\left[nW\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{\hat{c}_n^Q}{\sqrt{n}}\right)\right]-\mathbb{E}\left[n\hat{W}^Q\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{\hat{c}_n^Q}{\sqrt{n}}\right)+nM(\mu_0,\Sigma_0)\right]=o(1)$$

by the same argument as in the proof of Theorem 4.12. Then by Theorem 4.12,

$$\mathbb{E}\left[n\hat{W}^Q\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{\hat{c}_n^Q}{\sqrt{n}}\right)\right]-\mathbb{E}\left[nW^Q\left(\boldsymbol{\theta}_0+\frac{h}{\sqrt{n}},\boldsymbol{\pi}_0+\frac{\hat{c}_n^Q}{\sqrt{n}}\right)\right]=o(1)$$

and we conclude that

$$\mathbb{E}_{\tilde{\boldsymbol{\delta}}_n^Q}\left[V_n(\{y_i,z_i,x_i\}_{i=1}^n)\right]\geq\mathbb{E}_{\boldsymbol{\delta}^Q}\left[V_n^Q(D\mu_1)+M(\mu_0,\Sigma_0)\right]+o(n^{-1})$$

□

B Stochastic equicontinuity

Here we give sufficient conditions for Assumption 4.2. First, we give a definition of stochastic equicontinuity.

Definition B.1 (Stochastic equicontinuity)**:** *Let $\{A_n(\cdot),n\geq 1\}$ be a sequence of stochastic processes in a normed metric space Δ. We say $\{A_n(\cdot),n\geq 1\}$ is stochastically equicontinuous if for every $\varepsilon>0$ and $\eta>0$ there exists a $\iota>0$ such that*

$$\limsup_{n\to\infty}\mathbb{P}\left(\sup_{b,c\in\Delta:\|b-c\|<\iota}\|A_n(b)-A_n(c)\|>\eta\right)<\varepsilon.$$

Using the generative representation of Assumption 4.2, y_i and z_i are deterministic functions of x_i, ν_i, and ϵ_i for given $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$. Let the joint distribution of x_i, ν_i, and ϵ_i (equivalently x_i, z_i, and y_i) be denoted by $\mathbb{P}$. Expectations will be taken with respect to $\mathbb{P}$.

In what follows, we will refer to the following function classes:

$$\mathcal{F}_z=\{z(\cdot,\cdot;\boldsymbol{\delta}):\boldsymbol{\delta}\in\Delta\}$$

$$\mathcal{F}_\psi=\{\psi(y(z(\cdot,\cdot;\boldsymbol{\delta}),\cdot,\cdot;\boldsymbol{\theta}_0)\mid z(\cdot,\cdot;\boldsymbol{\delta})):\boldsymbol{\delta}\in\Delta\}$$

where

$$z_i = z(x_i, \nu_i; \boldsymbol{\delta})$$
$$y_i = y(z_i, x_i, \epsilon_i; \boldsymbol{\theta})$$

describe the data-generating process as in Assumption 4.2. For each $j \in \{1, \ldots, J\}$, let $\mathcal{F}_z^j$ be the set of functions returning the jth element of $z \in \mathcal{F}_z$.

For a real-valued function f, an (L_2) η-bracket is a pair of functions (f_j^m, f_j^e) (mnemonic: mean and error) such that $|f - f_j^m| \leq f_j^e$ and $\mathbb{E}[(f_j^e)^2]^{1/2} \leq \eta$[2]. For a class of real-valued functions $\mathcal{F}$, the bracketing number $N_{[]}(\eta, \mathcal{F}, L_2(\mathbb{P}))$ is the smallest number of η-brackets needed to cover $\mathcal{F}$. A set of brackets which covers $\mathcal{F}$ is called an η-bracketing set.

The bracketing entropy integral of a function class $\mathcal{F}$ is defined as

$$\mathcal{J}_{[]}(\mathcal{F}, L_2(\mathbb{P})) = \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}, L_2(\mathbb{P}))} d\eta.$$

B.1 Bracketing conditions for stochastic equicontinuity

Our first lemma controls the the bracketing number of the score under a simplicity condition on the treatment assignment and a smoothness condition on the score function.

Lemma B.2: *Suppose $N_{[]}(\eta, \mathcal{F}_z^j, L_2(\mathbb{P})) \leq N < \infty$ $j = 1, \ldots J$. Suppose that for every $z \in \mathcal{Z}$ and $\iota > 0$, there exists a constant C such that*

$$\left(\mathbb{E} \sup_{\tilde{z}: \|\tilde{z} - z\| \leq \iota} \left| \psi(y(z, x, \epsilon; \boldsymbol{\theta}_0) \mid z, x) - \psi(y(\tilde{z}, x, \epsilon; \boldsymbol{\theta}_0) \mid \tilde{z}, x) \right|^2 \right)^{1/2} \leq C\iota \tag{13}$$

Then $N_{[]}(JC\eta, \mathcal{F}_\psi, L_2(\mathbb{P})) \leq N^J$.

Condition (13) is a type of Lipschitz continuity condition on the score function. It is weaker than a typical Lipschitz condition in that it allows for some discontinuities, such as indicator functions of the form $\mathbb{1}[g(z) \leq \epsilon]$ which are common in economic choice models. Andrews (1994) calls this a type IV class (equation 5.3) and shows how to control its bracketing entropy and gives other

[2]This is the definition used by Andrews (1994), which differs slightly from that of Van der Vaart and Wellner (2013).

examples of functions satisfying this condition. Moreover, products and sums of these types of functions also have finite entropy integrals. Later, we will verify that this condition holds for the model of Section 5.

Proof. For each $j \in \{1, \ldots, J\}$, let $\{z^m_{s,j}, z^e_{s,j}\}_{s=1}^N$ be an η-bracketing set for $\mathcal{F}_z^j$. Construct $\{z^m_s, z^e_s\}_{s=1}^{N^J}$ by taking combinations of element-wise η-bracketing sets so that for any $z(\cdot,\cdot;\boldsymbol{\delta}) \in \mathcal{F}_z$, there is some s such that $|z - z^m_s| \leq z^e_s$ element-wise and $\mathbb{E}[(z^e_s)^2]^{1/2} \leq \eta$ element-wise.

Define $\{\psi^m_s, \psi^e_s\}_{s=1}^{N^J}$ by

$$\psi^m_s(x,\nu,\epsilon) = \psi(y(z^m_s(\nu), x, \epsilon; \boldsymbol{\theta}_0) \mid z^m_s(\nu), x)$$
$$\psi^e_s(x,\nu,\epsilon) = \sup_{\tilde{z}:\|\tilde{z} - z^m_s(x,\nu)\| \leq \|z^e_s(x,\nu)\|} \left|\psi(y(\tilde{z}, x, \epsilon; \boldsymbol{\theta}_0) \mid \tilde{z}, x) - \psi(y(z^m_s(\nu), x, \epsilon; \boldsymbol{\theta}_0) \mid z^m_s(\nu), x)\right|.$$

Let f be an arbitrary element of $\mathcal{F}_\psi$. Then f is of the form

$$f(x,\nu,\epsilon;\boldsymbol{\delta}) = \psi(y(z(x,\nu;\boldsymbol{\delta}), x, \epsilon; \boldsymbol{\theta}_0) \mid z(x,\nu;\boldsymbol{\delta}))$$

for some $z(\cdot,\cdot;\boldsymbol{\delta}) \in \mathcal{F}_z$. Let s be the index of the bracket containing $z(\cdot,\cdot;\boldsymbol{\delta})$. Then $|z(x,\nu;\boldsymbol{\delta}) - z^m_s(x,\nu)| \leq z^e_s(x,\nu)$ element-wise. This implies $\|z(x,\nu;\boldsymbol{\delta}) - z^m_s(x,\nu)\| \leq \|z^e_s(x,\nu)\|$ and therefore

$$\begin{aligned}|f(x,\nu,\epsilon;\boldsymbol{\delta}) - \psi^m_s(x,\nu,\epsilon)| &= \left|\psi(y(z(x,\nu;\boldsymbol{\delta}), x, \epsilon; \boldsymbol{\theta}_0) \mid z(x,\nu;\boldsymbol{\delta})) - \psi(y(z^m_s(x,\nu), x, \epsilon, \boldsymbol{\theta}_0) \mid z^m_s(x,\nu), x)\right| \\ &\leq \psi^e_s(x,\nu,\epsilon)\end{aligned}$$

