NBER WORKING PAPER SERIES

INFERRING WELFARE MAXIMIZING TREATMENT ASSIGNMENT UNDER BUDGET CONSTRAINTS

Debopam Bhattacharya
Pascaline Dupas

Working Paper 14447
http://www.nber.org/papers/w14447

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2008

Previously circulated under the title: "Nonparametric Inference on Efficient Treatment Assignment under Budget Constraints". We thank Andrew Chesher, Kei Hirano, Simon Lee as well as seminar participants at the Econometric Society winter meetings, Bristol ESRC Econometrics group meeting, Dartmouth, Oxford, UCL and the World Bank for useful comments. Dupas gratefully acknowledges funding from the Acumen Fund. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Inferring Welfare Maximizing Treatment Assignment under Budget Constraints
Debopam Bhattacharya and Pascaline Dupas
NBER Working Paper No. 14447
October 2008, Revised June 2009
JEL No. C01,C14,I38

ABSTRACT

This paper concerns the problem of allocating a binary treatment among a target population based on discrete and continuous observed covariates. The goal is to maximize the mean social utility of an eventual outcome when a budget constraint limits what fraction of the population can be treated. We propose a treatment allocation procedure based on sample data from randomized treatment assignment. We examine this procedure in the light of statistical decision theory and derive asymptotic frequentist properties of the allocation rule and the welfare generated from it. The resulting distribution theory is used to conduct inference on the welfare loss resulting from restricted covariate choice and on the dual value, i.e. the minimum resources needed to attain a specific average welfare via efficient treatment assignment. The methodology is applied to the optimal provision of anti-malaria bed net subsidies, using data from a randomized experiment conducted in western Kenya. We find that a government which can afford to distribute bed net subsidies to only 50% of its target population can, with an efficient allocation rule based on multiple covariates, increase bed-net use by 8 percentage points (25 percent) relative to random allocation and by 4 percentage points (11 percent) relative to one based on wealth only. Our methods do not rely on functional form assumptions and can be extended to situations encompassing conditional cash transfers, imperfect treatment take-up and spillover effects on non-eligibles.

Debopam Bhattacharya
Economics Department
Dartmouth College
301 Rockefeller Hall
Hanover, NH 03755
debopam.bhattacharya@dartmouth.edu

Pascaline Dupas
Department of Economics
UCLA
8283 Bunche Hall
Los Angeles, CA 90095
and NBER
pdupas@econ.ucla.edu

1 Introduction

Vulnerable populations in developing countries often lack access to critical health and educational facilities. Enhancing their access can generate both high private returns and, in many cases, significant positive externalities for society. Examples include improvement of female literacy rates or decreasing the incidence of infectious disease. These considerations often lead the governments to subsidize access to such key health and educational resources. However, such subsidizing efforts are also typically constrained by binding budget ceilings. When budgets are such that only a small fraction of a target population can receive a given subsidy, the eligibility rule used to decide who will receive the subsidy can have an important effect on the overall benefit arising from the subsidy program.

In this paper, we consider the problem of allocating a fixed amount of resources to a target population with the aim of maximizing the mean population outcome, and the dual problem of estimating the minimum cost of achieving a given mean outcome in the population by efficient targeting of a treatment. We set-up a statistical framework for studying this problem and apply it to design welfare-maximizing allocation of subsidies for an effective malaria control tool – insecticide-treated bed nets or ITNs– among households in a malaria-endemic region of Kenya. Our treatment of interest is making subsidized ITNs available to a section of this population and the welfare measure of interest is the fraction of households using an ITN.[1] We find that, if available resources allow us to treat only 50% of the target population, randomly allocating ITN subsidies is 19% (8 percentage points) less efficient than optimally allocating them based on a set of observed characteristics. Allocating the subsidies according to wealth only is 9% (4 percentage points) less efficient than allocating them based on a set of relevant covariates. Finally, allocating the subsidies based on all covariates but wealth is 10 to 16% (4 to 7 percentage points) less efficient than allocating them based on the complete set of relevant covariates.

Our paper contributes to a recent but steadily expanding literature in statistics and economics, discussed below, on how experimental evidence on treatment effect heterogeneity may be used to maximize gains from social programs. Our substantive contribution is to study such problems in the presence of aggregate budget constraints– an extremely common situation in real life but largely ignored in the treatment choice literature. The constraint makes the treatment assignment problem analytically very different from the unconstrained case and leads to novel inferential theory concerning a set of interesting parameters (denoted by γ and ρ below) which, to our knowledge, are new to the treatment effects literature.

Our proposed allocation rule is based on sample data from an experiment where the treatment was randomly assigned. From a decision theoretic angle, it can be interpreted as the Bayes rule corresponding to an uninformative prior under a nonparametric set-up, i.e., where no functional

[1] While incidence of malaria or wages lost due to malaria are the "ultimate" outcomes of interest, we concentrate on ITN use because researchers can observe it directly and thus avoid self-reporting errors, and also because it is a good proxy for health outcomes. Indeed, multiple trials have established that ITN use leads to significant reductions in morbidity and mortality, especially among young children (see Lengeler (2004) for a review).

form or distributional assumptions are made about the underlying random variables. The frequentist expected welfare resulting from our proposed rule depends on the marginal distribution of the conditional average treatment effect (CATE, henceforth). Specifically, when the budget limits the treatment fraction to $c \in (0,1)$, the welfare-maximizing treatment threshold and the resulting expected welfare approximate respectively the $(1-c)$th marginal quantile (denoted by γ) and the corresponding generalized Lorenz share (denoted by ρ) in the population distribution of the random variable $\theta(X)$ – where $\theta(x)$ represents the average treatment effect for the subpopulation whose value of the observed characteristic X is x. Given this, exact finite-sample inference for ρ and γ becomes analytically intractable– especially when X contains continuous components– and this leads us to asymptotic analysis.

The key technical challenge we face in conducting asymptotic inference on ρ and γ is that the *population* moment condition defining γ is a step-function in $\theta(\cdot)$, which invalidates the use of standard methods, e.g., Andrews (1994), Newey (1994) and Chen, Linton and van Keilegom (2003) (CLV, henceforth). We bypass this problem by using additional smoothing in defining the estimates and show that γ and ρ can be estimated at fast enough rates even if $\theta(\cdot)$ is left nonparametric. As a corollary, we also derive inference theory for the dual policy parameter, viz. the minimum fraction of the population which has to be treated in order to attain a target level of mean outcome. The value function for this dual problem is the inverse function of $\rho(\cdot)$.

The methods proposed here have wider applicability, beyond subsidy targeting in developing countries, to nearly any situation of constrained treatment assignment such as deciding eligibility rules for access to credit under aggregate fund constraints or allocating the unemployed to subsidized job-training programs when subsidy totals are limited by the government's budget outlay. We also discuss extensions of our methods to situations involving partial treatment take-up, the presence of spillover effects and, finally, the design of conditional cash-transfer programs which are becoming increasingly popular in developing countries.

The rest of the paper is organized as follows. Section 2 sets up the problem, introduces the estimands of interest and discusses the decision theoretic underpinnings of our methodology. Section 3 discusses where the present paper fits in the relevant literature in econometrics and development economics. Section 4 introduces the estimators and discusses some key issues regarding rate of convergence. Section 5 develops the relevant distribution theory and section 6 presents the benchmark case of parametric inference. Section 7 discusses extensions to (i) partial take-up of subsidy, (ii) the design of conditional cash transfer programs (iii) optimal targeting in presence of spillover effects and (iv) formulation of the best *linear* allocation rule. Section 8 presents the application to the welfare-maximizing allocation of bed nets in Kenya and section 9 concludes. The appendix contains some illustrative order of magnitude calculations alluded to in the text as well as proofs of all the theorems. The words "optimal" and "efficient" are used in the text to mean "welfare-maximizing" and do not have any "minimum variance" connotation.

2 Formulation of the Problem

2.1 Set-up

Let Y denote an household-level outcome which can be either binary or continuous and let S denote a binary treatment whose value can be affected directly by policy. Let X denote observed covariates which includes both discrete and continuous components and U denote unobserved determinants of Y. In the bed-net example, analyzed below in detail, the population of interest is rural households of western Kenya. We have a simple random sample drawn from two districts in Western Kenya. Each household is an observation. Y is a binary outcome which equals 1 if the household owns and uses a bed net. X is the presence of a child under 10, the wealth per capita and ownership of a bank account, while U represents unobserved determinants of Y. The treatment, i.e., $S = 1$, is offering a highly subsidized bed net to the household. Y_1 and Y_0 are the value of the outcome Y with and without the treatment, respectively, i.e., $Y = SY_1 + (1 - S) Y_0$.

Let $\phi(x, s)$ denote the expected outcome at $S = s$ for households with $X = x$: i.e. if an household with characteristic $X = x$ is randomly selected from the population and assigned a value s of S, then its expected outcome is $\phi(x, s)$. If S is independent of U conditional on X as in a randomized trial (the case studied here), then a nonparametric regression of Y on X for households with $S = s$ in the sample can be used to recover this function.

We will first consider an idealized version of a social planner's problem where $\phi(x, s)$ and the marginal distribution of X with support $\mathcal{X}$ are known to the planner. The planner faces a constraint on what fraction of households can be administered the treatment ($S = 1$). Suppose this fraction is c. We define the planner's idealized problem as the choice of a set $A \subset \mathcal{X}$ such that if an household's value of X is in this set, then the planner assigns that person to the treatment and not otherwise. We will assume that the planner wants to maximize mean outcome.[2] Then the planner's problem is

$$\max_{A \subset \mathcal{X}} \int_{x \in \mathcal{X}} \left[\phi(x, 1) 1(x \in A) + \phi(x, 0) 1(x \notin A)\right] dF(x)$$

subject to

$$c = \int_{x \in \mathcal{X}} 1(x \in A) \, dF(x) . \tag{1}$$

Obviously, the budget constraint will hold with equality at the optimum. It is also intuitive that the optimal set A will include those x's where $\phi(x, 1)$ is "large" relative to $\phi(x, 0)$. The following proposition formalizes this intuition. We will use the notation $\theta(x)$ to mean $\phi(x, 1) - \phi(x, 0)$, to be called the conditional average treatment effect (CATE) henceforth. Here the reader may note that while we allow for negative program impacts (i.e. $\Pr[\theta(X) < 0] \geq 0$), the present problem is interesting only if $\Pr(\theta(X) > 0) > c$; otherwise, the efficient assignment rule would be to give treatment to everybody whose average treatment effect is positive.

[2] More generally, if the planner is interested in maximizing (a possibly covariate weighted) outcome utility, then $\phi(x, 1)$ represents the expected value of the planner's utility defined on outcomes for individuals with $X = x$ and $S = 1$.

Proposition 1 *The solution to the planner's problem*

$$\max_{A\subset\mathcal{X}}\int_{x\in\mathcal{X}}\left[\phi\left(x,1\right)1\left(x\in A\right)+\phi\left(x,0\right)1\left(x\notin A\right)\right]dF\left(x\right)$$

subject to

$$c=\int_{x\in\mathcal{X}}1\left(x\in A\right)dF\left(x\right)$$

is of the form $A^{*}=\{x:\theta\left(x\right)>\gamma\}$ *where* $\theta\left(x\right)\equiv\phi\left(x,1\right)-\phi\left(x,0\right)$ *and* γ *satisfies*

$$c=\int_{x\in\mathcal{X}}1\left(\theta\left(x\right)>\gamma\right)dF\left(x\right).$$

Proof. Appendix ■

For the optimal choice of A, the value function, capturing the maximal gains from covariate based allocation, will be

$$\begin{aligned}\rho\left(c\right) &= \int_{x\in\mathcal{X}}\left[\phi\left(x,1\right)1\left\{\theta\left(x\right)>\gamma\left(c\right)\right\}+\phi\left(x,0\right)1\left\{\theta\left(x\right)\leq\gamma\left(c\right)\right\}\right]dF\left(x\right)\\ &= \int_{x\in\mathcal{X}}\phi\left(x,1\right)dF\left(x\right)-\int_{x\in\mathcal{X}}\theta\left(x\right)\times1\left\{\theta\left(x\right)\leq\gamma\left(c\right)\right\}dF\left(x\right). \qquad (2)\end{aligned}$$

The above proposition implies that one can solve for $\gamma\left(c\right)$ from

$$c=\int_{x\in\mathcal{X}}1\left\{\theta\left(x\right)>\gamma\left(c\right)\right\}dF\left(x\right). \qquad (3)$$

The above equation shows that $\gamma\left(c\right)$ is the $(1-c)$th quantile for the marginal distribution of the CATE, i.e., the random variable $\theta\left(X\right)$. Let us denote the population c.d.f. of this distribution by $G\left(\cdot\right)$. The corresponding value function from (2) can be written as

$$\rho\left(c\right)=E\left[\phi\left(X,1\right)\right]-\int_{z\in\Theta}\left[z\times1\left\{z\leq\gamma\left(c\right)\right\}\right]dG\left(z\right),$$

where $\int_{z\in\Theta}\left[z\times1\left\{z\leq\gamma\left(c\right)\right\}\right]dG\left(z\right)$ is the generalized Lorenz share of $\theta\left(X\right)$, corresponding to the percentile $(1-c)$ and Θ is the support of $\theta\left(X\right)$.

It is worth stressing here that γ is the $(1-c)$th quantile in the distribution of the CATE and thus very different from the $(1-c)$th quantile treatment effect which has often been discussed in the treatment effect literature (c.f., Abadie (2002) and references therein). Similarly for ρ.

Remark: While we focus the current paper on the mean utility of outcome as the objective, the idea applies in principle to any functional of the overall outcome distribution. Let $F_1\left(\cdot|x\right)$ and $F_0\left(\cdot|x\right)$ denote the marginal C.D.F. of the outcome, conditional on X, under treatment and no treatment respectively. These marginals can be identified using experimental data from randomized treatment allocation. If we fix the treatment set to be $A\subset\mathcal{X}$, then the C.D.F. of the overall outcome corresponding to the choice A is

$$G\left(\cdot;A\right)=\int_{x\in\mathcal{X}}\left[F_1\left(\cdot|x\right)\times1\left(x\in A\right)+F_0\left(\cdot|x\right)\times1\left(x\notin A\right)\right]dF\left(x\right).$$

If the planner wishes to maximize a functional $\mathcal{F}(\cdot)$ of the C.D.F. $G(\cdot; A)$, then the optimization problem reduces to

$$\max_{A \subset \mathcal{X}} \mathcal{F}(G(\cdot; A)) \text{ s.t. } c = \int_{x \in \mathcal{X}} 1(x \in A)\, dF(x).$$

For the mean utility case studied in this paper, $\mathcal{F}(G(\cdot; A)) \equiv \int U(w)\, dG(w; A)$, with $U(w) \equiv w$ denoting the mean outcome case. Similarly, $\mathcal{F}(G(\cdot; A)) \equiv G^{-1}(0.5; A)$ denotes the median maximization case. The form of the solution and the related distribution theory will of course change depending on the choice of $\mathcal{F}(\cdot)$.

Remark: Since our preferred analysis is fully nonparametric, it is not necessary to consider whether covariates should be entered as main effects only or via interactions. Secondly, since we will be ultimately concerned with the $\phi(x, \cdot)$'s– which are to be understood as expected outcome values at $X = x$, averaged over all other covariates– the issue of whether to "control for other covariates" is irrelevant here.

2.2 Parameters of interest

Treatment threshold: $\gamma(c)$ is a natural policy parameter of interest because it represents the treatment threshold for a specific budget c. Interestingly, it also equals $\rho'(c)$, which measures the shadow cost of the budget constraint, i.e., how much will the maximized expected outcome increase if the subsidy budget increases infinitesimally from c. Alternatively, $\gamma(c)$ measures the expected treatment effect on the "last" household made eligible for treatment under our budget-constrained rationing rule.

Value function: $\rho(c)$, the value function corresponding to the above optimization problem, represents the maximum mean outcome obtainable from a budget outlay of c. We consider ρ to be fundamentally a more important parameter than γ. It is useful for deciding on the budget outlay necessary for achieving a target mean level of outcome. It also represents a "first-best" scenario against which alternative suboptimal but easier-to-implement allocations can be compared.

Minimizing expenditure: The dual formulation of the problem is where the planner's objective is to achieve an expected outcome equal to b by allocating treatment based on covariates. The parameter of interest is the minimum amount of funds necessary to achieve b. This problem can be represented as

$$\min_{A \subset \mathcal{X}} \int_{x \in \mathcal{X}} 1\{x \in A\}\, dF(x) \tag{4}$$

subject to

$$\int_{x \in \mathcal{X}} [\phi(x, 1)\, 1(x \in A) + \phi(x, 0)\, 1(x \notin A)]\, dF(x) = b. \tag{5}$$

One can almost repeat the proof of proposition 1 to show that the optimal A will again be of the form $A^* = \{x : 1\{\theta(x) > \gamma(b)\}\}$ where $\gamma(b)$ is such that A^* satisfies (5). Note that by duality, the minimum value of (4) is simply $\rho^{-1}(b)$ where $\rho(\cdot)$ is defined in (2) and the inverse is well-defined because $\rho(\cdot)$ is monotone increasing. In particular, setting b equal to the currently observed mean outcome of an existing program, one can calculate how much resources could be saved by efficient allocation.

Restricted value function: Suppose $x_1 \subset x = (x_1, x_2)$ and consider situations where x_2 is an infeasible conditioner, either because conditioning on it is banned (e.g., x_2 is race)[3] or because observing it is costly (e.g., x_2 is income). Define

$$\xi(x_1, S) = E_{X_2|X_1=x_1}[\phi(x_1, x_2, S)]. \tag{6}$$

Then the covariate-restricted optimization problem becomes

$$\begin{aligned} &\max_{A \subset \mathcal{X}_1} \int_{x_1 \in \mathcal{X}_1} [\xi(x_1, 1) 1(x_1 \in A) + \xi(x_1, 0) 1(x_1 \notin A)] dF(x_1) \text{ s.t.} \\ c &= \int_{x_1 \in \mathcal{X}_1} 1(x_1 \in A) dF(x_1). \end{aligned}$$

Call the unrestricted maximum $\rho_{un}(c)$ and the restricted one, which conditions only on X_1, $\rho_{res}(c)$. The difference $\rho_{un}(c) - \rho_{res}(c)$ measures the welfare cost of these covariate restrictions on implementation. When gathering information on X_2 (e.g. income) is expensive and costs v per person, one can compare $\rho_{un}(c-v) - \rho_{res}(c)$ to decide on whether the extra survey cost for learning income is worthwhile to undertake. This is especially relevant for developing countries where the majority of hhds do not file tax returns, so that measuring wealth levels typically requires labor-intensive hhd surveys. Under-reporting of income and assets is also a common problem, especially if the population surveyed is aware of the existence of an eligibility threshold (Martinelli and Parker, 2007).

Note that all of the above are finite-dimensional parameters and therefore potentially estimable at the parametric rate. However, we will show below that although ρ (and its dual) is indeed estimable at parametric rates under appropriate conditions, the same does not appear to hold for γ.

2.3 Feasible policy and decision-theoretic issues

The "population problem" described above is not feasible in general because the distributions of (Y_1, X) and (Y_0, X) will typically be unknown to the planner. The feasible version of the problem can be studied via a decision theoretic approach as follows. Let $z_n = \{(Y_i, S_i, X_i), i = 1, 2, ...n\}$ be the data and let $Q_{Y_1,X|Z_n}(\cdot, \cdot|z_n)$ denote the planner's subjective probability distribution for the random variables (Y_1, X), given the data z_n. Similarly for (Y_0, X). The planner's treatment choice problem can now be formulated as

$$\max_{A(z_n)} \left\{ \begin{array}{c} \int U(y_1) 1(x \in A(z_n)) dQ_{Y_1,X|Z_n}(y_1, x|z_n) \\ + \int U(y_0) 1(x \notin A(z_n)) dQ_{Y_0,X|Z_n}(y_0, x|z_n) \end{array} \right\} \tag{7}$$

subject to

$$c = \int 1(x \in A(z_n)) dQ_{X|Z_n}(x|z_n).$$

The solution to this problem is exactly analogous to the solution implied by proposition 1 where all expectations and quantiles are now taken w.r.t. the subjective distribution $Q_{\cdot,\cdot|Z_n}$ rather than

[3] See Pope and Sydnor (2007) for some other relevant issues regarding legal restrictions on profiling.

the unknown population distributions. Now, the question is: what should the planner use as a subjective distribution $Q_{\cdot,\cdot|Z_n}$?

We focus on the case where the planner makes the "natural" choice, viz., he uses the empirical distribution $F_n(\cdot)$ of the observed data Z_n as $Q_{\cdot,\cdot|Z_n}$.[4] Since no distributional or functional form assumptions are made in our set-up, this corresponds to the predictive distribution of the random variables (Y_1, X) and (Y_0, X) in a nonparametric Bayesian approach under a certain uniformative prior (see appendix section 10.1). Thus "parameter uncertainty" is accounted for here via its effect on the planner's subjective expected utility, as in standard Bayesian decision theory. We will call this resulting $F_n(\cdot)$-based rule the empirical welfare maximizing rule (EWM). It is analogous to the conditional empirical success rule that Manski (2004) uses (without specifying an explicit decision theoretic justification) for the unconstrained problem under a known covariate distribution. The resulting solution from our EWM rule is the sample counterpart of the rule given in proposition 1 above.

Now, the welfare W resulting from the EWM rule is a random variable, in an ex ante, i.e., frequentist sense and, following a Manski (2004)-type analysis, one may try to bound its exact finite-sample *frequentist* expectation $\bar{W} \equiv E(W)$, where

$$W \equiv \phi(X,1)\, 1\left\{\hat{\theta}(X) > \hat{\gamma}(c)\right\} + \phi(X,0)\, 1\left\{\hat{\theta}(X) \leq \hat{\gamma}(c)\right\}.$$

The exact finite sample distribution of W, owing to its dependence on $\hat{\theta}(\cdot)$ and $\hat{\gamma}$, is analytically intractable, especially when X has continuous components. So we do not attempt to construct Manski-type finite-sample bounds here. Instead, we focus on the limiting value of this expectation as $n \to \infty$, which (under regularity conditions) will equal the same ρ that was defined w.r.t. the infeasible problem. This alternative interpretation of ρ– i.e., the limiting value of the (frequentist) expected welfare arising from the EWM rule– is probably the more relevant one in the context of the underlying decision problem.

Specifically, we construct a frequentist confidence interval for ρ, using the asymptotic distribution of the *estimated* value function– to be denoted by $\hat{\rho}_{EWM}$– arising from our EWM rule. This would, in turn, let us construct asymptotic frequentist confidence intervals (CI, henceforth) for both the dual value and the welfare loss from restricted covariate choice under EWM rule. The asymptotic approach is helpful for handling nonparametric regression functionals $\hat{\theta}(x)$ when X includes continuous components without making arbitrary functional form assumptions (e.g., $E[Y_1|X,S]$ has a probit form). The asymptotic approximation is also likely to be accurate in our application because we have a fairly large sample size (about 1000) relative to the effective number of parameters (1 for the value function). Indeed, sections 4, 5 and 6 below show that the dimension of conditioning covariates does not appear in the relevant asymptotic distributions. This distribution theory, though much harder to derive than in the unconstrained case, is easy to work with and can be used to construct large-sample frequentist CI for ρ and related functionals. But before

[4] Since X is allowed to have continuous components here, "using $F_n(\cdot)$" should be understood as inclusive of smoothing where necessary.

we develop this theory in detail, it would be interesting to consider how the EWM compares with alternative decision making approaches.

Toward that end, note that since the value function from the EWM rule and that from a general Bayes procedure (i.e. under a general prior) are both functions of z_n, one can compare them from a purely frequentist perspective, as in the predictive inference literature, c.f. Smith (1998), Cox (1975). One way to do this here is to focus on how fast the difference in the respective value functions $\hat{\rho}_{Bayes}$ and $\hat{\rho}_{EWM}$ falls with the sample size n, in terms of frequentist probability and it turns out in general that they will differ by $O_p\left(\frac{1}{n}\right)$ (see appendix sec 10.1 for a demonstration). Thus the asymptotic frequentist CI for $p\lim_{n\to\infty}\hat{\rho}_{Bayes}$ which equals $p\lim_{n\to\infty}\hat{\rho}_{EWM}$– both equalling ρ– will not differ numerically from each other up to first order because, as we will show below in theorem 4,

$$\hat{\rho}_{EWM} - p\lim_{n\to\infty}\hat{\rho}_{EWM} = O_p\left(\frac{1}{\sqrt{n}}\right) >> O_p\left(\frac{1}{n}\right)$$

for large n. Thus, not only does the EWM approach correspond to a Bayesian one with an uninformative prior in a nonparametric setting, the Bayesian (under other priors) and EWM distinction is also unimportant in regards to the (frequentist) asymptotic CI, on which we focus our analysis. Nonetheless, from a finite-sample inference perspective, it might be instructive to compare the Bayesian CI's under alternative informative priors and with the EWM ones, using smaller and smaller subsets of our sample data. This exercise is reserved for future research.

Some non-subjective alternatives to the Bayesian approach like the minmax regret criterion (MRC) have been proposed in the literature (c.f., Savage (1951), Manski (2004, 2005)). The MRC seems somewhat unsuited for our problem because our constraint is in an expectations form where the expectation is taken w.r.t. the unknown distribution of the covariates. A subjective expectations approach can handle this quite naturally. However, as a general issue, it would be interesting to explore MRC based treatment choice analysis under budget constraints– a problem that we also leave for future research.[5] A different but related issue, not addressed in the paper, is whether the planner, in addition to being risk-averse and Bayesian, should also be *averse* to parameter uncertainty beyond its effect on the subjective expected utility. This distinction between ambiguity aversion and Bayesian risk-aversion has been researched in economic theory (c.f., Ahn et al (2009) for a summary) and has been addressed by Manski (2005) in the context of unconstrained treatment choice. We provide a brief discussion in the appendix section 10.2.

