NBER WORKING PAPER SERIES

NONPARAMETRIC INFERENCE ON EFFICIENT 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

Nonparametric Inference on Efficient Treatment Assignment under Budget Constraints
Debopam Bhattacharya and Pascaline Dupas
NBER Working Paper No. 14447
October 2008
JEL No. C01,C14,I38

ABSTRACT

We consider the problem of efficiently allocating a binary treatment among a target population based on a set of discrete and continuous observed characteristics. The goal is to maximize the population mean of an eventual outcome when a budget constraint limits what fraction of the population can be treated. Using sample data resulting from randomized treatment allocation, the ATE conditional on covariates (CATE) is nonparametrically estimated in a first step. The optimal treatment threshold and resulting value function, which are non-smooth functionals of the CATE, are estimated based on sample realizations of the estimated CATE. We derive large-sample distribution theory for these estimates and for the estimated dual value, i.e. the minimum resources needed to attain a specific average outcome via efficient treatment assignment. These inferential methods are applied to the optimal provision of anti-malaria bed nets, using data from a randomized experiment conducted in western Kenya. We find that a government which can afford to distribute subsidized bed nets 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 can be extended to infer optimal design of eligibility in conditional cash transfer programs.

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

Nonparametric Inference on Efficient Treatment Assignment under Budget Constraints

Debopam Bhattacharya — Dartmouth College
Pascaline Dupas — UCLA & NBER

First version: October 10, 2007. This draft August 8, 2008.

Abstract

We consider the problem of efficiently allocating a binary treatment among a target population based on a set of discrete and continuous observed characteristics. The goal is to maximize the population mean of an eventual outcome when a budget constraint limits what fraction of the population can be treated. Using sample data resulting from randomized treatment allocation, the ATE conditional on covariates (CATE) is nonparametrically estimated in a first step. The optimal treatment threshold and resulting value function, which are non-smooth functionals of the CATE, are estimated based on sample realizations of the *estimated* CATE. We derive large-sample distribution theory for these estimates and for the estimated dual value, i.e. the minimum resources needed to attain a specific average outcome via efficient treatment assignment. These inferential methods are applied to the optimal provision of anti-malaria bed nets, using data from a randomized experiment conducted in western Kenya. We find that a government which can afford to distribute subsidized bed nets 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 can be extended to infer optimal design of eligibility in conditional cash transfer programs.

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 incidence of infectious diseases. These considerations often lead governments and private charities in developing countries 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. Even when budget constraints are not explicitly binding, as in many high-income countries, efficient use of available resources for treating vulnerable sections of society is still an important policy objective.

This paper considers the problem of allocating a fixed amount of resources to a target population with the aim of maximizing the mean outcome across members of the population, and the dual problem of estimating the minimum cost of achieving a given mean outcome in the population by proper targeting of a treatment. We show how, in the presence of observationally heterogeneous treatment effects, experimental data on a representative sample of the population can be used to infer the optimal treatment assignment rule. We apply this methodology to design optimal allocation of an effective malaria control tool – insecticide-treated bed nets – among households in a malaria-endemic region of Kenya. Our treatment of interest is making subsidized bednets available to a section of this population and the outcome of interest is the mean effective usage rate (the share of households using a bed net). We find that, if available resources allow us to treat only 50% of the target population, randomly allocating bed nets is 19% (8 percentage points) less efficient than optimally allocating them based on a set of observed characteristics. Allocating the bed nets according to wealth only is 9% (4 percentage points) less efficient than allocating them based on a set of relevant covariates. Finally, allocating the bed nets 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.

From the perspective of econometric methodology, our analysis is based on functionals of the marginal distribution of the conditional average treatment effect (CATE, henceforth). In particular, we show that when the budget limits the fraction of the treated to $c \in (0, 1)$, the optimal treatment threshold γ and the resulting value function ρ are equivalent respectively to the $(1 - c)$th marginal quantile and Lorenz share for the random variable $\theta(X)$ – where $\theta(x)$ represents the average treatment effect for the subpopulation whose value of the observable characteristic X is x. The parameters ρ and γ, which appear to be new to the treatment effects literature, can be expressed as solutions to semiparametric moment conditions involving the infinite-dimensional initial parameter $\theta(\cdot)$, typically a function of both discrete and continuous covariates. However, the functional form of $\theta(\cdot)$ is not known but estimated in an initial step; the estimated γ and ρ are based on the empirical distribution of the *estimated* $\theta(\cdot)$. The key technical challenge in conducting inference on ρ and γ is that the population moment condition defining γ is a step-function in $\theta(\cdot)$. Consequently, neither the classic semiparametric methods, c.f. Pakes and Pollard (1989), Andrews (1994) or Newey (1994), nor its recent extensions to nonsmooth sample moments (but smooth population moments), c.f. Ai and Chen (2003) or Chen, Linton and van Keilegom (2003) (CLV, henceforth), can be used here directly. 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. Since γ and ρ are functionals of the single-dimensional index $\theta(\cdot)$, the convergence rate of their estimates will not depend on the dimension of the continuous components of X.[1] Moreover, ρ but not necessarily γ can be estimated at the parametric rate under a set of regularity conditions. We will argue below that the value function ρ is often a more interesting policy parameter than the treatment threshold γ. The relatively fast rate for ρ means that we can estimate such policy-relevant scalar parameters well with comparatively small sample sizes without making any ad-hoc parametric assumptions on the data generating process. 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 simply the inverse of $\rho(\cdot)$, which is monotone increasing as a function of c.

On a broader substantive level, this paper suggests and describes how a government may use experimental (pilot) data to infer the eligibility rule that will generate the maximum possible benefit from a program before rolling out the program at a large scale. While generating and collecting experimental data on the effectiveness of an intervention and how it varies across possible beneficiaries has so far been limited to medical interventions, 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. Because the goal of impact evaluations is often to identify simply whether a program works, the parameter of interest for the evaluators is typically the average treatment effect (ATE). For this reason, experiments are typically not designed to precisely estimate interaction coefficients, i.e. how the effect of the program varies by

[1] For deriving these rates, we will require that the estimate of $\theta(.)$ converges in sup norm faster than $n^{1/4}$ and this requires a sample size that is large relative to the dimension of X.

observable covariates. We show here that relatively precise inference on the value of the optimal allocation rule is possible even when the experimental sample size does not permit very precise estimation of the (observable) heterogeneity in the treatment effect (i.e. γ and ρ can be estimated at faster rates than the function $\theta(\cdot)$, itself). Experimental data generated by a pilot program can thus be used not only to estimate whether a program is worth scaling-up, but also to infer how the program should be scaled-up. The methods developed here have wider applicability, beyond subsidy targeting in developing countries, to any situation of constrained treatment assignment. Examples include assigning patients to expensive surgical procedures, deciding eligibility rules for access to credit or allocating the unemployed to job-training programs.

The rest of the paper is organized as follows. Section 2 sets up the problem, introduces the parameters of interest and discusses the relation of the present paper to the relevant literature in econometrics and development economics. Section 3 introduces the estimators and discusses some key issues regarding rate of convergence for the parameters of interest. Section 4 develops the relevant distribution theory. Section 5 presents the benchmark case of parametric inference. Section 6 presents the application to the optimal allocation of bed nets in Kenya. Section 7 discusses an extension to the design of conditional cash transfer programs and section 8 concludes. All proof are collected in the appendix.

2 Formulation of the Problem

2.1 Set-up

Let Y denote an individual level outcome and let S denote a binary treatment whose level can be affected directly by policy. Let X denote observed covariates and U denote unobserved determinants of Y. In the bed-net example, analyzed below in details, 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 equal to 1 if the household owns and uses a bed net. X is the presence of a child under 10 in the household, the household's wealth per capita and ownership of a bank account, while U represents unobserved determinants of take-up. $S = 1$ denotes offering a highly subsidized bed net to the household.

Let $\phi(x, s)$ denote the expected outcome at $S = s$ for individuals with $X = x$: i.e. if an individual with characteristic $X = x$ is randomly selected from the population and assigned a value s of S, then her 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 individuals with $S = s$ in the sample can be used to recover this function.[2] This paper considers the case where X includes both discrete and continuous variables, S is binary and allocation to the treatment was randomized at the individual level.

We will be primarily concerned with a social planner's problem which is as follows. The planner faces a constraint on what fraction of individuals can be administered the treatment ($S = 1$). Suppose this fraction is c and let $\mathcal{X}$ denote the support of X. We define the planner's problem as the choice of a set $A \subset \mathcal{X}$ such that if an individual'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.[3] 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

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

It is obvious that 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 $\phi(x,1) - \phi(x,0)$.

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

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

subject to

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

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

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

[2] Otherwise, $\phi(x, \cdot)$ has to be identified by either using IV based methods or by assuming unconfounded (conditional on covariates) treatment assignments. Bhattacharya, Chandra and Chen (2007) investigate this case in the context of assigning a continuous treatment.

[3] 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 $X = x$.

Proof. Appendix ■

Note that the problem is interesting only if $c < \Pr(\theta(X) > 0)$; otherwise, the optimal assignment rule would be to give treatment to everybody whose average treatment effect is positive.

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

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

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

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

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

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

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

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, e.g. 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" individual 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 for several reasons. First, it is useful for deciding on the budget outlay necessary for achieving a target mean level of outcome (more on this below). Second, any choice of conditioning covariates is likely to be controversial in real life and some could be politically or legally infeasible. Some relevant covariates are also costly and/or difficult to measure. For example, in poorer areas in developing countries where the majority of households do not file tax returns, measuring income or wealth levels typically requires labor-intensive household surveys. Under-reporting of income and assets is a common problem, especially if the population surveyed is aware of the existence of an eligibility threshold (Martinelli and Parker, 2007). The unrestricted value function therefore represents a "first-best" scenario against which alternative allocations which are feasible and/or based on easily measurable covariates can be compared. The (unrestricted) γ may be less relevant relevant from a policy perspective if such an unrestricted allocation will never be feasible in real life.

Equivalent expenditure: The dual formulation of the optimal allocation problem is as follows. Suppose 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 dual problem can be represented as

$$\min_{A\subset\mathcal{X}}\int_{x\in\mathcal{X}}1\left\{x\in A\right\}dF\left(x\right) \tag{3}$$

subject to

$$\int_{x\in\mathcal{X}}\left[\phi\left(x,1\right)1\left(x\in A\right)+\phi\left(x,0\right)1\left(x\notin A\right)\right]dF\left(x\right)=b. \tag{4}$$

One can almost repeat the proof of proposition 1 to show that the optimal A will again be of the form $A^*=\{x:1\left\{\theta\left(x\right)>\gamma\left(b\right)\right\}\}$ where $\gamma\left(b\right)$ is such that A^* satisfies (4). Note that by duality, the minimum value of (3) is simply $\rho^{-1}\left(b\right)$ where $\rho\left(\cdot\right)$ is defined in (2) and the inverse is well-defined because $\rho\left(\cdot\right)$ 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 optimal 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 or because observing it is costly. Define

$$\xi\left(x_1,S\right)=E_{X_2|X_1=x_1}\left[\phi\left(x_1,x_2,S\right)\right]. \tag{5}$$

Then the optimization problem becomes

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

Call the unrestricted maximum $\rho_{un}\left(c\right)$ and the restricted one, which conditions only on X_1, $\rho_{res}\left(c\right)$. The difference $\rho_{un}\left(c\right)-\rho_{res}\left(c\right)$ measures the efficiency cost of these restrictions on implementation. When gathering information on X_1 (e.g. income) is expensive, one can compare the above efficiency cost against the cost of gathering information on X_1 to decide on whether the extra survey cost is worthwhile to undertake.