where the inequality follows from the definition of ψ^e_s. Further,

$$\begin{aligned}&\mathbb{E}[|\psi^e_s(x,\nu,\epsilon)|^2]^{1/2} \\ &= \mathbb{E}\left[\sup_{\tilde{z}:\|\tilde{z} - z^m_s(x,\nu)\| \leq \|z^e_s(x,\nu)\|} \left|\psi(y(z^m_s(x,\nu), x, \epsilon; \boldsymbol{\theta}_0) \mid z^m_s(x,\nu)) - \psi(y(\tilde{z}, x, \epsilon; \boldsymbol{\theta}_0) \mid \tilde{z}, x)\right|^2\right]^{1/2} \\ &\leq \mathbb{E}\left[C^2\|z^e_s(x,\nu)\|^2\right]^{1/2} \\ &= C\left(\sum_{j=1}^J \mathbb{E}|z^e_{s,j}(\nu)|^2\right)^{1/2}\end{aligned}$$

$$\leq C \sum_{j=1}^{J} \left(\mathbb{E}|z^e_{s,j}(\nu)|^2 \right)^{1/2}$$

$$\leq CJ\eta$$

where the second line follows from the definition of ψ^e_s, the third follows from the Lipschitz condition (13), the fourth is a simple algebraic manipulation, the fifth follows from convexity of the square root function, and the last follows from the fact that $\{z^e_{s,j}\}_{s=1}^N$ is an η-bracketing set for $\mathcal{F}^j_z$. Therefore, $\{\psi^m_s, \psi^e_s\}_{s=1}^{N^K}$ is a $JC\eta$-bracketing of $\mathcal{F}\psi$. □

We can now show that $\mathcal{J}_{[]}(\mathcal{F}_\psi, L_2(\mathbb{P}))$ is finite and therefore $A_n(\cdot)$ is stochastically equicontinuous.

Theorem B.3: *Suppose $\mathcal{J}_{[]}(\mathcal{F}^j_z, L_2(\mathbb{P}))$ is finite for each j. Moreover, suppose ψ satisfies the Lipschitz condition (13). Then $J_{[]}(\mathcal{F}_\psi, L_2(\mathbb{P}))$ is finite. Suppose further that $\mathcal{F}_z$ has envelope F_z and $\mathbb{E}F_z^{2+\iota} < \infty$ for some $\iota > 0$. Then $A_n(\cdot)$ is stochastically equicontinuous.*

Proof. The first part is straightforward consequence of Lemma B.2. Specifically,

$$\begin{aligned}
\mathcal{J}_{[]}(\mathcal{F}, L_2(\mathbb{P})) &= \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}, L_2(\mathbb{P}))} d\eta \\
&\leq \int_0^\infty \sqrt{\log \prod_{j=1}^{J} N_{[]}\left(\frac{\eta}{JC}, \mathcal{F}^j_z, L_2(\mathbb{P})\right)} d\eta \\
&= \int_0^\infty \sqrt{\sum_{j=1}^{J} \log N_{[]}(\eta, \mathcal{F}^j_z, L_2(\mathbb{P}))} d\eta \\
&\leq \sum_{j=1}^{J} \mathcal{J}_{[]}(\mathcal{F}^j_z, L_2(\mathbb{P})) \\
&< \infty.
\end{aligned}$$

The second part of the proposition follows from Andrews (1994) Theorem 4. □

B.2 Verifying conditions for stochastic equicontinuity

While the bracketing conditions of Theorem B.3 are lower-level than stochastic equicontinuity, they are not immediately interpretable. Here we show that simple policy classes such as those used in

the Progresa application satisfy these conditions.

Proposition B.4: *Suppose each element of $z(\cdot,\cdot;\boldsymbol{\delta})$ is of the form*

$$z(x,\nu,\boldsymbol{\delta}) = \delta_1'x + \delta_2'x \times \mathbb{1}[\rho(\delta_3'x) \geq \nu]$$

where $\boldsymbol{\delta}_j$ are subvectors of $\boldsymbol{\delta}$, and $\rho : \mathbb{R} \mapsto [0,1]$ is a strictly increasing Lipschitz-continuous function, and ν has a uniform distribution on $[0,1]$ independent of x. If $\mathbb{E}\|x\|^4 < \infty$ and Δ is a bounded subset of Euclidean space, then $\mathcal{F}_z$ has finite bracketing entropy.

Proof. We first establish that the simple functions used to construct z satisfy an L_4 Lipschitz condition similar to (13). Let $\mathcal{F}_1$ be the class of functions of the form $z(x;\boldsymbol{\delta}) = \boldsymbol{\delta}'x$. Then for $z \in \mathcal{F}_1$,

$$\begin{aligned}
\mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \left| z(x;\boldsymbol{\delta}) - z(x;\tilde{\boldsymbol{\delta}}) \right|^4 &= \mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \left| \boldsymbol{\delta}'x - \tilde{\boldsymbol{\delta}}'x \right|^4 \\
&\leq \mathbb{E}(\iota^2\|x\|^2)^2 \\
&= \iota^4 \mathbb{E}\|x\|^4
\end{aligned}$$

and therefore by Theorem 5 of Andrews (1994), $\mathcal{F}_1$ has finite L_4 bracketing entropy.

Now consider the class of functions $\mathcal{F}_2$ of the form $z(x,\nu;\boldsymbol{\delta}) = \mathbb{1}[\rho(\boldsymbol{\delta}'x) \geq \nu]$. Without loss of generality we will let ρ have Lipschitz constant 1. Then for $z \in \mathcal{F}_2$,

$$\begin{aligned}
&\mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \left| z(x,\nu;\boldsymbol{\delta}) - z(x,\nu;\tilde{\boldsymbol{\delta}}) \right|^4 \\
&= \mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \left| \mathbb{1}[\rho(\boldsymbol{\delta}'x) \geq \nu] - \mathbb{1}[\rho(\tilde{\boldsymbol{\delta}}'x) \geq \nu] \right|^4 \\
&= \mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \left| \mathbb{1}[\rho(\boldsymbol{\delta}'x) \geq \nu > \rho(\tilde{\boldsymbol{\delta}}'x)] + \mathbb{1}[\rho(\tilde{\boldsymbol{\delta}}'x) \geq \nu > \rho(\boldsymbol{\delta}'x)] \right|^4 \\
&= \mathbb{E} \sup_{\tilde{\boldsymbol{\delta}}:\|\tilde{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\leq\iota} \mathbb{1}[\rho(\boldsymbol{\delta}'x) \geq \nu > \rho(\tilde{\boldsymbol{\delta}}'x)] + \mathbb{1}[\rho(\tilde{\boldsymbol{\delta}}'x) \geq \nu > \rho(\boldsymbol{\delta}'x)] \\
&\leq \mathbb{E} \sup_{\tilde{\rho}:\|\tilde{\rho}-\rho(\boldsymbol{\delta}'x)\|\leq\iota\|x\|} \mathbb{1}[\rho(\boldsymbol{\delta}'x) \geq \nu > \tilde{\rho}] + \mathbb{1}[\tilde{\rho} \geq \nu > \rho(\boldsymbol{\delta}'x)] \\
&\leq \mathbb{E}\mathbb{1}[|\rho(\boldsymbol{\delta}'x) - \nu| \leq \iota\|x\|]
\end{aligned}$$

$$= \mathbb{E}[\mathbb{P}(|\rho(\boldsymbol{\delta}'x) - \nu| \leq \iota\|x\| \mid x)]$$
$$\leq \mathbb{E}[2\iota\|x\|]$$

and therefore by Theorem 5 of Andrews (1994), $\mathcal{F}_2$ has finite L_4 bracketing entropy.

We may now follow Theorem 6 of Andrews (1994) to show that the form of z under consideration has finite L_2 bracketing entropy. Having established that these simple functions have finite L_4 bracketing entropy, we first show that the product of these functions has finite L_2 bracketing entropy.