3 Related Literature and Contributions

Our paper contributes to a relatively recent and growing literature on treatment choice. For related works in econometrics see Chamberlain (2000), Chamberlain and Imbens (2003), Dehejia (2001),

[5] One alternative is to assume that the planner knows the population marginal of X, which may hold in some situations. Another, less attractive, approach might be to redefine the objective as mean outcome net of costs. But this requires putting a controversial monetary value on possibly nonmonetary outcomes. Besides, a (roughly) fixed budget appears to be the more realistic set-up for subsidy disbursement in poor countries, c.f., reference [36] below.

Manski (2001, 2004, 2005) and Hirano and Porter (2008). In labor economics see Berger, Black and Smith (2001), Frolich (2006), Behnke, Frolich and Lechner (2008), Lechner and Smith (2007), the papers in Eberts, O'Leary and Wandner (ed., 2002) and references therein. In medical statistics, see Gunter, Zhu and Murphy (2007) and Collins, Murphy, Nair and Strecher (2007). The present paper differs from the above works substantively as it studies welfare maximizing allocations under aggregate budget constraints–which, to our knowledge, has not been explored analytically in the literature.[6] Such constraints make the problem both substantively more realistic and analytically richer.

A small set of recent papers have addressed the related problem of input allocation in production processes–c.f., Graham, Imbens and Ridder (2005, 2006) and Bhattacharya (2006). The present paper differs from the above ones in that it analyzes efficient allocation based on both discrete and continuous conditioners, which makes the problem nonparametric in a nontrivial way. Deriving the asymptotic properties of the relevant estimates requires independent analysis owing to the lack of smoothness of the corresponding population moment conditions with respect to the underlying infinite-dimensional parameters. In particular, methods described in Newey-McFadden's Handbook of Econometrics chapter (NM, henceforth) or in CLV are not directly applicable here.

Recently, Hahn, Hirano and Karlan (2007) have considered the problem of designing an experiment with a view to minimize the variance of the estimated unconditional ATE, estimated from it. Their goal is therefore fundamentally different from the present paper. In principle, one could construct an HHK (2007) type experimental design for efficient estimation of the parameters we introduce in the present paper.

In ongoing work, Bhattacharya, Chandra and Chen (2007) are investigating optimal covariate-based allocation of a *continuous* resource, e.g., Medicare spending on heart-attack patients, using observational data and instrumental variations. There the distribution theories are very different due to endogeneity and more structure is needed on the underlying production function to guarantee unique solutions to a planner's optimization problem.

A few recent studies have used experimental data to estimate the parameters of dynamic structural models of behavior and utilized the estimates to simulate the effects of counterfactual policy interventions (c.f. Attanasio, Meghir and Santiago (2006) and Duflo, Hanna and Ryan (2007)). Mahajan, Tarozzi, Yoong and Blackburn (2008) discuss identification and estimation of a static structural model of ITN adoption using observational data alone and use the estimated parameters to perform counterfactual policy analysis. Todd and Wolpin (2006, 2007) discuss the estimation of structural models of behavior using pre-program data and compare predictions of their estimated model with subsequent experimental data. In contrast, we propose here a methodology through which experimental data can be used directly to infer the welfare-maximizing targeting of programs under budget constraints.

[6] Manski (2005) studies planning problems which satisfy 'separability' and specifically mentions (page 10–11) budget constraints as a situation where separability is violated and, consequently, not studied by him.

4 Estimation

We now formally define our estimates corresponding to the EWM approach described in section 2.3. Suppose $X \equiv \left(X^d, X^c\right)$ where X^d contains the discrete components of X and X^c is a p-variate vector of the continuous components of X with support $\mathcal{X}^c$ and density $f(\cdot)$. Let $K(\cdot)$ be any standard density kernel and σ_n a sequence of bandwidths converging to zero at an appropriate rate, to be specified later, as $n \to \infty$. Define the preliminary quantities

$$\begin{aligned}
\hat{\mu}\left(X_i\right) &= \frac{1}{n-1}\sum_{j\neq i}\frac{y_i s_i}{\sigma_n^p}K\left(\frac{X_j^c - X_i^c}{\sigma_n}\right)1\left(X_j^d = X_i^d\right) \\
\hat{\nu}\left(X_i\right) &= \frac{1}{n-1}\sum_{j\neq i}\frac{y_i\left\{1-s_i\right\}}{\sigma_n^p}K\left(\frac{X_j^c - X_i^c}{\sigma_n}\right)1\left(X_j^d = X_i^d\right) \\
\hat{\pi}\left(X_i\right) &\equiv \frac{1}{n-1}\sum_{j\neq i}\frac{s_i}{\sigma_n^p}K\left(\frac{X_j^c - X_i^c}{\sigma_n}\right)1\left(X_j^d = X_i^d\right) \\
\hat{\delta}(X_i) &\equiv \frac{1}{n-1}\sum_{j\neq i}\frac{1-s_i}{\sigma_n^p}K\left(\frac{X_j^c - X_i^c}{\sigma_n}\right)1\left(X_j^d = X_i^d\right).
\end{aligned}$$

Now $\hat{\theta}\left(X_i\right)$ can be defined in terms of the above quantities as

$$\hat{\theta}\left(X_i\right) = \frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)} - \frac{\hat{\nu}\left(X_i\right)}{\hat{\delta}(X_i)}.$$

The natural estimates of our parameters of interest would have been given by solutions to the equations

$$\begin{aligned}
0 &= 1 - c - \frac{1}{n}\sum_{i=1}^{n} 1\left\{\hat{\theta}\left(X_i\right) \leq \hat{\gamma}\right\}, \\
0 &= \hat{\rho} - \frac{1}{n}\sum_{i=1}^{n}\hat{\phi}\left(X_i, 1\right) + \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}\left(X_i\right) \times 1\left\{\hat{\theta}\left(X_i\right) \leq \hat{\gamma}\right\}.
\end{aligned}$$

Notice that the first sample moment condition above is not differentiable in either $\hat{\theta}(\cdot)$ or in $\hat{\gamma}$, so that usual first-order expansions cannot be used. More interestingly, it turns out that even the population analog of the first moment condition is not differentiable in the nonparametric component. Indeed, the analogous population moment conditions are given by

$$\begin{aligned}
0 &= 1 - c - \int_{x\in\mathcal{X}} 1\left\{\theta\left(x\right) \leq \gamma\right\} dF\left(x\right), \\
0 &= \rho - E\left[\phi\left(X, 1\right)\right] + \underbrace{\int \theta\left(x\right) 1\left\{\theta\left(x\right) \leq \gamma\right\} dF\left(x\right)}_{\zeta},
\end{aligned}$$

where $\theta(\cdot)$ and $\phi(\cdot)$ should be thought of as preliminary parameters which are estimated in a nonparametric first-step. Now notice that the first moment condition is differentiable in the scalar γ if $\theta(X)$ has a density but not functionally differentiability in $\theta(\cdot)$, owing to the presence of the indicator. This makes it infeasible to directly apply the methods of e.g. CLV which requires differentiability of all the population moment conditions with respect to both the finite and the infinite dimensional parameters.

So we use further smoothing to construct our estimators. Suppose that $\theta(X)$ is bounded between $[-M, M]$ on the support of X. Then choose a symmetric (about zero) kernel $L(\cdot)$ with bounded support, w.l.o.g. $[-1, 1]$, the corresponding C.D.F. kernel $\bar{L}(t) = \int_{-1}^{t} L(s)\,ds$ for each $t \in [-1, 1]$ and a sequence of bandwidths h_n converging (slowly) to zero as $n \to \infty$. The C.D.F. kernel simply converts the indicator function $1\{\theta(x) \leq \gamma\}$ to a smooth function that changes smoothly from 0 to 1 as $\theta(x) - \gamma$ changes sign from positive to negative in finite samples but approaches the indicator as $n \to \infty$.

Now define $\hat{\gamma}$, and $\hat{\rho}$ by

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)-\{1-c\}\right\} &= 0,^{7} \\
\hat{\rho}-\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}(X_i,1)+\hat{\theta}(X_i)\left\{1-\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)\right\}\right] &= 0. \qquad (8)
\end{aligned}$$

For future use, also define

$$\begin{aligned}
\hat{\zeta} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}(X_i)\times\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right), \\
\hat{E}[\phi(X,1)] &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mu}(X_i)}{\hat{\pi}(X_i)},
\end{aligned}$$

so that $\hat{\rho} = \hat{E}[\phi(X,1)] - \hat{\zeta}$.

The smoothing applied in (8) is similar in spirit to Horowitz's (1992) analysis of the smoothed maximum score. But in that problem, the finite-dimensional parameter of interest does not explicitly depend on any infinite-dimensional underlying parameter. In contrast, here the key parameters of interest, viz., γ and ρ, are based on the infinite-dimensional component $\theta(\cdot)$ through population moments that are not smooth in $\theta(\cdot)$. Thus the present estimators lie at the intersection of classical 2-step semiparametric estimators and smoothing-based estimators for countering non-differentiability. This makes both the results and the proofs substantially different from both strands of the literature.

5 Large sample theory

The discrete regressors will not play any substantive roles in our analysis; so we will drop them in our proofs and put them back into our final results at the end. In our proofs, the notation $\tilde{\theta}(x)$ and $\tilde{\gamma}$ will be used to denote values intermediate between $\hat{\theta}(x)$ and $\theta(x)$ and $\hat{\gamma}$ and γ, respectively;

[7] Note that under the resulting rule, the budget constraint will be satisfied only in an approximate sense and, in general, not exactly.

M_1 and $M(x)$ will denote a bounded positive constant and a uniformly bounded positive function respectively, whose actual values may be different in different places. The latter will be used in the expressions for upper bounds for various quantities which appear in the proof. Let

$$\pi(x) \equiv \Pr(s = 1|x = X), \ \mu(x) \equiv E(Y|S = 1, X = x) \times \pi(x)$$

and

$$\delta(x) \equiv \Pr(s = 0|x = X), \ \nu(x) \equiv E(Y|S = 0, X = x) \times \delta(x).$$

Assumptions

A0(i) (Y_i, X_i, S_i) $i = 1, 2, ...n$ is a random sample, $\theta(X)$ is continuously distributed.

A0(ii) S is randomly allocated[8] so that

$$\begin{aligned} \frac{\mu(x)}{\pi(x)} - \frac{\nu(x)}{\delta(x)} &= E(Y|S = 1, X = x) - E(Y|S = 0, X = x) \\ &= E(Y(1)|X = x) - E(Y(0)|X = x) \\ &\equiv ATE(x) \end{aligned}$$

where $Y(1)$ and $Y(0)$ are the conventional notations for the outcome with and without treatment respectively for an household.

A1 Conditional on every value x^d assumed by the discrete regressors, the support $\mathcal{X}^c$ of the continuous components X^c is a p-dimensional compact set and the density of X^c satisfies that $f(x) \geq \delta > 0$ for all $x \in X^c$. Furthermore, the density is q-times continuously differentiable with the derivatives uniformly bounded on $\mathcal{X}^c$.

A2 For some $M > 0$, $\theta(x) \in [-M, M]$ for every $x \in \mathcal{X}$.

A3 $K(\cdot)$ is an qth order p-dimensional bounded kernel, with $q > p$ and the bandwidth sequence σ_n satisfying (i) $\sigma_n \to 0$ (ii) $\sqrt{n}\sigma_n^q \to 0$.

A4 The kernel $\bar{L}(\cdot)$ is uniformly bounded with a bandwidth sequence $h_n \to 0$ and $nh_n \to \infty$.

Assumptions A0 (i) and (ii) define the set-up. A1 and A2 are somewhat restrictive but are routinely assumed (c.f. Hirano, Imbens and Ridder (2003), assumption 2). In fact, we can simply redefine the problem such that we are designing allocations based only on those values of X where they hold. Assumption A3 part (i) is standard. Assumption A3 part (ii) is an "undersmoothing" requirement, which is commonly used in semiparametric problems for bias removal; it is also a key condition for assumption B10 below (c.f. NM, lemma 8.10).

[8] It is enough to assume that S is allocated randomly conditional on X, for the theory to go through. But in our empirical example and, indeed, most real life social experiments, treatment is in fact fully randomized.

5.1 Consistency of $\hat{\gamma}$

Let $\hat{f}_\theta(u)$ and $f_\theta(u)$ denote respectively the estimated and the true density of $\theta(X)$ at u.

B1. $\sup_{x\in\mathcal{X}}\left|\hat{\theta}(x)-\theta(x)\right| = O_p\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}$.

B2. $\sup_{u\in[-M,M]}\left|\hat{f}_\theta(u)-f_\theta(u)\right| = o_p(1)$

B3 (i) The first derivative of kernel $\bar{L}(\cdot)$, denoted by L, is also uniformly bounded.

B4. (i) $h_n\to 0$, $nh_n\to\infty$, $\sqrt{n}h_n^2\to\infty$ and $n^{1/4}\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}\to 0$.

Sufficient low level conditions for B1 and B2 are fairly standard. In particular, for B1 c.f. Hansen (2008). For B2, c.f. Pagan and Ullah (1999) theorem 2.8.

We are now ready to state and prove the first consistency result with one additional assumption.

B5. The density of $\theta(X)$ is strictly positive on an open set containing γ

Theorem 1 *Under assumptions A0-A3, A4(i), B1, B2, B3(i) and B4(i) and B5, we have that*

$$\hat{\gamma}-\gamma = o_p(1)$$

Proof. Appendix ■

5.2 Distribution Theory for $\hat{\gamma}$

Assume that $\bar{L}(\cdot)$ is differentiable and let

$$\hat{f}_{\hat{\theta}}(t) = \frac{1}{nh_n}\sum_{i=1}^{n} L\left(\frac{t-\hat{\theta}(X_i)}{h_n}\right).$$

The asymptotic behavior of $\hat{f}_{\hat{\theta}}(t)$ will be useful for our distribution theories. Toward that end, add to the above assumptions that:

A4 (ii) The kernel $\bar{L}(\cdot)$ has two derivatives which are also uniformly bounded.

B4 (ii) $\frac{1}{h_n^2}\times\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}\to 0$.

The following first-order expansion for $\hat{\gamma}$ will be used for deriving the distribution theory for $\hat{\gamma}$:

$$\begin{aligned}
&(\hat{\gamma}-\gamma)\\
=\;& \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\underbrace{\left\{F_\theta(\gamma)-\frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{u-\theta(X_i)}{h_n}\right)\right\}}_{T_{1n}}\\
&+\left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\underbrace{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\bar{L}\left(\frac{u-\theta(X_i)}{h_n}\right)-\bar{L}\left(\frac{u-\hat{\theta}(X_i)}{h_n}\right)\right)\right]}_{T_{2n}}.
\end{aligned}$$

The proof will proceed in three steps: step 1 is that the multiplier $\left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}$ converges in probability to $\{f_\theta(\gamma)\}^{-1}$. Step 2 is that the term T_{1n} will be $O_p\left(\frac{1}{\sqrt{n}}\right)$. Finally in step 3 we will show, using

U-statistic type decompositions, that the term T_{2n} will be $O_p\left(\frac{1}{\sqrt{nh_n}}\right)$. Thus, we will eventually get that $\sqrt{nh_n}\left(\hat{\gamma}-\gamma\right)$ will converge to normal distribution.

The following additional assumptions will be used in the proof.

B7. For some $r \geq 2$, the density of $\theta\left(X\right)$ is $\left(r-1\right)$ times continuously differentiable, the derivative is bounded and Lipschitz in a neighborhood of γ and $nh_n^{2r+1} \rightarrow \lambda < \infty$. Denote the above derivative at γ by $f_\theta^{(r-1)}\left(\gamma\right)$.

B8. $\frac{\ln n}{\sqrt{n}\sigma_n^p h_n^{3/2}} \rightarrow 0$ and $\sigma_n^{2q}\frac{\sqrt{n}}{h_n^{3/2}} \rightarrow 0$

B9. $L\left(\cdot\right)$ is symmetric around zero and has bounded support $[-1,1]$, is of order r and $\int_{-\infty}^{\infty} L^2\left(u\right)du = \int_{-1}^{1} L^2\left(u\right)du < \infty$.

B10. $Var(Y|S=1)$ and $Var\left(Y|S=0\right)$ are finite.

B11. $\sqrt{n}\sup_{x\in\mathcal{X}}\left\|\left\{\hat{\mu}\left(x\right)-\mu(x)\right\}\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\| = o_p\left(1\right)$ and $\sqrt{n}\sup_x\left\|\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\|^2 = o_p\left(1\right)$.

Assumption B11 is also a well-known requirement for $\sqrt{n}$ -normality for semiparametric estimators (c.f. NM, section 8.3).

Theorem 2 *Under assumptions A0-A4 and B1-B11, we have that*

$$\sqrt{nh_n}\left(\hat{\gamma}-\gamma\right) \xrightarrow{d} N\left(\beta, \frac{\tau^2\left(\gamma\right)+\omega^2\left(\gamma\right)}{f_\theta\left(\gamma\right)}\int_{-1}^{1} L^2\left(u\right)du\right),$$

where

$$\begin{aligned}
\tau^2\left(\gamma\right) &= E\left\{\left\{\frac{\delta\left(X\right)Y\left(1-S\right)-\nu\left(X\right)\left(1-S\right)}{\delta^2\left(X\right)}f\left(X\right)\right\}^2 |\theta\left(X\right)=\gamma\right\} \\
\omega^2\left(\gamma\right) &= E\left\{\left\{\frac{\pi\left(X\right)YS-\mu\left(X\right)S}{\pi^2\left(X\right)}f\left(X\right)\right\}^2 |\theta\left(X\right)=\gamma\right\} \\
\beta &= \left(-1\right)^{r+1}\frac{\sqrt{\lambda}}{r!}\times f_\theta^{(r-1)}\left(\gamma\right)\int_{-1}^{1} u^r L\left(u\right)du.
\end{aligned}$$

Proof. Appendix ■

Incorporating the discrete regressors back into the analysis is straightforward. Let $X = \left(X^c, X^d\right)$, let the discrete regressor (vector) X^d assumes values $a_1, ..., a_J$ and suppose $f_{X^c|X^d=a_j}\left(x|a_j\right)$ denotes the density of X^c, conditional on $X^d = a_j$. Then we simply replace

$$\begin{aligned}
\tau^2\left(\gamma\right) &= E\left\{\left\{\frac{\delta\left(X\right)Y\left(1-S\right)-\nu\left(X\right)\left(1-S\right)}{\delta^2\left(X\right)}f\left(X^c, X^d\right)\right\}^2 |\theta\left(X\right)=\gamma\right\} \\
\omega^2\left(\gamma\right) &= E\left\{\left\{\frac{\pi\left(X\right)YS-\mu\left(X\right)S}{\pi^2\left(X\right)}f\left(X^c, X^d\right)\right\}^2 |\theta\left(X\right)=\gamma\right\},
\end{aligned}$$

where $f\left(x^c, x^d\right) \equiv \sum_{j=1}^{J} f_{X^c|X^d=a_j}\left(x^c|a_j\right)1\left(x^d=a_j\right)$.

5.3 Consistency for $\hat{\rho}$

Theorem 3 *Under assumptions A0-A4 and B1-B11, we have that*

$$\hat{\rho} - \rho = o_p(1).$$

Proof. Appendix ■

5.4 Distribution theory for $\hat{\rho}$

Recall that $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} \hat{\phi}(X_i, 1) - \hat{\zeta}$. The first term will be analyzed via lemma 3 in the appendix and the following expansion will be used to show that $\hat{\zeta}$ is asymptotically $\sqrt{n}$-normal:

$$\begin{aligned}
& \hat{\zeta} - \zeta \\
= & \underbrace{\frac{1}{n}\sum_{i=1}^{n} \theta(X_i)\bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right) - \zeta}_{T_{1n}} \\
& \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i) - \theta(X_i)\right\}\left\{\bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right) - \frac{1}{h_n}\theta(X_i)L\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\}}_{T_{2n}} \\
& \underbrace{+(\hat{\gamma} - \gamma)\frac{1}{nh_n}\sum_{i=1}^{n}\theta(X_i)L\left(\frac{\gamma - \theta(X_i)}{h_n}\right)}_{T_{3n}} + R_n.
\end{aligned}$$

The proof will show that R_n is $o_p\left(\frac{1}{\sqrt{n}}\right)$ and T_{1n}, T_{2n} and T_{3n} are all $O_p\left(\frac{1}{\sqrt{n}}\right)$.

The following additional assumptions will be used.

B4 (iii) $\frac{\sqrt{n}}{h_n^2} \times \left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2} + \sigma_n^q\right\}^2 \to 0$ which is implied by $\frac{\sqrt{n}}{h_n^2} \times \sigma_n^{2q} \to 0$ and $\left(\frac{\ln n}{\sqrt{n}\sigma_n^p h_n^2}\right) \to 0$.

B12. $nh_n^6 \to \infty$, r of assumption B7 is at least 4 and $nh_n^{2r} \to 0$.

Theorem 4 *Under assumptions A0-A5, B1-B12,*

$$\begin{aligned}
& \sqrt{n}(\hat{\rho} - \rho) \\
= & -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\psi_{1i} + \psi_{2i} + \psi_{3i} - \psi_{4i}\} \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{E(S|X_i) \times Y_i S_i - E(SY|X_i) \times S_i}{\{E(S|X_i)\}^2} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\psi_{1i} &= \gamma\{F_\theta(\gamma) - 1(\theta(X_j) \le \gamma)\} \\
\psi_{2i} &= \theta(X_i) \times 1\{\theta(X_i) \le \gamma\} - \zeta \\
\psi_{3i} &= 1(\theta(X_i) \le \gamma) \times \frac{\pi(X_i)Y_i S_i - \mu(X_i)S_i}{\pi^2(X_i)} \times f_X(X_i) \\
\psi_{4i} &= 1(\theta(X_i) \le \gamma) \times \frac{\delta(X_i)Y_i(1 - S_i) - \nu(X_i)(1 - S_i)}{\delta^2(X_i)} \times f_X(X_i).
\end{aligned}$$

It follows by an ordinary CLT (under standard second moment restrictions) that $\sqrt{n}\left(\hat{\rho}-\rho\right)$ will be mean-zero normal.

Proof. The proof works by showing that

$$\sqrt{n}\left\{\hat{\zeta}-\zeta\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\psi_{1i}+\psi_{2i}+\psi_{3i}-\psi_{4i}\right\}+o_p\left(1\right),$$

where the ψ terms are described above and then combining this with an asymptotic expansion of $\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}\left(X_i,1\right)$. Details are in the appendix. ■

To incorporate the discrete regressors back into the analysis, we simply replace the terms $f_X\left(X_i\right)$ in ψ_{3i} and ψ_{4i} by $f\left(X_i^c,X_i^d\right)\equiv\sum_{j=1}^{J}f_{X^c|X^d=a_j}\left(X_i^c|a_j\right)1\left(X_i^d=a_j\right)$, where $f_{X^c|X^d=b}\left(a|b\right)$ denotes the density of X^c at a, conditional on $X^d=b$.

The final variance can be consistently estimated using sample cross-products, under standard conditions for the WLLN.

Remark 1 *It may be noted here that the estimation error in $\hat{\theta}\left(\cdot\right)$ affects the distribution of $\hat{\rho}$ through the terms ψ_{3i} and ψ_{4i}.*

5.5 Distribution theory for dual

Recall that the value function for the dual problem $\delta\left(b\right)$ represents the smallest fraction of households who have to be assigned to treatment (optimally) to guarantee that the expected mean outcome is at least b. In other words, $\rho\left[\delta\left(b\right)\right]$ equals b, where $\delta\left(b\right)$ plays the role of c in the primal problem. From a standard first-order expansion argument, it follows that

$$\sqrt{n}\left(\hat{\delta}\left(b\right)-\delta\left(b\right)\right)=-\frac{\sqrt{n}\left(\hat{\rho}\left(\delta\left(b\right)\right)-\rho\left\{\delta\left(b\right)\right\}\right)}{\rho'\left\{\delta\left(b\right)\right\}}+o_p\left(1\right),$$

where $\rho\left\{\delta\left(b\right)\right\}=b$. Since $\rho\left(c\right)=E\left\{\phi\left(X,1\right)\right\}-\int_{-\infty}^{G^{-1}(1-c)}tdG\left(t\right)$, it follows that $\rho'\left(c\right)$ equals $G^{-1}\left(1-c\right)$ which is simply $\gamma\left(c\right)$. Replacing, we get that

$$\sqrt{n}\left(\hat{\delta}\left(b\right)-\delta\left(b\right)\right)=-\frac{\sqrt{n}\left(\hat{\rho}\left(\delta\left(b\right)\right)-\rho\left\{\delta\left(b\right)\right\}\right)}{\gamma\left\{\delta\left(b\right)\right\}}+o_p\left(1\right),$$

from which the asymptotic normality of $\sqrt{n}\left(\hat{\delta}\left(b\right)-\delta\left(b\right)\right)$ follows.