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 Related Literature and Contributions

This paper contributes to the new and growing literature on treatment choice (c.f. Dahejia (2001), Manski (2004), Hirano and Porter (2006) in econometrics and the related problem of optimal allocation of inputs in production processes (c.f. Graham, Imbens and Ridder (2005, 2006) and Bhattacharya (2006)). The present paper differs from the above works substantively as it studies optimal allocation under budget constraints– a problem that leads to interesting economic parameters that are apparently new to the econometrics literature. Analytically, the paper differs from Graham, Imbens and Ridder (2006) and Bhattacharya (2006) in that it analyzes optimal allocation rules based on both discrete and continuous conditioners. This makes the problem nonparametric in a nontrivial way. Furthermore, 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 appear to be 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 method is based on covariate-based treatment assignment and uses data from a pilot experiment which is run prior to the main experiment. The goal of HHK 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 a working paper, 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. Analytically, that problem differs significantly from the present paper because distribution theories are very different under endogeneity and more structure is needed on the underlying production function to guarantee unique solutions to a planner's optimization problem.

More broadly, the present paper proposes a new use of experimental data on social programs. So far, experimental data have typically been collected and used to measure the impact of a program and determine whether the program is worth its cost or not. A few recent studies have also used experimental data to estimate the parameters of dynamic structural models 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). On the other hand, 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 new methodology through which experimental data can be used directly to infer optimal targeting of programs. As randomized trials of social programs (e.g. PROGRESA in Mexico) become more common in both developed and developing countries, the methodology we propose will help governments and aid-agencies roll out positive-impact programs via efficient allocation rules. The present paper also discusses, albeit briefly, how analogous methods can be used to design optimal eligibility in conditional cash-transfer programs, which have gained popularity in a large number of central and south American countries.

3 Estimation

Now we define our estimates formally. 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 X and density $f(\cdot)$. First define the quantities

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

Then $\hat{\theta}(X_i)$ is defined as

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

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

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

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

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

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. For each $t \in [-A, A]$, choose a symmetric (about zero) kernel $L(\cdot)$ with bounded support $[-1,1]$, the corresponding C.D.F. kernel $\bar{L}(t) = \int_{-1}^{t} L(s)\, ds$ and a sequence of bandwidth h_n converging (slowly) to zero as $n \to \infty$. Now define $\hat{\gamma}$, and $\hat{\rho}$ by

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\left\{\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right) - \{1-c\}\right\} &= 0, \\ \frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(X_i,0)\,\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right) + \hat{\phi}(X_i,1)\left\{1-\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)\right\} - \hat{\rho} &= 0. \end{aligned} \tag{6}$$

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 (6) is similar in spirit to Horowitz's (1992) analysis of 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.

4 Large sample theory

The discrete regressors will not play any substantive roles in our analysis; so we will drop them in our discussion from now on and put them back into our final results at the end. Every condition we use will have to hold conditional on each specific value assumed by the discrete regressors. 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_0, respectively; 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 would be used in the expressions for upper bounds for various quantities which appear in the proof.

Assumptions

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

A0(ii) S is randomly allocated 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 individual.

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 $-1 < -A \leq \theta(x) \leq A < 1$ 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(i) 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 is somewhat restrictive but is routinely assumed (c.f. Hirano, Imbens and Ridder (2003), assumption 2). If this fails, we can simply redefine the problem such that we are designing allocations based only on those values of X where this condition holds. Assumption A2 is standard and furthermore, like in the case of assumption A1, we can redefine the problem for those values of X where this condition holds. Assumption A3 (i) is standard. Assumption A3 (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).

4.1 Consistency of $\hat{\gamma}$

The following lemma will be useful in several proofs below. We will introduce several high level assumptions before invoking the lemma.

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

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.

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

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

Proof. Appendix ■

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_0

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

$$\hat{\gamma} - \gamma_0 = o_p(1)$$

Proof. Appendix ■

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

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

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

Proof. Appendix ■

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

$$\begin{aligned} & (\hat{\gamma} - \gamma_0) \\ = & \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_{\theta}(\gamma_0) - \frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{u - \theta(X_i)}{h_n}\right)\right\} \\ & + \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\bar{L}\left(\frac{u - \theta(X_i)}{h_n}\right) - \bar{L}\left(\frac{u - \hat{\theta}(X_i)}{h_n}\right)\right)\right]. \end{aligned}$$

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 $\sqrt{nh_n}(\hat{\gamma} - \gamma_0)$ will converge to normal distribution.

The following additional assumptions will be used in the proof.

B7. For some $r \geq 2$, the density of $\theta(X)$ is $(r-1)$ times continuously differentiable, the derivative is bounded and Lipschitz in a neighborhood of γ_0 and $nh_n^{2r+1} \to \lambda < \infty$. Denote the above derivative at γ_0 by $f_\theta^{(r-1)}(\gamma_0)$.

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

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

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

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

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*

$$\sqrt{nh_n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N\left(\beta, \frac{\tau^2(\gamma_0) + \omega^2(\gamma_0)}{f_\theta(\gamma_0)} \int_{-1}^{1} L^2(u)\,du\right),$$

where

$$\begin{aligned}
\tau^2(\gamma_0) &= E\left\{\left\{\frac{\delta(X)Y(1-S) - \nu(X)(1-S)}{\delta^2(X)} f(X)\right\}^2 |\theta(X) = \gamma_0\right\} \\
\omega^2(\gamma_0) &= E\left\{\left\{\frac{\pi(X)YS - \mu(X)S}{\pi^2(X)} f(X)\right\}^2 |\theta(X) = \gamma_0\right\} \\
\beta &= (-1)^{r+1}\frac{\sqrt{\lambda}}{r!} \times f_\theta^{(r-1)}(\gamma_0) \int_{-1}^{1} u^r L(u)\,du.
\end{aligned}$$

Proof. Appendix ■

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

$$\begin{aligned}
\tau^2(\gamma_0) &= E\left\{\left\{\frac{\delta(X)Y(1-S) - \nu(X)(1-S)}{\delta^2(X)} f(X^c, X^d)\right\}^2 |\theta(X) = \gamma_0\right\} \\
\omega^2(\gamma_0) &= E\left\{\left\{\frac{\pi(X)YS - \mu(X)S}{\pi^2(X)} f(X^c, X^d)\right\}^2 |\theta(X) = \gamma_0\right\},
\end{aligned}$$

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

4.3 Consistency for $\hat{\rho}$

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

$$\hat{\zeta} - \zeta_0 = o_p(1).$$

Proof. Appendix ■

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

Recall that $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} \hat{\phi}(X_i, 1) - \hat{\zeta}$. We will analyze the first term using the following lemma and then $\hat{\zeta}$, using theorem 2.

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

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

Proof. Appendix ■

The final step is to derive the large sample distribution of $\hat{\zeta}$, for which the following expansion will be used.

$$\begin{aligned}
& \hat{\zeta} - \zeta_0 \\
= & \underbrace{\frac{1}{n} \sum_{i=1}^{n} \theta(X_i) \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) - \zeta_0}_{T_{1n}} \\
& \underbrace{+\frac{1}{n} \sum_{i=1}^{n} \left\{\hat{\theta}(X_i) - \theta(X_i)\right\} \left\{\bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) - \frac{1}{h_n} \theta(X_i) L\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\}}_{T_{2n}} \\
& \underbrace{+(\hat{\gamma} - \gamma_0) \frac{1}{nh_n} \sum_{i=1}^{n} \theta(X_i) L\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)}_{T_{3n}} + R_n
\end{aligned}$$

The proof will work by showing 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,*

$$\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\psi_{1i}+\psi_{2i}+\psi_{3i}-\psi_{4i}\right\}+o_p(1),$$

where

$$\begin{aligned}
\psi_{1i} &= \gamma_0\left\{F_\theta(\gamma_0)-1\left(\theta(X_j)\le\gamma_0\right)\right\}\\
\psi_{2i} &= \theta(X_i)\times 1\left\{\theta(X_i)\le\gamma_0\right\}-\zeta_0\\
\psi_{3i} &= 1\left(\theta(X_i)\le\gamma_0\right)\times\frac{\pi(X_i)Y_iS_i-\mu(X_i)S_i}{\pi^2(X_i)}\times f_X(X_i)\\
\psi_{4i} &= 1\left(\theta(X_i)\le\gamma_0\right)\times\frac{\delta(X_i)Y_i(1-S_i)-\nu(X_i)(1-S_i)}{\delta^2(X_i)}\times f_X(X_i).
\end{aligned}$$

It follows by an ordinary CLT (under standard second moment restrictions) that $\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\}$ will be mean-zero normal.

Proof. Appendix ■

To incorporate the discrete regressors back into the analysis, we simply replace the terms $f_X(X_i)$ 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 $f_{X^c|X^d=b}(a|b)$ 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.

Corollary 5 *From the previous lemma and the theorem, it follows that*

$$\begin{aligned}
&\sqrt{n}\left(\hat{\rho}-\rho_0\right)\\
=\;& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\psi_{1i}+\psi_{2i}+\psi_{3i}-\psi_{4i}\right\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{E(S|X_i)\times Y_iS_i-E(SY|X_i)\times S_i}{\left\{E(S|X_i)\right\}^2}+o_p(1).
\end{aligned}$$

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

The variances can be estimated by the average of squares of the terms in the linear expansions above and that this estimate will be consistent follows from the standard WLLN.

4.5 Distribution theory for dual

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

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

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

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

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

Remark 2 *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 $\theta(\cdot)$ and also that $\theta(\cdot)$ is unknown. If $\theta(\cdot)$ were known, then realizations of $\theta(X)$ would be observed and so its estimated quantile would be $\sqrt{n}$-normal. Conversely, if the moment condition were smooth and $\theta(\cdot)$ 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 $\gamma=G(\theta)^{-1}(1-c)$ and $\rho=\int_{1-c}^{1}G(\theta)^{-1}(u)\,du$ where $G(\theta)$ represents the c.d.f. of $\theta(X)$. 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 (c.f., proof of lemma 3 below). 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_0 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 $(\hat{\gamma}-\gamma_0)$ is given by

$$\beta = (-1)^{r+1}\frac{h_n^r}{r!}\times f_\theta^{(r-1)}(\gamma_0)\int_{-1}^{1}u^r L(u)\,du + o(h_n^r).$$

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

$$h_n^{2r}\times\underbrace{\left[\frac{f_\theta^{(r-1)}(\gamma_0)}{r!}\int_{-1}^{1}u^r L(u)\,du\right]^2}_{C}+\frac{1}{nh_n}\underbrace{\left[\frac{\tau^2(\gamma_0)+\omega^2(\gamma_0)}{f_\theta(\gamma_0)}\int_{-1}^{1}L^2(u)\,du\right]}_{B},$$

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.

5 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 parametric analysis has the obvious limitation that it is susceptible to mis-specification of the functional form and thus may lead to a suboptimal value function. In our application we show the results for both parametric and nonparametric specifications and estimate the efficiency loss arising from the potential mis-specification of the parametric model.