Suppose $z(x, \nu; \boldsymbol{\delta}) = \delta_1'x \times \mathbb{1}[\rho(\delta_2'x) \geq \nu]$. Let z_1^m, z_1^e and z_2^m, z_2^e be η-brackets for $\delta_1'x$ and $\mathbb{1}[\rho(\delta_2'x) \geq \nu]$ respectively. Then

$$\begin{aligned}
|z - z_1^m z_2^m| &\leq |\delta_1'x\mathbb{1}[\rho(\delta_2'x) \geq \nu] - z_1^m \mathbb{1}[\rho(\delta_2'x) \geq \nu]| \\
&\quad + |z_1^m \mathbb{1}[\rho(\delta_2'x) \geq \nu] - z_1^m z_2^m| \\
&\leq |z_1^e|F_2 + F_1|z_2^e|
\end{aligned}$$

where F_1 and F_2 are L_4-integrable envelopes for $\mathcal{F}_1$ and $\mathcal{F}_2$ (which exist becase Δ is bounded and the indicator is bounded by 1). Therefore

$$\begin{aligned}
\mathbb{E}[|z - z_1^m z_2^m|^2]^{1/2} &\leq \mathbb{E}[|z_1^e|^2 F_2^2]^{1/2} + \mathbb{E}[F_1^2|z_2^e|^2]^{1/2} \\
&\leq \mathbb{E}[|z_1^e|^4]^{1/4}\mathbb{E}[F_2^4]^{1/4} + \mathbb{E}[F_1^4]^{1/4}\mathbb{E}[|z_2^e|^4]^{1/4} \\
&\leq 2C\eta
\end{aligned}$$

for some constant C. In particular, this means that a $2C\eta$-bracketing set for the product of functions in $\mathcal{F}_1$ and $\mathcal{F}_2$ can be constructed by taking combinations of η-brackets for functions in $\mathcal{F}_1$ and $\mathcal{F}_2$. That is,

$$N_{[]}(2C\eta, \mathcal{F}_1 \times \mathcal{F}_2, L_2(P)) \leq N_{[]}(\eta, \mathcal{F}_1, L_2(P))N_{[]}(\eta, \mathcal{F}_2, L_2(P)).$$

Next, we show that sums of functions with finite bracketing entropy have finite bracketing entropy. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be classes of functions with finite bracketing entropy. Let f_1 and f_2 be

functions in $\mathcal{G}_1$ and $\mathcal{G}_2$. If f_1^m, f_1^e and f_2^m, f_2^e are η-brackets for f_1 and f_2 respectively, then

$$|f_1 + f_2 - f_1^m - f_2^m| \leq f_1^e + f_2^e$$

and

$$\mathbb{E}[(f_1^e + f_2^e)^2]^{1/2} \leq \mathbb{E}[(f_1^e)^2]^{1/2} + \mathbb{E}[(f_2^e)^2]^{1/2} \leq 2\eta$$

so

$$N_{[]}(2\eta, \mathcal{F}_1 + \mathcal{F}_2, L_2(P)) \leq N_{[]}(\eta, \mathcal{F}_1, L_2(P)) N_{[]}(\eta, \mathcal{F}_2, L_2(P)).$$

Together, we have

$$\begin{aligned}
\mathcal{J}_{[]}(\mathcal{F}_1 + \mathcal{F}_1 \times \mathcal{F}_2, L_2(P)) &= \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}_1 + \mathcal{F}_1 \times \mathcal{F}_2, L_2(P))} d\eta \\
&= \int_0^\infty \sqrt{\log N_{[]}(2\eta, \mathcal{F}_1 + \mathcal{F}_1 \times \mathcal{F}_2, L_2(P))} d\eta \\
&\leq \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}_1, L_2(P)) + \log N_{[]}(\eta, \mathcal{F}_1 \times \mathcal{F}_2, L_2(P))} d\eta \\
&\leq \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}_1, L_2(P)) + \log N_{[]}(\eta, \mathcal{F}_1 \times \mathcal{F}_2, L_2(P))} d\eta \\
&\leq \mathcal{J}_{[]}(\mathcal{F}_1, L_2(P)) + \int_0^\infty \sqrt{\log N_{[]}(2C\eta, \mathcal{F}_1 \times \mathcal{F}_2, L_2(P))} d\eta \\
&\leq \mathcal{J}_{[]}(\mathcal{F}_1, L_2(P)) + \int_0^\infty \sqrt{\log N_{[]}(\eta, \mathcal{F}_1, L_2(P)) + \log N_{[]}(\eta, \mathcal{F}_2, L_2(P))} d\eta \\
&\leq 2\mathcal{J}_{[]}(\mathcal{F}_1, L_2(P)) + \mathcal{J}_{[]}(\mathcal{F}_2, L_2(P)) \\
&< \infty
\end{aligned}$$

and any z of the form given in the proposition is in $\mathcal{F}_1 + \mathcal{F}_1 \times \mathcal{F}_2$. □

The class of policies described in Proposition B.4 contains the Progresa treatment function as a special case, which we now recall.

Example B.5 (Progresa treatment)**:** In our school subsidy application, the subsidy is of the form

$$z(x, \nu; \boldsymbol{\delta}) = \delta_1' x \times \mathbb{1}[\delta_2 \geq \nu]$$

which is a special case of this proposition. The proposition also allows for the "control" group to recieve a baseline subsidy amount which differs from that of the treatment group and for the probability of treatment to depend on covariates. Moreover, the distinction between x and ν is not important for this proposition, allowing for additive random variation in the subsidy value. ◊

By Theorem B.3, it remains to verify that the score of the model used in Section 5 satisfies the Lipschitz condition (13).

Example B.6 (Progresa score)**:** In the school choice model we use, the score is

$$\psi(y \mid z, x) = (y(z, x, \epsilon) - q(z, x)) \frac{\nabla_{\boldsymbol{\theta}} q(z, x)}{q(z, x)(1 - q(z, x))}$$

where $q(z, x) = P[y = 1 \mid z, x]$ is the probability of attending school. In our dynamic model $q(z, x)$ is defined recursively so it is difficult to verify the Lipschitz condition, but it can be verified for functions like logit and probit where we let ϵ be uniform and

$$y(z, x, \epsilon) = \mathbb{1}[\epsilon \leq q(z, x)]$$

◊

C Approximating nonlinear constraints

In this section we present a method for approximating the value function in the presence of nonlinear constraints. In the proof of Lemma 4.11, we showed that

$$V_n^Q(D\mu_1) = \tilde{V}_n^Q(D\mu_1) + o(n^{-1})$$

where

$$\begin{aligned} \tilde{V}_n^Q(D\mu_1) = \max_{\boldsymbol{\pi}} \quad & (\boldsymbol{\pi} - \boldsymbol{\pi}_0)' D(\mu_1 - \boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\pi}_0)'[H + \boldsymbol{\lambda}_0' \nabla^2_{\boldsymbol{\pi}\boldsymbol{\pi}} g(\boldsymbol{\pi}_0)](\boldsymbol{\pi} - \boldsymbol{\pi}_0) \\ & \text{s.t.} \\ & \boldsymbol{\pi}' \nabla g_j(\boldsymbol{\pi}_0) = 0, j \in \mathcal{J}_1 \end{aligned}$$

$$\boldsymbol{\pi}'\nabla g_j(\boldsymbol{\pi}_0) < 0, j \in \mathcal{J}_2.$$

In particular, $\tilde{V}_n^Q$ characterizes the second-order directional Hadamard derivative of the Gaussian value function at $\boldsymbol{\theta}_0$ in the sense that

$$V_n^G(\mu_1, \Sigma_1) - V_n^G(\mu_0, \Sigma_0)$$

It is tempting to approximate $\tilde{V}_n^Q$ by plugging in $(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\pi}}_0, \hat{\boldsymbol{\lambda}}_0)$ in place of $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0, \boldsymbol{\lambda}_0)$. However, this in general may not be a consistent estimate of $\tilde{V}_n^Q$. This is because $\tilde{V}_n^Q$ may not be twice (fully) differentiable at $\boldsymbol{\theta}_0$, and so $\sqrt{n}$-consistency of $(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\pi}}_0, \hat{\boldsymbol{\lambda}}_0)$ does not imply that the corresponding estimate of the second-order directional derivative is consistent. This problem is clearly articulated in Fang and Santos (2018), which studies properties of the delta method and bootstrap under failure of first-order full differentiability.

We propose an alternative estimator for $\tilde{V}_n^Q$ that retains the computational convenience of the quadratic programming formulation, but remains consistent even when $\tilde{V}_n^Q$ is not twice differentiable at $\boldsymbol{\theta}_0$. For this section, will make the dependence of $\tilde{V}_n^Q$ on the reference parameter $\boldsymbol{\theta}_0$ explicit by writing

$$\tilde{V}_n^Q(D\mu_1; \boldsymbol{\theta}_0)$$

and keep in mind that $\boldsymbol{\pi}_0$ and $\boldsymbol{\lambda}_0$ are determined by $\boldsymbol{\theta}_0$.