Remark : The qualitative difference between the asymptotic distributions of $\hat{\gamma}$ and $\hat{\rho}$ is somewhat intriguing. It is caused jointly by the facts that the moment condition defining γ is nonsmooth in $\theta\left(\cdot\right)$ and also that $\theta\left(\cdot\right)$ is unknown. If $\theta\left(\cdot\right)$ were known, then realizations of $\theta\left(X\right)$ would be observed and so its estimated quantile would be$\sqrt{n}$-normal. Conversely, if the moment condition were smooth and $\theta\left(\cdot\right)$ unknown, then a CLV– type analysis would lead to $\sqrt{n}$-normality for $\hat{\gamma}$ under regularity conditions. One way to interpret the difference between the asymptotic distributions of $\hat{\gamma}$ and $\hat{\rho}$ is to note that $\gamma=G\left(\theta\right)^{-1}\left(1-c\right)$ and $\rho=\int_{1-c}^{1}G\left(\theta\right)^{-1}\left(u\right)du$ where $G\left(\theta\right)$ represents the c.d.f. of $\theta\left(X\right)$. This suggests that γ is the value at a point of a nonparametric function while ρ is its integral. Thus γ is somewhat analogous to the value of a demand function at a price whereas ρ is akin to the (approximate) consumer surplus (c.f. NM (1994), page 2195) calculated from that demand curve. So it is likely that $\hat{\gamma}$ would behave like a purely nonparametric estimator whereas $\hat{\rho}$ behaves like a parametric one. However, we recognize that this analogy is not perfect because $G(\theta)^{-1}(\cdot)$ is not a standard density or conditional mean function, since $\theta(\cdot)$ is unknown. It is also interesting to observe that the mean of $\theta(X)$ is estimable at the parametric rate, i.e., $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i)-E[\theta(X)]\right\}=O_p(1)$, which can be shown by using U-statistic type results. This may suggest that a quantile of $\theta(X)$ should also be estimable at the parametric rate. But this assertion remains to be either proved or disproved. What we have shown so far is that there exists *one* estimator of γ that converges slower than the parametric rate while the corresponding estimator for ρ has the $\sqrt{n}$-normal distribution, asymptotically.

Bias Removal: Notice that we have always used bias-removal in our analysis above. This is not necessary and may, in fact increase the MSE for γ estimation. From the proof of theorem 2, it is easy to see that if the density $f_\theta(\cdot)$ has bounded derivatives up to order $(r-1)$, then the bias of $(\hat{\gamma}-\gamma)$ is given by

$$\beta=(-1)^{r+1}\frac{h_n^r}{r!}\times f_\theta^{(r-1)}(\gamma)\int_{-1}^{1}u^r L(u)\,du+o(h_n^r).$$

Using the formula for the variance, one gets that the MSE is given by

$$h_n^{2r}\times\underbrace{\left[\frac{f_\theta^{(r-1)}(\gamma)}{r!}\int_{-1}^{1}u^r L(u)\,du\right]^2}_{C}+\frac{1}{nh_n}\underbrace{\left[\frac{\tau^2(\gamma)+\omega^2(\gamma)}{f_\theta(\gamma)}\int_{-1}^{1}L^2(u)\,du\right]}_{B},$$

implying an MSE minimizing bandwidth choice of $h_n=\lambda^* n^{-\frac{1}{2r+1}}$, where $\lambda^*=\left(\frac{C}{2rB^2}\right)^{\frac{1}{2r+1}}$. Horowitz (1992) calculates analogous quantities for his smoothed maximum score estimator and discusses both estimation of λ^* and adjusts the asymptotic theory of the eventual estimators to allow for an estimated λ^*.

The above choice of h_n does not work for theorem 4 because (c.f. step 6A in the proof) for this choice of h_n, we have that $\sqrt{n}h_n^r=O\left(n^{\frac{1}{2(2r+1)}}\right)$ which blows up to $+\infty$ and so we cannot have a $\sqrt{n}$-rate for $\hat{\zeta}$ and thus for $\hat{\rho}$. So we need to choose h_n to be smaller than the one that is MSE-optimal for γ.

6 Parametric Analysis

It is useful to compare our results from a nonparametric analysis to a benchmark parametric model which is easier to estimate and thus potentially more useful for applied work. The plug-in approach in the parametric case *does not* have a Bayesian interpretation, unlike in the nonparametric case and is not justifiable, in general, from a decision theoretic standpoint. However, the difference in value function from a plug-in approach and a Bayesian approach will still be of the order $O_p\left(\frac{1}{n}\right)$ and higher order corrections can be made to the plug-in approach that yield similar $O_p\left(\frac{1}{n}\right)$ improvements in finite samples (see appendix). Nonetheless, the parametric approach has the insurmountable limitation that it is susceptible to mis-specification of functional form, leading to a suboptimal value function, even in the limit. In our application we show the results for both parametric and nonparametric specifications and estimate the asymptotic welfare loss arising from the potential mis-specification of the parametric model.

To get an idea for the distribution theory, suppose $\theta(x)$ is parametrically specified as $G(x,\beta)$, where $G(\cdot)$ is known; typically β (the so-called "pseudo-true value") can be estimated at parametric rates using, say, GMM. For estimation of γ and ρ resulting from the plug-in approach, we will still use smoothing with the c.d.f. kernel $\bar{L}(\cdot)$ to handle the nonsmoothness, e.g.,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{\hat{\gamma}-G\left(X_i,\hat{\beta}\right)}{h_n}\right)-\{1-c\}\right\}=0$$

For some specific functional forms of $G(\cdot,\cdot)$, e.g., a linear one, the function $h(\beta)=\int_{-M}^{M}1\{G(x,\beta)\leq\gamma\}\,dF(x)$ may be differentiable in β and then no smoothing would be necessary; but smoothing-based methods are more generally applicable and so we focus on that.

The key result is that both γ and ρ can be estimated at the $\sqrt{n}$-rate. To see this, recall the asymptotic expansion for $\hat{\gamma}$:

$$\begin{aligned}
&\sqrt{n}\left(\hat{\gamma}-\gamma\right)\\
=\;&\left\{\hat{f}_{\hat{\theta}}\left(\tilde{\gamma}\right)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{F_{\theta}\left(\gamma\right)-\bar{L}\left(\frac{\gamma-G\left(X_i,\beta\right)}{h_n}\right)\right\}\\
&+\left\{\hat{f}_{\hat{\theta}}\left(\tilde{\gamma}\right)\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma-G\left(X_i,\beta\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma-G\left(X_i,\hat{\beta}\right)}{h_n}\right)\right]\right\}.
\end{aligned}$$

Using similar steps as in the proof of theorem 2 in the appendix, the first term is asymptotically normal with mean equal to

$$\lim_{n\rightarrow\infty}\sqrt{n}h_n^r\times\left[\frac{(-1)^{r+1}f_{\theta}^{(r-1)}\left(\gamma\right)\times\int_{-1}^{1}u^r L\left(u\right)du}{r!}\right]+o\left(\sqrt{n}h_n^r\right)$$

which is finite if $\lim_{n\rightarrow\infty}\sqrt{n}h_n^r<\infty$.

As for the second term, (and this is what makes $\hat{\gamma}$ a $\sqrt{n}$-consistent estimator in the parametric case) notice that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma-G\left(X_i,\beta\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma-G\left(X_i,\hat{\beta}\right)}{h_n}\right)\right]\\
=\;&\sqrt{n}\left(\hat{\beta}-\beta\right)'\frac{1}{nh_n}\sum_{i=1}^{n}\nabla G\left(X_i,\beta\right)L\left(\frac{\gamma-G\left(X_i,\beta\right)}{h_n}\right)\\
&+T_n,
\end{aligned}$$

where

$$|T_n| \leq M \frac{n \left\| \hat{\beta} - \beta \right\|^2}{2\sqrt{n} h_n^2} \frac{1}{n} \sum_{i=1}^{n} M_1(X_i) L' \left(\frac{G\left(X_i, \tilde{\beta}\right) - \gamma}{h_n} \right),$$

with M a fixed positive constant and $M_1(X)$ a uniformly bounded function. Since $\sqrt{n}\left(\hat{\beta} - \beta\right) = O_p(1)$, by assumptions B4(i) and A4 (ii), the RHS of the previous display goes to zero if $nh_n^4 \to \infty$. Then we have that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\bar{L}\left(\frac{\gamma - G(X_i, \beta)}{h_n} \right) - \bar{L}\left(\frac{\gamma - G\left(X_i, \hat{\beta}\right)}{h_n} \right) \right] \\ = & \left[\sqrt{n}\left(\hat{\beta} - \beta\right)' \nabla G(X_i, \beta) \right] \times f_{G(X,\beta)}(\gamma) + o_p(1). \end{aligned}$$

This implies that $\sqrt{n}(\hat{\gamma} - \gamma)$ will converge to a zero mean normal if $nh_n^{2r} \to 0$ and $nh_n^4 \to \infty$ and when the density of $G(X, \beta)$ has uniformly bounded derivatives up to order $(r-1)$ where $r \geq 3$. The result for $\hat{\rho}$ will follow.

7 Extensions

7.1 Partial subsidy and partial take-up

In-kind transfer programs are sometimes characterized by two features– viz., (i) vouchers are distributed among eligible households and these have to be redeemed for getting the product and (ii) the amount of subsidy is not large enough that obtaining the product and selling it in the open market afterwards is profitable for every eligible household. The first feature implies that voucher cashing may not translate 1-for-1 into using the product– because households can obtain the product and sell it outside. The second feature implies that voucher cashing itself can be less than 100%, which loosens the planner's budget constraint. The voucher allocation problem now becomes

$$\max_{A \subset \mathcal{X}} \int_{x \in \mathcal{X}} [\phi(x, 1) 1(x \in A) + \phi(x, 0) 1(x \notin A)] \, dF(x)$$

subject to the budget constraint

$$c = \int_{x \in \mathcal{X}} h(x) \times 1(x \in A) \, dF(x),$$

where $h(x)$ denotes the probability that an x-type household cashes the voucher upon getting it (though not necessarily uses the good itself) and $\phi(x, \cdot)$ has the same interpretation as in section 2. The solution will have the same form as described in proposition 1, i.e., $A^* = \{x : \phi(x, 1) - \phi(x, 0) > \gamma\}$, but γ will now be determined by

$$c = \int_{x \in \mathcal{X}} h(x) \times 1(\phi(x, 1) - \phi(x, 0) > \gamma) \, dF(x),$$

rather than (3). The $h(\cdot)$ function can be identified from experimental data and inference theory for this version of the problem can be determined analogously using C.D.F. type smoothing.

7.2 Conditional cash-transfer programs

In some government programs, transfers are contingent both on the household's characteristics *as well as* its having attained the outcome of interest. Such programs are currently being implemented in at least 16 developing countries (c.f., the website "go.worldbank.org/BWUC1CMXM0") in Asia and in south and central America. These programs typically pay a transfer only if the household sends its children to school and pays regular visits to health clinics for preventive care. For such behavior-contingent transfers, the budget constraint changes because transfers are paid only when the desired outcome is realized.

Consider the set-up where the target outcome is binary (e.g. children attending school) and covariates X with support $\mathcal{X}$ can include both discrete and continuous components. Now the set A will represent "eligibility for being offered the program". The eventual outcome, denoted by Y, is the joint occurrence of (an eligible) household participating in the program and sending its children to school. Transfers are made if and only if the household is both eligible (i.e., its value of X lies in A) *and the outcome* $Y = 1$ *is realized*. In this case, $\phi(x, s)$ will denote the probability that $Y = 1$ for a randomly picked x-type household when offered the treatment $s \in \{0, 1\}$. Notice that the relevant policy in this case is deciding whom to offer the program and so identifying $\phi(x, s)$ will not require any corrections for nonrandom take-up as long as the program was *offered* purely randomly. This is in contrast to identifying the mean effect of *participation* in the program.

Now the planner's problem becomes one of determining "optimal eligibility", viz.

$$\max_{A \subset \mathcal{X}} \int_{x \in \mathcal{X}} \left[\phi(x, 1)\, 1(x \in A) + \phi(x, 0)\, 1(x \notin A)\right] dF(x)$$

subject to the budget constraint

$$c = \int_{x \in \mathcal{X}} \phi(x, 1) \times 1(x \in A)\, dF(x),$$

which differs from (1) because a transfer is made here only when the outcome $Y = 1$ is attained. Simple algebra shows that this optimization problem is equivalent to

$$\min_{A \subset \mathcal{X}} \int_{x \in \mathcal{X}} \phi(x, 0) \times 1(x \in A)\, dF(x) \text{ s.t. } \int_{x \in \mathcal{X}} \phi(x, 1) \times 1(x \in A)\, dF(x) = c,$$

implying a solution of the form

$$A^* = \{x \in \mathcal{X} : \phi(x, 0) \leq \alpha\}, \text{ with } \int_{x \in \mathcal{X}} \phi(x, 1) \times 1(\phi(x, 0) \leq \alpha)\, dF(x) = c,$$

and a corresponding value function

$$\mu = c + E\left[\phi(X, 0) \times 1\{\phi(X, 0) > \alpha\}\right].$$

The analogous estimates $\hat{\alpha}$ and $\hat{\mu}$ can be obtained via c.d.f. type smoothing.

7.3 Optimizing under externalities

Some recent program evaluation studies have found that treatments effects on treatment-eligibles can have spillover effects on behavioral outcomes of non-eligibles living in the same village or locality, c.f., Attanasio, Meghir and Santiago (2006), Angelucci and De Giorgi (forthcoming) or Sobel (2006) and references therein. Such externalities can significantly boost the eventual efficacy of prevention programs against infectious diseases, such as the use of anti-malarial ITNs considered in our application below. In that context, spillover can work through two mutually reinforcing channels: one, ITN use by subsidized households leads to increased ITN take-up and use by neighbors via peer effects (Dupas, 2009); and two, ITN use by a subsidized household protects neighboring households against malaria by reducing the chance of transmission of the parasite (Hawley et al., 2003).

One can modify our analysis to utilize gains from such spillover.[9] Suppose spillover occurs only within villages. Then one can consider optimization at the village level only and allocate the treatment within the village by randomization. A more efficient scheme is to optimize based on both village and household level characteristics, which is what we describe now. Let X denote household characteristics (including those that are common to all households in the village) and Π denote the distribution of all combinations of X in the village from which the household comes. For example, if $x^i \equiv(x_1^i, x_2^i)$=(have child under 10, income below poverty line), then π^i=$(\pi_{11}^i, \pi_{10}^i, \pi_{01}^i, \pi_{00}^i)$, where π_{jk}^i is the fraction of households in i's village with $X_1 = j$ and $X_2 = k$.

Let $\phi(x, s, w)$ denote the expected outcome of an x-type household, receiving treatment $s \in \{0, 1\}$ when the fraction of households getting treatment in its village is $w \in [0, 1]$. Spillovers are captured by $\frac{\partial \phi(\cdot,\cdot,w)}{\partial w} \neq 0$. Now consider a generic treatment choice rule where a household i is treated only if $\pi^i \in B$ and $x_i \in A\left(\pi^i\right)$. So if $\pi^i \notin B$, then no household in i's village gets the treatment and if $\pi^i \in B$, $x_i \notin A\left(\pi^i\right)$ then i will not get the treatment but a fraction $p\left(A\left(\pi^i\right)\right) \equiv \sum \pi_{jk}^i \geq 0$ in its village will, where the last sum is over those combinations of X that are in A.

In the above example, a possible candidate pair (not necessarily optimal) could be of the form: $B = \{\pi : \pi_1 + \pi_2 \geq 0.5\}$ and $A(\pi) = \{\{1, 1\} \cup \{1, 0\}$ if $\pi_1 < 0.5\}$ and $A(\pi) = \{\{1, 1\}$ if $\pi_1 \geq 0.5\}$. For this case, $p(A(\pi)) = \pi_1 + \pi_2$ if $\pi_1 + \pi_2 \geq 0.5$ and $\pi_1 < 0.5$; $p(A(\pi)) = \pi_1$ if $\pi_1 + \pi_2 \geq 0.5$ and $\pi_1 \geq 0.5$; and $p(A(\pi)) = 0$ if $\pi_1 + \pi_2 < 0.5$. In other words, if the proportion of households who have a child under 10 and are poor is above 50%, treating all of them would be enough to generate positive spillovers onto other households in the village and therefore treating those with children under 10 but who are not poor is not necessary. But if the proportion of households who have a child under 10 and are poor is not high enough to generate enough positive spillover onto others, treating those households with a child under 10 but are not poor will be necessary.

Under a rule $\{B, A(\cdot)\}$, the expected outcome of a household with $X = x$ and $\Pi = \pi$ will be

[9] See pre-exisiting work by Graham, Imbens and Ridder (2009, section 6) on the related but different theme of designing optimal peer-group formation to maximize gains from peer-effects.

given by

$$\phi\left(x,1,p\left(A\left(\pi\right)\right)\right)\times 1\left\{\pi\in B,x\in A\left(\pi\right)\right\}+\phi\left(x,0,p\left(A\left(\pi\right)\right)\right)\times 1\left\{\pi\in B,x\notin A\left(\pi\right)\right\}$$
$$+\phi\left(x,0,0\right)\times 1\left\{\pi\notin B\right\}$$

Denoting by $F_{X,\Pi}\left(\cdot,\cdot\right)$ the joint distribution of (X,Π) in the population, the "idealized" allocation problem becomes

$$\max_{B,A(.)}\int\left[\begin{array}{c}\phi\left(x,1,p\left(A\left(\pi\right)\right)\right)\times 1\left\{\pi\in B,x\in A\left(\pi\right)\right\}\\ +\phi\left(x,0,p\left(A\left(\pi\right)\right)\right)\times 1\left\{\pi\in B,x\notin A\left(\pi\right)\right\}\\ +\phi\left(x,0,0\right)\times 1\left\{\pi\notin B\right\}\end{array}\right]dF_{X,\Pi}\left(x,\pi\right)$$

s.t.

$$\int 1\left\{\pi\in B,x\in A\left(\pi\right)\right\}dF_{X,\Pi}\left(x,\pi\right)=c.$$

This problem is substantively different from the previous ones due to the presence of A inside the $\phi\left(\cdot\right)$ function. An additional technical complication is that $\pi\left(\cdot\right)$ will be infinite dimensional if X is allowed to be continuous. For these reasons, this case requires independent analysis and is left to future research.

7.4 Best Linear Rule

The rule described above requires computing $\theta\left(x\right)$ nonparametrically for every household, which can be onerous for a practitioner. An alternative that is easier to implement is an index-based rule of the form: treat if $x'\beta>\gamma$ and not otherwise. This problem can be solved by noting that for the optimal choice of β, the index $x'\beta$ should sort the data in the same way as $\theta\left(x\right)$. This preserves the ranking of households based on $\theta\left(X\right)$ and thus allots treatment to those who benefit the most from it. So one can estimate a suitably "normalized" β by

$$\hat{\beta}=\arg\max_{\beta}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\begin{array}{c}1\left(\hat{\theta}\left(x_i\right)>\hat{\theta}\left(x_j\right)\right)1\left(x_i'\beta>x_j'\beta\right)\\ +1\left(\hat{\theta}\left(x_i\right)<\hat{\theta}\left(x_j\right)\right)1\left(x_i'\beta<x_j'\beta\right)\end{array}\right\}$$

and the threshold γ can be estimated from a smoothed version of $c=\frac{1}{n}\sum_{i=1}^{n}1\left(x_i'\hat{\beta}>\hat{\gamma}\right)$.This objective function differs from Han's maximum rank correlation estimator's because $\hat{\theta}\left(\cdot\right)$ is nonparametrically estimated here. An alternative is to approximate $\theta\left(\cdot\right)$ directly by a linear index $x'\delta$ where δ is estimated by an OLS regression of $\hat{\theta}\left(x_i\right)$ on x_i, i.e., $\hat{\delta}=\left(\sum_{i=1}^{n}x_ix_i'\right)^{-1}\left(\sum_{i=1}^{n}x_i\hat{\theta}\left(x_i\right)\right)$. This estimator may be analyzed via U-statistic techniques.

8 Application to bednet provision

8.1 Background

We now apply our inference method to the allocation of subsidies for long-lasting insecticide-treated nets (ITNs) to households, using experimental evidence from Kenya.

The rationale for public funding of ITNs comes from their proven efficacy in reducing the burden of malaria through the presence of both large private and large social returns to ITN use. ITNs have been shown to reduce overall child mortality by up to 38 percent in regions of Africa where malaria is the leading cause of death among children under 5.[10] ITN coverage protects pregnant women and their children from the serious detrimental effects of maternal malaria. In addition, ITN use can help avert some of the substantial direct costs of treatment and the indirect costs of malaria infection on lost income.[11] Lucas (2007) estimates that, alone, the gains to education of a malaria-free environment more than compensate for the cost of an ITN. Costing $5 - $7 a net, however, ITNs are not affordable to most families (Cohen and Dupas, 2007; Dupas, forthcoming). For this reason, there is a large consensus that ITNs should be fully subsidized (WHO, 2007; Sachs, 2005).

Teklehaimanot, McCord and Sachs (2007) estimate that providing one free long-lasting ITN for every two at-risk persons in sub-Saharan Africa would amount to 2.5 billion dollars. The funds committed by governments and donor agencies for ITNs have not yet reached that amount, however. For example, the Government of Kenya estimates that around 1 million pregnant women are in need of an ITN every year, but their budget will allow them to provide only 0.5 million nets per year to pregnant women over the next 5 years (Kenya Round 7 Proposal, 2007).

Under such a budget constraint, the question of how to allocate the available ITNs among households becomes an important policy question. If the treatment effect (the health impact of getting a subsidized ITN) is exactly the same for everyone in the population, then all possible allocations will lead to the same overall gains. However, when there is heterogeneity in the treatment effect (e.g. the health impact of getting a subsidized ITN varies with observed covariates, such as socioeconomic status, presence of children in the household, etc.), the gains can be maximized by a covariate-based allocation. While the health impact of using an ITN might be homogenous, the health impact of getting a highly subsidized ITN might vary across covariates since usage rates (conditional on having a net) are likely to vary across covariates. For example, households who can afford to purchase an ITN in the absence of any subsidy (because they have access to credit or are wealthy enough) will not benefit from the treatment very much (i.e. their $\phi(x, 0)$ will be large and thus for them the difference $\phi(x, 1) - \phi(x, 0)$ is likely to be small). Likewise, since young children are the most vulnerable to the disease, households without young children might not benefit much from the treatment (i.e. their $\phi(x, 1)$ will be small and thus the difference $\phi(x, 1) - \phi(x, 0)$ is likely to be small). For these reasons, the treatment effect is likely to vary across observed covariates such as wealth, access to financial services, and the presence of young children. An allocation rule that takes into account such heterogeneity could potentially generate important welfare gains.

[10] See Lengeler (2004) for a review. Earlier estimates of ITN use on reductions in child mortality from a randomized trial in Gambia were as high as 60 percent, but most estimates from randomized trials in Africa are closer to 20 percent.

[11] Ettling et al. (1994) find that poor households in a malaria-endemic area of Malawi spend roughly 28 percent of their cash income treating malaria episodes.

8.2 Design

For this application we use data from a randomized experiment conducted with rural households in Western Kenya in 2007 (Dupas, forthcoming). The price at which a household could purchase an ITN varied in steps of \$0.50 from \$0 (a free ITN) to \$4, and households were randomly assigned to a price. People had three months to redeem the voucher entitling them to an ITN at the assigned price. In this application, we consider two groups: households that faced a very low (highly subsidized) price (\$0 or \$0.50) and households that faced a high price of \$2 or more. Table 1 presents summary statistics on the 985 households that form the sample used in the analysis. The take-up rate of the ITN subsidy was 84% in the low price group and 16% in the high price group. Conditional on take-up, the usage rate was slightly higher in the low price group than in the high price group (70% versus 58%), leading to unconditional usage rates of 61% and 7%, respectively. In what follows, we consider the low price group as the treatment group and the high price group constitutes the control. The treatment is thus “having access to a low-price ITN”. One may note that the short-run take-up in the low price group was not 100%, since some of the "treated" had to pay a small fee (i.e., \$0.50) to access the net–i.e., the subsidy was not 100% for everyone. This situation corresponds to our discussion in section 7.1. In the analysis below, however, we assume that take-up was 100% in the treatment group. In other words, we consider that those who did not take-up the subsidy cost as much to the government as those who took-up the subsidy but didn't use their net. There are two reasons behind this approximation. First, the take-up was 100% for households that were offered the net for free. We could thus have restricted our definition of the "treatment group" to those that were randomly assigned a free net and we would be in the exact case described above. We could not do that for sample size reasons, however. Second, the price of \$0.50 was also quite low and we believe that if people had more than three months to redeem their voucher, the take-up would potentially have reached 100% at this price. In particular, since the take-up at intermediate prices (e.g., \$1) was not negligible, people who did not take-up the \$0.50 subsidy could simply have taken it up and sold their voucher to an ineligible household for a profit. While this did not happen in the pilot experiment we use here, it would likely happen for a large government program implemented nation-wide on a continuous basis. That said, to extrapolate our results to lower subsidy rates (higher prices), one would need to follow the procedure outlined in section 7.1.

Table 2 presents suggestive evidence of heterogeneity in the treatment effect. The table shows the results of an OLS regression of ITN usage on the treatment, three covariates, and the interactions between the treatment and the covariates. The covariates are: a binary variable equal to 1 if the household includes at least one child under 10; the natural log of the value of the household's wealth per capita; and a binary variable equal to 1 if the household owns a bank account. The first covariate (presence of a child) was chosen as an indicator of the private returns to using a bed net (since young children are the most vulnerable to malaria). The two other covariates were chosen as proxies for socioeconomic status and ability to pay. They were measured through a baseline survey administered through household visits. In particular, wealth per capita was measured as follows: households were asked to list all their assets (including animal assets) and to estimate their resale value. The combined value of all assets was then divided by household size to obtain the "wealth per capita" indicator. The treatment was randomized at the household-level so no clustering correction is needed. We find that having a higher wealth per capita correlates with a higher ITN usage rate in the absence of treatment, and the treatment effect appears significantly higher for households with a child under 10 and significantly lower for households that own a bank account. An F-test of the joint significance of the three interaction terms is significant at 10% which, superficially, suggests that a covariate-based allocation may lead to important welfare gains.