As an illustration, consider the linear parametric form, i.e.

$$y=\beta_0+x'\beta_1+(\delta_0+x'\delta_1)S-u$$

implying a conditional ATE given by

$$\begin{aligned}\theta(x) &= \beta_0+x'\beta_1+(\delta_0+x'\delta_1)-(\beta_0+x'\beta_1)\\ &= \delta_0+x'\delta_1.\end{aligned}$$

The optimal treatment rule is now described by the threshold γ defined by

$$c = \Pr\left[\theta\left(x\right) \geq \gamma\right]$$

yielding a value function

$$\rho\left(c\right) = E\left\{\begin{array}{c}\left[\beta_0 + x'\beta_1 + \left(\delta_0 + x'\delta_1\right)\right] \times 1\left(\theta\left(x\right) \geq \gamma\right) \\ +\left(\beta_0 + x'\beta_1\right) \times 1\left(\theta\left(x\right) < \gamma\right)\end{array}\right\}.$$

More generally, suppose $\theta\left(x\right)$ is parametrically specified as $G\left(x, \beta\right)$, where $G\left(\cdot\right)$ is known; typically β can be estimated at parametric rates using, say, GMM. For estimation of γ and ρ, we will still use smoothing with bandwidth sequence h_n and the c.d.f. kernel $\bar{L}\left(\cdot\right)$ to handle the nonsmoothness. For some specific functional forms of $G\left(\cdot, \cdot\right)$, e.g., a linear one, the function $h\left(\beta\right) = \int 1\left\{G\left(x, \beta\right) \leq \gamma\right\} dF\left(x\right)$ 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 distribution theory for $\hat{\rho}\left(c\right)$ and $\hat{\gamma}\left(c\right)$ corresponding to a parametric specification of $\theta\left(\cdot\right)$ is a simpler version of the nonparametric case. In particular, we will get that both γ and ρ can be estimated at the $\sqrt{n}$-rate. The details are as follows. Recall the asymptotic expansion for $\hat{\gamma}$:

$$\begin{aligned}
& \sqrt{n}\left(\hat{\gamma} - \gamma_0\right) \\
= \quad & \left\{\hat{f}_{\hat{\theta}}\left(\tilde{\gamma}\right)\right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{F_\theta\left(\gamma_0\right) - \bar{L}\left(\frac{\gamma_0 - G\left(X_i, \beta_0\right)}{h_n}\right)\right\} \\
& + \left\{\hat{f}_{\hat{\theta}}\left(\tilde{\gamma}\right)\right\}^{-1} \left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\bar{L}\left(\frac{\gamma_0 - G\left(X_i, \beta_0\right)}{h_n}\right) - \bar{L}\left(\frac{\gamma_0 - G\left(X_i, \hat{\beta}\right)}{h_n}\right)\right]\right\}.
\end{aligned}$$

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

$$\lim_{n \to \infty} \sqrt{n} h_n^r \times \left[\frac{(-1)^{r+1} f_\theta^{(r-1)}\left(\gamma_0\right) \times \int_{-1}^{1} u^r L\left(u\right) du}{r!}\right] + o\left(\sqrt{n} h_n^r\right)$$

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

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

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma_0-G\left(X_i, \beta\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma_0-G\left(X_i, \hat{\beta}\right)}{h_n}\right)\right] \\
= \; & \sqrt{n}\left(\hat{\beta}-\beta\right)^{\prime} \frac{1}{n h_n} \sum_{i=1}^{n} \nabla G\left(X_i, \beta\right) L\left(\frac{\gamma_0-G\left(X_i, \beta_0\right)}{h_n}\right) \\
& +T_n,
\end{aligned}
$$

where

$$
\left|T_n\right| \leq M \frac{n\left\|\hat{\beta}-\beta\right\|^2}{2 \sqrt{n} h_n^2} \frac{1}{n} \sum_{i=1}^{n} M_1\left(X_i\right) L^{\prime}\left(\frac{G\left(X_i, \tilde{\beta}\right)-\gamma_0}{h_n}\right),
$$

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 $n h_n^4 \rightarrow \infty$. Then we have that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma_0-G\left(X_i, \beta_0\right)}{h_n}\right)-\bar{L}\left(\frac{\gamma_0-G\left(X_i, \hat{\beta}\right)}{h_n}\right)\right] \\
= \; & \left[\sqrt{n}\left(\hat{\beta}-\beta\right)^{\prime} \nabla G\left(X_i, \beta_0, \delta_0\right)\right] \times f_{G(X, \beta)}\left(\gamma_0\right)+o_p(1) .
\end{aligned}
$$

This implies that $\sqrt{n}\left(\hat{\gamma}-\gamma_0\right)$ will converge to a zero mean normal if $n h_n^{2r} \rightarrow 0$ and $n h_n^4 \rightarrow \infty$ and when the density of $G\left(X, \beta_0\right)$ has uniformly bounded derivatives up to order $(r-1)$ where $r \geq 3$. The result for $\hat{\rho}$ will follow.

6 Application to bednet provision

6.1 Background

We now apply this inference method to the optimal allocation of heavily subsidized 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.[4]

[4] 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.

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.[5] 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 (Dupas, 2008; Cohen and Dupas, 2007). 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 person 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 observable 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 observable 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 efficiency gains.

[5] 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.

6.2 Design

For this application we use data form a randomized experiment conducted with rural households in Western Kenya in 2007 (Dupas, 2008). The price at which household could purchase an ITN varied from \$0 (a free ITN) to \$4, and households were randomly assigned to a 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 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 coup constitutes the control. The treatment is thus "having access to a low-price ITN" (note that the take-up in the low price group was not 100%, since some of the "treated" had to pay a small fee to access the net. In such case, the expected cost of giving eligibility to a group of size N is lower than N times the unit cost of the treatment. For treatments that do not require cost-sharing, however, the take-up is likely to be close to 100%).

Table 2 presents evidence of heterogeneity in the treatment effect. The table shows the results of an OLS regression of ITN on 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 combined was then divided by household size to obtain the "wealth per capita" indicator. The treatment was randomized at the individual-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 rejects the null hypothesis. This suggests that a covariate-based allocation will lead to important efficiency gains.

6.3 Analysis

6.3.1 Non-Parametric Analysis: Choice of Kernels and Bandwidths

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

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

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: $c = 50\%$ (Panel A) and $c = 25\%$ (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.

6.3.2 Conditional ATE

The nonparametric estimate of the CATE $\hat{\theta}(x) = \hat{\phi}(x_i, 1) - \hat{\phi}(x_i, 0)$ was computed corresponding to two bandwidths $\sigma_n = 0.3$ and $\sigma_n = 0.4$. The parametric estimate of $\theta(X)$ was computed as $\hat{\theta}(x) = \left(\hat{\delta}_0 + x'\hat{\delta}_1\right)$, where $\hat{\delta}_0$ and $\hat{\delta}_1$ are OLS estimates in the regression (presented in Table 2):

$$y_i = \beta_0 + x_i'\beta_1 + \delta_0 Treatment_i + x_i'\delta_1 Treatment_i + \varepsilon_i.$$

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 (ii) above.

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.

6.3.3 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.

6.3.4 Efficiency 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 efficiency 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 nonparametric 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 3.

We compute the standard errors of the efficiency losses generated by the three suboptimal allocations over a range of budget levels in Table 3. 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 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 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 efficiency 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 3 suggest that the efficiency 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 efficiency gain of a means-tested allocation compared to other allocation schemes might not be worth its cost.

6.3.5 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.

7 Extension: conditional cash-transfer programs

In some government programs, transfers can be, and often 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. The larger ones among these include Oportunidades, previously known as PROGRESA, in Mexico and the Bolsa Escola in Brazil. 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. However, methods analogous to those developed above can be used to devise optimal design of such behavior-contingent transfers, as follows.

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 as solutions to

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\{c-\hat{\phi}(X_i,1)\times\bar{L}\left(\frac{\hat{\alpha}-\hat{\phi}(X_i,0)}{h_n}\right)\right\}&=0,\\ c+\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\phi}(X_i,0)\times\left(1-\bar{L}\left(\frac{\hat{\alpha}-\hat{\phi}(X_i,0)}{h_n}\right)\right)\right\}-\hat{\mu}&=0.\end{aligned}$$

In future work, we intend to explore large sample theory for these estimates and apply them to study optimal eligibility rules, using data from the Oportunidades program in Mexico.

8 Conclusion

This paper considered a social planner's problem of allocating a binary treatment among a target population based on observed characteristics in the presence of budget constraints.

Outcome data from a randomized allocation of treatment to a representative sample are used to estimate the average treatment effect $\theta(\cdot)$ conditional on covariates X and the marginal distribution of $\theta(X)$ in the population. This distribution is used to design an optimal targeting rule which maximizes a mean outcome. In this rule, the optimal treatment threshold γ and the corresponding value function ρ equal respectively a quantile and the corresponding Lorenz share in the population distribution of $\theta(X)$. We show that γ can be consistently estimated and ρ can be estimated both consistently and at the parametric rate, even when $\theta(\cdot)$ is nonparametrically estimated. This result holds even though the population moment conditions defining the finite-dimensional parameters (γ, ρ) are not differentiable in $\theta(\cdot)$, so that existing methods for semiparametric moment condition models cannot be applied here.

From a broader substantive standpoint, this paper contributes to a nascent literature on the possible uses that can be made of experimental data in designing optimal policies. We suggest how governments may use experimental (pilot) data to infer the participation eligibility rule that will generate the maximum possible benefit from a program before rolling out the program on a larger scale. Applying our method to experimental data on the provision of anti-malaria bed nets in western Kenya, we find that a government which can afford to distribute subsidized bed nets 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. Future work will extend these methods to the design of optimal eligibility in conditional cash-transfer programs, which have gained popularity in a large number of central and south American countries.

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 3
Allocation Efficiency

(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 4
Dual Problem: Cost of Reaching a Target Outcome

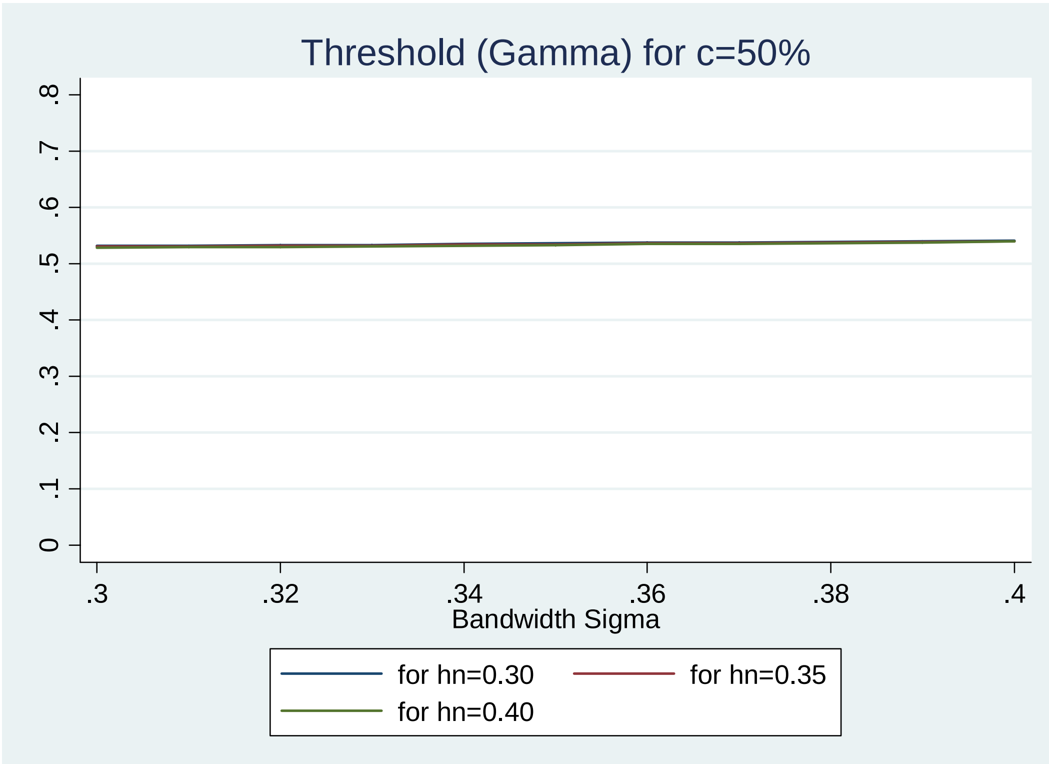
(1)	(2)	(3)		(4)		(5)	
Objective Function: Target ρ(*c*)	Unrestricted case: Share of population that needs to be treated to reach this target	Restricted Cases: Additional share that needs to be treated to achieve the target when conditioning on:					
		All covariates except wealth		*Wealth only*		*Nothing (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. Compared to the unrestricted case, an additional 0.059 of the population needs to be treated if the conditioning is based on all covariates except wealth (column 3). Compared to the unrestricted case, 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 (compared to the unrestricted case) 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