Let $\Theta_\vee$ be the set of $\boldsymbol{\theta}_0$ such that $\tilde{V}_n^Q(D\mu_1; \boldsymbol{\theta}_0)$ is not twice differentiable. Let ω_n be a sequence of positive numbers satisfying

$$\omega_n \to 0 \quad \text{and} \quad \omega_n\sqrt{n} \to \infty$$

Define $\hat{\boldsymbol{\theta}}_\vee$ as the point in $\Theta_\vee$ that is closest to $\hat{\boldsymbol{\theta}}_0$; that is,

$$\hat{\boldsymbol{\theta}}_\vee \in \text{argmin}_{\boldsymbol{\theta} \in \Theta_\vee} \|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}\|.$$

Define the estimator of the value function as

$$\hat{V}_n^Q(D\mu_1;\hat{\boldsymbol{\theta}}_0) = \begin{cases} \tilde{V}_n^Q(D\mu_1;\hat{\boldsymbol{\theta}}_\vee) & \text{if } \|\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_\vee\| \leq \omega_n \\ \tilde{V}_n^Q(D\mu_1;\hat{\boldsymbol{\theta}}_0) & \text{otherwise} \end{cases}$$

To interpret, we are conducting a pre-test for whether $\boldsymbol{\theta}_0$ is a point of nondifferentiability. If we cannot reject, we use the directional derivative $\tilde{V}_n^Q$ at $\hat{\boldsymbol{\theta}}_\vee$. If we can reject, we use the derivative at $\hat{\boldsymbol{\theta}}_0$. The rate conditions on ω_n ensure that the type 1 error rate tends to zero. If $\boldsymbol{\theta}_0$ is a point of nondifferentiability, then the probability of using the correct derivative tends to one. If $\boldsymbol{\theta}_0$ is not a point of nondifferentiability, then as n grows, we correctly identify this situation with probability tending to one, allowing us to use the correct derivative. This reasoning follows that of Example 2.1 in Fang and Santos (2018).

We now make this reasoning precise.

Theorem C.1: *Suppose $\hat{V}_n^Q$ is constructed as above. Otherwise, maintain the assumptions of Theorem 4.13. Then*

$$\mathbb{E}_{\boldsymbol{\delta}_n^Q}\left[V_n^Q(D\mu_1;\boldsymbol{\theta}_0)\right] - \mathbb{E}_{\hat{\boldsymbol{\delta}}_n^Q}\left[\hat{V}_n^Q(D\mu_1;\hat{\boldsymbol{\theta}}_0)\right] = o(n^{-1})$$

Proof. Conditional on h and scaled by n, the difference in welfare is given by

$$nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) - n\hat{W}\left(\hat{\boldsymbol{\theta}}_0 + \frac{h}{\sqrt{n}}, \hat{\boldsymbol{\pi}}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right)$$

We cannot immediately bound this from below as in the proof of Theorem 4.12 because the feasible sets of the two optimization problems are different. Instead, we show consistency in cases.

Case 1: Suppose $\boldsymbol{\theta}_0$ is a point of nondifferentiability. Then for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\|\hat{V}_n^Q(Dh) - V_n^Q(Dh)\| > \epsilon\right) &\leq \mathbb{P}\left(\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\| > \omega_n\right) \\ &= \mathbb{P}\left(\|\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)\| > \omega_n\sqrt{n}\right) \to 0 \end{aligned}$$

because $\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)$ is tight and $\omega_n\sqrt{n} \to \infty$.

Case 2: Suppose $\boldsymbol{\theta}_0$ is not a point of nondifferentiability. Without loss of generality assume n

large enough that $\|\hat{\boldsymbol{\theta}}_{\mathrm{V}} - \boldsymbol{\theta}_0\| \geq \omega_n$ For any $\epsilon > 0$,

$$\mathbb{P}\left(\|\hat{V}_n^Q(Dh) - V_n^Q(Dh)\| > \epsilon\right) \leq \mathbb{P}\left(\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\| > \omega_n\right) + \mathbb{P}\left(\|\hat{V}(Dh) - V(Dh)\| > \epsilon \text{ and } \|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\| \leq \omega_n\right)$$

The first term goes to zero by consistency of $\hat{\boldsymbol{\theta}}_0$. To show that the second term goes to zero, note that for n large enough, $\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\| \leq \omega_n$ implies that $\hat{V}_n^Q(Dh; \hat{\boldsymbol{\theta}}_0)$ has the correct set of active constraints, which all hold with strict complementary slackness. This means $\hat{V}_n^Q(Dh)$ is characterized by first-order conditions, and consistency follows from the implicit function theorem.

We have thus established that

$$nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) - n\hat{W}\left(\hat{\boldsymbol{\theta}}_0 + \frac{h}{\sqrt{n}}, \hat{\boldsymbol{\pi}}_0 + \frac{\hat{c}_n^Q}{\sqrt{n}}\right) = o_p(1)$$

and the expectation is $o(n^{-1})$ because regret is uniformly integrable. □

D Robust Bayes

D.1 Kullback-Leibler sets of priors

For the robust Bayes procedure of Section 6.2, we propose solving the value function

$$\begin{aligned} \tilde{V}_{\mathrm{kl}}\left(D\mu_1, D\Sigma_1 D'\right) = \max_{\boldsymbol{\pi}} \quad & \log \mathbb{E}\left[\exp\left(-\frac{1}{\kappa} W_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right) \mid \hat{\boldsymbol{\theta}}_1\right] \\ \text{s.t.} \quad & g_\infty(\boldsymbol{\pi}) \leq 0 \end{aligned} \tag{14a}$$

where $\boldsymbol{\theta} \sim N(D\mu_1, D\Sigma_1 D')$ given $\hat{\boldsymbol{\theta}}_1$. The optimal design solves

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \log \mathbb{E}\left[\exp\left(-\frac{1}{\kappa}\tilde{V}_{\mathrm{kl}}\left(D\mu_1, D\Sigma_1 D'\right)\right)\right] \\ \text{s.t.} \quad & \\ & f(\boldsymbol{\delta}) \leq 0 \end{aligned} \tag{14b}$$

where $\boldsymbol{\theta} \sim N(\mu_0, \Sigma_0)$. This problem is similar to (6a)-(6b) with a modified objective function, and can be solved by similar methods.

Proposition D.1: *The solution to (8a)-(8b) is given by (14a)-(14b).*

Proof. We begin with the terminal decision problem

$$\min_{\boldsymbol{\pi}} \max_{q} \quad \mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right] + \kappa \int \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta}))} q(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

s.t.

$$g_\infty(\boldsymbol{\pi}) \leq 0$$

$$\boldsymbol{\theta} \sim q(\boldsymbol{\theta})$$

and consider the inner minimzation over q. Let $p_1(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta}))$ be the posterior under the reference prior. We define q_1^* by its likelihood ratio

$$\frac{q_1^*(\boldsymbol{\theta})}{p_1(\boldsymbol{\theta})} = \frac{\exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right)}{\int \exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right) p_1(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

and observe that for any q, the objective function can be written

$$\begin{aligned}
&\mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right] + \kappa \int \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta}))} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right) q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \kappa \int \log \frac{q(\boldsymbol{\theta})}{p_1(\boldsymbol{\theta}))} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right) q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \kappa \int \log \frac{q(\boldsymbol{\theta})}{q_1^*(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \kappa \int \log \frac{q_1^*(\boldsymbol{\theta})}{p_1(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right) q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \kappa \int \log \frac{q(\boldsymbol{\theta})}{q_1^*(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \kappa \int \log \frac{\exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right)}{\int \exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right) p_1(\boldsymbol{\theta}) d\boldsymbol{\theta}} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \kappa \int \log \frac{q(\boldsymbol{\theta})}{q_1^*(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} - \kappa \log \int \exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right) p_1(\boldsymbol{\theta}) d\boldsymbol{\theta}
\end{aligned}$$

The first term is nonnegative and zero if and only if $q = q_1^*$. The second term does not depend on q. We conclude that the optimal q is q_1^*, and therefore the optimal value function is characterized by the posterior under the reference prior, which is Gaussian with mean μ_1 and covariance Σ_1. Thus the terminal decision problem is equivalent to

$$\min_{\boldsymbol{\pi}} \quad -\kappa \log \mathbb{E}\left[\exp\left(-\frac{1}{\kappa} R_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi})\right)\right]$$

s.t.

$$g_\infty(\boldsymbol{\pi}) \leq 0$$

$$D\boldsymbol{\theta} \sim N(D\mu_1, D\Sigma_1 D')$$

The optimal value of which we denote $\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')$.