8.3 Analysis

8.3.1 Non-Parametric Analysis: Choice of Kernels and Bandwidths

For bias-removal, we use the higher order kernels corresponding to $r = 4$ and $q = 3$, viz.,

$$\begin{aligned} K(s) &= 0.5 \times \left(3 - s^2\right) \times \phi(s), \\ \bar{L}(s) &= \frac{15}{32}\left(\frac{7}{5}s^5 - \frac{10}{3}s^3 + 3s + \frac{16}{15}\right) \times 1(-1 \leq s \leq 1) + 1(s > 1), \end{aligned}$$

where $\phi(\cdot)$ is the standard normal density. Two bandwidths are needed for the non-parametric estimation: the bandwidth σ_n in the estimation of the conditional ATE $\theta(X)$, and the bandwidth h_n in the smoothing correction. Figure 1 graphs how the estimated treatment threshold $\hat{\gamma}$ (Panel A) and value function $\hat{\rho}$ (Panel B) vary with h_n for a range of possible σ_n. We find that both estimates are insensitive to the choice of h_n. They are also quite stable over a large range of σ_n. In Figure 2, we present $\hat{\gamma}$ and $\hat{\rho}$ for two budget constraint levels: $c = 0.5$ (Panel A) and $c = 0.25$ (Panel B). The stability of $\hat{\rho}$ over a reasonable range of bandwidths suggests that the choice of bandwidths should have little effect on the nonparametric estimates of the value function.

Figure 3 graphs a leave-one-out cross validation criterion function for $\theta(x)$. The function is plotted over the range $\sigma_n \in [0.3, 0.4]$, which correspond roughly to $n^{-1/6}$ and $n^{-1/8}$, respectively. The function seems to dip around $\sigma_n = 0.33$. Given the small sensitivity of our estimates of ρ and, to a certain extent, γ to the choice of σ_n, we show the results for both $\sigma_n = 0.3$ and $\sigma_n = 0.4$. We use $h_n = 0.35$; recall that the results seem very insensitive to the choice of h_n for a given choice of σ_n.

To see how our results are affected by choice of a higher order kernel, we also repeat and report part of the analysis for a standard normal kernel. The results are numerically not very different and do not imply any substantively different conclusion. Since we have a large sample size– close to 200 for each cell defined by the discrete regressors– and a single continuous regressor, we prefer the higher order kernels for our analysis.

8.3.2 Conditional ATE

The nonparametric estimate of the CATE $\hat{\theta}(x) = \hat{\phi}(x, 1) - \hat{\phi}(x, 0)$ was computed corresponding to two bandwidths $\sigma_n = 0.3$ and $\sigma_n = 0.4$. The parametric estimate of $\theta(X)$ was computed as

$\hat{\theta}(x) = \left(\hat{\delta} + x'\hat{\delta}_1\right)$, where $\hat{\delta}$ and $\hat{\delta}_1$ are OLS estimates in the regression (presented in Table 2):

$$y_i = \beta_0 + X_i'\beta_1 + \delta_0 S_i + X_i'\delta_1 \times S_i + \varepsilon_i.$$

Figure 4 graphs the kernel density of the conditional ATE $\theta(X)$ computed with the two proposed bandwidths. Observations with X such that $\theta(X)$ is below -0.2 or above 0.9 were discarded in accordance with assumption A2 above. These cutoffs were the 1 and 99 percentile values in the empirical distribution of $\theta(X)$.

Figure 5 presents the c.d.f. of the conditional ATE $\theta(X)$ computed both parametrically and nonparametrically. The stepwise shape for the c.d.f. in the parametric model is essentially due to the binary nature of two of the three covariates since the interaction of the treatment with wealth appears to be nearly zero in the parametric case.

8.3.3 Unrestricted and Restricted Value Functions

In what follows, we compare the "first best" allocation (the unrestricted case, in which the allocation is based on all three covariates) with three "restricted" cases: (i) basing the allocation on the first two covariates only, leaving out wealth, which is typically harder to observe without conducting expensive household surveys; **(ii)** means-testing where the allocation is based only on wealth– which is extremely common in both developed and developing countries, and (iii) purely random allocation which is not covariate-based at all. Notice that in the random allocation case, the estimated value function is linear in c:

$$\hat{\rho}(c) = \frac{1}{n}\sum_{i=1}^{n}\left\{c \times \hat{\phi}(X_i, 1) + (1-c) \times \hat{\phi}(X_i, 0)\right\}.$$

Figure 6 graphs the parametric and nonparametric estimates for the treatment threshold $\gamma(c)$ and the value function $\rho(c)$ in the unrestricted case. The nonparametric estimates seem very stable over the two choices of bandwidth. The nonparametric estimates of the unrestricted value function are higher than the parametric estimates.

Panel A of Figure 7 graphs the estimates of the value function $\rho(c)$ when conditioning is done on wealth but no other covariates and Panel B of Figure 7 graphs the estimates of $\rho(c)$ when the allocation is purely random.

8.3.4 Welfare Losses

Representing all four cases (unrestricted allocation, allocation on all covariates but wealth, allocation based on wealth only, and random allocation) on the same graph helps visualize the welfare loss when the optimal allocation is not implementable, as well as the gains from means-testing compared to non-wealth based allocations. Figure 8 combines the parametric estimates of the value function $\rho(c)$ for all four cases in Panel A and the nonparametric estimates in Panel B. In contrast to the parametric estimates, the non-parametric estimates suggest that means-testing is a clear "second best", generating a higher mean outcome than random allocation does. The parametric estimates for the means-tested case is visually indistinguishable from the random allocation case– a fact more clearly depicted in Table 3a.

We compute the standard errors of the welfare losses generated by the three suboptimal allocations over a range of budget levels in Table 3a. Panel A presents the parametric estimates and Panel B the nonparametric estimates. Panel C presents the differences between the parametric and nonparametric estimates. As noted in Figures 6B and 7A, the estimates of the unrestricted value function are significantly different between the parametric and the nonparametric analyses (column 2, Panel C). The non-parametric estimates are overall quite robust to the choice of bandwidth σ.

The estimated inefficiency of basing the allocation on all covariates but wealth is between 11% (for $\sigma = 0.3$) and 15% (for $\sigma = 0.4$) when the budget allows us to treat 25% of the population (Panel B, column 3). This means that $\rho_{res}(25)$ is 3 to 4 percentage points lower than $\rho_{un}(25)$. (Note that the gap between the two non-parametric estimates comes from the gap in the estimates of $\rho_{un}(25)$. The gap in the estimates of $\rho_{un}(25)$ is less than 1 percentage point, but off of a base of 0.25 it amounts to close to 4 percent.)

The inefficiency of basing the allocation on wealth only is estimated at 7%-8% (Panel B, column 4) when the budget allows to treat 25% of the population. This means that $\rho_{res}(25)$ is 2 percentage points lower than $\rho_{un}(25)$. When estimated non-parametrically, the welfare loss due to random allocation is higher, at 20% (5 pp) for $\sigma = 0.3$ and 18% (4pp) for $\sigma = 0.4$ (Table 3, column 5).

Overall, the estimates presented in Table 3a suggest that the welfare costs of restricted allocation schemes can be substantial. In the Kenyan context analyzed here, we also find that means-testing only does not generate a much higher outcome than an allocation based on covariates other than wealth. Depending on the cost of collecting information on households' assets (or other proxies for wealth), which typically requires labor- and time-consuming household survey efforts in countries where too few people pay taxes for the tax returns to be informative, the welfare gain of a means-tested allocation compared to other allocation schemes might not be worth its cost.

Table 3b reports analogous results for a standard normal kernel which is not a higher-order kernel. The results are numerically somewhat different but do not imply any substantively different conclusion. But, given our sample sizes, we report the other results only for the higher-order kernel.

8.3.5 Dual Problem

In Table 4 we report the minimum resources needed to attain a certain expected outcome: we compute the share of the population that needs to be treated in order to achieve a given target value function by allocating treatment based on all three covariates (column 2). We then calculate the additional resources that are needed when the optimal, unrestricted allocation is not possible, and the allocation is instead based on all covariates except wealth (column 3), only on wealth (column 4) or the allocation is purely random (column 5). The nonparametric estimates with the bandwidth $\sigma = 0.4$ suggest that an allocation based on all covariates but wealth requires treating an additional. 8.7 percentage points of the population compared to the optimal allocation in order to reach a mean usage rate $\rho = 0.40$ (Panel B2, column 3). An allocation based on wealth only would require treating an additional 9 percentage points of the population compared to the optimal allocation (Panel B2, column 4). The additional spending is even higher when the allocation is purely random: an extra 12.4 percentage points of the population need to be treated to reach the target usage rate, compared to the optimal allocation (Panel B2, column 5).

Allocation rules based on wealth only ("means-testing") are very common in developed countries, e.g. housing benefits; food stamps or Medicaid in the US, but less so in developing countries where wealth or income data are not easily verifiable due to the absence of tax records. By comparing these estimates of the minimum resources needed to attain a certain expected outcome across restricted cases (means-testing only vs. "all but wealth" and random allocations), one can judge whether it is worth collecting the data needed to means-test.

9 Conclusion

In this paper, we have considered a social planner's problem of allocating a binary treatment among a target population based on observed covariates in the presence of budget constraints. The paper proposes a simple allocation rule based on sample data from an experiment, where the treatment is randomly allocated and examines this rule from a decision theoretic perspective. The paper then derives and uses large-sample frequentist properties of these rules to infer (i) the expected welfare from the rule, (ii) the minimum cost of attaining a specific average welfare– i.e., the dual and (iii) the welfare loss corresponding to restricted covariate choice. These methods are applied to data on the provision of anti-malaria bed nets in western Kenya. The empirical findings are that a government which can afford to distribute bed net subsidies to only 50% of its target population can, if using an allocation rule based on multiple covariates, increase actual bed-net coverage by 8 percentage points (19%) relative to random allocation and by 4 percentage points (9%) relative to an allocation scheme based on wealth only.

This paper has left several related topics to future research. One is to extend the methods to the design of conditional cash-transfer programs, which have gained popularity in a large number of central and south American countries. A second topic, outlined in section 7.3 above, concerns how treatment externalities can be incorporated into an analysis of efficient treatment assignment. Deriving the welfare-maximizing linear rule, as discussed in section 7.4, would also be a theoretically interesting and practically relevant exercise as would be the extension to multiple treatments. Finally, from an applications point of view, it would be interesting to compare parametric and non-parametric Bayesian solutions and the minmax regret solutions with the EWM methods proposed here, using smaller subsets of our sample.

A caveat to our analysis is the implicit assumption that the covariate distributions are not affected by the targeting strategy used. This may be violated if the population composition changes in response to changes in the targeting rule, e.g., switching subsidy eligibility towards families with children in a district may see an influx of families with children from neighboring districts, thereby altering the marginal distribution of covariates. Such migration is plausible only when the size of the transfer is high enough relative to migration costs and thus quite unlikely at the usual scale of in-kind transfer schemes present in the developing world today. But for larger sized transfers, this caveat can potentially be an important one.

We would like to end with the observation that in development circles, there has been a recent push for more experimental evidence on the impact of social programs, as part of a general effort to improve the effectiveness of aid (Duflo, Kremer and Glennerster, 2006). For example, the World Bank recently launched the DIME initiative, an effort to increase the number of Bank-funded projects with impact evaluation components. We believe that as randomized trials of social programs, e.g., Oportunidades (PROGRESA) in Mexico, become more common in both developed and developing countries, our methodology will become increasingly relevant in helping governments and aid-agencies roll out positive-impact programs via efficient allocation rules.

References

[1] Abadie, A. (2002): Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models, *Journal of the American Statistical Association*, vol. 97, no. 457, March 2002, pp. 284-92.

[2] Ahn, D., Choi, S., Kariv, S. and Gale, D. (2009): Estimating Ambiguity Aversion in a Portfolio Choice Experiment, mimeo., UC Berkeley.

[3] Andrews, D. W. K. (1994): Empirical Process methods on Econometrics, *Handbook of Econometrics*, vol. 4, Elsevier.

[4] Angelucci, M. and De Giorgi, G. (forthcoming): Indirect Effects of an Aid Program: How do Cash Transfers Affect Non-Eligibles' Consumption?, *The American Economic Review*.

[5] Attanasio, Orazio, Costas Meghir and Santiago (2006): "Education Choices in Mexico. Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," MIMEO, UCL.

[6] Behnke, Stephanie, Markus Frölich and Michael Lechner (2008): "Targeting Labour Market Programmes – Results from a Randomized Experiment." IZA Discussion Paper, No. 3085.

[7] Berger, Mark, Dan Black and Jeffrey Smith (2001): "Evaluating Profiling as a Means of Allocating Government Services," in Michael Lechner and Friedhelm Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 59-84.

[8] Bhattacharya, D. (2006): "Inferring Optimal Peer Allocation using Experimental Data," forthcoming, *Journal of the American Statistical Association*.

[9] Bhattacharya, D., A. Chandra and X. Chen (2007): Nonparametric analysis of optimal healthcare expenditure using observational data (in preparation).

[10] Chamberlain, G. (2000): Decision theory in Econometrics, Journal of Econometrics.

[11] Chamberlain, G. and Imbens, G. (2003): Nonparametric Applications of Bayesian Inference, Journal of Business and Economic Statistics, vol. 21, no. 1, pp. 12-18.

[12] (CLV) Chen, X., O. Linton and I. van Keilegom (2003): Estimation of semiparametric models when the criterion function is not smooth, *Econometrica* 71, pp. 1591–1608.

[13] Cohen, Jessica and Pascaline Dupas (2007): "Free distribution or cost-sharing? Evidence from a randomized malaria experiment in Kenya". NBER WP#14406

[14] Collins, L.M., S.A. Murphy, V. Nair & V. Strecher (2007): A Strategy for Optimizing and Evaluating Behavioral Interventions , *Annals of Behavioral Medicine*. 30:65-73.

[15] Cox, David (1975): "Prediction intervals and empirical Bayes confidence intervals," in *Perspectives in Probability and Statistics*, (ed., J.Gani), Applied Probability Trust, Sheffield, England.

[16] Dehejia, Rajeev H (2005): Program Evaluation as a decision Problem, *Journal of Econometrics*, vol. 125, no. 1-2, pp. 141-73.

[17] Duflo Esther, Rema Hanna and Stephen Ryan (2007). "Monitoring Works: Getting Teachers to Come to School", mimeo, MIT.

[18] Duflo, Esther, Michael Kremer and Rachel Glennerster (2006). "Using Ramdomization in Development Economics Research: A Tool Kit". NBER Working Paper No. T0333

[19] Dupas, Pascaline (forthcoming). "What matters (and what does not) in households' decision to invest in malaria prevention?" *American Economic Review Papers and Proceedings*.

[20] Dupas, Pascaline (2009). "Short-Run Subsidies and Long-Term Adoption of New Health Products: Evidence from a Field Experiment". mimeo, UCLA.

[21] Eberts, Randall W., Christopher J. O'Leary, and Stephen A. Wandner, (eds. 2002): *Targeting Employment Services*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

[22] Ettling M, et al. (1994): "Economic Impact of Malaria in Malawian Households." *Tropical Medicine and Parasitology*. 45: 74-79.

[23] Ferguson, T (1973): "A Bayesian Analysis of some Nonparametric problems," *Annals of Statistics*, vol 1., pp 209-30.

[24] Frolich, M. (2003): Programme Evaluation and Treatment Choice, Lecture Notes in Economics and Mathematical Systems, Vol. 524. Heidelberg: Springer.

[25] Frolich, M. (2006): "Statistical treatment choice: an application to active labour market programmes," IZA Discussion Paper No. 2187.

[26] Graham, B., G. Imbens and G. Ridder (2005): "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis," manuscript.

[27] Graham, B., G. Imbens and G. Ridder (2006): "Complementarity and the Optimal Allocation of Inputs," manuscript.

[28] Graham, B., G. Imbens and G. Ridder (2009): Measuring the average outcome and inequality effects of segregation in the presence of social spillovers, manuscript.

[29] Gunter,L. , J. Zhu and S. A. Murphy (2007). "Variable Selection for Optimal Decision Making, Proceedings of the 11th Conference on Artificial Intelligence in Medicine," AIME 2007, LNCS/LNAI 4594, 149-154.

[30] Hahn, J., K. Hirano and D. Karlan (2007): "Adaptive Experimental Design using the Propensity Score", manuscript.

[31] Hansen, B. (2008): "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, pp. 726-748.

[32] Härdle, Wolfgang; Hall, Peter; Marron, J. S. (1988): "How far are automatically chosen regression smoothing parameters from their optimum?," *Journal of the American Statistical Association* 83, no. 401, 86–101.

[33] Hawley, William A. et al. (2003): "Community-Wide Effects of Permethrin-Treated Bed Nets on Child Mortality and Malaria Morbidity in Western Kenya." *American Journal of Tropical Medicine and Hygiene*. 68(Suppl. 4): 121-127.

[34] Hirano, K. and J. Porter (2008): "Asymptotics for Statistical Treatment Rules", forthcoming, Econometrica.

[35] Hirano, K., G. Imbens, G. Ridder (2003): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica* 71, 1161-1189.

[36] Horowitz, J (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, Vol. 60, No. 3, 505-531.

[37] Kenya Round 7 Proposal in response to the Global Fund to fight AIDS, Tuberculosis and Malaria 7th call for proposals. Available online at: http://www.theglobalfund.org/en/files/apply/call7/notapproved/7KENM_1527_0_full.pdf

[38] Lechner, Michael and Jeffrey Smith (2007): "What is the Value Added by Case Workers?", *Labour Economics* 14(2): 135-151.

[39] Lengeler, Christian (2004): "Insecticide-treated bed nets and curtains for preventing malaria". Cochrane Database Syst Rev 2004; 2:CD000363.

[40] Lucas, Adrienne (2007): "Economic Effects of Malaria Eradication: Evidence from the Malarial Periphery". Mimeo, Wellesley College.

[41] Mahajan, A., A. Tarozzi, J. Yoong & B. Blackburn (2009): "Bednets, Information and Malaria in Orissa," mimeo.

[42] Manski, C. (2001): "Designing Programs for Heterogeneous Populations: the value of covariate Information," *American Economic Review*, 91(2), pp. 103-6.

[43] Manski, C. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, vol. 72, no. 4, pp. 1221-46.

[44] Manski, C. (2005): Social choice with partial knowledge of treatment response, Princeton University Press.

[45] Martinelli, Cesar and Susan W. Parker (2007). "Deception and Misreporting in a Social Program". Mimeo.

[46] (NM) Newey, W. and McFadden, D. (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, vol IV, pp 2113-2245, Elsevier Science B.V. Amsterdam.

[47] Pagan, A. and Aman Ullah (1999): *Nonparametric Econometrics*, Cambridge University Press.

[48] Pope, Devin G and Justin Sydnor (2007): "Implicit Statistical Discrimination in Profiling Models", mimeo., Wharton School.

[49] Sachs, Jeffrey (2005): *The End of Poverty: Economic Possibilities for Our Time*, New York: Penguin Press.

[50] Savage, L. (1951): The Theory of Statistical Decision, Journal of the American Statistical Association, Vol. 46, No. 253, pp. 55-67

[51] Schervish, Mark, J. (1997): *Theory of Statistics*, Springer-Verlag, New York.

[52] Smith, Richard, L. (1998): "Bayesian and Frequentist Approaches to Parametric Predictive Inference", in J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (eds), *Bayesian Statistics*, vol. 6, Oxford University Press. downloadable at http://www.unc.edu/depts/statistics/postscript/rs/predv3.ps

[53] Sobel, M. E. (2006): "What Do Randomized Studies of Housing Mobility Demonstrate?", *Journal of the American Statistical Association*, 101(476), pp. 1398-1407.

[54] Teklehaimanot, Awash, Gordon C. McCord and Jeffrey D. Sachs (2007): "Scaling Up Malaria Control in Africa: An Economic and Epidemiological Assessment," NBER Working Papers 13664.

[55] Todd, Petra and Kenneth Wolpin (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, Vol. 96, No.5, 1384-1417.

[56] WHO (2007): WHO Global Malaria Programme: Position Statement on ITNs. http://www.who.int/malaria/docs/itn/ITNspospaperfinal.pdf

Table 1
Summary Statistics

	Sample Mean
Treatment	0.16
	(0.36)
Outcome = 1 (All)	0.16
	(0.36)
Outcome = 1 (Treatment Group)	0.61
	(0.49)
Outcome = 1 (Control Group)	0.07
	(0.26)
Has a child under 10 years of age	0.55
	(0.50)
Household Size	7.01
	(2.63)
Household's Wealth in US$, per capita	44
	(28)
Owns a Bank Account	0.13
	(0.34)
Observations (households)	985

Table 2
Treatment Effects

Dependent Variable	Outcome
Treatment	0.455
	(0.312)
Has a child under 10 years of age	0.018
	(0.021)
Treatment X Has a child under 10 years of age	0.102
	(0.054)*
Log Wealth per Capita	0.024
	(0.017)
Treatment X Log Wealth per Capita	0.007
	(0.040)
Has a bank account	0.052
	(0.031)*
Treatment X Has a bank account	-0.178
	(0.105)*
Constant	-0.13
	(0.129)
Observations	985
R-Squared	0.30
Joint F-Test for three interaction terms	2.15
Prob > F	0.092

Standard Deviations in parentheses. Household-level data collected in Western Kenya in 2007. "Treatment" is a dummy equal to 1 if the household received a coupon for a bed net to be purchased at a low price ($0 or $0.50), and 0 if the household received a coupon for a bed net to be purchased at a price of $2 or above. Outcome = 1 only if (1) the household has redeemed the coupon and (2) the household had started using the bed net at the time of the follow-up visit.

Table 3a
Allocation Efficiency
(with higher order kernels)

(1)	(2)	(3)	(4)	(5)
		Restricted Cases: Efficieny Loss as a share of ρ(c)		
Population share c that the program can afford to treat	Value Function ρ(c): Unrestricted Case	*All covariates except wealth*	*Wealth only*	*Nothing (random assignment)*
Panel A: Parametric Estimates				
0.00	0.08			
0.25	0.22	0.00	0.08	0.08
	(0.01) ***	(0.02)	(0.04) *	(0.05) *
0.50	0.37	0.00	0.09	0.08
	(0.02) ***	(0.01)	(0.05) *	(0.04) *
Panel B: Non-Parametric Estimates				
B1. Bandwidth σ = 0.3				
0.00	0.08			
0.25	0.26	0.15	0.08	0.20
	(0.01) ***	(0.05) ***	(0.04) **	(0.04) ***
0.50	0.42	0.16	0.09	0.19
	(0.03) ***	(0.05) ***	(0.04) ***	(0.04) ***
B2. Bandwidth σ = 0.4				
0.00	0.08			
0.25	0.25	0.11	0.07	0.18
	(0.01) ***	(0.05) **	(0.04)	(0.04) ***
0.50	0.41	0.10	0.11	0.16
	(0.02) ***	(0.04) **	(0.03) ***	(0.03) ***
Panel C: Differences between Non-Parametric (bandwdith σ = 0.3) and Parametric Estimates				
0.25	0.04	0.14	0.00	0.13
	(0.01) **	(0.05) ***	(0.06)	(0.05) ***
0.50	0.05	0.16	0.01	0.11
	(0.02) **	(0.04) ***	(0.05)	(0.04) **

Unrestricted case: conditioning on all 3 covariates available (presence of a child under 5, bank account ownership and normal log of value of household's wealth per capita.) Standard errors in parentheses, significant at 1% (***), 5%(**), 10% (*) levels.

The table reads as follows: (Panel B1, second row): by treating a share 0.25 of the population, a value function of 0.26 will be reached if the allocation can be based on all covariates (unrestricted case, column 2). In the presence of restrictions on what the conditioning can be based on, the efficiency of targetting decreases. The value function will be 15 % lower than in the unrestricted case if the allocation conditions on everything but wealth; it will be 8% lower if it conditions only on wealth (column 4), and 20% lower if the allocation is random (column 5).

Table 3b
Allocation Efficiency
(with standard normal kernels)

(1)	(2)	(3)	(4)	(5)
		Restricted Cases: Efficieny Loss as a share of ρ(c)		
Population share c that the program can afford to treat	Value Function ρ(c): Unrestricted Case	*All covariates except wealth*	*Wealth only*	*Nothing (random assignment)*
Panel A: Parametric Estimates				
0.00	0.08			
0.25	0.22	0.00	0.08	0.08
	(0.01) ***	(0.02)	(0.04) **	(0.04) *
0.50	0.37	0.00	0.09	0.08
	(0.03) ***	(0.01)	(0.04) **	(0.04) *
Panel B: Non-Parametric Estimates				
B1. Bandwidth σ = 0.3				
0.00	0.08			
0.25	0.25	0.10	0.08	0.16
	(0.01) ***	(0.03) ***	(0.04) **	(0.04) ***
0.50	0.40	0.06	0.08	0.14
	(0.03) ***	(0.03) **	(0.03) **	(0.04) ***
B2. Bandwidth σ = 0.4				
0.00	0.08			
0.25	0.24	0.06	0.08	0.13
	(0.01) ***	(0.03) **	(0.04) **	(0.04) ***
0.50	0.38	0.03	0.07	0.10
	(0.03) ***	(0.02)	(0.04) **	(0.04) ***
Panel C: Differences between Non-Parametric (bandwdith σ = 0.3) and Parametric Estimates				
0.25	0.02	0.10	0.00	0.08
	(0.01) **	(0.04) ***	(0.05)	(0.04) **
0.50	0.02	0.06	0.00	0.05
	(0.01)	(0.03) **	(0.04)	(0.03) *

Unrestricted case: conditioning on all 3 covariates available (presence of a child under 5, bank account ownership and normal log of value of household's wealth per capita.) Standard errors in parentheses, significant at 1% (***), 5%(**), 10% (*) levels.