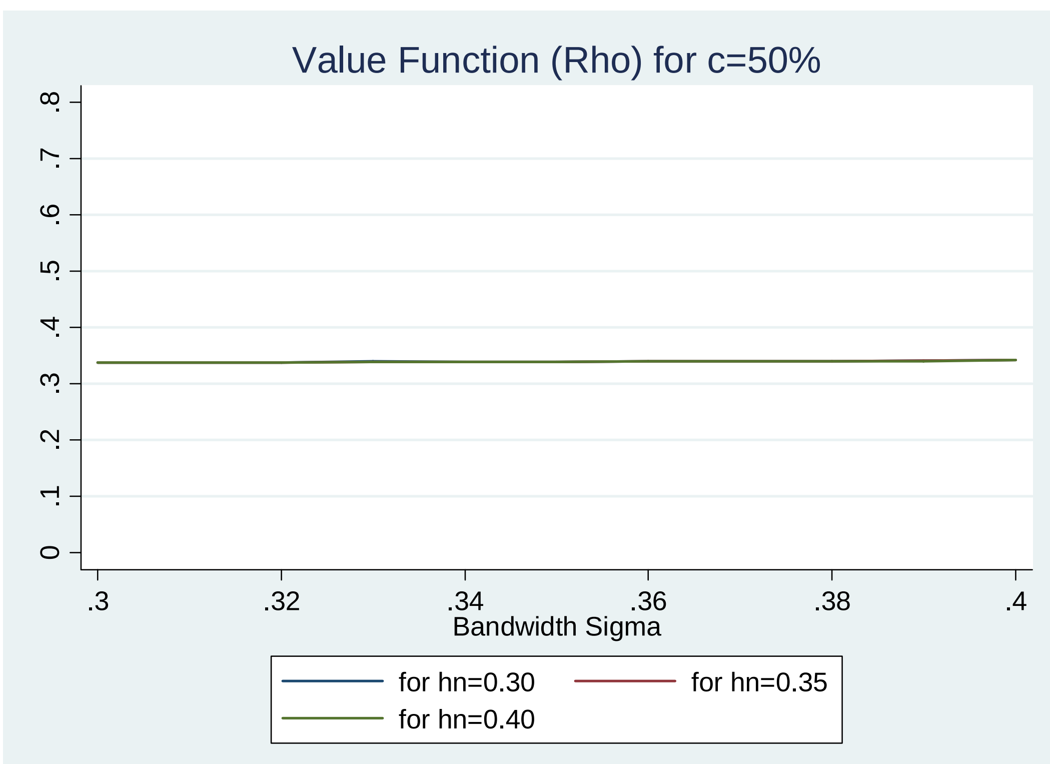
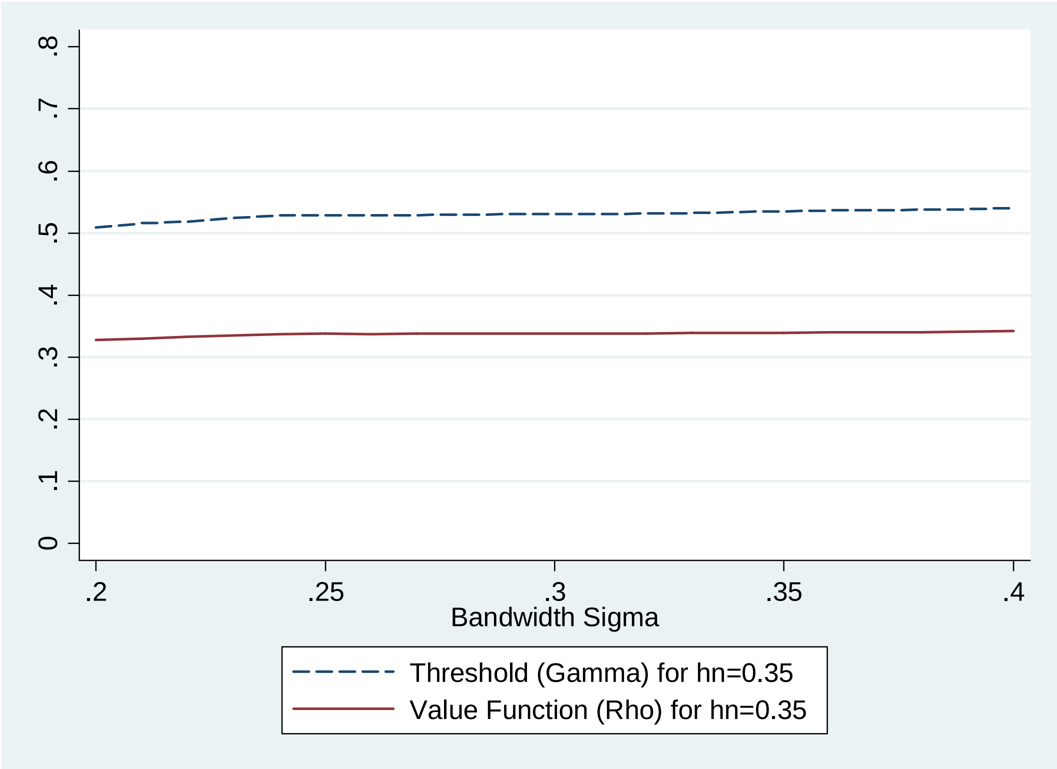


Figure 2

Panel A. γ(c) and ρ(c) when c= 0.50



Panel B. γ(c) and ρ(c) when c= 0.25

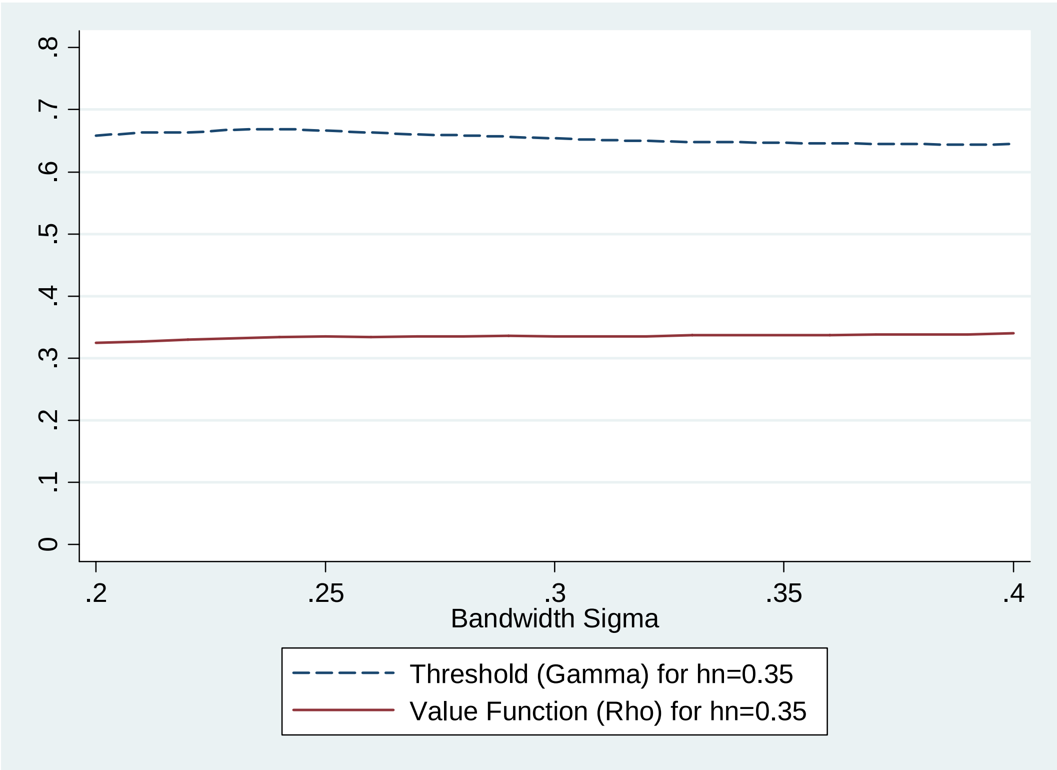


Figure 3

Leave-one-out Cross-Validation Criterion

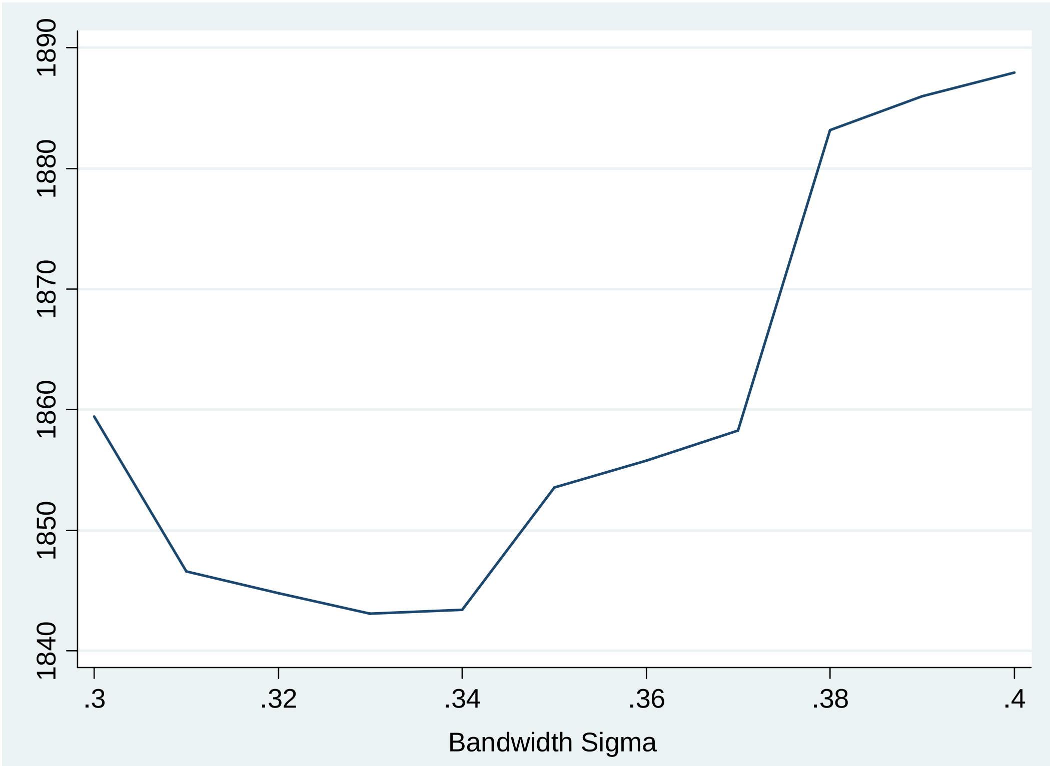


Figure 4

Kernel Density of Estimates of Conditional ATE θ(X)