We now consider the decision problem in the first period, which is

$$\begin{aligned}
\min_{\boldsymbol{\delta}} \max_{q} \quad & \mathbb{E}\left[\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')\right] + \kappa \int \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
& \text{s.t.} \\
& f(\boldsymbol{\delta}) \leq 0 \\
& \hat{\boldsymbol{\theta}} \sim dN(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}, J(\boldsymbol{\delta})) q(\boldsymbol{\theta})
\end{aligned}$$

and μ_1, Σ_1 are given by Bayesian updating under the Gaussian reference prior p. As before, we define

$$\frac{q^*(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} = \frac{\exp\left(-\frac{1}{\kappa}\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')\right)}{\int \exp\left(-\frac{1}{\kappa}\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')\right) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

and perform the same calculations as above to show that the optimal q is q^* and the objective is

$$\begin{aligned}
\min_{\boldsymbol{\delta}} \quad & -\kappa \log \mathbb{E}\left[\exp\left(-\frac{1}{\kappa}\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')\right)\right] \\
& \text{s.t.} \\
& f(\boldsymbol{\delta}) \leq 0 \\
& \hat{\boldsymbol{\theta}} \sim N(\mu_0, \Sigma_0 + J(\boldsymbol{\delta})^{-1})
\end{aligned}$$

where again the law of motion is given by Bayesian updating under the reference prior. □

D.2 Robustness to nuisance parameters

We propose an alternative value function which ignores both estimates of the nuisance parameter and the prior on the nuisance parameter. We use $(\tilde{\mu}_1, \tilde{\Sigma}_1)$ to denote values of the posterior mean and covariance of $D\boldsymbol{\theta}$. We imagine an experimenter who observes only $D\hat{\boldsymbol{\theta}}$, and updates the posterior

mean and covariance of $D\boldsymbol{\theta}$ according to

$$\begin{aligned}\tilde{\mu}_1 &= \left((D\Sigma_0 D)^{-1} + nDJ(\boldsymbol{\delta})D'\right)^{-1}\left((D\Sigma_0 D)^{-1}D\mu_0 + nDJ(\boldsymbol{\delta})D'D\hat{\boldsymbol{\theta}}\right)\\ \tilde{\Sigma}_1 &= \left((D\Sigma_0 D)^{-1} + nDJ(\boldsymbol{\delta})D'\right)^{-1}.\end{aligned} \tag{15}$$

That is, the experimenter updates the posterior mean and covariance as if the nuisance parameter were not present. For any particular prior q on the full vector $\boldsymbol{\theta}$, this update will not generally coincide with the correct Bayesian updating formula for the full posterior. However, Theorem D.2 shows that this updating formula coincides with that of the least favorable prior in $\mathcal{Q}$.

We therefore propose the following value function:

$$\tilde{V}_\perp\left(\tilde{\mu}_1, \tilde{\Sigma}_1\right) = \max_{\boldsymbol{\pi}} \quad \mathbb{E}[W_\infty(D\boldsymbol{\theta}, \boldsymbol{\pi}) \mid D\hat{\boldsymbol{\theta}}_1] \quad \text{s.t.} \quad g_\infty(\boldsymbol{\pi}) \leq 0 \tag{16a}$$

where $D\boldsymbol{\theta} \sim N(\tilde{\mu}_0, \tilde{\Sigma}_0)$ given $\hat{\boldsymbol{\theta}}_1$. The optimal design solves

$$\max_{\boldsymbol{\delta}} \quad \mathbb{E}\left[\tilde{V}_\perp\left(\tilde{\mu}_1, \tilde{\Sigma}_1\right)\right] \quad \text{s.t.} \quad f(\boldsymbol{\delta}) \leq 0 \tag{16b}$$

where $D\boldsymbol{\theta} \sim N(D\mu_0, D\Sigma_0 D')$ and the law of motion is described by (15). This problem is just as easy to solve as (6a)-(6b). In fact, the terminal value function $\tilde{V}_\perp$ is exactly the same as the value function in the limit experiment V_∞. The difference is that when optimizing over the design $\boldsymbol{\delta}$, the experimenter ignores possible realizations of the nuisance parameter and need only observe $D\hat{\boldsymbol{\theta}}_1$.

Proposition D.2: *The solution to (9a)-(9b) is given by (16a)-(16b).*

Before proving the theorem, we establish some notation. We will write any $q \in \mathcal{P}$ as

$$q(\boldsymbol{\theta}) = q_D(D\boldsymbol{\theta})q_\perp(D^\perp\boldsymbol{\theta} \mid D\boldsymbol{\theta}).$$

Define

$$\overline{D} = \begin{bmatrix} D \\ D^\perp \end{bmatrix}$$

and let Ω be

$$\Omega = \overline{D} J(\boldsymbol{\delta})^{-1} \overline{D}' = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

Where $\Omega_{11} = DJ(\boldsymbol{\delta})^{-1} D'$, etc. We likewise define the prior mean and variance on $\overline{D}\boldsymbol{\theta}$ as

$$m = \overline{D}\mu_0$$
$$S = \overline{D}\Sigma_0\overline{D}'$$

We also partition S conformably as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

and to avoid confusion use m^+ and S^+ to denote the posterior. We define the orthogonalization matrix M as follows, and also give its inverse for convenience.

$$M = \begin{bmatrix} I & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I \end{bmatrix} \qquad\qquad M^{-1} = \begin{bmatrix} I & 0 \\ \Omega_{21}\Omega_{11}^{-1} & I \end{bmatrix}.$$

Our first result gives a particular prior that factors the same way as the likelihood.

Lemma D.3: *If $S_{21} = \Omega_{21}\Omega_{11}^{-1} S_{11}$, then the posterior on $D\boldsymbol{\theta}$ does not depend on $D^\perp\hat{\boldsymbol{\theta}}$ or the prior on $D^\perp\boldsymbol{\theta}$. That is,*

$$m^+ = (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}\Omega_{11}^{-1} D\hat{\boldsymbol{\theta}} + (S_{11}^{-1} + \Omega_{11}^{-1})^{-1} S_{11}^{-1} m_1$$
$$S_{11}^+ = (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}$$

Proof. Consider a prior of the form

$$\overline{D}\boldsymbol{\theta} \sim N\left(\begin{bmatrix} D\mu_0 \\ D^\perp\mu_0 \end{bmatrix}, \begin{bmatrix} S_{11} & S_{11}\Omega_{11}^{-1}\Omega_{12} \\ \Omega_{21}\Omega_{11}^{-1} S_{11} & S_{22} \end{bmatrix} \right)$$

which is in $\mathcal{P}$. Consider the posterior variance on $M\overline{D}\boldsymbol{\theta}$:

$$\begin{aligned} MS^+M' &= M(S^{-1}+\Omega^{-1})^{-1}M' \\ &= [(MSM')^{-1}+(M\Omega M')^{-1}]^{-1} \end{aligned}$$

Then

$$MS = \begin{bmatrix} S_{11} & S_{11}\Omega_{11}^{-1}\Omega_{12} \\ 0 & S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix} \qquad M\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix}$$

$$MSM' = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix} \qquad M\Omega M' = \begin{bmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix}$$

and

$$(MSM')^{-1}+(M\Omega M')^{-1} = \begin{bmatrix} S_{11}^{-1}+\Omega_{11}^{-1} & 0 \\ 0 & (S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12})^{-1}+(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{bmatrix}$$

$$MS^+M' = \begin{bmatrix} (S_{11}^{-1}+\Omega_{11}^{-1})^{-1} & 0 \\ 0 & [(S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12})^{-1}+(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}]^{-1} \end{bmatrix}$$

is the posterior variance on $M\overline{D}\boldsymbol{\theta}$. Next, we calculate S^+, the posterior variance on $\overline{D}\boldsymbol{\theta}$.

$$M^{-1}MS^+M' = \begin{bmatrix} (S_{11}^{-1}+\Omega_{11}^{-1})^{-1} & 0 \\ \Omega_{21}\Omega_{11}^{-1}(S_{11}^{-1}+\Omega_{11}^{-1})^{-1} & [(S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12})^{-1}+(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}]^{-1} \end{bmatrix}$$

$$M^{-1}MS^+M'M'^{-1} = \begin{bmatrix} (S_{11}^{-1}+\Omega_{11}^{-1})^{-1} & (S_{11}^{-1}+\Omega_{11}^{-1})^{-1}\Omega_{11}^{-1}\Omega_{12} \\ \Omega_{21}\Omega_{11}^{-1}(S_{11}^{-1}+\Omega_{11}^{-1})^{-1} & [(S_{22}-\Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12})^{-1}+(\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}]^{-1} \end{bmatrix}$$