The table reads as follows: (Panel B1, second row): by treating a share 0.25 of the population, a value function of 0.26 will be reached if the allocation can be based on all covariates (unrestricted case, column 2). In the presence of restrictions on what the conditioning can be based on, the efficiency of targetting decreases. The value function will be 15 % lower than in the unrestricted case if the allocation conditions on everything but wealth; it will be 8% lower if it conditions only on wealth (column 4), and 20% lower if the allocation is random (column 5).

Table 4
Dual Problem: Cost of Reaching a Target Outcome
(with higher order kernels)

(1)	(2)		(3)		(4)		(5)	
Objective Function: Target $\rho(c)$	Unrestricted case: Share of population that needs to be treated to reach this target		Restricted Cases: Additional share that needs to be treated to achieve the target when conditioning on: *All covariates except wealth*		*Wealth only*		*Nothing (random assignment)*	
	Panel A: Parametric Computation							
0.250	0.291		0.001		0.039		0.039	
	(0.026)	***	(0.009)		(0.020)	*	(0.023)	*
0.400	0.552		0.000		0.057		0.058	
	(0.047)	***	(0.010)		(0.036)		(0.035)	*
	Panel B: Non-Parametric Computation							
B1. Bandwidth σ = 0.3								
0.250	0.235		0.059		0.033		0.095	
	(0.018)	***	(0.025)	**	(0.018)	*	(0.024)	***
0.400	0.463		0.136		0.069		0.147	
	(0.037)	***	(0.041)	***	(0.048)		(0.038)	***
B2. Bandwidth σ = 0.4								
0.250	0.247		0.047		0.031		0.083	
	(0.019)	***	(0.021)	**	(0.020)		(0.021)	***
0.400	0.486	***	0.087		0.090		0.124	
	(0.035)	***	(0.033)	***	(0.029)	***	(0.033)	***
	Panel C: Differences between Non-Parametric (bandwdith σ = 0.3) and Parametric Estimates							
0.250	-0.056		0.058		-0.006		0.056	
	(0.023)		(0.022)	***	(0.028)		(0.023)	**
0.400	-0.089		0.136		0.012		0.089	
	(0.037)		(0.038)	***	(0.055)		(0.038)	**

Standard errors in parentheses, significant at 1% (***), 5%(**), 10% (*) levels.
The table reads as follows: (Panel B1, row 1): to reach a target value function of 0.250, a share 0.235 of the population needs to be treated if the allocation can be based on all covariates (unrestricted case, column 2). In the presence of restrictions on what the conditioning can be based on, the efficiency of targetting decreases. An additional 0.059 of the population needs to be treated if the conditioning is based on all covariates except wealth (column 3). An additional 0.033 of the population needs to be treated if the conditioning is based on wealth only (column 4). If the allocation is purely random, an additional 0.095 of the population needs to be treated to achieve the 0.250 target value function (column 5).

Figure 1
Sensitivity of γ and ρ to the Choice of Bandwidths

Panel A. Threshold γ — Panel B. Value Function ρ

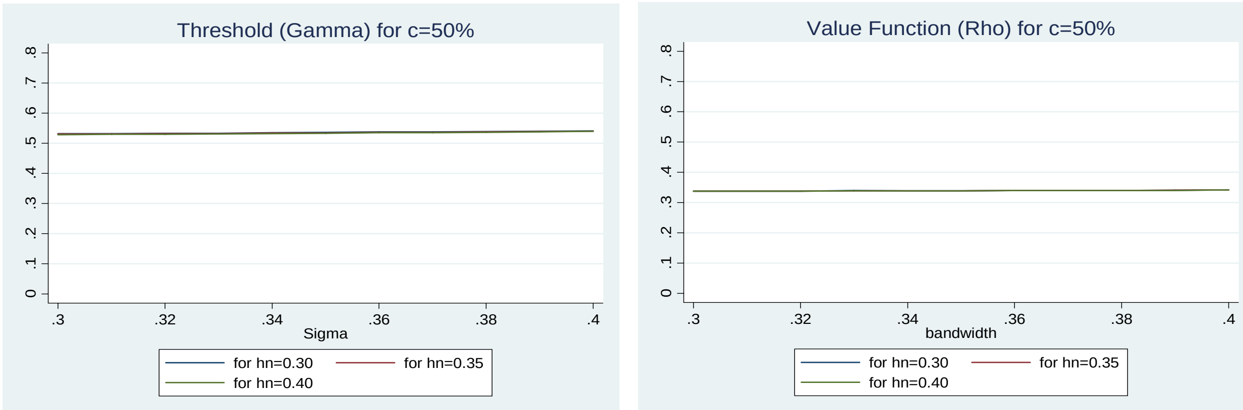


Figure 2

Panel A. γ(c) and ρ(c) when c= 0.50 — Panel B. γ(c) and ρ(c) when c= 0.25

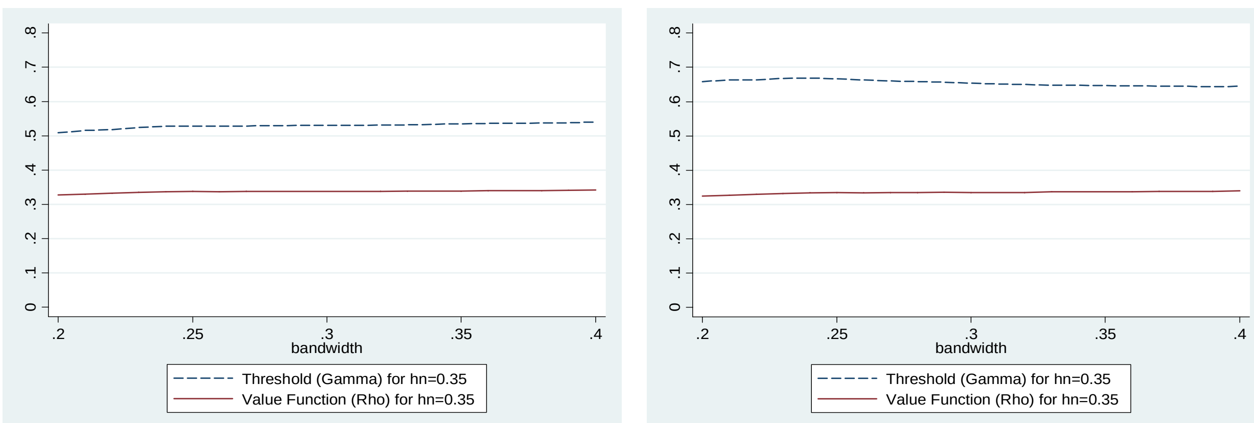


Figure 3
Leave-one-out Cross-Validation Criterion

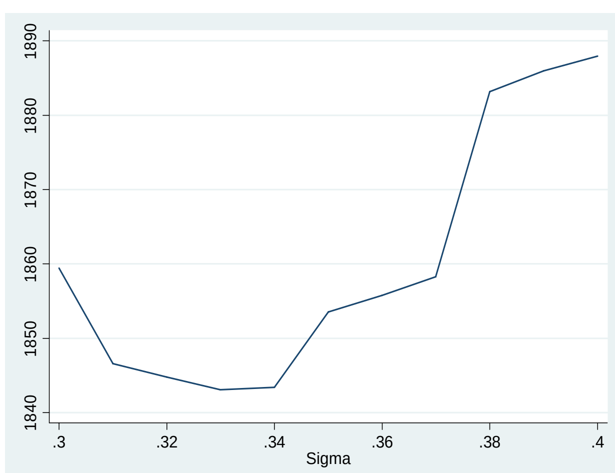


Figure 4

Kernel Density of Estimates of Conditional ATE θ(X)

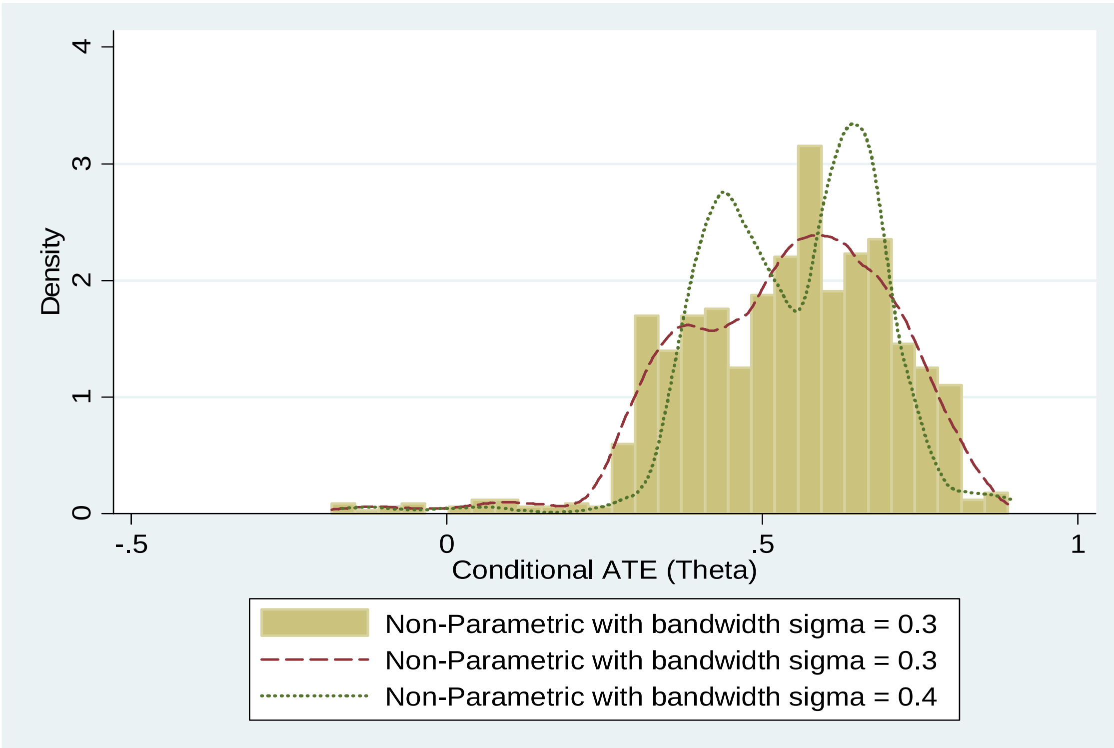


Figure 5

Cumulative Distribution Function of Conditional ATE θ(X)

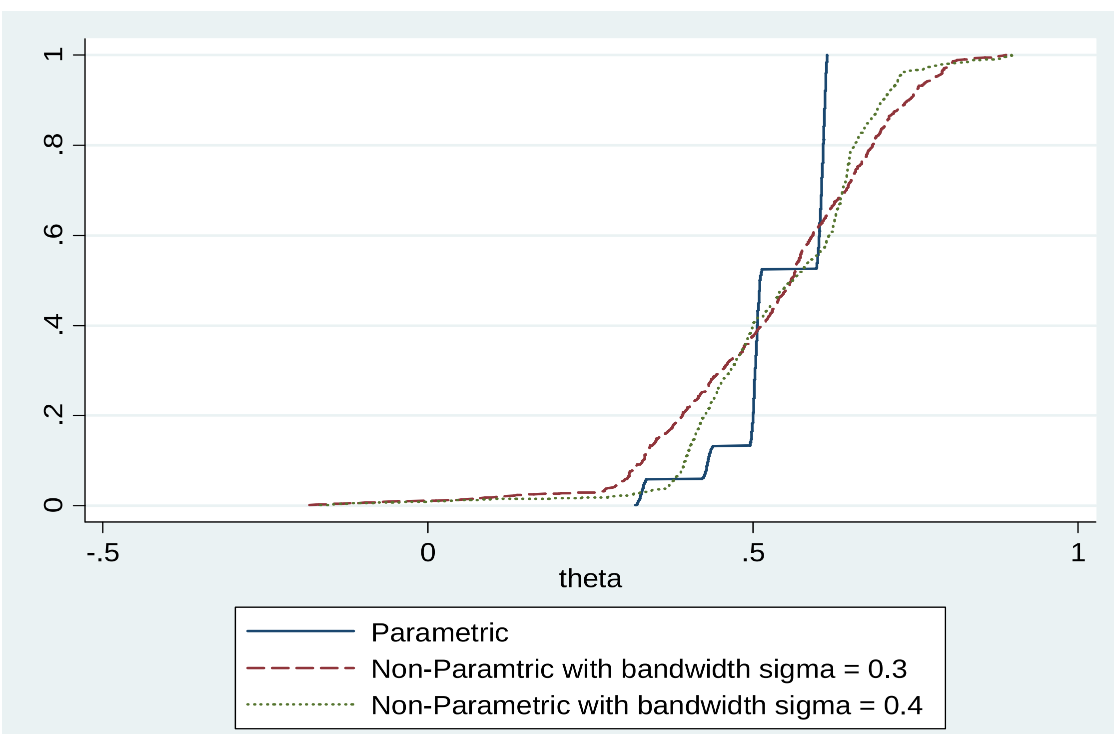
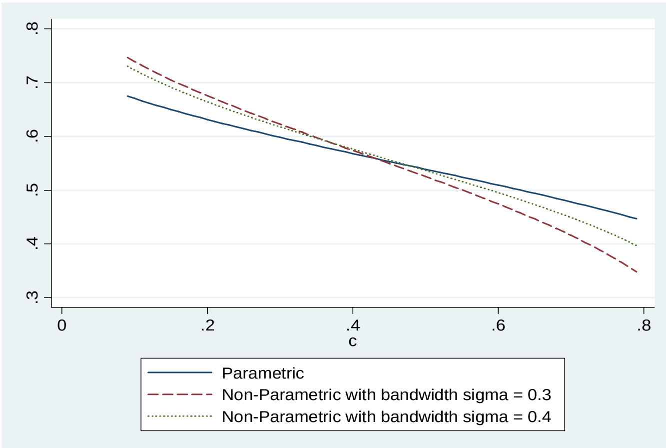


Figure 6: Unrestricted Case: Conditioning on all observables

Panel A. Threshold γ(c)



Panel B. Value Function ρ(c)

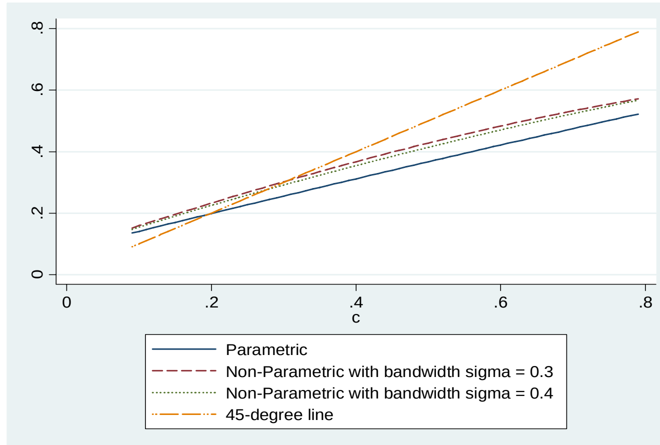
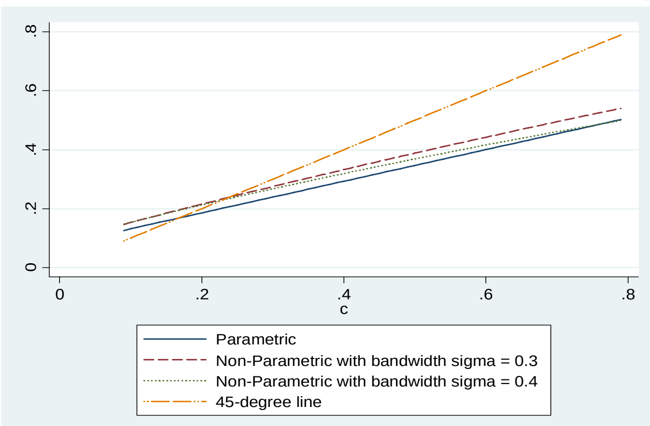


Figure 7 : Value Function ρ(c) in Restricted Case and Random Allocation Case

Panel A. Restricted Case: Conditioning on Wealth Only



Panel B. Random Allocation (no conditioning)

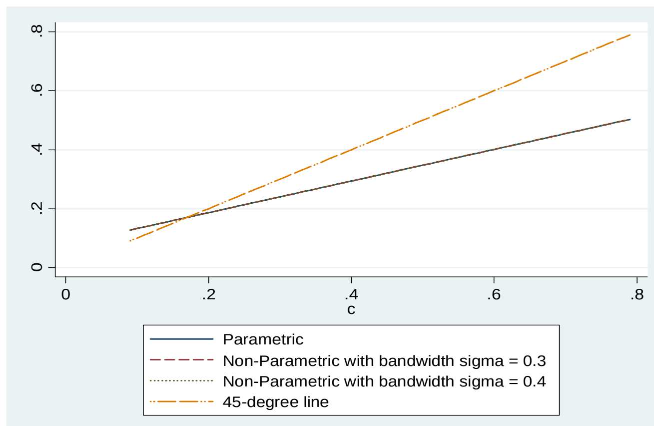
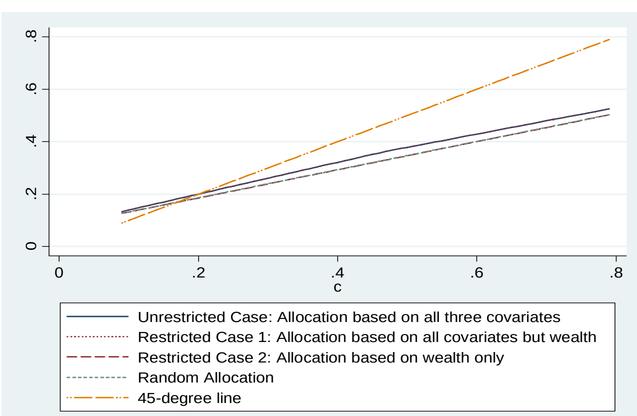
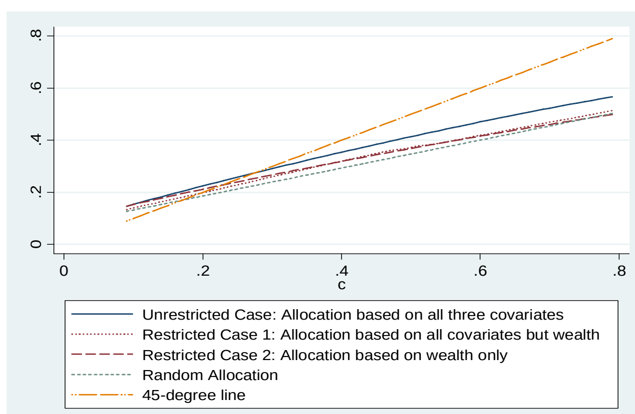


Figure 8: Value Function ρ(c), Parametric vs. Non-Parametric Estimation

Panel A. Parametric Estimation



Panel B. Non-Parametric Estimation (Bandwidth σ=0.4)



10 Appendix

10.1 Illustrative comparison of EWM and Bayes risks

Consider a standard decision problem where the loss function is given by $L(\delta, x)$ with δ denoting the action to be taken and x the value of the random variable X which will occur after the action is taken.

We first show that in both the parametric and nonparametric case, the frequentist risk of the EWM rule and that of the Bayes rule will differ by $O_p\left(\frac{1}{n}\right)$.

Nonparametric case: Suppose that the distribution of X is not assumed to belong to a specific parametric family. In this case, the EWM approach results in the expected risk

$$R_{EWM}(\delta|z_n) = \int_{\mathcal{X}} L(\delta, x)\, dF_n(x)$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x)$. In contrast, for a nonparametric Bayes (npb) solution, assume a Dirichlet process prior (Ferguson (1973)) with base measure $H(.)$ and concentration parameter α on the space of distribution functions for X.[12] This DP prior will result in the following predictive distribution (c.f., Schervish (1997), prop 1.98) for a "new" draw:

$$G_n(x) = \frac{\alpha}{\alpha+n} H(x) + \frac{n}{\alpha+n} \times \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x).$$

The difference between $F_n(x)$ and $G_n(x)$ is thus

$$|G_n(x) - F_n(x)| = \frac{\alpha}{\alpha+n}\left|H(x) - \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x)\right| = O_p\left(\frac{1}{n}\right),$$

uniformly in x. Furthermore, in this case, as $\alpha \to 0$, i.e., the prior becomes more diffuse, the predictive distribution tends to the empirical distribution.

Now, if the loss function is uniformly bounded on the action space, then a DCT will imply that

$$\sup_{\delta} |R_{EWM}(\delta|z_n) - R_{Bayes}(\delta|z_n)| = O_p\left(\frac{1}{n}\right). \tag{9}$$

This will imply that the difference in value functions, viz., $|\min_\delta R_{EWM}(\delta|z_n) - \min_\delta R_{Bayes}(\delta|z_n)|$ is also $O_p\left(\frac{1}{n}\right)$. To see why, suppose $\delta^* = \arg\min_\delta R_{EWM}(\delta|z_n)$ and $\tilde{a} = \arg\min_\delta R_{Bayes}(\delta|z_n)$. Now consider a sequence b_n such that $nb_n \to \infty$ as $n \to \infty$. Then

$$\begin{aligned}
& \Pr\left[R_{EWM}(\delta^*|z_n) - R_{Bayes}(\tilde{a}|z_n) > b_n\right] \\
\leq\; & \Pr\left[R_{EWM}(\tilde{a}|z_n) - R_{Bayes}(\tilde{a}|z_n) > b_n\right] \\
\leq\; & \Pr\left[\sup_{\delta} |R_{EWM}(\delta|z_n) - R_{Bayes}(\delta|z_n)| > b_n\right] \to 0 \text{ by (9).}
\end{aligned}$$

[12] For actual nonparametric Bayesian analysis, one may use other, e.g., Gaussian process or perturbed DP priors owing to the well-known property that the DP assigns probability 1 to a set of discrete distributions (c.f., Schervish (1997), page 56). We use the DP for our illustrations here to ease the exposition.

A similar reasoning works for $\Pr\left[R_{Bayes}\left(\tilde{a}|z_n\right) - R_{EWM}\left(\delta^*|z_n\right) > b_n\right]$. Since this argument holds for any sequence b_n of larger order than $\frac{1}{n}$, the conclusion follows.

Parametric case: Suppose the distribution of X is known to be $N(\theta, 1)$ and one has data on past IID realizations of X, i.e. the data are $z_n = (x_1, ...x_n)$. Letting $\phi(\cdot)$ denote the standard normal density, the risk function is $R(\delta; \theta) = \int L(\delta, x)\,\phi(x - \theta)\,dx$. Since θ is unknown, we consider the plug-in and the Bayes approximation to this risk based on z_n, as follows. It is worth noting that in the parametric case, a plug-in approach does not coincide with a Bayesian approach under an uninformative prior.

An (MLE) plug-in rule will result in the approximation

$$R_{plug-in}(\delta|z_n) = \int L(\delta, x)\,\phi(x - \bar{x})\,dx. \tag{10}$$

This function, though easier to compute, does not take into account finite-sample parameter uncertainty in θ.

On the other hand, in the Bayesian approach, a $N\left(\mu, \tau^2\right)$ prior for θ will result in the posterior risk

$$R_{Bayes}(\delta|z_n) = \int L(\delta, x)\left[\int_{-\infty}^{\infty} \phi(x - \theta)\,g\left(\theta; n, \mu, \tau^2, \bar{x}\right)d\theta\right]dx,$$

where

$$g\left(\theta; n, \mu, \tau^2, \bar{x}\right) \equiv \sqrt{n + \tau^{-2}} \times \phi\left(\left(\frac{1}{n\tau^2 + 1}\mu + \frac{n\tau^2}{n\tau^2 + 1}\bar{x} - \theta\right) \times \sqrt{n + \tau^{-2}}\right)$$

is the posterior density of θ satisfying

$$\int_{-\infty}^{\infty} g\left(\theta; n, \mu, \tau^2, \bar{x}\right)d\theta = 1,$$

and the resulting (posterior) predictive distribution of a new X, given the past data, is

$$h\left(x|z_n, \mu, \tau^2\right) \equiv \int_{-\infty}^{\infty} \phi(x - \theta)\,g\left(\theta; n, \mu, \tau^2, \bar{x}\right)d\theta.$$

If n is small, i.e. the posterior variance $\left(n + \tau^{-2}\right)^{-1}$ is large, so that there is more (posterior) parameter uncertainty in θ, then the predictive distribution of a new X will be more dispersed. Under convexity of $L(\delta, \cdot)$, this will lead to larger Bayes risk for every δ. As $\tau \to \infty$, i.e., the prior becomes more diffuse, $h\left(x|z_n, \mu, \tau^2\right) \to \sqrt{n} \times \phi\left(\sqrt{n}(x - \bar{x})\right)$ and

$$R_{Bayes}(\delta|z_n) \to \int L(\delta, x)\sqrt{n} \times \phi\left(\sqrt{n}(x - \bar{x})\right)dx$$

which differs from $R_{plug-in}(\delta|z_n)$ in (10), i.e., flat priors do not lead to the plug-in risk.