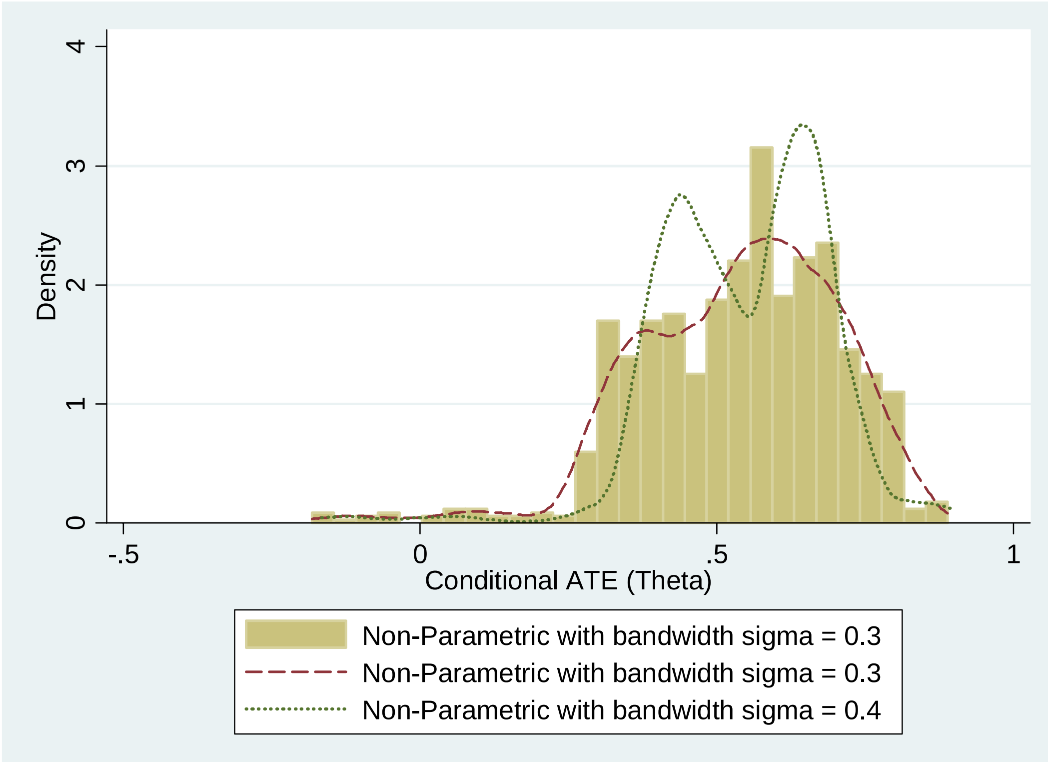


Figure 5

Cumulative Distribution Function of Conditional ATE θ(X)

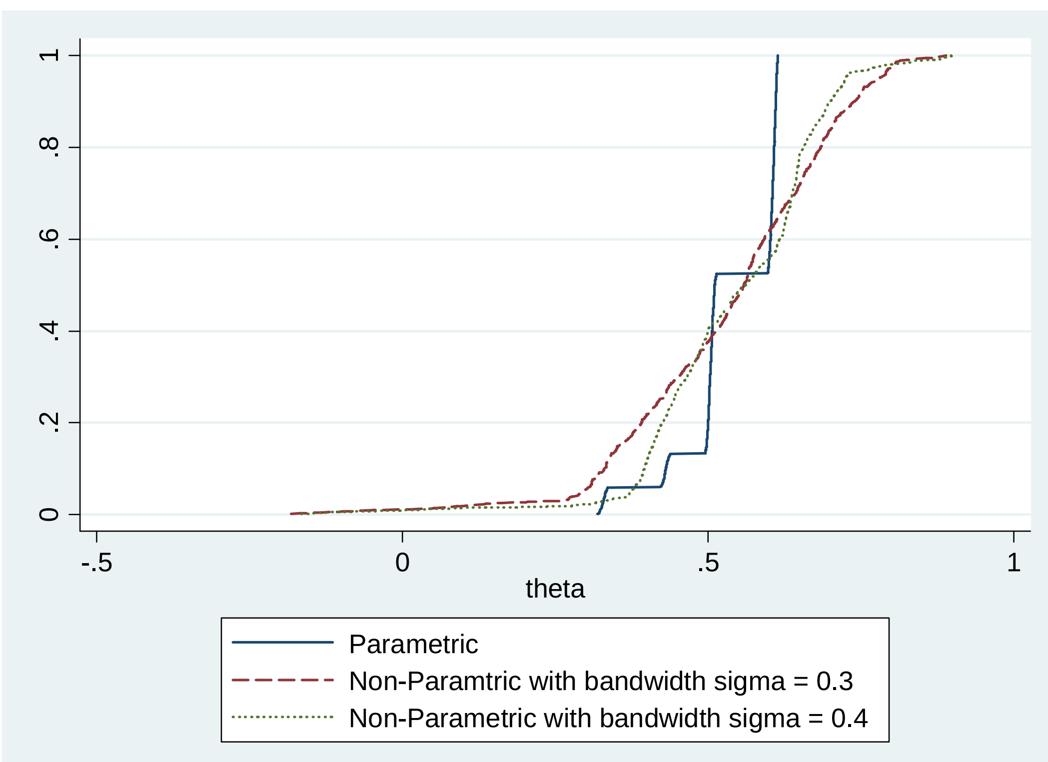
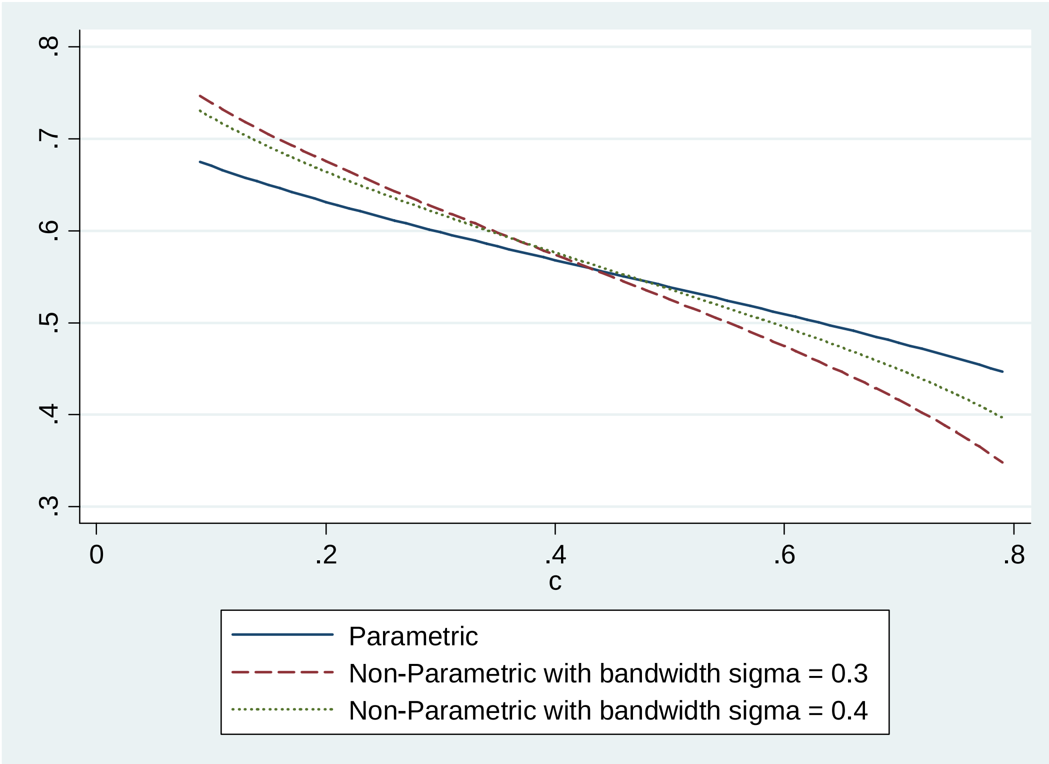


Figure 6

Panel A. Threshold γ(c), Unrestricted Case



Panel B. Value Function ρ(c), Unrestricted Case

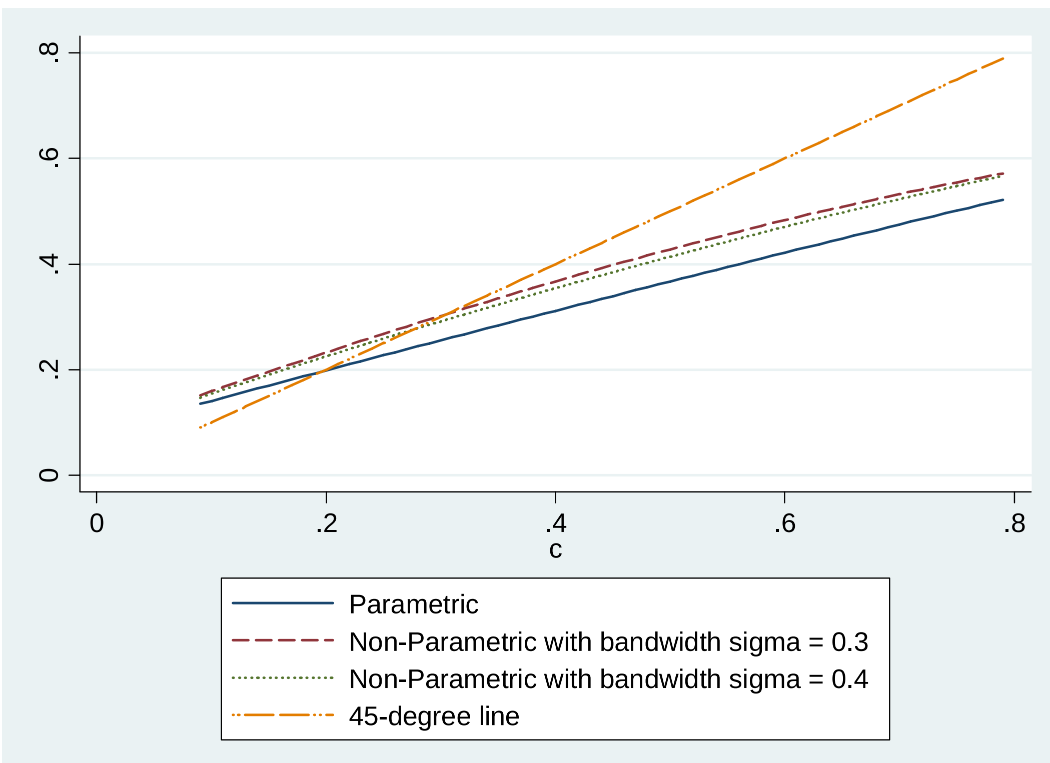
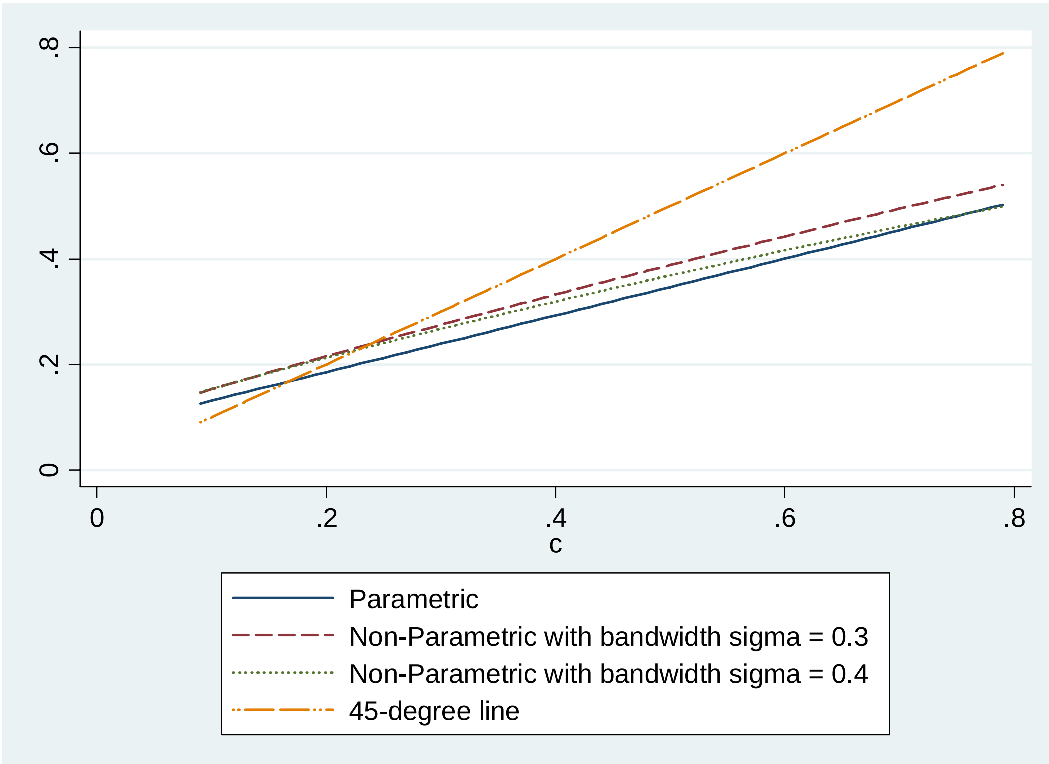