We now compute m^+, the posterior mean on $\overline{D}\boldsymbol{\theta}$. This is given by

$$m^+ = S^+\Omega^{-1}\overline{D}\hat{\boldsymbol{\theta}} + S^+S^{-1}m$$

where

$$S^+\Omega^{-1} = S^+M'(M\Omega M')^{-1}$$
$$= \begin{bmatrix} (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}\Omega_{11}^{-1} & 0 \\ S_{21}^+\Omega_{11}^{-1} & S_{22}^+(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{bmatrix}$$

and

$$S^+S^{-1} = \begin{bmatrix} (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}S_{11}^{-1} & 0 \\ S_{21}^+S_{11}^{-1} & S_{22}^+(S_{22} - \Omega_{21}\Omega_{11}^{-1}S_{11}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{bmatrix}$$

Since the top right entry of both $S^+\Omega^{-1}$ and S^+S^{-1} is zero,

$$m_1^+ = (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}\Omega_{11}^{-1}D\hat{\boldsymbol{\theta}} + (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}S_{11}^{-1}m_1$$
$$S_{11}^+ = (S_{11}^{-1} + \Omega_{11}^{-1})^{-1}$$

which does not depend on $D^\perp\hat{\boldsymbol{\theta}}$ or the prior on $D^\perp\boldsymbol{\theta}$. □

We use the above result to characterize the least favorable prior in the terminal period.

Lemma D.4:

$$\min_{\boldsymbol{\pi}} \max_{q\in\mathcal{P}} \quad \mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right]$$

Proof. Consider a prior q^* where $\overline{D}\boldsymbol{\theta} \sim N(m, S)$ where S is of the form specified in Lemma D.3. Let $\boldsymbol{\pi}^*$ solve the terminal value function when the prior is q^*,

$$\min_{\boldsymbol{\pi}} \quad \mathbb{E}_{q^*}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right]$$
$$\text{s.t.}$$
$$g_\infty(\boldsymbol{\pi}) \leq 0$$
$$D\boldsymbol{\theta} \sim N(m^+, S^+)$$

where m^+ and S^+ are calculated in the proof of Lemma D.3.

Because $q^* \in \mathcal{P}$, we have for any $\boldsymbol{\pi}$

$$\begin{aligned}\min_{\boldsymbol{\pi}} \sup_{q \in \mathcal{P}} \quad \mathbb{E}_q \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi} \right) \right] &\geq \min_{\boldsymbol{\pi}} \quad \mathbb{E}_{q^*} \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi} \right) \right] \\ &= \mathbb{E}_{q^*} \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right]\end{aligned}$$

where in the first line we have left the constraints implicit. Likewise,

$$\min_{\boldsymbol{\pi}} \sup_{q \in \mathcal{P}} \quad \mathbb{E}_q \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi} \right) \right] \leq \sup_{q \in \mathcal{P}} \quad \mathbb{E}_q \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right]$$

We defined $\boldsymbol{\pi}^*$ so that it only depends on m^+ and S^+, which by Lemma D.3 only depend on the prior on $D\boldsymbol{\theta}$ and the observed $D\hat{\boldsymbol{\theta}}$. Therefore, for any $q \in \mathcal{P}$,

$$\begin{aligned}\mathbb{E}_q \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right] &= \int \int R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) dN(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}, J(\boldsymbol{\delta})) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int \int \int \int R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \times \\ &\qquad dN(MD^\perp \hat{\boldsymbol{\theta}}; MD^\perp \boldsymbol{\theta}, (M\Omega M')_{22}) \times \\ &\qquad dN(MD\hat{\boldsymbol{\theta}}; MD\boldsymbol{\theta}, (M\Omega M')_{11}) \times \\ &\qquad q_\perp(D^\perp \boldsymbol{\theta} \mid D\boldsymbol{\theta}) d(D^\perp \boldsymbol{\theta}) \times \\ &\qquad q_D(D\boldsymbol{\theta}) d(D\boldsymbol{\theta}) \\ &= \int R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) dN(MD\hat{\boldsymbol{\theta}}; MD\boldsymbol{\theta}, (M\Omega M')_{11}) q_D(D\boldsymbol{\theta}) d(D\boldsymbol{\theta}) \\ &= \int R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) dN(D\hat{\boldsymbol{\theta}}; D\boldsymbol{\theta}, \Omega_{11}) q_D(D\boldsymbol{\theta}) d(D\boldsymbol{\theta}) \\ &= \mathbb{E}_{q^*} \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right]\end{aligned}$$

where the second line follows from the fact that $(M\Omega M')$ is diagonal and so $MD\boldsymbol{\theta}$ and $MD^\perp \boldsymbol{\theta}$ are independent, the third line is due to $\boldsymbol{\pi}^*$ not depending on $D^\perp \boldsymbol{\theta}$ or $D^\perp \hat{\boldsymbol{\theta}}$, the fourth uses properties of M shown above, and the last is because all $q \in \mathcal{P}$, including q^*, have the same marginal over $D\boldsymbol{\theta}$. Combining the two previous displays, we have

$$\sup_{q \in \mathcal{P}} \quad \mathbb{E}_q \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right] = \mathbb{E}_{q^*} \left[R_\infty \left(D\boldsymbol{\theta}, \boldsymbol{\pi}^* \right) \right]$$

We have shown that

$$\min_{\boldsymbol{\pi}} \max_{q \in \mathcal{P}} \quad \mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right] = \mathbb{E}_{q^*}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}^*\right)\right]$$

and therefore q^* is the least favorable prior and $\boldsymbol{\pi}^*$ is the minimax policy. □

Now we prove Proposition D.2.

Proof. In the terminal period, the experimenter solves

$$\begin{aligned}
&\min_{\boldsymbol{\pi}} \max_{q \in \mathcal{P}} \quad \mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right] \\
&\text{s.t.} \\
&g_\infty(\boldsymbol{\pi}) \leq 0 \\
&\boldsymbol{\theta} \sim q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, J(\boldsymbol{\delta}))
\end{aligned}$$

which, by Lemma D.4, is equivalent to

$$\begin{aligned}
\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D') = \min_{\boldsymbol{\pi}} \quad & \mathbb{E}\left[R_\infty\left(D\boldsymbol{\theta}, \boldsymbol{\pi}\right)\right] \\
\text{s.t.} \\
& g_\infty(\boldsymbol{\pi}) \leq 0 \\
& D\boldsymbol{\theta} \sim N(D\mu_1, D\Sigma_1 D')
\end{aligned}$$

In the first period, the experimenter solves

$$\begin{aligned}
&\min_{\boldsymbol{\delta}} \max_{q \in \mathcal{P}} \quad \mathbb{E}\left[\tilde{V}_{\text{kl}}(D\mu_1, D\Sigma_1 D')\right] \\
&\text{s.t.} \\
&f(\boldsymbol{\delta}) \leq 0 \\
&D\mu_1 = ((D\Sigma_1 D')^{-1} + (DJ(\boldsymbol{\delta})D')^{-1})^{-1}(DJ(\boldsymbol{\delta})D')^{-1} D\hat{\boldsymbol{\theta}} \\
&\quad + ((D\Sigma_1 D')^{-1} + (DJ(\boldsymbol{\delta})D')^{-1})^{-1}(D\Sigma_1 D')^{-1} D\mu_0 \\
&D\Sigma_1 D' = ((D\Sigma_1 D')^{-1} + (DJ(\boldsymbol{\delta})D')^{-1})^{-1}
\end{aligned}$$

$$\hat{\boldsymbol{\theta}} \sim N(\mu_0, \Sigma_0 + J(\boldsymbol{\delta})^{-1})$$

Since the evolution of $D\mu_1$ and $D\Sigma_1 D'$ does not depend on the prior on $D^{\perp}\boldsymbol{\theta}$ or $D^{\perp}\hat{\boldsymbol{\theta}}$, the maximization over q and may be dropped and only $D\hat{\boldsymbol{\theta}}$ need be observed. □

E Multi-wave experiments

The restriction to the Gaussian estimate $\hat{\boldsymbol{\theta}}_t$ is justified by the following extension of Theorem 4.7. As in the single-period case, we say $(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T, c)$ is an adaptive design and policy in the limit experiment if

$$\begin{aligned}\boldsymbol{\delta}_t &= \boldsymbol{\delta}_t(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_{t-1}, A_1(\boldsymbol{\delta}_1), \ldots, A_t(\boldsymbol{\delta}_{t-1}), U)\\ c &= c(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T, A_1(\boldsymbol{\delta}_1), \ldots, A_t(\boldsymbol{\delta}_T), U)\end{aligned}$$

where A_t is the limit of the score process in wave t and U is a uniform random variable independent of A_t for all t.