As $n \to \infty$, of course, the two risks will get closer in probability, calculated w.r.t. the population distribution of z_n. To see the rate of approach as $n \to \infty$ for a fixed τ, let $E_{\theta|z_n}(\cdot)$ denote expectation w.r.t. the posterior, viz., $\theta|z_n \sim N\left(\frac{1}{n\tau^2+1}\mu + \frac{n\tau^2}{n\tau^2+1}\bar{x}, \left\{n + \tau^{-2}\right\}^{-1}\right)$ and observe that by a standard mean value expansion,

$$\begin{aligned}
& \left|\int_{-\infty}^{\infty}\left[\phi\left(x-\theta\right)-\phi\left(x-\bar{x}\right)\right]g\left(\theta;n,\mu,\tau^{2},\bar{x}\right)d\theta\right| \\
= & \left|\int_{-\infty}^{\infty}\left[\left(\bar{x}-\theta\right)\phi'\left(x-\bar{x}\right)+\frac{1}{2}\phi''\left(x-\bar{x}\right)\left(\bar{x}-\theta\right)^{2}+o\left(\left(\bar{x}-\theta\right)^{2}\right)\right]g\left(\theta;n,\mu,\tau^{2},\bar{x}\right)d\theta\right| \\
= & O\left(\left|\phi'\left(x-\bar{x}\right)E_{\theta|z_{n}}\left(\bar{x}-\theta\right)+\frac{1}{2}\phi''\left(x-\bar{x}\right)E_{\theta|z_{n}}\left(\bar{x}-\theta\right)^{2}\right|\right) \\
= & O\left(\left|\phi'\left(x-\bar{x}\right)\frac{1}{n\tau^{2}+1}\left\{\mu-\bar{x}\right\}+\frac{1}{2}\phi''\left(x-\bar{x}\right)\frac{1}{n+\tau^{-2}}\right|\right) \\
= & O_{p}\left(\frac{1}{n}\right)\text{, uniformly in } x
\end{aligned}$$

where the "P" in "$O_p(\cdot)$" corresponds to the population distribution of $\bar{x}$ and uniformity follows because the functions ϕ' and ϕ'' are bounded uniformly on $\mathbb{R}$. Now, the conclusion follows from similar reasoning as the nonparametric case above.

Higher-order correction: Higher-order improvements in a frequentist sense and analogous to those in predictive inference can be achieved in the parametric case as follows. Suppose in the normal example above, our goal is to know the expected loss from action δ, i.e.,

$$G\left(\delta,\theta\right)\overset{def}{=}E\left[L\left(\delta,X_{n+1}\right)|X_{n+1}\sim N\left(\theta,1\right)\right]=\int L\left(\delta,x\right)\phi\left(x-\theta\right)dx.$$

Since we do not know θ, we can use $\bar{x}$ to approximate it. Now, the question is whether we can find a function $g\left(\cdot,X_1,...X_n\right)$ such that $E_{|\theta}\left\{\int L\left(\delta,x\right)g\left(x,X_1,...X_n\right)dx\right\}$ is a better approximation to $G\left(\delta,\theta\right)$ than the approximation $E_{\bar{x}|\theta}\left[\int L\left(\delta,x\right)\phi\left(x-\bar{x}\right)dx\right]$. To find such a $g\left(\cdot\right)$, notice that

$$\phi\left(x-\bar{x}\right)-\phi\left(x-\theta\right)=-\phi'\left(x-\theta\right)\times\left(\bar{x}-\theta\right)+\frac{1}{2}\phi''\left(x-\theta\right)\times\left(\bar{x}-\theta\right)^{2}+o\left(\left(\bar{x}-\theta\right)^{2}\right).$$

Taking expectation w.r.t. the population distribution of $\bar{x}$, under regularity conditions,

$$E_{\bar{x}|\theta}\left[\int L\left(\delta,x\right)\phi\left(x-\bar{x}\right)dx\right]=\int L\left(\delta,x\right)\phi\left(x-\theta\right)dx+\frac{1}{2}Var_{\bar{x}|\theta}\left(\bar{x}\right)\times\int\phi''\left(x-\theta\right)L\left(\delta,x\right)dx+o\left(\frac{1}{n}\right), \tag{11}$$

where $Var\left(\bar{x}\right)=O\left(\frac{1}{n}\right)$. So a better approximation (up to order $\frac{1}{n}$) is provided by

$$\int L\left(\delta,x\right)\left[\phi\left(x-\bar{x}\right)-\frac{1}{2n}\times\phi''\left(x-\bar{x}\right)\right]dx,$$

and choosing δ to minimize this risk will result in a smaller frequentist risk in finite samples. Cox (1975) discusses analogous higher order corrections for predictive inference problems and notes the typically small $O\left(\frac{1}{n}\right)$ order of adjustment that results from it.

10.2 Parameter uncertainty and ambiguity aversion

Suppose X has a discrete distribution with known support $\{a_1,...a_M\}$ and associated unknown probabilities $\theta=\{\theta_1,...\theta_M\}$. This may be viewed as the nonparametric case with discrete random variables. Suppose the prior distribution of θ is given by the Dirichlet $\pi(\theta) \propto \Pi_{m=1}^{M}\theta_m^{-1}$, which turns out to be an uniformative prior (c.f., Chamberlain and Imbens (2003)). Then the posterior is given by $\bar{\pi}(\theta|z_n) \propto \Pi_{m=1}^{M}\theta_m^{n_m-1}$, where n_m denotes the number of times a_m is realized in the IID sample $z_n \equiv \{x_1, ..., x_n\}$, with $\sum_{m=1}^{M} n_m = n$, the sample size. Then the Bayesian posterior risk is given by

$$\begin{aligned}
\int \left\{\sum_{m=1}^{M} L(\delta, a_m)\theta_m\right\} d\bar{\pi}(\theta|z_n) &= \sum_{m=1}^{M} L(\delta, a_m) \underbrace{\int \theta_m d\bar{\pi}(\theta|z_n)}_{predictive\ \Pr(X=a_m|z_n)} \\
&= \sum_{m=1}^{M} L(\delta, a_m) \times \frac{n_m}{n} = \int L(\delta, x)\, dF_n(x),
\end{aligned}$$

i.e., the EWM criterion. This equality essentially results from the underlying risk-function, i.e., $\sum_{m=1}^{M} L(\delta, a_m)\theta_m$ being linear in the unknown parameter θ. A new independent draw X_{n+1} has support $\{a_1, ...a_M\}$ with associated predictive probabilities $\left\{\frac{n_1}{n}, ...\frac{n_M}{n}\right\}$. Its variance is given by

$$\begin{aligned}
Var(X_{n+1}|z_n) &= E_{\bar{\pi}(\theta|z_n)}\{Var(X_{n+1}|\theta)\} + Var_{\bar{\pi}(\theta|z_n)}\{E(X_{n+1}|\theta)\} \\
&= \underbrace{E_{\bar{\pi}(\theta|z_n)}\left\{\sum_{j=1}^{M} a_j^2\theta_j - \left(\sum_{j=1}^{M} a_j\theta_j\right)^2\right\}}_{\text{conditional uncertainty}} + \underbrace{Var_{\bar{\pi}(\theta|z_n)}\left\{\sum_{j=1}^{M} a_j\theta_j\right\}}_{\text{parameter uncertainty}} \\
&= E_{\bar{\pi}(\theta|z_n)}\left\{\sum_{j=1}^{M} a_j^2\theta_j\right\} - \left(E_{\bar{\pi}(\theta|z_n)}\left\{\sum_{j=1}^{M} a_j\theta_j\right\}\right)^2 \\
&= \sum_{j=1}^{M} a_j^2\frac{n_j}{n} - \left\{\sum_{j=1}^{M} a_j\frac{n_j}{n}\right\}^2.
\end{aligned}$$

Thus the two sources of uncertainty (conditional and parametric) in the predictive distribution of X_{n+1} are treated symmetrically in how they affect the expected utility from the action δ.

In contrast, an ambiguity averse planner's objective would be of the form

$$\int V\left\{\sum_{m=1}^{M} L(\delta, a_m)\theta_m\right\} d\bar{\pi}(\theta|z_n),$$

where $V(\cdot)$ is concave and reflects aversion to ambiguity about θ, c.f. Ahn et al (2009).

10.3 Proofs of theorems

In the proofs below, CMT will denote continuous mapping theorem and DCT the Lebesgue dominated convergence theorem.

Proposition 1:

Proof. Note that for a generic set A, the objective function equals

$$\int_{x\in\mathcal{X}} [\phi(x,1) - \phi(x,0)]\, 1(x \in A)\, dF(x) + \int_{x\in\mathcal{X}} \phi(x,0)\, dF(x), \tag{12}$$

and the second term does not depend on A. So in the proof below, we will simply refer to the first term as the objective function.

Note that the objective function for a generic choice set A can be written as

$$\begin{aligned}
& \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)>\gamma\} dF(x) + \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)\le\gamma\} dF(x) \\
= & \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)>\gamma\} dF(x) + \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)\le\gamma\} dF(x) \\
& - \int_{x\in\mathcal{X}} [\theta(x)] 1(x\notin A) 1\{\theta(x)>\gamma\} dF(x) - \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)>\gamma\} dF(x) \\
& + \int_{x\in\mathcal{X}} [\theta(x)] 1\{\theta(x)>\gamma\} dF(x). \\
= & \int_{x\in\mathcal{X}} [\theta(x)] 1(x\in A) 1\{\theta(x)\le\gamma\} dF(x) - \int_{x\in\mathcal{X}} [\theta(x)] 1(x\notin A) 1\{\theta(x)>\gamma\} dF(x) \\
& + \int_{x\in\mathcal{X}} [\theta(x)] 1\{\theta(x)>\gamma\} dF(x). \qquad (13)
\end{aligned}$$

Now, the first term in the previous display is bounded above by

$$\gamma \int_{x\in\mathcal{X}} 1(x\in A) 1\{\theta(x)\le\gamma\} dF(x), \qquad (14)$$

while the second term, without the negative sign, is strictly bounded below by

$$\gamma \int_{x\in\mathcal{X}} 1(x\notin A) 1\{\theta(x)>\gamma\} dF(x). \qquad (15)$$

Now from the budget constraint, we have that

$$\begin{aligned}
c &= \int_{x\in\mathcal{X}} 1(x\in A) dF(x) \\
&= \int_{x\in\mathcal{X}} \left[\begin{array}{c} 1(x\in A) 1\{\theta(x)\le\gamma\} \\ +1(x\in A) 1\{\theta(x)>\gamma\} \end{array}\right] dF(x)
\end{aligned}$$

and

$$c = \int_{x\in\mathcal{X}} 1\{\theta(x)>\gamma\} = \int_{x\in\mathcal{X}} [1(x\notin A) 1\{\theta(x)>\gamma\} + 1(x\in A) 1\{\theta(x)>\gamma\}] dF(x)$$

whence it follows that

$$\gamma \int_{x\in\mathcal{X}} 1(x\in A) 1\{\theta(x)\le\gamma\} dF(x) = \gamma \int_{x\in\mathcal{X}} 1(x\notin A) 1\{\theta(x)>\gamma\} dF(x). \qquad (16)$$

It follows from (13), (14), (15), (16) that the objective function in (12) is bounded above by

$$\int_{x\in\mathcal{X}} [\theta(x)] 1\{\theta(x)>\gamma\} dF(x) + \int_{x\in\mathcal{X}} \phi(x,0) dF(x),$$

which corresponds to setting $A = \{x\in X : \theta(x)>\gamma\}$ with $\gamma\equiv\gamma(c)$ satisfying

$$c = \int_{x\in\mathcal{X}} 1\{\theta(x)>\gamma(c)\} dF(x).$$

■

Lemma 1 *Under assumptions A0-A3, A4(i), B1, B2, B3(i) and B4(i),*

$$\sup_{t\in[-M,M]}\left|\hat{F}_{\hat{\theta}}(t)-F_{\theta}(t)\right|\xrightarrow{P}0.$$

Proof. Observe that

$$\begin{aligned}
&\hat{F}_{\hat{\theta}}(t)-F_{\theta}(t)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{t-\hat{\theta}(X_i)}{h_n}\right)-F_{\theta}(t)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{t-\hat{\theta}(X_i)}{h_n}\right)-\frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{t-\theta(X_i)}{h_n}\right)+\frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{t-\theta(X_i)}{h_n}\right)-1(\theta(X_i)\le t)\right\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\{1(\theta(X_i)\le t)-F_{\theta}(t)\}\\
=\;&\frac{1}{nh_n}\sum_{i=1}^{n}L\left(\frac{t-\tilde{\theta}(X_i)}{h_n}\right)\left\{\theta(X_i)-\hat{\theta}(X_i)\right\}+\frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{t-\theta(X_i)}{h_n}\right)-1(\theta(X_i)\le t)\right\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\{1(\theta(X_i)\le t)-F_{\theta}(t)\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{t\in[-M,M]}\left|\hat{F}_{\hat{\theta}}(t)-F_{\theta}(t)\right|\\
\le\;&\sup_{t\in[-M,M]}\left|\frac{1}{n}\sum_{i=1}^{n}\{1(\theta(X_i)\le t)-F_{\theta}(t)\}\right|\\
&+\sup_{t\in[-M,M]}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{t-\theta(X_i)}{h_n}\right)-1(\theta(X_i)\le t)\right\}\right|\\
&+\frac{1}{n^{1/4}h_n}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[-M,M]}\left|L\left(\frac{t-\tilde{\theta}(X_i)}{h_n}\right)\right|\right)\times\left\{n^{1/4}\sup_{a}\left|\theta(a)-\hat{\theta}(a)\right|\right\}
\end{aligned}$$

By assumption B3(i) (i.e. $L(\cdot)$ is uniformly bounded), assumption B4(i) (i.e. $nh_n^4\to\infty$) and assumption B1, the third term is $o_p(1)$. The first term is $o_p(1)$ by the standard Glivenko-Cantelli theorem. The second term is $o_p(1)$ by Horowitz (1992), lemma 4 under assumptions about $\bar{L}$ and that $\theta(X)$ has a Lebesgue density which is uniformly bounded above (analogous to his proof that $\lim_{\alpha\to0}\Pr(|b'x|<\alpha)$, here we have that

$$\begin{aligned}
\lim_{\alpha\to0}\Pr(|t-\theta(X)|<\alpha)&=\lim_{\alpha\to0}\Pr(-\alpha<t-\theta(X)<\alpha)\\
&=\lim_{\alpha\to0}\Pr(t-\alpha<\theta(X)<t+\alpha)\\
&=\lim_{\alpha\to0}[F_{\theta}(t+\alpha)-F_{\theta}(t-\alpha)]\\
&\le2\lim_{\alpha\to0}\left\{\alpha\times\sup_{s\in\mathbb{R}}[f_{\theta}(s)]\right\}=0,
\end{aligned}$$

and the rest of the proof is identical to Horowitz lemma 4). ∎

Theorem 1:

Proof. Fix $\varepsilon > 0$. Then $F_\theta(\gamma+\varepsilon) - 1 + c > 0$ and $1 - c - F_\theta(\gamma-\varepsilon) > 0$, by assumption (B5). Therefore, we have that

$$\begin{aligned}
\Pr(|\hat{\gamma}-\gamma| > \varepsilon) &\leq \Pr(\hat{\gamma} > \gamma+\varepsilon) + \Pr(\hat{\gamma} < \gamma-\varepsilon) \\
&\leq \Pr\left(\hat{F}_{\hat{\theta}}(\hat{\gamma}) > \hat{F}_{\hat{\theta}}(\gamma+\varepsilon)\right) + \Pr\left(\hat{F}_{\hat{\theta}}(\hat{\gamma}) < \hat{F}_{\hat{\theta}}(\gamma-\varepsilon)\right) \\
&= \Pr\left(1-c > \hat{F}_{\hat{\theta}}(\gamma+\varepsilon)\right) + \Pr\left(1-c < \hat{F}_{\hat{\theta}}(\gamma-\varepsilon)\right) \\
&\leq \Pr\left(F_\theta(\gamma+\varepsilon) - 1 + c < F_\theta(\gamma+\varepsilon) - \hat{F}_{\hat{\theta}}(\gamma+\varepsilon)\right) \\
&\quad + \Pr\left(1 - c - F_\theta(\gamma-\varepsilon) < \hat{F}_{\hat{\theta}}(\gamma-\varepsilon) - F_\theta(\gamma-\varepsilon)\right) \\
&\leq \Pr\left(F_\theta(\gamma+\varepsilon) - 1 + c < \sup_{t\in[-M,M]}\left|\hat{F}_{\hat{\theta}}(t) - F_\theta(t)\right|\right) \\
&\quad + \Pr\left(1 - c - F_\theta(\gamma-\varepsilon) < \sup_{t\in[-M,M]}\left|\hat{F}_{\hat{\theta}}(t) - F_\theta(t)\right|\right)
\end{aligned}$$

both of which converge to zero by lemma 1.

∎

The following lemma shows that $\hat{f}_{\hat{\theta}}(\cdot)$ converges to $f_\theta(\cdot)$ in probability, uniformly on the support of $\theta(X)$.

Lemma 2 *Under assumptions A0-A4 and B1-B5,*

$$\sup_{u\in[-M,M]}\left|\hat{f}_{\hat{\theta}}(u) - f_\theta(u)\right| = o_p(1).$$

Proof. Observe that

$$\hat{f}_{\hat{\theta}}(u) - f_\theta(u) = \frac{1}{nh_n}\sum_{i=1}^{n} L\left(\frac{u-\hat{\theta}(X_i)}{h_n}\right) - f_\theta(u)$$

By triangle inequality,

$$\sup_{u\in[-M,M]}\left|\hat{f}_{\hat{\theta}}(u) - f_\theta(u)\right| \leq \sup_{u\in[-M,M]}\left|\hat{f}_\theta(u) - f_\theta(u)\right| + \sup_{u\in[-M,M]}\left|\hat{f}_{\hat{\theta}}(u) - \hat{f}_\theta(u)\right|.$$

The first term is $o_p(1)$ under assumption B2. As for the second term, notice that

$$\begin{aligned}
\left|\hat{f}_{\hat{\theta}}(t) - \hat{f}_\theta(t)\right| &= \left|\frac{1}{nh_n}\sum_{i=1}^{n}\left\{L\left(\frac{t-\hat{\theta}(X_i)}{h_n}\right) - L\left(\frac{t-\theta(X_i)}{h_n}\right)\right\}\right| \\
&= \left|\frac{1}{nh_n^2}\sum_{i=1}^{n} L'\left(\frac{t-\tilde{\theta}(X_i)}{h_n}\right)\left\{\hat{\theta}(X_i) - \theta(X_i)\right\}\right| \\
&\leq \frac{\sup_x\left|\hat{\theta}(x) - \theta(x)\right|}{h_n^2}\frac{1}{n}\sum_{i=1}^{n}\left|L'\left(\frac{t-\tilde{\theta}(X_i)}{h_n}\right)\right| \\
&= O_p\left(\frac{1}{h_n^2}\times\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2} + \sigma_n^q\right\}\right),
\end{aligned}$$

by assumptions B2 and B3. Therefore by assumption B4, we get the conclusion. ∎

Theorem 2:

To derive the distribution theory for $\hat{\gamma}$, we will use the following first-order approximation

$$F_\theta(\gamma) = 1 - c = \hat{F}_{\hat{\theta}}(\hat{\gamma}) = \hat{F}_{\hat{\theta}}(\gamma) + (\hat{\gamma} - \gamma)\hat{f}_{\hat{\theta}}(\tilde{\gamma})$$

where $\tilde{\gamma}$ is intermediate between $\hat{\gamma}$ and γ. This gives us the following expansion for $\hat{\gamma}$.

$$\begin{aligned}
& (\hat{\gamma} - \gamma) \\
= & \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma) - \hat{F}_{\hat{\theta}}(\gamma)\right\} \\
= & \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma) - \frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{\gamma - \hat{\theta}(X_i)}{h_n}\right)\right\} \\
= & \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma) - \frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\} \\
& + \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right) - \bar{L}\left(\frac{\gamma - \hat{\theta}(X_i)}{h_n}\right)\right]\right\}. \qquad (17)
\end{aligned}$$

Proof. Step 1. We first show that

$$\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\gamma) \xrightarrow{P} 0. \qquad (18)$$

$$\begin{aligned}
\left|\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\gamma)\right| & \leq \left|\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\tilde{\gamma})\right| + \left|f_\theta(\tilde{\gamma}) - f_\theta(\gamma)\right| \\
& \leq \underbrace{\sup_{s\in[-M,M]}\left|\hat{f}_{\hat{\theta}}(s) - f_\theta(s)\right|}_{o_p(1)\text{, by lemma 2}} + \underbrace{\left|f_\theta(\tilde{\gamma}) - f_\theta(\gamma)\right|}_{o_p(1)\text{ by CMT and theorem 1}} \\
& = o_p(1).
\end{aligned}$$

Step 2: We will show that

$$\sqrt{nh_n}\left\{F_\theta(\gamma) - \bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\} = \beta + o_p(1). \qquad (19)$$

Observe that

$$\begin{aligned}
T_n & = \frac{1}{n}\sum_{i=1}^{n}\left\{F_\theta(\gamma) - \bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\} \\
& = \frac{1}{n}\sum_{i=1}^{n}\left\{F_\theta(\gamma) - 1(\theta(X_i) \leq \gamma)\right\} + \frac{1}{n}\sum_{i=1}^{n}\left\{1(\theta(X_i) \leq \gamma) - \bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\} \\
& \equiv T_{2n} - T_{1n}. \qquad (20)
\end{aligned}$$

Now,

$$\sqrt{nh_n}T_{2n} = \sqrt{h_n} \times \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{F_\theta(\gamma) - 1(\theta(X_i) \leq \gamma)\right\}}_{O_p(1)} = o_p(1).$$

We will show that

$$E\left(\sqrt{nh_n}T_{1n}-\beta\right)^2 = h_n Var\left(\sqrt{n}T_{1n}\right)+\left\{E\left(\sqrt{nh_n}T_{1n}-\beta\right)\right\}^2 \to 0 \tag{21}$$

and thus

$$\sqrt{nh_n}T_{1n}-\beta = o_p\left(1\right). \tag{22}$$

Now,

$$\begin{aligned}
& Var\left(\sqrt{n}T_{1n}\right) \\
= \;& Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\le\gamma\right)-\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}\right) \\
= \;& Var\left\{1\left(\theta\left(X_i\right)\le\gamma\right)-\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
= \;& E\left\{1\left(\theta\left(X_i\right)\le\gamma\right)-\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}^2-\left\{F_\theta\left(\gamma\right)-E\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}\right\}^2
\end{aligned} \tag{23}$$

Observe that

$$\begin{aligned}
& E\left(1\left(\theta\left(X_i\right)\le\gamma\right)-\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right)^2 \\
= \;& \int_{-M}^{M}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)-1\left(s\le\gamma\right)\right\}^2 f_\theta\left(s\right)ds \\
= \;& \int_{-M}^{\gamma}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)-1\left(s\le\gamma\right)\right\}^2 f_\theta\left(s\right)ds \\
& +\int_{\gamma}^{A}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)-1\left(s\le\gamma\right)\right\}^2 f_\theta\left(s\right)ds \\
= \;& \int_{-M}^{\gamma}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)-1\right\}^2 f_\theta\left(s\right)ds+\int_{\gamma}^{A}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)\right\}^2 f_\theta\left(s\right)ds
\end{aligned}$$

and both of the terms in the previous display converge to zero by the DCT since $\lim_{a\to\infty}\bar{L}\left(a\right)=1=1-\lim_{a\to-\infty}\bar{L}\left(a\right)$.

Next,

$$\begin{aligned}
& F_\theta\left(\gamma\right)-E\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
= \;& F_\theta\left(\gamma\right)-\int_{-M}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)f_\theta\left(s\right)ds \\
= \;& F_\theta\left(\gamma\right)-\int_{-M}^{\gamma}\bar{L}\left(\frac{\gamma-s}{h_n}\right)f_\theta\left(s\right)ds-\int_{\gamma}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)f_\theta\left(s\right)ds \\
= \;& \int_{-M}^{\gamma}\left[1\left(s\le\gamma\right)-\bar{L}\left(\frac{\gamma-s}{h_n}\right)\right]f_\theta\left(s\right)ds-\int_{\gamma}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)f_\theta\left(s\right)ds \\
\to \;& 0,\text{ by the DCT.}
\end{aligned}$$

Thus, from (23), we have that

$$Var\left(\sqrt{n}T_{1n}\right)\to 0 \text{ as } n\to\infty. \tag{24}$$

Next, consider

$$
\begin{aligned}
E\left(T_{1n}\right) &= E\left\{1\left(\theta\left(X_i\right)\leq\gamma\right)-\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&= \left\{F_\theta\left(\gamma\right)-\int_{-M}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)f_\theta\left(s\right)ds\right\} \\
&= \left\{F_\theta\left(\gamma\right)-\bar{L}\left(\frac{\gamma-s}{h_n}\right)F_\theta\left(s\right)|_{-M}^{M}-\frac{1}{h_n}\int_{-M}^{M}F_\theta\left(s\right)L\left(\frac{\gamma-s}{h_n}\right)ds\right\} \\
&= \left\{F_\theta\left(\gamma\right)-\int_{\frac{\gamma-M}{h_n}}^{\frac{\gamma+M}{h_n}}F_\theta\left(\gamma-uh_n\right)L\left(u\right)du\right\} \\
&= \left(-1\right)^{r+1}\frac{h_n^r}{r!}\times f_\theta^{(r-1)}\left(\gamma\right)\times\int_{-1}^{1}u^rL\left(u\right)du+o\left(h_n^r\right), \text{ by assumption B7.}
\end{aligned}
$$

This implies that

$$
\begin{aligned}
E\left(\sqrt{nh_n}T_{1n}\right) &= \left(-1\right)^{r+1}\frac{\sqrt{n}h_n^{r+1/2}}{r!}\times f_\theta^{(r-1)}\left(\gamma\right)\times\int_{-1}^{1}u^rL\left(u\right)du+o\left(h_n^r\right) \\
&\to \beta, \text{ by assumption B7.}
\end{aligned}
\tag{25}
$$

Now, (24) and (25) imply (21) and thus (22).