Figure 7

Panel A. Value Function ρ(c), Restricted Case: Conditioning on Wealth Only



Panel B. Value Function ρ(c), Random Allocation

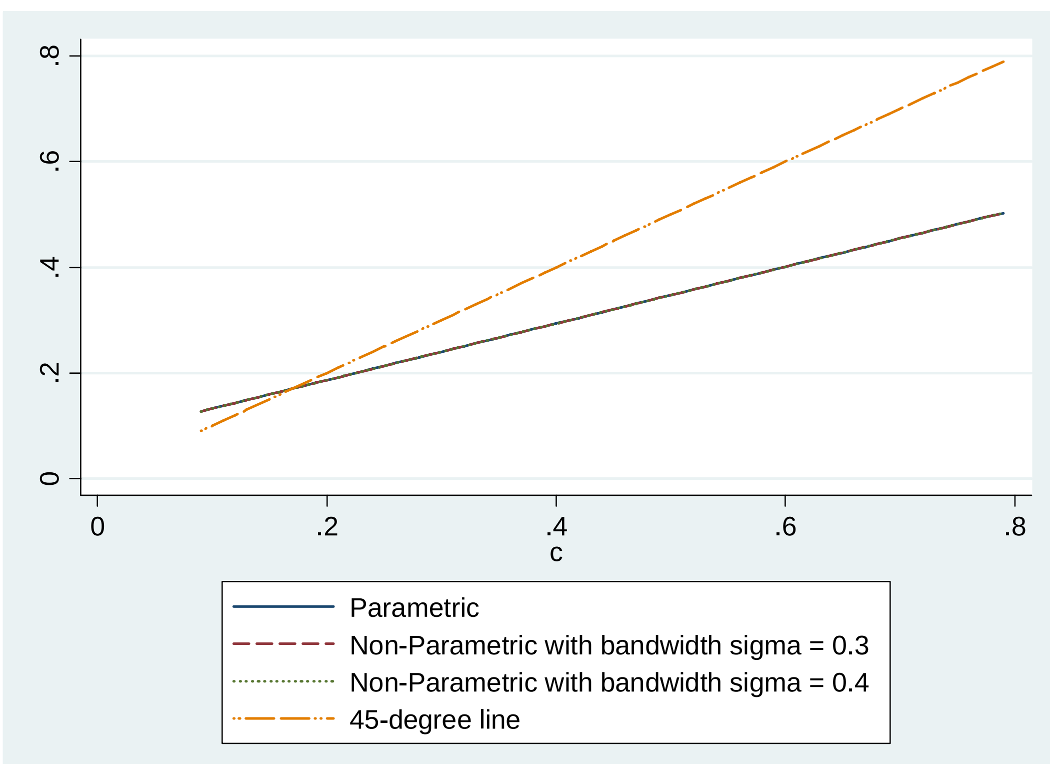
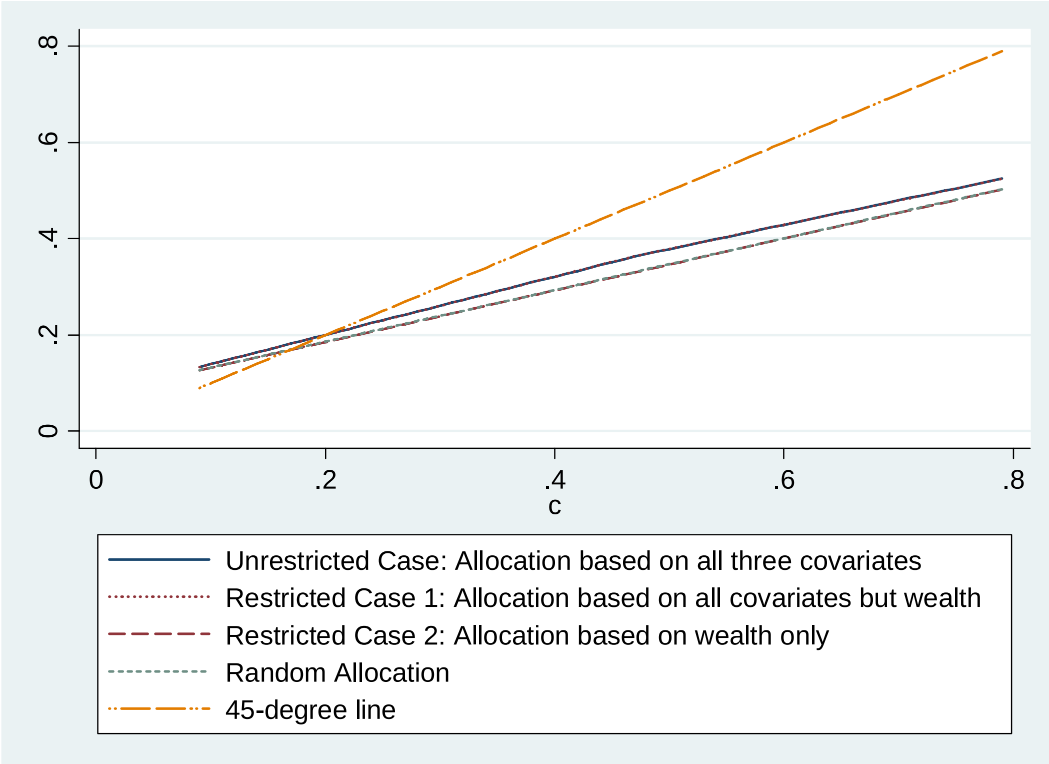
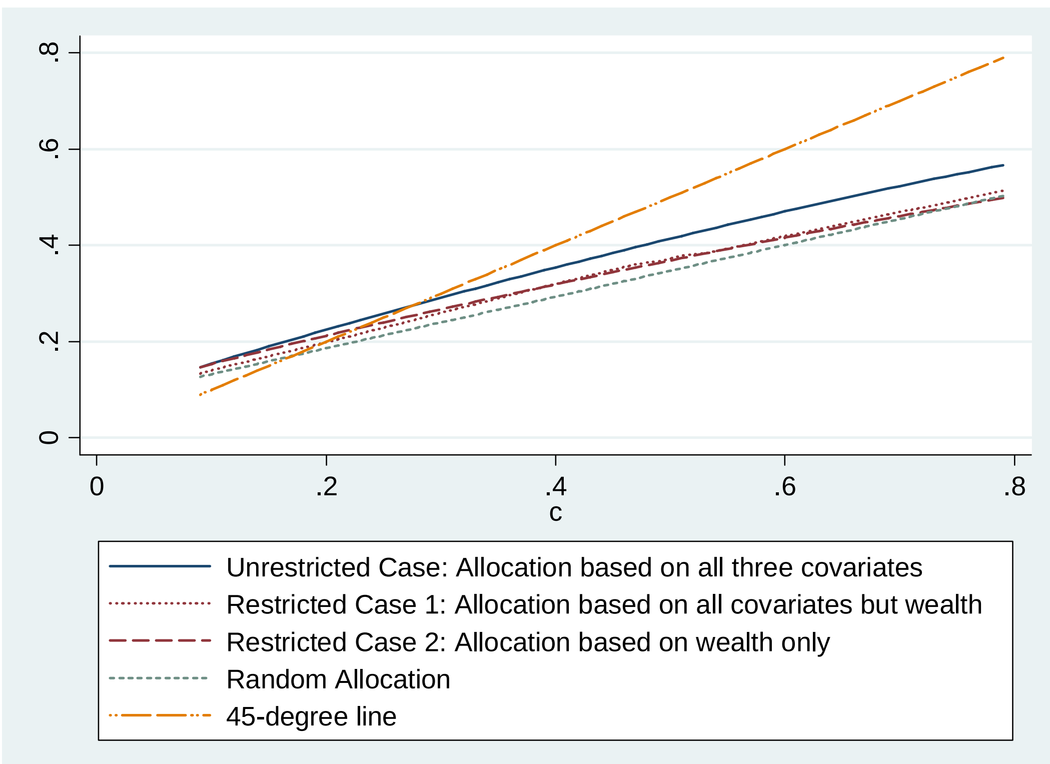


Figure 8

Panel A. Value Function ρ(c), Parametric Model



Panel B. Value Function ρ(c), Non-Parametric Analysis (Bandwidth σ=0.4)



References

[1] Ai, C and Chen, X. (2003): Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. Econometrica, Vol 71, No 6, pp. 1795-1843.

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

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

[4] Bhattacharya, D. (2007): Inference on inequality from household survey data, Journal of Econometrics, Volume 137, Issue 2, April 2007, Pages 674-707.

[5] Bhattacharya, D. (2006): Inferring Optimal Allocation of Peers using Experimental Data, forthcoming in the Journal of the American Statistical Association.

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

[7] (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.

[8] Cohen, Jessica and Pascaline Dupas (2007): Free distribution or cost-sharing? Evidence from a randomized malaria experiment in Kenya. BREAD Working Paper No. 169.

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

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

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

[12] Dupas, Pascaline (2008): Pricing strategies in scaling-up access to long-lasting insecticide-treated nets: Evidence from a randomized voucher experiment in Western Kenya. Mimeo, UCLA.

[13] Easterly, William (2001). The Elusive Quest for Growth. MIT Press.

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

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

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

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

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

[19] Härdle, Wolfgang; Hall, Peter; Marron, J. S. (1988): How far are automatically chosen regression smoothing parameters from their optimum?, J. Amer. Statist. Assoc. 83, no. 401, 86–101.

[20] Hirano, K. and J. Porter (2005): Asymptotics for Statistical Treatment Rules. Manuscript.

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

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

[23] 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

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

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

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

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

[28] (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.

[29] Pakes, A. and Pollard, D. (1989): Simulation and the Asymptotics of Optimization Estimators, Econometrica, vol. 57(5), pages 1027-57.

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

[31] Saaty, T. (1981): Modern nonlinear equations, Dover Publications, New York.

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

[33] 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, National Bureau of Economic Research, Inc.

[34] 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.