Lemma E.1: *Suppose Assumptions 4.1 and 4.2 hold. Assume that* $0 < \lim_{n\to\infty} n_t/n < 1$ *for all* t. *For any convergent sequence of designs and policies* $(\boldsymbol{\delta}_{n1}, \ldots, \boldsymbol{\delta}_{nT}, \boldsymbol{\pi}_n)$ *such that the vector*

$$\left(\boldsymbol{\delta}_{n1}, \ldots, \boldsymbol{\delta}_{nT}, \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)\right)$$

converges in distribution under $\boldsymbol{\theta}_0$, *there exists an adaptive design and policy* $(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T, c)$ *in the limit experiment such that*

$$\left(\boldsymbol{\delta}_{n1}, \ldots, \boldsymbol{\delta}_{nT}, \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0)\right) \overset{h}{\rightsquigarrow} \left(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T, c\right)$$

where $\overset{h}{\rightsquigarrow}$ *denotes convergence in distribution under the sequence* $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h/\sqrt{n}$.

Proof. This proof is similar to the proof of Theorem 4.7. Let

$$p_{n,\boldsymbol{\theta}}(\boldsymbol{\delta}) = \prod_{t=1}^{T} \prod_{i \in \mathcal{I}_t} p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}) p_{z|x}(z_i \mid x_i; \boldsymbol{\delta}_t) p_x(x_i)$$

be the density of the data generated by the dynamic experiment. We first note that the likelihood ratio factors so that

$$\log \frac{p_{n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_T)}{p_{n,\boldsymbol{\theta}_0}(\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_T)} = \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} \log \frac{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0 + h/\sqrt{n})}{p_{y|z,x}(y_i \mid z_i, x_i; \boldsymbol{\theta}_0)}$$

By Lemma A.1, each element of the sum over t converges in probability (as a process in Δ) to

$$\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_t} h'\psi_i(\cdot) - \frac{1}{2}h'J(\cdot)h$$

As shown in the proof of Theorem 4.7, the weak limit of

$$A_{nt}(\cdot) = \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_t}\psi_i(\cdot)$$

under $\boldsymbol{\theta}_0$ is a Gaussian process A_t with mean zero and covariance

$$\operatorname{Cov}\left(A_t(b), A_t(c)\right) = \operatorname{Cov}\left(\psi_i(b), \psi_i(c)\right)$$

for all $b, c \in \Delta$.

By assumption, $(\boldsymbol{\delta}_{n1},\ldots,\boldsymbol{\delta}_{nT},\boldsymbol{\pi}_n)$ all converge marginally in distribution under $\boldsymbol{\theta}_0$. This implies that there exists a subsequence along which

$$\begin{aligned}&\left(\boldsymbol{\delta}_{n1},\ldots,\boldsymbol{\delta}_{nT},\sqrt{n}(\boldsymbol{\pi}_n-\boldsymbol{\pi}_0),A_{n1}(\cdot),\ldots,A_{nT}(\cdot)\right)\\ &\overset{\boldsymbol{\theta}_0}{\rightsquigarrow}\left(\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_T,c,A_1(\cdot),\ldots,A_T(\cdot)\right)\end{aligned}$$

and therefore

$$\begin{aligned}&\left(\boldsymbol{\delta}_{n1},\ldots,\boldsymbol{\delta}_{nT},\sqrt{n}(\boldsymbol{\pi}_n-\boldsymbol{\pi}_0)\log\frac{p_{1,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta}_{n1})}{p_{1,n,\boldsymbol{\theta}_0}(\boldsymbol{\delta}_{n1})},\ldots,\log\frac{p_{T,n,\boldsymbol{\theta}_0+h/\sqrt{n}}(\boldsymbol{\delta}_{nT})}{p_{T,n,\boldsymbol{\theta}_0}(\boldsymbol{\delta}_{nT})}\right)\\ &\overset{\boldsymbol{\theta}_0}{\rightsquigarrow}\left(\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_T,c,h'A_1(\boldsymbol{\delta}_1)-\frac{1}{2}h'J(\boldsymbol{\delta}_1)h,\ldots,h'A_T(\boldsymbol{\delta}_T)-\frac{1}{2}h'J(\boldsymbol{\delta}_T)h\right)\end{aligned}$$

By Le Cam's third lemma, we obtain the limiting distribution of $(\boldsymbol{\delta}_{n1},\ldots,\boldsymbol{\delta}_{nT},\boldsymbol{\pi}_n)$ under local alternatives along the subsequence. Since the full sequence converges in distribution under local

alternatives, this is also the limiting distribution of the full sequence. For any Borel set B, the limiting distribution is given by

$$L_h(B) = \mathbb{E}_{\boldsymbol{\theta}_0} \mathbb{1}\left[\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T, c\right] \exp\left(\sum_t h' A_t(\boldsymbol{\delta}_t) - \frac{1}{2} h' J(\boldsymbol{\delta}_t) h\right)$$

The construction of $(\boldsymbol{\delta}_1^h, \ldots, \boldsymbol{\delta}_T^h, c^h)$ in the limit experiment with this distribution follows the same steps as in the proof of Theorem 4.7. □

The quadratic approximation of the terminal value function is justified by Theorem 4.12. This leads to the same optimality guarantees as in the single-period case.

Theorem E.2: *In addition to the assumptions of Lemma E.1, suppose Assumptions 4.6, 4.9, and 4.10 hold. If $c_n = \sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}_0))$ is a sequence of feasible policies in the finite-sample experiment which is uniformly bounded in probability under $\boldsymbol{\theta}_0$, then*

$$\limsup_{n\to\infty} \mathbb{E}\left[nW\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n}{\sqrt{n}}\right)\right] \leq \mathbb{E}\left[nW^Q\left(\boldsymbol{\theta}_0 + \frac{h}{\sqrt{n}}, \boldsymbol{\pi}_0 + \frac{c_n^Q}{\sqrt{n}}\right) + M(\mu_0, \Sigma_0)\right]$$

Moreover, this upper bound is attained by implemnting the feasible analog $(\hat{\boldsymbol{\delta}}_{n,1}^Q, \ldots, \hat{\boldsymbol{\delta}}_{n,T}^Q, \hat{c}_n^Q)$ as in Theorem 4.13.

The proof follows that of Theorem 4.13 and is omitted.

F Details of the Progresa application

Table 2 shows the values of the pilot estimates $\hat{\boldsymbol{\theta}}_0$ for the Progresa application. The pilot estimate of $\hat{bm}\pi_0$ is obtained using the Ipopt nonlinear solver which uses an interior point algorithm. The estimates $(\hat{C}, \hat{D}, \hat{H})$ are then be computed by automatic differentiation to define W_n^Q. Algorithm 1 describes the optimization algorithm used to construct the estimate of V_n^{GQ}. After computing V_n^{GQ}, we compute

$$\mathbb{E}_{\boldsymbol{\delta}}\left[V_n^{GQ}(\hat{D}\mu_1) \mid \hat{\boldsymbol{\theta}}_0\right]$$