Step 3: We will now analyze the second term in (17):

$$
S_n=\frac{1}{n}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma-\hat{\theta}\left(X_i\right)}{h_n}\right)\right]du,
$$

using U-statistic type decompositions to show that

$$
\begin{aligned}
\sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{j=1}^{n}\left\{\left[\lambda_{1n}\left(Z_j\right)-E\left\{\lambda_{1n}\left(Z_j\right)\right\}\right]-\left[\lambda_{2n}\left(Z_j\right)-E\left\{\lambda_{2n}\left(Z_j\right)\right\}\right]\right\} \\
&\quad +o_p\left(1\right) \\
&\xrightarrow{d} N\left(0,\eta^2\right),
\end{aligned}
\tag{26}
$$

where the triangular arrays $\lambda_{1n}\left(Z_j\right)$, $\lambda_{2n}\left(Z_j\right)$ and the constant $\eta^2>0$, will be specified below.

To that end observe that

$$
\begin{aligned}
\sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma-\hat{\theta}\left(X_i\right)}{h_n}\right)\right] \\
&= \frac{\sqrt{h_n}}{\sqrt{n}h_n}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right) \\
&\quad +\frac{\sqrt{h_n}}{2\sqrt{n}h_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}^2L'\left(\frac{\tilde{\theta}\left(X_i\right)-\gamma}{h_n}\right).
\end{aligned}
$$

The second term in absolute value has an expectation which is of the order of

$$
\sup_{x\in\mathcal{X}}\left|\theta\left(x\right)-\hat{\theta}\left(x\right)\right|^2\frac{\sqrt{n}}{h_n^{3/2}}=O_p\left\{\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}^2\frac{\sqrt{n}}{h_n^{3/2}}\right\}\to 0, \text{ by assumption B8.}
$$

Thus we get that

$$
\begin{aligned}
\sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)-\bar{L}\left(\frac{\hat{\theta}\left(X_i\right)-\gamma}{h_n}\right)\right] \\
&= \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)-\hat{\theta}\left(X_i\right)\right\}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+o_p\left(1\right) \\
&= -\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+o_p\left(1\right).
\end{aligned}
$$

Now, note that

$$
\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)=\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\}-\left\{\frac{\hat{\nu}\left(X_i\right)}{\hat{\delta}(X_i)}-\frac{\nu\left(X_i\right)}{\delta\left(X_i\right)}\right\} \tag{27}
$$

We will simply work with the first term because the proof is exactly analogous for the second term and show that

$$
\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)=O_p\left(1\right).
$$

Step 3A: Now,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) \\
=\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)-\mu(X_i)}{\pi\left(X_i\right)}\right\}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\mu\left(X_i\right)}{\pi\left(X_i\right)}\frac{\hat{\pi}\left(X_i\right)-\pi(X_i)}{\pi\left(X_i\right)}\right\}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left\{\hat{\mu}\left(X_i\right)-\mu(X_i)\right\}\left\{\hat{\pi}\left(X_i\right)-\pi(X_i)\right\}}{\pi\left(X_i\right)\hat{\pi}\left(X_i\right)}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu(X_i)\left\{\hat{\pi}\left(X_i\right)-\pi(X_i)\right\}^2}{\pi^2\left(X_i\right)\hat{\pi}\left(X_i\right)}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right).
\end{aligned} \tag{28}
$$

The last two terms in absolute value have expectations that are bounded above by a positive scalar times $\sqrt{n}\sup_x\left\|\left\{\hat{\mu}\left(x\right)-\mu(x)\right\}\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\|$ and $\sqrt{n}\sup_x\left\|\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\|^2$, respectively and these are both $o_p\left(1\right)$ under standard conditions (c.f. NM, section 8.3) which is assumption B11 above.

Now, the first two terms in (28) add up to

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\pi\left(X_i\right)\hat{\mu}\left(X_i\right)-\mu\left(X_i\right)\hat{\pi}\left(X_i\right)}{\pi^2\left(X_i\right)}\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) \\
= \; & \sqrt{n}\frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right)\times\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)\right] \\
\equiv \; & \sqrt{n}\frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}w_n\left(Z_i,Z_j\right) \\
= \; & \underbrace{\frac{1}{\sqrt{n}\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}\left[w_n\left(Z_i,Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_i\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)+E\left(w_n\left(Z_i,Z_j\right)\right)\right]}_{U_{1n}} \\
& \underbrace{+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)\right)\right]}_{U_{2n}} \\
& \underbrace{+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(w_n\left(Z_i,Z_j\right)|Z_i\right).}_{U_{3n}} \qquad (29)
\end{aligned}
$$

Step 3B: We first show that

$$U_{3n}=o_p\left(1\right). \qquad (30)$$

Notice that

$$
\begin{aligned}
& E\left[\frac{\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma}{h_n}\right)}{\pi^2\left(X_i\right)}\times\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right)|Z_i\right] \\
& \overset{L.I.E.}{=} E\left[\frac{\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma}{h_n}\right)}{\pi^2\left(X_i\right)}\times\left\{\pi\left(X_i\right)\frac{\mu\left(X_j\right)}{f\left(X_j\right)}-\mu\left(X_i\right)\frac{\pi\left(X_j\right)}{f\left(X_j\right)}\right\}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right)|Z_i\right] \\
= \; & \frac{\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma}{h_n}\right)}{\pi^2\left(X_i\right)}\times\int\frac{\pi\left(X_i\right)\mu\left(x\right)-\mu\left(X_i\right)\pi\left(x\right)}{f\left(x\right)}\frac{1}{\sigma_n^p}K\left(\frac{x-X_i}{\sigma_n}\right)f\left(x\right)dx \\
= \; & \frac{\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma}{h_n}\right)}{\pi^2\left(X_i\right)}\times\int\left[\pi\left(X_i\right)\mu\left(X_i+u\sigma_n\right)-\mu\left(X_i\right)\pi\left(X_i+u\sigma_n\right)\right]K\left(u\right)du \\
& \overset{A1}{=} H\left(X_i\right)\times\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)\times O\left(\sigma_n^q\right),
\end{aligned}
$$

for some uniformly bounded function H by assumption. Therefore,

$$U_{3n}=O\left(\sigma_n^q\right)\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H\left(X_i\right)\times\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)=O_p\left(\sqrt{n}\sigma_n^q\right)=o_p\left(1\right)$$

by assumption B8.

Step 3C: The term

$$U_{1n}=\frac{1}{\sqrt{n}\,(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left[w_n\left(Z_i,Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_i\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)+E\left(w_n\left(Z_i,Z_j\right)\right)\right]$$

can be analyzed using essentially the steps of Powell, Stoker and Stock (1989), lemma 3.1, whence one can conclude that

$$E\left(U_{1n}^2\right)=o\left(1\right) \tag{31}$$

The key step is to show that

$$E\left(w_n^2\left(Z_i,Z_j\right)\right)=o\left(n\right).$$

Observe that

$$\begin{aligned}
& n^{-1}E\left(w_n^2\left(Z_i,Z_j\right)\right)\\
= \;& n^{-1}E\left\{\frac{1}{\pi^4\left(X_i\right)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}^2\frac{1}{\sigma_n^{2p}}K^2\left(\frac{X_j-X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)\right]^2\right\}\\
= \;& n^{-1}E\left\{\begin{array}{c}\frac{1}{\pi^4(X_i)}\frac{1}{\sigma_n^{2p}}K^2\left(\frac{X_j-X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma}{h_n}\right)\right]^2\\ \times\left\{\pi^2\left(X_i\right)E\left(Y^2S|X_j\right)+\mu^2(X_i)E\left(S|X_j\right)-2\pi\left(X_i\right)\mu\left(X_i\right)E\left(YS|X_j\right)\right\}\end{array}\right\}\\
= \;& \frac{1}{n\sigma_n^p h_n^2}\int\left\{\begin{array}{c}\frac{1}{\pi^4(X)}K^2\left(u\right)\times\left[L\left(\frac{\theta(x)-\gamma}{h_n}\right)\right]^2\\ \times\left\{\begin{array}{c}\pi^2\left(x\right)E\left(Y^2S|X=x+u\sigma_n\right)\\ +\mu^2(x)E\left(S|X=x+u\sigma_n\right)\\ -2\pi\left(x\right)\mu\left(x\right)E\left(YS|X=x+u\sigma_n\right)\end{array}\right\}\end{array}\right\}f_X\left(x\right)f_X\left(x+u\sigma_n\right)dudx\\
= \;& O\left(\frac{1}{n\sigma_n^p h_n^2}\right)\to 0 \text{ which is implied by B8.}
\end{aligned}$$

Step 3D: Now consider the term

$$U_{2n}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)\right)\right]$$

Observe that

$$
\begin{aligned}
& E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) \\
= \; & E\left\{\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right)\right]\right]|Y_j,S_j,X_j\right\} \\
= \; & \int\left[\frac{1}{\pi^2\left(x\right)}\left\{\pi\left(x\right)Y_jS_j-\mu\left(x\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+x}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(x\right)-\gamma}{h_n}\right)\right]\right]f\left(x\right)dx \\
= \; & \int\left[\frac{\pi\left(X_j+u\sigma_n\right)Y_jS_j-\mu\left(X_j+u\sigma_n\right)S_j}{\pi^2\left(X_j+u\sigma_n\right)}K\left(u\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j+u\sigma_n\right)-\gamma}{h_n}\right)\right]\right]f\left(X_j+u\sigma_n\right)du \\
= \; & \left[\frac{1}{\pi^2\left(X_j\right)}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j\right)-\gamma}{h_n}\right)\right]\right]\int K\left(u\right)du+O\left(\sigma_n^q\right) \\
= \; & \left[\underbrace{\left[\frac{1}{\pi^2\left(X_j\right)}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right)\right]}_{W(Z_j)}\times\underbrace{\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j\right)-\gamma}{h_n}\right)\right]}_{V_{1n}(\theta(X_j))}\right] \\
& +O\left(\sigma_n^q\right).
\end{aligned}
$$

Notice that

$$
\begin{aligned}
& E\left\{W\left(Z_j\right)V_n\left(\theta\left(X_j\right)\right)\right\} \\
= \; & E\left\{V_n\left(\theta\left(X_j\right)\right)E\left(W\left(Z_j\right)|X_j\right)\right\} \\
= \; & E\left\{\frac{V_n\left(\theta\left(X_j\right)\right)f\left(X_j\right)}{\pi^2\left(X_j\right)}E\left(\left\{\pi\left(X_j\right)E\left\{Y_jS_j|X_j\right\}-\mu\left(X_j\right)E\left(S|X_j\right)\right\}\right)\right\} \\
= \; & E\left\{\frac{V_n\left(\theta\left(X_j\right)\right)f\left(X_j\right)}{\pi^2\left(X_j\right)}\times 0\right\}=0.
\end{aligned}
$$

Now

$$
\begin{aligned}
Var\left(U_{2n}\right) &= Var\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)\right)\right]\right\} \\
&= Var\left\{\left[E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)\right)\right]\right\} \\
&= Var\left\{E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)\right\} \\
&= E\left(W\left(Z_j\right)V_{1n}\left(\theta\left(X_j\right)\right)+O\left(\sigma_n^q\right)\right)^2-O\left(\sigma_n^{2q}\right) \\
&= E\left\{W^2\left(Z_j\right)V_{1n}^2\left(\theta\left(X_j\right)\right)\right\}+O\left(\sigma_n^{2q}\right) \\
&= O\left(E\left\{W^2\left(Z_j\right)V_{1n}^2\left(\theta\left(X_j\right)\right)\right\}\right).
\end{aligned}
$$

Now, let $\omega^2(s) = E\left\{W^2(Z_j)\,|\theta(X_j) = s\right\}$. Then

$$\begin{aligned}
& E\left\{W^2(Z_j)\,V_{1n}^2(\theta(X_j))\right\} \\
= & \int_{-M}^{M} \omega^2(s)\left[\frac{1}{h_n}L\left(\frac{s-\gamma}{h_n}\right)\right]^2 f_\theta(s)\,ds \\
= & \frac{1}{h_n}\int_{\frac{-M-\gamma}{h_n}}^{\frac{M-\gamma}{h_n}} \omega^2(\gamma+uh_n)\,L^2(u)\,f_\theta(\gamma+uh_n)\,du \\
= & \frac{1}{h_n}\omega^2(\gamma)\,f_\theta(\gamma)\int_{\frac{-M-\gamma}{h_n}}^{\frac{M-\gamma}{h_n}} L^2(u)\,du + \text{terms of smaller order.}
\end{aligned}$$

This implies that

$$\begin{aligned}
Var\left(\sqrt{h_n}U_{2n}\right) &= Var\left(\sqrt{\frac{h_n}{n}}\sum_{j=1}^{n}\left[E\left(w_n(Z_i,Z_j)\,|Z_j\right) - E\left(w_n(Z_i,Z_j)\right)\right]\right) \\
&\to \omega^2(\gamma)\,f_\theta(\gamma)\int_{-\infty}^{\infty} L^2(u)\,du. \qquad (32)
\end{aligned}$$

Now we will apply the Liapunov condition and use the Lindeberg CLT for triangular arrays. Consider the array

$$R_{nj} = \frac{\sqrt{h_n}}{\sqrt{n}}\left[E\left(w_n(Z_i,Z_j)\,|Z_j\right) - E\left(w_n(Z_i,Z_j)\right)\right],$$

which is independent across j and $E(R_{nj}) = 0$. Let $U_n = \sum_{j=1}^{n} R_{nj}$. Then

$$\begin{aligned}
E\left(U_n^2\right) &= \sum_{j=1}^{n} E\left(R_{nj}^2\right) \\
&= \frac{h_n}{n}\sum_{j=1}^{n} E\left(E\left(w_n(Z_i,Z_j)\,|Z_j\right) - E\left(w_n(Z_i,Z_j)\right)\right)^2 \\
&= \frac{h_n}{n}\sum_{j=1}^{n} Var\left(W(Z_j)\,V_{1n}(\theta(X_j))\right) + o(1) \\
&= \frac{h_n}{h_n}\omega^2(\gamma)\,f_\theta(\gamma)\int_{-\infty}^{\infty} L^2(u)\,du + o(1) \\
&\to \omega^2(\gamma)\,f_\theta(\gamma)\int_{-\infty}^{\infty} L^2(u)\,du.
\end{aligned}$$

by (32). To apply the Liapunov condition, observe that for any $\varepsilon > 0$,

$$\begin{aligned}
\sum_{j=1}^{n} E\left|R_{nj}\right|^{2+\varepsilon} &= n\left(\frac{h_n}{n}\right)^{\frac{2+\varepsilon}{2}} E\left|W(Z_j)\,V_{1n}(\theta(X_j))\right|^{2+\varepsilon} \\
&= O\left(n\left(\frac{h_n}{n}\right)^{\frac{2+\varepsilon}{2}}\frac{1}{h_n^{1+\varepsilon}}\right) \\
&= O\left((h_n n)^{-\varepsilon/2}\right) \to 0.
\end{aligned}$$

Thus the Liapunov condition holds and applying the Lindeberg CLT, we get that

$$U_n = \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) - E\left(w_n\left(Z_i, Z_j\right)\right)\right] \xrightarrow{d} N\left(0, \omega^2\left(\gamma\right) f_\theta\left(\gamma\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right). \quad (33)$$

Putting together (30), (31), (32) and (32), we get that

$$\begin{aligned} & \sqrt{\frac{h_n}{n}} \sum_{i=1}^{n} \left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)} - \frac{\mu(X_i)}{\pi\left(X_i\right)}\right\} \left[\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma}{h_n}\right)\right] \\ = & \sqrt{h_n} U_{1n} + \sqrt{h_n} U_{2n} + \sqrt{h_n} U_{3n} \\ = & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) - E\left(w_n\left(Z_i, Z_j\right)\right)\right] + o_p\left(1\right) \\ = & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[\lambda_{1n}\left(Z_j\right) - E\left\{\lambda_{1n}\left(Z_j\right)\right\}\right] \\ & \xrightarrow{d} N\left(0, \omega^2\left(\gamma\right) f_\theta\left(\gamma\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right), \end{aligned}$$

where

$$\begin{aligned} & \lambda_{1n}\left(Z_j\right) \\ = & \left[\frac{f\left(X_j\right)}{\pi^2\left(X_j\right)} \left\{\pi\left(X_j\right) Y_j S_j - \mu\left(X_j\right) S_j\right\}\right] \frac{1}{h_n} L\left(\frac{\theta\left(X_j\right) - \gamma}{h_n}\right) \\ & \text{and} \\ \omega^2\left(s\right) = & E\left\{\left\{\frac{\pi\left(X\right) Y S - \mu\left(X\right) S}{\pi^2\left(X\right)} f\left(X\right)\right\}^2 |\theta\left(X\right) = s\right\}. \end{aligned} \quad (34)$$

Similarly, we will get that

$$\begin{aligned} & \sqrt{\frac{h_n}{n}} \sum_{i=1}^{n} \left\{\frac{\hat{\nu}\left(X_i\right)}{\hat{\delta}(X_i)} - \frac{\nu\left(X_i\right)}{\delta\left(X_i\right)}\right\} \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma}{h_n}\right) \\ = & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \lambda_{2n}\left(Z_j\right) \xrightarrow{d} N\left(0, \tau^2\left(\gamma\right) f_\theta\left(\gamma\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right), \end{aligned}$$

where

$$\begin{aligned} & \lambda_{2n}\left(Z_j\right) \\ = & \left[\frac{f\left(X_j\right)}{\delta^2\left(X_j\right)} \left\{\delta\left(X_j\right) Y_j \left(1 - S_j\right) - \nu\left(X_j\right)\left(1 - S_j\right)\right\}\right] \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma}{h_n}\right) \\ & \text{and} \\ \tau^2\left(s\right) = & E\left\{\left\{\frac{\delta\left(X\right) Y\left(1 - S\right) - \nu\left(X\right)\left(1 - S\right)}{\delta^2\left(X\right)} f\left(X\right)\right\}^2 |\theta\left(X\right) = s\right\}. \end{aligned} \quad (35)$$

Thus we get that

$$\sqrt{n h_n} S_n = \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left\{\lambda_{1n}\left(Z_j\right) - \lambda_{2n}\left(Z_j\right)\right\} + o_p\left(1\right) \xrightarrow{d} N\left(0, \eta^2\right),$$

which establishes (26).

To get the expression for η^2, note further that

$$\begin{aligned}
& E\left\{\lambda_{1n}\left(Z_j\right)\lambda_{2n}\left(Z_j\right)\right\} \\
= & E\left\{\frac{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j}{\pi^2\left(X_j\right)}\times\frac{\delta\left(X_j\right)Y_j\left(1-S_j\right)-\nu\left(X_j\right)\left(1-S_j\right)}{\delta^2\left(X_j\right)}\times f^2\left(X_j\right)L\left(\frac{\theta\left(X_j\right)-\gamma}{h_n}\right)^2\right\}
\end{aligned}$$

Now,

$$\begin{aligned}
& \left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}\times\left\{\delta\left(X_j\right)Y_j\left(1-S_j\right)-\nu\left(X_j\right)\left(1-S_j\right)\right\} \\
= & S_j\left(1-S_j\right)\times\left\{\pi\left(X_j\right)Y_j-\mu\left(X_j\right)\right\}\times\left\{\delta\left(X_j\right)Y_j-\nu\left(X_j\right)\right\} \\
= & 0,
\end{aligned}$$

since $S_j\left(1-S_j\right)=0$ for every j. Therefore, $E\left\{\lambda_{1n}\left(Z_j\right)\lambda_{2n}\left(Z_j\right)\right\}=0$. Moreover, $E\left(\lambda_{1n}\left(Z_j\right)\right)=0$. Therefore, $cov\left(\lambda_{1n}\left(Z_j\right),\lambda_{2n}\left(Z_j\right)\right)=0$. This implies that

$$\eta^2=\left\{\tau^2\left(\gamma\right)+\omega^2\left(\gamma\right)\right\}\times f_\theta\left(\gamma\right)\int_{-\infty}^{\infty}L^2\left(u\right)du, \tag{36}$$

where

$$\begin{aligned}
\tau^2\left(\gamma\right) &= E\left\{\left\{\frac{\delta\left(X\right)Y\left(1-S\right)-\nu\left(X\right)\left(1-S\right)}{\delta^2\left(X\right)}f\left(X\right)\right\}^2|\theta\left(X\right)=\gamma\right\} \\
\omega^2\left(\gamma\right) &= E\left\{\left\{\frac{\pi\left(X\right)YS-\mu\left(X\right)S}{\pi^2\left(X\right)}f\left(X\right)\right\}^2|\theta\left(X\right)=\gamma\right\}.
\end{aligned}$$

Now put together (18), (19), (26) and (36) to conclude from (17) that

$$\begin{aligned}
\sqrt{nh_n}\left(\hat{\gamma}-\gamma\right) &= \frac{1}{f_\theta\left(\gamma\right)}\frac{\sqrt{h_n}}{\sqrt{n}}\sum_{j=1}^{n}\left\{\lambda_{1n}\left(Z_j\right)-\lambda_{2n}\left(Z_j\right)\right\}+o_p\left(1\right) \\
& \stackrel{d}{\to} N\left(0,\frac{\tau^2\left(\gamma\right)+\omega^2\left(\gamma\right)}{f_\theta\left(\gamma\right)}\int_{-\infty}^{\infty}L^2\left(u\right)du\right).
\end{aligned}$$

■

Lemma 3 *Under assumptions A0-A4 and B1-B11,*

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\hat{\phi}\left(X_j,1\right)-\phi\left(X_j,1\right)\right] \\
= & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right)\frac{1}{\pi^2\left(X_j\right)}+o_p\left(1\right) \\
= & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{1}{\left\{E\left(S|X_j\right)\right\}^2}\left\{E\left(S|X_j\right)\times Y_jS_j-E\left(SY|X_j\right)\times S_j\right\}+o_p\left(1\right).
\end{aligned}$$

Proof. Note that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\hat{\phi}\left(X_j,1\right)-\phi\left(X_j,1\right)\right] \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\} \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)-\mu(X_i)}{\pi\left(X_i\right)}\right\}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\mu\left(X_i\right)}{\pi\left(X_i\right)}\frac{\hat{\pi}\left(X_i\right)-\pi(X_i)}{\pi\left(X_i\right)}\right\} \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left\{\hat{\mu}\left(X_i\right)-\mu(X_i)\right\}\left\{\hat{\pi}\left(X_i\right)-\pi(X_i)\right\}}{\pi\left(X_i\right)\hat{\pi}\left(X_i\right)} \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu(X_i)\left\{\hat{\pi}\left(X_i\right)-\pi(X_i)\right\}^2}{\pi^2\left(X_i\right)\hat{\pi}\left(X_i\right)}. \qquad (37)
\end{aligned}
$$

The last two terms are bounded above by a positive scalar times $\sqrt{n}\sup_x\left\|\left\{\hat{\mu}\left(x\right)-\mu(x)\right\}\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\|$ and $\sqrt{n}\sup_x\left\|\left\{\hat{\pi}\left(x\right)-\pi(x)\right\}\right\|^2$, respectively and these are both $o_p\left(1\right)$ under assumption B11 above. Thus we only need to show that the sum of the first two terms in (37) is asymptotically equivalent to ($\sqrt{n}$ times) a centered sample average.

Now, the first two terms in (37) add up to

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\pi\left(X_i\right)\hat{\mu}\left(X_i\right)-\mu\left(X_i\right)\hat{\pi}\left(X_i\right)\right\} \\
= & \sqrt{n}\frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j}{\pi^2\left(X_i\right)}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right) \\
\equiv & \sqrt{n}\frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}w\left(Z_i,Z_j,\sigma_n\right) \\
= & \underbrace{\frac{1}{\sqrt{n}\left(n-1\right)}\sum_{i=1}^{n}\sum_{j\neq i}\left[w\left(Z_i,Z_j,\sigma_n\right)-E\left(w\left(Z_i,Z_j,\sigma_n\right)|Z_i\right)-E\left(w\left(Z_i,Z_j,\sigma_n\right)|Z_j\right)+E\left(w\left(Z_i,Z_j,\sigma_n\right)\right)\right]}_{U_{1n}} \\
& \underbrace{+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w\left(Z_i,Z_j,\sigma_n\right)|Z_j\right)-E\left(w\left(Z_i,Z_j,\sigma_n\right)\right)\right]}_{U_{2n}} \\
& \underbrace{+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(w\left(Z_i,Z_j,\sigma_n\right)|Z_i\right).}_{U_{3n}}
\end{aligned}
$$

We will show that

$$
\begin{aligned}
E\left(U_{1n}\right)^2 &= o\left(1\right), \qquad (38) \\
U_{2n} &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{E\left(S|X_j\right)\times Y_jS_j-E\left(SY|X_j\right)\times S_j\right\}+o_p\left(1\right), \qquad (39) \\
U_{3n} &= o_p\left(1\right). \qquad (40)
\end{aligned}
$$

Observe that

$$
\begin{aligned}
& E\left[\frac{\pi\left(X_i\right) Y_j S_j-\mu\left(X_i\right) S_j}{\pi^2\left(X_i\right)} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) | Z_i\right] \\
& \stackrel{L.I.E.}{=} E\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right) \frac{\mu\left(X_j\right)}{f\left(X_j\right)}-\mu\left(X_i\right) \frac{\pi\left(X_j\right)}{f\left(X_j\right)}\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) | Z_i\right] \\
= & \frac{1}{\pi^2\left(X_i\right)} \int\left[\pi\left(X_i\right) \mu(x)-\mu\left(X_i\right) \pi(x)\right] \frac{1}{\sigma_n^p} K\left(\frac{x-X_i}{\sigma_n}\right) d x \\
= & \frac{1}{\pi^2\left(X_i\right)} \int\left[\pi\left(X_i\right) \mu\left(X_i+u \sigma_n\right)-\mu\left(X_i\right) \pi\left(X_i+u \sigma_n\right)\right] K(u) d u \\
& \stackrel{A1}{=} H\left(X_i\right) \times O\left(\sigma_n^q\right),
\end{aligned}
$$

for some uniformly bounded function H by assumption. Therefore, $U_{3n}=O_p\left(\sqrt{n}\sigma_n^q\right)=o_p(1)$ by assumption A3 and this establishes (40).