[35] Webster, Jane, Jo Lines, and Lucy Smith. 2007. "Protecting all Pregnant Women and Children Under Five Years Living in Malaria Endemic Areas in Africa with Insecticide Treated Mosquito Nets." World Health Organization Working Paper. http://www.who.int/malaria/docs/VulnerableGroupsWP.pdf

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

9 Appendix (Proofs)

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

$$\int_{x\in\mathcal{X}} [\phi(x,1)-\phi(x,0)]\,1(x\in A)\,dF(x)+\int_{x\in\mathcal{X}}\phi(x,0)\,dF(x)\,, \tag{7}$$

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

$$\begin{aligned}
&\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)>\gamma\}\,dF(x)+\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)\le\gamma\}\,dF(x)\\
=&\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)>\gamma\}\,dF(x)+\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)\le\gamma\}\,dF(x)\\
&-\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\notin A)\,1\{\theta(x)>\gamma\}\,dF(x)-\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)>\gamma\}\,dF(x)\\
&+\int_{x\in\mathcal{X}}[\theta(x)]\,1\{\theta(x)>\gamma\}\,dF(x).\\
=&\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\in A)\,1\{\theta(x)\le\gamma\}\,dF(x)-\int_{x\in\mathcal{X}}[\theta(x)]\,1(x\notin A)\,1\{\theta(x)>\gamma\}\,dF(x)\\
&+\int_{x\in\mathcal{X}}[\theta(x)]\,1\{\theta(x)>\gamma\}\,dF(x).
\end{aligned} \tag{8}$$

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

$$\gamma\int_{x\in\mathcal{X}}1(x\in A)\,1\{\theta(x)\le\gamma\}\,dF(x)\,, \tag{9}$$

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

$$\gamma\int_{x\in\mathcal{X}}1(x\notin A)\,1\{\theta(x)>\gamma\}\,dF(x)\,. \tag{10}$$

Now from the budget constraint, we have that

$$\begin{aligned}
c&=\int_{x\in\mathcal{X}}1(x\in A)\,dF(x)\\
&=\int_{x\in\mathcal{X}}\begin{bmatrix}1(x\in A)\,1\{\theta(x)\le\gamma\}\\+1(x\in A)\,1\{\theta(x)>\gamma\}\end{bmatrix}dF(x)
\end{aligned}$$

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

$$\gamma \int_{x\in\mathcal{X}} 1\left(x\in A\right) 1\left\{\theta\left(x\right)\leq\gamma\right\} dF\left(x\right) = \gamma \int_{x\in\mathcal{X}} 1\left(x\notin A\right) 1\left\{\theta\left(x\right)>\gamma\right\} dF\left(x\right). \tag{11}$$

It follows from (8), (9), (10), (11) that the objective function in (7) is bounded above by

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

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

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

■

Lemma 1:

We want to show that

$$\sup_t \left|\hat{F}_{\hat{\theta}}\left(t\right)-F_{\theta}\left(t\right)\right| = o_p\left(1\right).$$

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

Therefore,

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

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 $\lim_{\alpha\to 0}\Pr(|b'x|<\alpha)$, here we have that

$$\begin{aligned}
\lim_{\alpha\to 0}\Pr\left(\left|t-\theta(X)\right|<\alpha\right) &= \lim_{\alpha\to 0}\Pr\left(-\alpha<t-\theta(X)<\alpha\right) \\
&= \lim_{\alpha\to 0}\Pr\left(t-\alpha<\theta(X)<t+\alpha\right) \\
&= \lim_{\alpha\to 0}\left[F_{\theta}(t+\alpha)-F_{\theta}(t-\alpha)\right] \\
&\leq 2\lim_{\alpha\to 0}\left\{\alpha\times\sup_{s\in\mathbb{R}}\left[f_{\theta}(s)\right]\right\}=0,
\end{aligned}$$

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

Theorem 1:

Proof. Fix $\varepsilon>0$. Then $F_{\theta}(\gamma_0+\varepsilon)-1+c>0$ and $1-c-F_{\theta}(\gamma_0-\varepsilon)>0$, by assumption (B5). Therefore, we have that

$$\begin{aligned}
\Pr\left(\left|\hat{\gamma}-\gamma_0\right|>\varepsilon\right) &\leq \Pr\left(\hat{\gamma}>\gamma_0+\varepsilon\right)+\Pr\left(\hat{\gamma}<\gamma_0-\varepsilon\right) \\
&\leq \Pr\left(\hat{F}_{\hat{\theta}}(\hat{\gamma})>\hat{F}_{\hat{\theta}}(\gamma_0+\varepsilon)\right)+\Pr\left(\hat{F}_{\hat{\theta}}(\hat{\gamma})<\hat{F}_{\hat{\theta}}(\gamma_0-\varepsilon)\right) \\
&= \Pr\left(1-c>\hat{F}_{\hat{\theta}}(\gamma_0+\varepsilon)\right)+\Pr\left(1-c<\hat{F}_{\hat{\theta}}(\gamma_0-\varepsilon)\right) \\
&\leq \Pr\left(F_{\theta}(\gamma_0+\varepsilon)-1+c<F_{\theta}(\gamma_0+\varepsilon)-\hat{F}_{\hat{\theta}}(\gamma_0+\varepsilon)\right) \\
&\quad +\Pr\left(1-c-F_{\theta}(\gamma_0-\varepsilon)<\hat{F}_{\hat{\theta}}(\gamma_0-\varepsilon)-F_{\theta}(\gamma_0-\varepsilon)\right) \\
&\leq \Pr\left(F_{\theta}(\gamma_0+\varepsilon)-1+c<\sup_{t\in[-A.A]}\left|\hat{F}_{\hat{\theta}}(t)-F_{\theta}(t)\right|\right) \\
&\quad +\Pr\left(1-c-F_{\theta}(\gamma_0-\varepsilon)<\sup_{t\in[-A.A]}\left|\hat{F}_{\hat{\theta}}(t)-F_{\theta}(t)\right|\right)
\end{aligned}$$

both of which converge to zero by lemma 1.

■

Lemma 2:

Proof.

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

By triangle inequality,

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

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

$$F_\theta(\gamma_0) = 1 - c = \hat{F}_{\hat{\theta}}(\hat{\gamma}) = \hat{F}_{\hat{\theta}}(\gamma_0) + (\hat{\gamma}-\gamma_0)\hat{f}_{\hat{\theta}}(\tilde{\gamma})$$

where $\tilde{\gamma}$ is intermediate between $\hat{\gamma}$ and γ_0. This gives us the following expansion for $\hat{\gamma}$.

$$\begin{aligned}
&(\hat{\gamma}-\gamma_0) \\
=\;& \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma_0) - \hat{F}_{\hat{\theta}}(\gamma_0)\right\} \\
=\;& \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma_0) - \frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{\gamma_0-\hat{\theta}(X_i)}{h_n}\right)\right\} \\
=\;& \left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{F_\theta(\gamma_0) - \frac{1}{n}\sum_{i=1}^{n}\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)\right\} \\
&+\left\{\hat{f}_{\hat{\theta}}(\tilde{\gamma})\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right) - \bar{L}\left(\frac{\gamma_0-\hat{\theta}(X_i)}{h_n}\right)\right]\right\}. \qquad (12)
\end{aligned}$$

Proof. Step 1. We first show that

$$\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\gamma_0) \xrightarrow{P} 0. \tag{13}$$

$$\begin{aligned}
\left|\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\gamma_0)\right| &\leq \left|\hat{f}_{\hat{\theta}}(\tilde{\gamma}) - f_\theta(\tilde{\gamma})\right| + \left|f_\theta(\tilde{\gamma}) - f_\theta(\gamma_0)\right| \\
&\leq \underbrace{\sup_{s\in[-A.A]}\left|\hat{f}_{\hat{\theta}}(s) - f_\theta(s)\right|}_{o_p(1),\text{ by lemma 2}} + \underbrace{\left|f_\theta(\tilde{\gamma}) - f_\theta(\gamma_0)\right|}_{o_p(1)\text{ by CMT and theorem 1}} \\
&= o_p(1).
\end{aligned}$$

Step 2: We will show that

$$\sqrt{nh_n}\left\{F_\theta(\gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\} = \beta + o_p(1). \tag{14}$$

Observe that

$$\begin{aligned}
T_n &= \frac{1}{n}\sum_{i=1}^{n}\left\{F_\theta(\gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{F_\theta(\gamma_0) - 1(\theta(X_i) \leq \gamma_0)\right\} + \frac{1}{n}\sum_{i=1}^{n}\left\{1(\theta(X_i) \leq \gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\} \\
&\equiv T_{2n} - T_{1n}.
\end{aligned} \tag{15}$$

Now,

$$\sqrt{nh_n}T_{2n} = \sqrt{h_n} \times \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{F_\theta(\gamma_0) - 1(\theta(X_i) \leq \gamma_0)\right\}}_{O_p(1)} = o_p(1).$$

We will show that

$$E\left(\sqrt{nh_n}T_{1n} - \beta\right)^2 = h_n Var\left(\sqrt{n}T_{1n}\right) + \left\{E\left(\sqrt{nh_n}T_{1n} - \beta\right)\right\}^2 \to 0 \tag{16}$$

and thus

$$\sqrt{nh_n}T_{1n} - \beta = o_p(1). \tag{17}$$

Now,

$$\begin{aligned}
&Var\left(\sqrt{n}T_{1n}\right) \\
&= Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1(\theta(X_i) \leq \gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\}\right) \\
&= Var\left\{1(\theta(X_i) \leq \gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\} \\
&= E\left\{1(\theta(X_i) \leq \gamma_0) - \bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\}^2 - \left\{F_\theta(\gamma_0) - E\left\{\bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right)\right\}\right\}^2
\end{aligned} \tag{18}$$

Observe that

$$
\begin{aligned}
& E\left(1\left(\theta\left(X_i\right) \leq \gamma_0\right)-\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right)^2 \\
= & \int_{-A}^{A}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)-1\left(s \leq \gamma_0\right)\right\}^2 f_\theta(s)\, ds \\
= & \int_{-A}^{\gamma_0}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)-1\left(s \leq \gamma_0\right)\right\}^2 f_\theta(s)\, ds \\
& +\int_{\gamma_0}^{A}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)-1\left(s \leq \gamma_0\right)\right\}^2 f_\theta(s)\, ds \\
= & \int_{-A}^{\gamma_0}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)-1\right\}^2 f_\theta(s)\, ds+\int_{\gamma_0}^{A}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)\right\}^2 f_\theta(s)\, ds
\end{aligned}
$$

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

Next,

$$
\begin{aligned}
& F_\theta\left(\gamma_0\right)-E\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
= & F_\theta\left(\gamma_0\right)-\int_{-A}^{A} \bar{L}\left(\frac{\gamma_0-s}{h_n}\right) f_\theta(s)\, ds \\
= & F_\theta\left(\gamma_0\right)-\int_{-A}^{\gamma_0} \bar{L}\left(\frac{\gamma_0-s}{h_n}\right) f_\theta(s)\, ds-\int_{\gamma_0}^{A} \bar{L}\left(\frac{\gamma_0-s}{h_n}\right) f_\theta(s)\, ds \\
= & \int_{-A}^{\gamma_0}\left[1\left(s \leq \gamma_0\right)-\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)\right] f_\theta(s)\, ds-\int_{\gamma_0}^{A} \bar{L}\left(\frac{\gamma_0-s}{h_n}\right) f_\theta(s)\, ds \\
\rightarrow & \ 0, \text{ by the DCT.}
\end{aligned}
$$

Thus, from (18), we have that

$$
Var\left(\sqrt{n} T_{1n}\right) \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{19}
$$

Next, consider

$$
\begin{aligned}
E\left(T_{1n}\right) & = E\left\{1\left(\theta\left(X_i\right) \leq \gamma_0\right)-\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
& = \left\{F_\theta\left(\gamma_0\right)-\int_{-A}^{A} \bar{L}\left(\frac{\gamma_0-s}{h_n}\right) f_\theta(s)\, ds\right\} \\
& = \left\{F_\theta\left(\gamma_0\right)-\bar{L}\left(\frac{\gamma_0-s}{h_n}\right) F_\theta(s)\Big|_{-A}^{A}-\frac{1}{h_n} \int_{-A}^{A} F_\theta(s) L\left(\frac{\gamma_0-s}{h_n}\right) ds\right\} \\
& = \left\{F_\theta\left(\gamma_0\right)-\int_{\frac{\gamma_0-A}{h_n}}^{\frac{\gamma_0+A}{h_n}} F_\theta\left(\gamma_0-u h_n\right) L(u)\, du\right\} \\
& = (-1)^{r+1} \frac{h_n^r}{r!} \times f_\theta^{(r-1)}\left(\gamma_0\right) \times \int_{-1}^{1} u^r L(u)\, du+o\left(h_n^r\right), \text{ by assumption B7.}
\end{aligned}
$$

This implies that

$$\begin{aligned} E\left(\sqrt{nh_n}T_{1n}\right) &= (-1)^{r+1}\frac{\sqrt{n}h_n^{r+1/2}}{r!} \times f_\theta^{(r-1)}(\gamma_0) \times \int_{-1}^{1} u^r L(u)\,du + o(h_n^r) \\ &\to \beta \text{, by assumption B7.} \end{aligned} \tag{20}$$

Now, (19) and (20) imply (16) and thus (17).