Table 2: Preliminary estimates $\boldsymbol{\theta}_0$ from pilot data

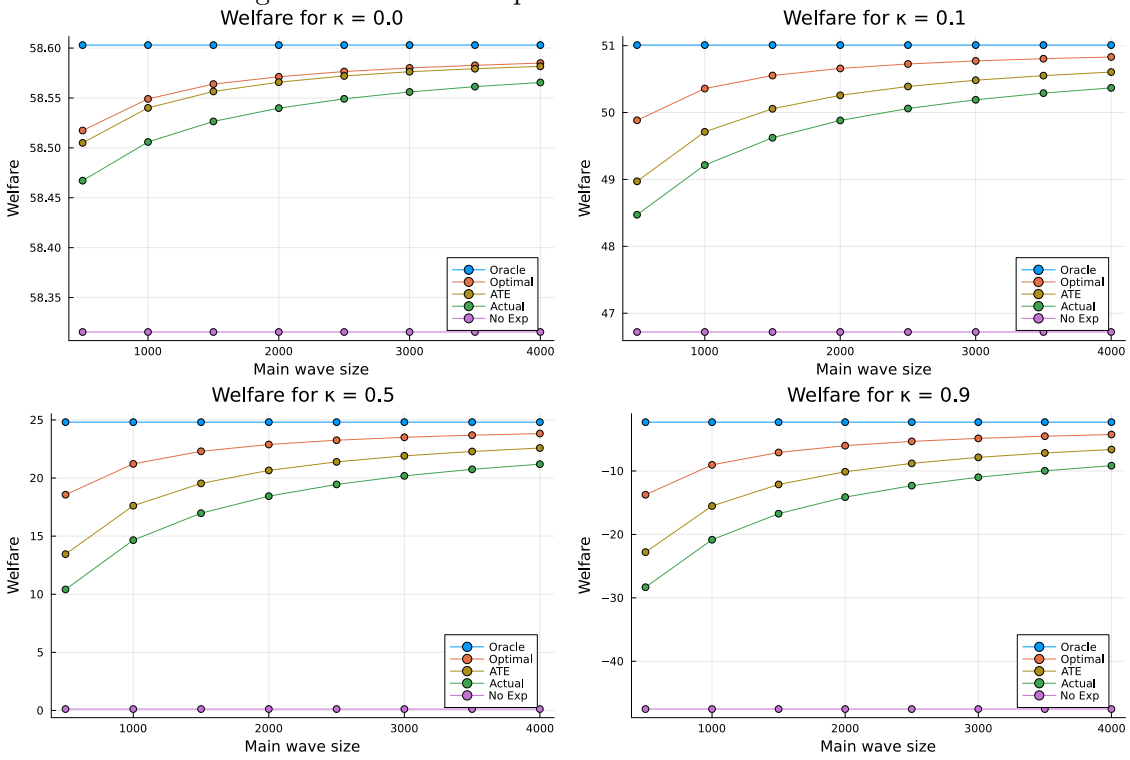
	Estimate	Standard Error
Constant	1.74	1.77
Fem	1.45	2.61
Grade	1.39	0.21
Grade × Fem	-0.06	0.3
Age	-0.59	0.15
Age × Fem	-0.03	0.22
Secondary School	-6.96	1.82
Fem × Secondary School	-1.72	2.68
Wage	-6.53	2.56
Wage × Fem	-4.27	5.0
Wage × Secondary School	5.01	3.01
Wage × Fem × Secondary School	6.52	5.47
Subsidy	-1.81	5.8
Subsidy × Fem	11.82	9.33
Subsidy × Secondary School	5.07	6.14
Subsidy × Fem × Secondary School	-11.16	9.86
Terminal value primary	-0.57	1.24
Terminal value secondary	0.56	1.38
β	0.95	
N	500	

Note: Standard errors are conditional on β, which is calibrated to match Attanasio, Meghir, and Santiago (2012).

for any $\boldsymbol{\delta}$ by Monte Carlo integration over the distribution of μ_1 conditional on $\hat{\boldsymbol{\theta}}_0$. This is a differentiable function of $\boldsymbol{\delta}$, and so we can solve for the optimal $\boldsymbol{\delta}$ again using Ipopt.

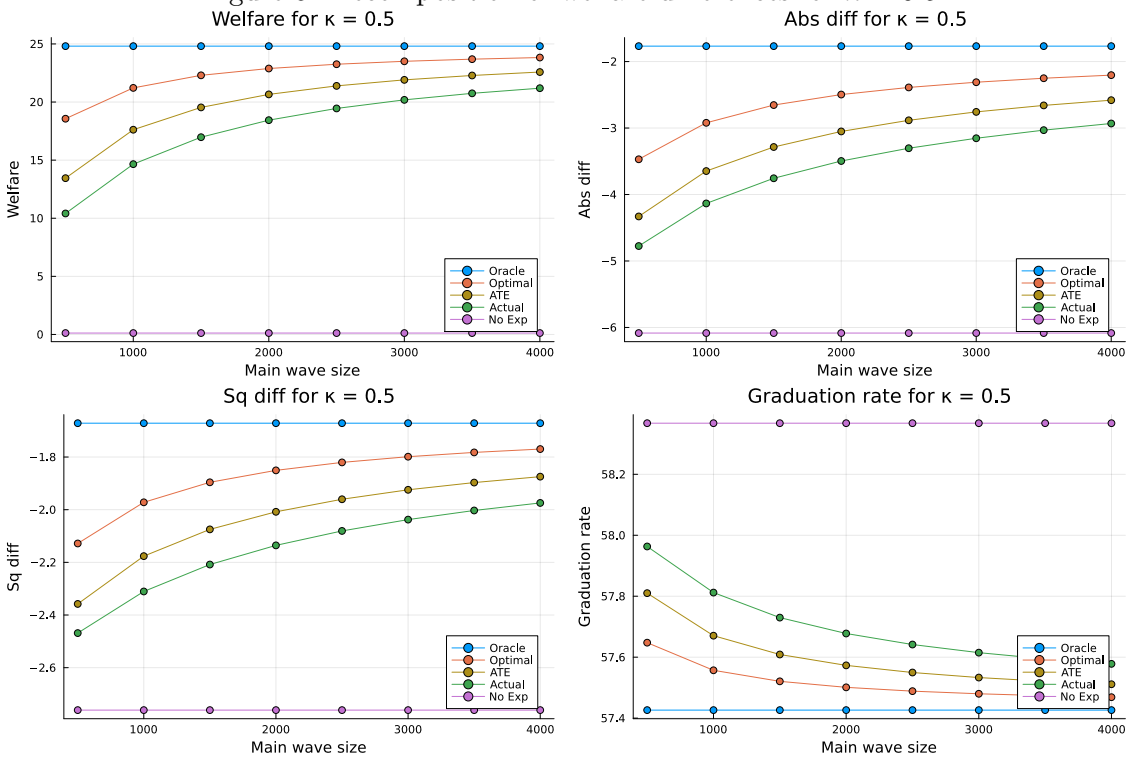
Results for various levels of κ are shown in Figure F. In addition to the designs shown in Figure 3, we also show the performance of the design which maximizes the precision of the average effect of the original Progresa subsidy schedule. In Figure F, we show how the increase in welfare is driven by the increase in graduation rates versus the reduction in the gender gap. In fact, the improvement in the gender gap entirely drives the increase in welfare for $\kappa = 0.5$, and graduation rates actually decrease very slightly with the sample size. However, if we were to set $\kappa = 1$, then the welfare function is not strictly concave, violating Assumption 4.10.

Figure 5: Welfare comparison for different values of κ



Note: The expected welfare from the optimal experiment for different values of κ and a variety of values of n_1. "Oracle" refers to the expected welfare from the infeasible, optimal policy given the true parameter values. "Optimal" refers to the expected welfare from running the optimal experiment. "ATE" refers to the expected welfare from running an experiment which maximizes the precision of the estimated average treatment effect of the original Progresa subsidy schedule. "Progresa" refers to the expected welfare from running the original Progresa experiment. "No Exp" refers to the expected welfare from using the pilot data only.

Figure 6: Decomposition of welfare differences for $\kappa = 0.5$



Note: The expected welfare from the optimal experiment when $\kappa = 0.5$ is shown for a variety of values of n_1, broken down into the contributions from the increase in average graduation rates and the decrease in gender disparities in graduation rates. "Oracle" refers to the expected welfare from the infeasible, optimal policy given the true parameter values. "Optimal" refers to the expected welfare from running the optimal experiment. "ATE" refers to the expected welfare from running an experiment which maximizes the precision of the estimated average treatment effect of the original Progresa subsidy schedule. "Progresa" refers to the expected welfare from running the original Progresa experiment. "No Exp" refers to the expected welfare from using the pilot data only.

Algorithm 1: Optimization Algorithm with ADAM

Input: Initial parameters of V, prior (μ_0, Σ_0), learning rate, number of epochs, tolerance
Output: Trained model V, final r^2 value
Initialize `opt_state` using ADAM with learning rate 0.01;
Set `nepoch` to 10000;
Set tolerance `tol` to 0.999;
for $i = 1$ *to* `nepoch` **do**
    if i mod $100 = 0$ **then**
        Draw sample $\mathrm{D}\mu$ from $\mathtt{N}(D\mu_0, D\Sigma_0 D')$;
        Solve $\mathtt{QP}(\mathrm{D}\mu)$ to get v ;
        Compute predicted values `V_fit` from model $V(D\mu)$;
        Compute $r = \mathtt{cor}(v, V\_fit)$;
        if $r >$ `tol` **then**
            break;
    Compute gradients `grads` of V;
    Update parameters V with `opt_state` and `grads`;