Next observe that

$$
\begin{aligned}
& E\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right) Y_j S_j-\mu\left(X_i\right) S_j\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) | Z_j\right] \\
= & \int \frac{1}{\pi^2(x)}\left\{\pi(x) Y_j S_j-\mu(x) S_j\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-x}{\sigma_n}\right) f(x) d x \\
= & \int \frac{1}{\pi^2\left(X_j+u \sigma_n\right)}\left\{\pi\left(\left(X_j+u \sigma_n\right)\right) Y_j S_j-\mu\left(\left(X_j+u \sigma_n\right)\right) S_j\right\} K(u) f\left(\left(X_j+u \sigma_n\right)\right) d u \\
= & \left\{\pi\left(X_j\right) Y_j S_j-\mu\left(X_j\right) S_j\right\} f\left(X_j\right) \int \frac{1}{\pi^2\left(X_j+u \sigma_n\right)} K(u) d u \\
& +\sigma_n\left\{\pi^{\prime}\left(X_j\right) Y_j S_j-\mu^{\prime}\left(X_j\right) S_j\right\} f\left(X_j\right) \int \frac{1}{\pi^2\left(X_j+u \sigma_n\right)} K(u) u d u \\
& +\ldots.+\sigma_n^q\left\{\pi^{(q)}\left(X_j\right) Y_j S_j-\mu^{(q)}\left(X_j\right) S_j\right\} f\left(X_j\right) \int \frac{1}{\pi^2\left(X_j+u \sigma_n\right)} K(u) u^q d u \\
= & \left\{\pi\left(X_j\right) Y_j S_j-\mu\left(X_j\right) S_j\right\} \frac{f\left(X_j\right)}{\pi^2\left(X_j\right)}+O\left(\sigma_n^q\right) \\
= & \frac{1}{\left\{E\left(S | X_j\right)\right\}^2}\left\{E\left(S | X_j\right) \times Y_j S_j-E\left(S Y | X_j\right) \times S_j\right\}+O\left(\sigma_n^q\right),
\end{aligned}
$$

by a dominated convergence theorem, given the uniform boundedness of $\pi(\cdot)$. Together with assumption A3, we get (39).

One can establish (38) by essentially repeating the proof of Powell, Stoker and Stock (1989) lemma 3.1.

Combining (38), (39) and (40), we get that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\} \\
= & \frac{1}{\sqrt{n}} \sum_{j=1}^n\left\{\pi\left(X_j\right) Y_j S_j-\mu\left(X_j\right) S_j\right\} f\left(X_j\right) \frac{1}{\pi^2\left(X_j\right)}+o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{1}{\left\{E\left(S | X_j\right)\right\}^2}\left\{E\left(S | X_j\right) \times Y_j S_j-E\left(S Y | X_j\right) \times S_j\right\}+o_p(1) .
\end{aligned}
$$

■

Theorem 3:

Proof.

$$\begin{aligned}
\hat{\zeta}-\zeta &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}\left(X_i\right)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}\left(X_i\right)}{h_n}\right)-\zeta \\
&= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}\left(X_i\right)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}\left(X_i\right)}{h_n}\right)-\frac{1}{n}\sum_{i=1}^{n}\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)}_{T_{1n}} \\
&\quad \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\theta\left(X_i\right)\left[\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-1\left\{\theta\left(X_i\right)\leq\gamma\right\}\right]}_{T_{2n}} \\
&\quad \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)1\left\{\theta\left(X_i\right)\leq\gamma\right\}-\zeta\right\}.}_{=o_p(1),\text{ by standard WLLN.}}
\end{aligned}$$

Now,

$$\begin{aligned}
\left|T_{1n}\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)}{h_n}\right|\left|h_n\bar{L}\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)-\tilde{\theta}\left(X_i\right)L\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right| \\
&\quad +\frac{\left(\hat{\gamma}-\gamma\right)}{h_n}\times\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{\theta}\left(X_i\right)L\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right| \\
&\leq \frac{\sup\left|\hat{\theta}\left(x\right)-\theta\left(x\right)\right|}{h_n}\frac{1}{n}\sum_{i=1}^{n}\left|h_n\bar{L}\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)-\tilde{\theta}\left(X_i\right)L\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right| \\
&\quad +\left(\frac{nh_n\left(\hat{\gamma}-\gamma\right)^2}{nh_n^3}\right)^{1/2}\times\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{\theta}\left(X_i\right)L\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right|.
\end{aligned}$$

Since $L,\bar{L}$ are uniformly bounded, the above display is of the form

$$\leq\frac{\sup_{x\in\mathcal{X}}\left|\hat{\theta}\left(x\right)-\theta\left(x\right)\right|}{h_n}\times O_p\left(1\right)+\left(\frac{nh_n\left(\hat{\gamma}-\gamma\right)^2}{nh_n^3}\right)^{1/2}\times O_p\left(1\right).$$

Now, theorem 2 implies that $nh_n\left(\hat{\gamma}-\gamma\right)^2=O_p\left(1\right)$, Assumptions B1 and B4 (i) imply that $\frac{\sup\left|\hat{\theta}(x)-\theta(x)\right|}{h_n}=o_p\left(1\right)$ and that $nh_n^3\rightarrow\infty$. Thus we have that $T_{1n}=o_p\left(1\right)$.

As for T_{2n}, observe that since $\theta\left(\cdot\right)$ is uniformly bounded, by using steps exactly analogous to step 2 in the proof of theorem 2 (leading to (19)), we will get by the DCT that $T_{2n}=o_p\left(1\right)$.

Now combine with lemma 3 to conclude. ■

Theorem 4:

Proof. We will work with the following expansion

$$
\begin{aligned}
& \hat{\zeta}-\zeta \\
= \; & \underbrace{\frac{1}{n}\sum_{i=1}^{n}\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\zeta}_{T_{1n}} \\
& \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\frac{1}{h_n}\theta\left(X_i\right)L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}}_{T_{2n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma\right)\frac{1}{nh_n}\sum_{i=1}^{n}\theta\left(X_i\right)L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)}_{T_{3n}} \\
& \underbrace{-\frac{1}{4nh_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}^2\left\{-2h_nL\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)+\tilde{\theta}\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right\}}_{T_{4n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma\right)^2\times\frac{1}{4nh_n^2}\sum_{i=1}^{n}\theta\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)}_{T_{5n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma\right)\frac{1}{2nh_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left[h_nL\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)-\tilde{\theta}\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right]}_{T_{6n}}. \quad (41)
\end{aligned}
$$

Step 4: Under assumptions B1 and B8, the fourth term in (41) will be $o_p\left(\frac{1}{\sqrt{n}}\right)$ since $L'(\cdot)$ is assumed to be uniformly bounded in absolute value. As for the fifth term, observe by the previous theorem, that $\frac{(\hat{\gamma}-\gamma)^2}{h_n^2}=O_p\left(\frac{1}{nh_n^3}\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$ by assumption B12. So the fifth term in (41) will be $o_p\left(\frac{1}{\sqrt{n}}\right)$. That the sixth term is $o_p(1)$ follows from combining the two previous results.

Step 5: The multiplier for the third term in (41) equals

$$
\begin{aligned}
\frac{1}{nh_n}\sum_{i=1}^{n}\theta\left(X_i\right)L\left(\frac{\theta\left(X_i\right)-\gamma}{h_n}\right) & \rightarrow E\left(\theta\left(X_i\right)\frac{1}{h_n}\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right) \\
& = \gamma f_\theta(\gamma)+O\left(h_n^r\right) \\
& \rightarrow \gamma f_\theta(\gamma),
\end{aligned}
$$

which follows from the standard consistency proof for e.g. kernel density estimates.

Combining steps 4 and 5, we get that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta\right\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\zeta\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\frac{1}{h_n}\theta\left(X_i\right)L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+\sqrt{n}\left(\hat{\gamma}-\gamma\right)\times\gamma f_{\theta}\left(\gamma\right)+o_p\left(1\right).
\end{aligned} \tag{42}$$

Replacing in the previous display the asymptotic expansion of $(\hat{\gamma}-\gamma)$ from (17), we have that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta\right\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\zeta\right\}-\frac{\gamma}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma\right)-F_{\theta}\left(\gamma\right)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+\gamma\frac{1}{2\sqrt{n}h_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}^2 L'\left(\frac{\tilde{\theta}\left(X_i\right)-\gamma}{h_n}\right) \\
&+o_p\left(1\right).
\end{aligned} \tag{43}$$

The third term in (43) in absolute value is dominated by

$$\begin{aligned}
&\frac{\gamma}{2}\times\sup_u\left\{\hat{\theta}\left(u\right)-\theta\left(u\right)\right\}^2\times\frac{\sqrt{n}}{h_n^2}\frac{1}{n}\sum_{i=1}^{n}\left|L'\left(\frac{\tilde{\theta}\left(X_i\right)-\gamma}{h_n}\right)\right| \\
= \; & \frac{\gamma}{2}\times\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}^2\times\frac{\sqrt{n}}{h_n^2}\times O_p\left(1\right) \\
= \; & o_p\left(1\right), \text{ by assumption B4(iii).}
\end{aligned}$$

Thus from (43), we have that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta\right\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\zeta\right\}-\frac{\gamma}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma\right)-F_{\theta}\left(\gamma\right)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+o_p\left(1\right).
\end{aligned} \tag{44}$$

Step 6A: Consider the first term in (44)

$$
\begin{aligned}
&= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right]}_{T_{4n}} \\
&\quad \underbrace{+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}-\zeta\right]}_{T_{5n}=O_p(1),\text{ by CLT.}} \qquad (45)
\end{aligned}
$$

We will show that $T_{4n}=o_p\left(1\right)$ using the arguments similar to the ones used for showing (22).

Define $g_\theta\left(s\right)=sf_\theta\left(s\right)$ and $G_\theta\left(s\right)=\int_{-M}^{s}g_\theta\left(t\right)dt$. Then

$$
\begin{aligned}
E\left(T_{4n}\right) &= \sqrt{n}E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-G_\theta\left(\gamma\right)\right\} \\
&= \sqrt{n}\left\{\int_{-M}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)g_\theta\left(s\right)ds-G_\theta\left(\gamma\right)\right\} \\
&= \sqrt{n}\left[\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)G_\theta\left(s\right)\right\}|_{-M}^{M}-\int_{-M}^{M}\frac{1}{h_n}L\left(\frac{\gamma-s}{h_n}\right)G_\theta\left(s\right)ds-G_\theta\left(\gamma\right)\right] \\
&= \sqrt{n}\left[-\int_{\frac{\gamma-M}{h_n}}^{\frac{\gamma+M}{h_n}}L\left(u\right)G_\theta\left(\gamma-uh_n\right)du-G_\theta\left(\gamma\right)\right] \\
&= O\left(\sqrt{n}h_n^r\right)\to 0,\text{ by B12.} \qquad (46)
\end{aligned}
$$

Next, define $g_\theta\left(s\right)=s^2f_\theta\left(s\right)$ and $G_\theta\left(s\right)=\int_{-M}^{s}g_\theta\left(t\right)dt$. Then

$$
\begin{aligned}
& E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right\}^2 \\
= & \int_{-M}^{M}\left\{\bar{L}\left(\frac{\gamma-s}{h_n}\right)-1\left\{s<\gamma\right\}\right\}^2 g_\theta\left(s\right)ds \\
= & \int_{-M}^{M}\bar{L}^2\left(\frac{\gamma-s}{h_n}\right)g_\theta\left(s\right)ds+\int_{-M}^{\gamma}1\left\{s<\gamma\right\}g_\theta\left(s\right)ds \\
& -2\int_{-M}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)\times 1\left\{s<\gamma\right\}\times g_\theta\left(s\right)ds
\end{aligned}
$$

Using the DCT repeatedly (c.f. the steps leading to (20)), we get that

$$
\begin{aligned}
& E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right\}^2 \\
= & \int_{-M}^{M}\bar{L}^2\left(\frac{\gamma-s}{h_n}\right)g_\theta\left(s\right)ds+\int_{-M}^{\gamma}1\left\{s<\gamma\right\}g_\theta\left(s\right)ds \\
& -2\int_{-M}^{M}\bar{L}\left(\frac{\gamma-s}{h_n}\right)\times 1\left\{s<\gamma\right\}\times g_\theta\left(s\right)ds \\
\to & G_\theta\left(\gamma\right)+G_\theta\left(\gamma\right)-2G_\theta\left(\gamma\right)=0.
\end{aligned}
$$

This implies that for T_{4n} defined in (45),

$$\begin{aligned}
Var\left(T_{4n}\right) &= Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right]\right)\\
&= Var\left(\theta\left(X_i\right)\times\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right)\\
&\leq E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}\right\}^2\\
&\to 0. \qquad (47)
\end{aligned}$$

From (46) and (47), we get that $E\left(T_{4n}\right)^2\to 0$ and thus $T_{4n}=o_p(1)$.

Replacing in (44), we get that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta\right\}\\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma\right\}-\zeta\right]-\frac{\gamma}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma\right)-F_\theta\left(\gamma\right)\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}\\
&+o_p\left(\frac{1}{\sqrt{n}}\right). \qquad (48)
\end{aligned}$$

The final step is to analyze the third term in (48), using U-statistic type decompositions. First notice that analogous to (29) above, we have here that up to $o_p(1)$ terms:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta\left(X_i\right)}{h_n}\right)\right\}\\
= &\sqrt{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left[\begin{array}{c}\frac{1}{\pi^2(X_i)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right)\\ \times\left[\bar{L}\left(\frac{\gamma-\theta(X_i)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_i)}{h_n}\right)\right]\end{array}\right]\\
\equiv &\sqrt{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}w_n\left(Z_i,Z_j\right)\\
= &\underbrace{\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left[w_n\left(Z_i,Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_i\right)-E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)+E\left(w_n\left(Z_i,Z_j\right)\right)\right]}_{U_{1n}}\\
&\underbrace{+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i,Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i,Z_j\right)\right)\right]}_{U_{2n}}\\
&\underbrace{+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(w_n\left(Z_i,Z_j\right)|Z_i\right)}_{U_{3n}}.
\end{aligned}$$

It is straightforward (replace the kernel involving terms) to verify that we will get the same conclusion as (31) and (30) here. So we only perform the analysis for U_{2n}.

Using steps similar to the case for $\hat{\gamma}$, one gets that

$$
\begin{aligned}
& E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) \\
= \; & E\left\{\left[\begin{array}{c} \frac{1}{\pi^2(X_i)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+X_i}{\sigma_n}\right) \\ \times\left[\bar{L}\left(\frac{\gamma-\theta(X_i)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_i)}{h_n}\right)\right] \end{array}\right]|Y_j, S_j, X_j\right\} \\
= \; & \int\left[\begin{array}{c} \frac{1}{\pi^2(X)}\left\{\pi\left(x\right)Y_jS_j-\mu\left(x\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+x}{\sigma_n}\right) \\ \times\left[\bar{L}\left(\frac{\gamma-\theta(X_i)}{h_n}\right)+\left\{\gamma-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_i)}{h_n}\right)\right] \end{array}\right]f\left(x\right)dx \\
= \; & \int\left[\begin{array}{c} \frac{1}{\pi^2(X_j+u\sigma_n)}\left\{\pi\left(X_j+u\sigma_n\right)Y_jS_j-\mu\left(X_j+u\sigma_n\right)S_j\right\}K\left(u\right) \\ \times\left[\bar{L}\left(\frac{\gamma-\theta(X_j+u\sigma_n)}{h_n}\right)+\left\{\gamma-\theta\left(X_j+u\sigma_n\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_j+u\sigma_n)}{h_n}\right)\right] \end{array}\right]f\left(X_j+u\sigma_n\right)du \\
= \; & \left[\begin{array}{c} \frac{1}{\pi^2(X_j)}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right) \\ \times\left[\bar{L}\left(\frac{\gamma-\theta(X_j)}{h_n}\right)+\left\{\gamma-\theta\left(X_j\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_j)}{h_n}\right)\right] \end{array}\right]\int K\left(u\right)du+O\left(\sigma_n^q\right) \\
= \; & \left[\begin{array}{c} \underbrace{\left[\frac{\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}}{\pi^2\left(X_j\right)}f\left(X_j\right)\right]}_{W(Z_j)} \\ \times\underbrace{\left[\bar{L}\left(\frac{\gamma-\theta\left(X_j\right)}{h_n}\right)+\left\{\gamma-\theta\left(X_j\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta\left(X_j\right)}{h_n}\right)\right]}_{V_{2n}(\theta(X_j))} \end{array}\right]+O\left(\sigma_n^q\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right)-E\left(w_n\left(Z_i, Z_j\right)\right)\right] \\
= \; & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right)-W\left(Z_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma\right)\right] \\
& +\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{W\left(Z_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma\right)\right\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{W\left(Z_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma\right)\right\} \\
& +\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\underbrace{W\left(Z_j\right)\left[\begin{array}{c} \left\{\bar{L}\left(\frac{\gamma-\theta(X_j)}{h_n}\right)-1\left(\theta\left(X_j\right)\leq\gamma\right)\right\} \\ +\left\{\gamma-\theta\left(X_j\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_j)}{h_n}\right) \end{array}\right]}_{T_{nj}}. \qquad (49)
\end{aligned}
$$

Now, we will show that the second term in the previous display is $o_p(1)$. Recall the notation

$\omega^2(s) = E\left(W^2(Z_j)\,|\theta(X_j) = s\right)$ and thus

$$\begin{aligned}
& E\left(T_{nj}^2\right) \\
= & \int_{-M}^{M} \omega^2(s)\left[\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1(s \le \gamma)\right\} + \{\gamma - s\}\frac{1}{h_n}L\left(\frac{\gamma - s}{h_n}\right)\right]^2 f_\theta(s)\,ds \\
= & \int_{-M}^{M} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1(s \le \gamma)\right\}^2 f_\theta(s)\,ds \\
& + \int_{-M}^{M} \omega^2(s)\left(\frac{\gamma - s}{h_n}\right)^2 L^2\left(\frac{\gamma - s}{h_n}\right) f_\theta(s)\,ds \\
& + 2\int_{-M}^{M} \omega^2(s)\,\bar{L}\left(\frac{\gamma - s}{h_n}\right)\left(\frac{\gamma - s}{h_n}\right) L\left(\frac{\gamma - s}{h_n}\right) f_\theta(s)\,ds. \qquad (50)
\end{aligned}$$

The first term in (50) equals

$$\begin{aligned}
& \int_{-M}^{M} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1(s \le \gamma)\right\}^2 f_\theta(s)\,ds \\
= & \int_{-M}^{\gamma} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1(s \le \gamma)\right\}^2 f_\theta(s)\,ds + \int_{\gamma}^{M} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1(s \le \gamma)\right\}^2 f_\theta(s)\,ds \\
= & \int_{-M}^{\gamma} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right) - 1\right\}^2 f_\theta(s)\,ds + \int_{\gamma}^{M} \omega^2(s)\left\{\bar{L}\left(\frac{\gamma - s}{h_n}\right)\right\}^2 f_\theta(s)\,ds
\end{aligned}$$

and both of the terms in the previous display converge to zero by the DCT since $\lim_{a\to\infty}\bar{L}(u) = 1 = 1 - \lim_{a\to-\infty}\bar{L}(u)$. The second integral in (50) converges to zero by the DCT since $\lim_{u\to\pm\infty} u^2L^2(u) = 0$. The third integral in (50) also converges to zero by $\lim_{u\to\pm\infty} uL(u) = 0$ and the DCT. This implies that $E\left(T_{nj}^2\right) \to 0$ and thus

$$0 < Var\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} T_{nj}\right) = Var(T_{nj}) \le E\left(T_{nj}^2\right) \to 0.$$

Next,

$$\begin{aligned}
& \sqrt{n}E\left\{W(Z_j)\left[\left\{\bar{L}\left(\frac{\gamma - \theta(X_j)}{h_n}\right) - 1(\theta(X_j) \le \gamma)\right\} + \{\gamma - \theta(X_j)\}\frac{1}{h_n}L\left(\frac{\gamma - \theta(X_j)}{h_n}\right)\right]\right\} \\
= & \sqrt{n}E_{X_j}\left\{\underbrace{E\{W(Z_j)\,|X_j\}}_{=0} \times \left[\begin{array}{c} \left\{\bar{L}\left(\frac{\gamma - \theta(X_j)}{h_n}\right) - 1(\theta(X_j) \le \gamma)\right\} \\ + \{\gamma - \theta(X_j)\}\frac{1}{h_n}L\left(\frac{\gamma - \theta(X_j)}{h_n}\right) \end{array}\right]\right\} \\
= & 0.
\end{aligned}$$

So it follows that the second term in (49) is $o_p(1)$.

Thus we have that

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}(X_i)}{\hat{\pi}(X_i)} - \frac{\mu(X_i)}{\pi(X_i)}\right\}\left\{\bar{L}\left(\frac{\gamma - \theta(X_i)}{h_n}\right) + \{\gamma - \theta(X_i)\}\frac{1}{h_n}L\left(\frac{\gamma - \theta(X_i)}{h_n}\right)\right\} \\
= & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\{\pi(X_j)Y_jS_j - \mu(X_j)S_j\}}{\pi^2(X_j)}f_X(X_j) \times 1(\theta(X_j) \le \gamma)\right\} + o_p(1) \\
& \xrightarrow{d} N\left(0, \int_{-M}^{\gamma}\omega^2(s)\,f_\theta(s)\,ds\right),
\end{aligned}$$

where

$$\omega^2(a) = E\left\{\left[\frac{\{\pi(X_j) Y_j S_j - \mu(X_j) S_j\}}{\pi^2(X_j)} f_X(X_j)\right]^2 |\theta(X_j) = a\right\}$$

Using exactly analogous steps, we will also get that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\nu}(X_i)}{\hat{\delta}(X_i)} - \frac{\nu(X_i)}{\delta(X_i)}\right\}\left\{\bar{L}\left(\frac{\gamma-\theta(X_j)}{h_n}\right) + \{\gamma-\theta(X_j)\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_j)}{h_n}\right)\right\} \\
= \;&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\delta(X_j) Y_j(1-S_j) - \nu(X_j)(1-S_j)}{\delta^2(X)} f_X(X_j) \times 1(\theta(X_j) \le \gamma)\right\} + o_p(1). \\
&\xrightarrow{d} N\left(0, \int_{-M}^{\gamma} \tau^2(s) f_\theta(s)\, ds\right),
\end{aligned}$$

where

$$\tau^2(a) = E\left\{\left\{\frac{\delta(X) Y(1-S) - \nu(X)(1-S)}{\delta^2(X)} f(X)\right\}^2 |\theta(X) = a\right\}.$$

Finally, we get that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i) - \theta(X_i)\right\}\left\{\bar{L}\left(\frac{\gamma-\theta(X_j)}{h_n}\right) + \{\gamma-\theta(X_j)\}\frac{1}{h_n}L\left(\frac{\gamma-\theta(X_j)}{h_n}\right)\right\} \\
&\xrightarrow{d} N\left(0, \int_{-M}^{\gamma} \left\{\omega^2(s) + \tau^2(s)\right\} f_\theta(s)\, ds\right),
\end{aligned} \tag{51}$$

since the covariances will be zero (as can be easily seen from the asymptotic linear expansions because $S(1-S) = 0$).

Replacing in (42), we finally arrive at

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta} - \zeta\right\} \\
= \;&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\{\pi(X_j) Y_j S_j - \mu(X_j) S_j\}}{\pi^2(X_j)} f_X(X_j) \times 1(\theta(X_j) \le \gamma)\right\} \\
&-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\delta(X_j) Y_j(1-S_j) - \nu(X_j)(1-S_j)}{\delta^2(X)} f_X(X_j) \times 1(\theta(X_j) \le \gamma)\right\} \\
&+\gamma \times \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{F_\theta(\gamma) - 1(\theta(X_j) \le \gamma)\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\theta(X_i) \times 1\{\theta(X_i) < \gamma\} - \zeta\} + o_p(1).
\end{aligned} \tag{52}$$

Now, observe that

$$\begin{aligned}
&\hat{\rho} - \rho \\
= \;&\underbrace{\frac{1}{n}\sum_{j=1}^{n}\left[\hat{\phi}(X_j, 1) - \phi(X_j, 1)\right]}_{S_{1n}} + \underbrace{\frac{1}{n}\sum_{j=1}^{n}[\phi(X_j, 1) - E\{\phi(X_j, 1)\}]}_{S_{2n}} - \left\{\hat{\zeta} - \zeta\right\}.
\end{aligned}$$

By lemma 3, $S_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$, S_{2n} is a standard empirical process and so $O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\left\{\hat{\zeta} - \zeta\right\}$ is $O_p\left(\frac{1}{\sqrt{n}}\right)$ as shown as part of the proof of theorem 2. ■