Step 3: We will now analyze the second term in (12):

$$S_n = \frac{1}{n}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) - \bar{L}\left(\frac{\gamma_0 - \hat{\theta}(X_i)}{h_n}\right)\right] du,$$

using U-statistic type decompositions to show that

$$\begin{aligned} \sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{j=1}^{n}\left\{\left[\lambda_{1n}(Z_j) - E\{\lambda_{1n}(Z_j)\}\right] - \left[\lambda_{2n}(Z_j) - E\{\lambda_{2n}(Z_j)\}\right]\right\} \\ &\quad + o_p(1) \\ &\xrightarrow{d} N\left(0, \eta^2\right), \end{aligned} \tag{21}$$

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

To that end observe that

$$\begin{aligned} \sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) - \bar{L}\left(\frac{\gamma_0 - \hat{\theta}(X_i)}{h_n}\right)\right] \\ &= \frac{\sqrt{h_n}}{\sqrt{n}h_n}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i) - \theta(X_i)\right\} L\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) \\ &\quad + \frac{\sqrt{h_n}}{2\sqrt{n}h_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i) - \theta(X_i)\right\}^2 L'\left(\frac{\tilde{\theta}(X_i) - \gamma_0}{h_n}\right). \end{aligned}$$

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

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

Thus we get that

$$\begin{aligned} \sqrt{nh_n}S_n &= \frac{\sqrt{h_n}}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{L}\left(\frac{\theta(X_i) - \gamma_0}{h_n}\right) - \bar{L}\left(\frac{\hat{\theta}(X_i) - \gamma_0}{h_n}\right)\right] \\ &= \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left\{\theta(X_i) - \hat{\theta}(X_i)\right\} L\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) + o_p(1) \\ &= -\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i) - \theta(X_i)\right\} L\left(\frac{\gamma_0 - \theta(X_i)}{h_n}\right) + o_p(1). \end{aligned}$$

Now, note that

$$\hat{\theta}(X_i) - \theta(X_i) = \left\{ \frac{\hat{\mu}(X_i)}{\hat{\pi}(X_i)} - \frac{\mu(X_i)}{\pi(X_i)} \right\} - \left\{ \frac{\hat{\nu}(X_i)}{\hat{\delta}(X_i)} - \frac{\nu(X_i)}{\delta(X_i)} \right\} \tag{22}$$

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

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} \left\{ \hat{\theta}(X_i) - \theta(X_i) \right\} L\left(\frac{\gamma_0 - \theta(X_i)}{h_n} \right) = O_p(1).$$

Step 3A: Now,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\hat{\mu}(X_i)}{\hat{\pi}(X_i)} - \frac{\mu(X_i)}{\pi(X_i)} \right\} \frac{1}{h_n} L\left(\frac{\theta(X_i) - \gamma_0}{h_n} \right) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\hat{\mu}(X_i) - \mu(X_i)}{\pi(X_i)} \right\} \frac{1}{h_n} L\left(\frac{\theta(X_i) - \gamma_0}{h_n} \right) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\mu(X_i)}{\pi(X_i)} \frac{\hat{\pi}(X_i) - \pi(X_i)}{\pi(X_i)} \right\} \frac{1}{h_n} L\left(\frac{\theta(X_i) - \gamma_0}{h_n} \right) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\{\hat{\mu}(X_i) - \mu(X_i)\}\{\hat{\pi}(X_i) - \pi(X_i)\}}{\pi(X_i)\hat{\pi}(X_i)} \frac{1}{h_n} L\left(\frac{\theta(X_i) - \gamma_0}{h_n} \right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mu(X_i)\{\hat{\pi}(X_i) - \pi(X_i)\}^2}{\pi^2(X_i)\hat{\pi}(X_i)} \frac{1}{h_n} L\left(\frac{\theta(X_i) - \gamma_0}{h_n} \right).
\end{aligned} \tag{23}$$

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

Now, the first two terms in (23) add up to

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\pi\left(X_i\right) \hat{\mu}\left(X_i\right)-\mu\left(X_i\right) \hat{\pi}\left(X_i\right)}{\pi^2\left(X_i\right)} \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right) \\
= & \sqrt{n} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right) Y_j S_j-\mu\left(X_i\right) S_j\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) \times \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)\right] \\
\equiv & \sqrt{n} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} w_n\left(Z_i, Z_j\right) \\
= & \underbrace{\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}\left[w_n\left(Z_i, Z_j\right)-E\left(w_n\left(Z_i, Z_j\right) \mid Z_i\right)-E\left(w_n\left(Z_i, Z_j\right) \mid Z_j\right)+E\left(w_n\left(Z_i, Z_j\right)\right)\right]}_{U_{1n}} \\
& \underbrace{+\frac{1}{\sqrt{n}} \sum_{j=1}^{n}\left[E\left(w_n\left(Z_i, Z_j\right) \mid Z_j\right)-E\left(w_n\left(Z_i, Z_j\right)\right)\right]}_{U_{2n}} \\
& \underbrace{+\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left(w_n\left(Z_i, Z_j\right) \mid Z_i\right).}_{U_{3n}}
\end{aligned} \tag{24}
$$

Step 3B: We first show that

$$
U_{3n}=o_p(1) . \tag{25}
$$

Notice that

$$
\begin{aligned}
& E\left[\frac{\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)}{\pi^2\left(X_i\right)} \times\left\{\pi\left(X_i\right) Y_j S_j-\mu\left(X_i\right) S_j\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) \mid Z_i\right] \\
& \stackrel{L.I.E.}{=} E\left[\frac{\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)}{\pi^2\left(X_i\right)} \times\left\{\pi\left(X_i\right) \frac{\mu\left(X_j\right)}{f\left(X_j\right)}-\mu\left(X_i\right) \frac{\pi\left(X_j\right)}{f\left(X_j\right)}\right\} \frac{1}{\sigma_n^p} K\left(\frac{X_j-X_i}{\sigma_n}\right) \mid Z_i\right] \\
= & \frac{\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)}{\pi^2\left(X_i\right)} \times \int \frac{\pi\left(X_i\right) \mu(x)-\mu\left(X_i\right) \pi(x)}{f(x)} \frac{1}{\sigma_n^p} K\left(\frac{x-X_i}{\sigma_n}\right) f(x) d x \\
= & \frac{\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)}{\pi^2\left(X_i\right)} \times \int\left[\pi\left(X_i\right) \mu\left(X_i+u \sigma_n\right)-\mu\left(X_i\right) \pi\left(X_i+u \sigma_n\right)\right] K(u) d u \\
& \stackrel{A1}{=} H\left(X_i\right) \times \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right) \times O\left(\sigma_n^q\right),
\end{aligned}
$$

for some uniformly bounded function H by assumption. Therefore,

$$
U_{3n}=O\left(\sigma_n^q\right) \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H\left(X_i\right) \times \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)=O_p\left(\sqrt{n} \sigma_n^q\right)=o_p(1)
$$

by assumption B8.

Step 3C: The term

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

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

$$E\left(U_{1n}^2\right)=o\left(1\right) \tag{26}$$

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

$$\begin{aligned}
& n^{-1}E\left(w_n^2\left(Z_i,Z_j\right)\right)\\
= \;& n^{-1}E\left\{\frac{1}{\pi^4\left(X_i\right)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}^2\frac{1}{\sigma_n^{2p}}K^2\left(\frac{X_j-X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)\right]^2\right\}\\
= \;& n^{-1}E\left\{\begin{array}{c}\frac{1}{\pi^4(X_i)}\frac{1}{\sigma_n^{2p}}K^2\left(\frac{X_j-X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta(X_i)-\gamma_0}{h_n}\right)\right]^2\\ \times\left\{\pi^2\left(X_i\right)E\left(Y^2S|X_j\right)+\mu^2(X_i)E\left(S|X_j\right)-2\pi\left(X_i\right)\mu\left(X_i\right)E\left(YS|X_j\right)\right\}\end{array}\right\}\\
= \;& \frac{1}{n\sigma_n^p h_n^2}\int\left\{\begin{array}{c}\frac{1}{\pi^4(X)}K^2\left(u\right)\times\left[L\left(\frac{\theta(x)-\gamma_0}{h_n}\right)\right]^2\\ \times\left\{\begin{array}{c}\pi^2\left(x\right)E\left(Y^2S|X=x+u\sigma_n\right)\\ +\mu^2(x)E\left(S|X=x+u\sigma_n\right)\\ -2\pi\left(x\right)\mu\left(x\right)E\left(YS|X=x+u\sigma_n\right)\end{array}\right\}\end{array}\right\}f_X\left(x\right)f_X\left(x+u\sigma_n\right)dudx\\
= \;& O\left(\frac{1}{n\sigma_n^p h_n^2}\right)\to 0 \text{ which is implied by B8.}
\end{aligned}$$

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

$$
\begin{aligned}
& E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) \\
= \ & E\left\{\left[\frac{1}{\pi^2\left(X_i\right)}\left\{\pi\left(X_i\right)Y_jS_j-\mu\left(X_i\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+X_i}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right)\right]\right]|Y_j, S_j, X_j\right\} \\
= \ & \int\left[\frac{1}{\pi^2\left(x\right)}\left\{\pi\left(x\right)Y_jS_j-\mu\left(x\right)S_j\right\}\frac{1}{\sigma_n^p}K\left(\frac{-X_j+x}{\sigma_n}\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(x\right)-\gamma_0}{h_n}\right)\right]\right]f\left(x\right)dx \\
= \ & \int\left[\frac{\pi\left(X_j+u\sigma_n\right)Y_jS_j-\mu\left(X_j+u\sigma_n\right)S_j}{\pi^2\left(X_j+u\sigma_n\right)}K\left(u\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j+u\sigma_n\right)-\gamma_0}{h_n}\right)\right]\right]f\left(X_j+u\sigma_n\right)du \\
= \ & \left[\frac{1}{\pi^2\left(X_j\right)}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right)\times\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j\right)-\gamma_0}{h_n}\right)\right]\right]\int K\left(u\right)du+O\left(\sigma_n^q\right) \\
= \ & \left[\underbrace{\left[\frac{1}{\pi^2\left(X_j\right)}\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}f\left(X_j\right)\right]}_{W(Z_j)}\times\underbrace{\left[\frac{1}{h_n}L\left(\frac{\theta\left(X_j\right)-\gamma_0}{h_n}\right)\right]}_{V_{1n}(\theta(X_j))}\right] \\
& +O\left(\sigma_n^q\right).
\end{aligned}
$$

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 $\omega^2(s) = E\{W^2(Z_j)|\theta(X_j) = s\}$. Then

$$\begin{aligned}
& E\left\{W^2(Z_j)V_{1n}^2(\theta(X_j))\right\} \\
= & \int_{-A}^{A}\omega^2(s)\left[\frac{1}{h_n}L\left(\frac{s-\gamma_0}{h_n}\right)\right]^2 f_\theta(s)\,ds \\
= & \frac{1}{h_n}\int_{\frac{-A-\gamma_0}{h_n}}^{\frac{A-\gamma_0}{h_n}}\omega^2(\gamma_0+uh_n)L^2(u)f_\theta(\gamma_0+uh_n)\,du \\
= & \frac{1}{h_n}\omega^2(\gamma_0)f_\theta(\gamma_0)\int_{\frac{-A-\gamma_0}{h_n}}^{\frac{A-\gamma_0}{h_n}}L^2(u)\,du + \text{terms of smaller order.}
\end{aligned}$$

This implies that

$$\begin{aligned}
Var\left(\sqrt{h_n}U_{2n}\right) &= Var\left(\sqrt{\frac{h_n}{n}}\sum_{j=1}^{n}\left[E\left(w_n(Z_i,Z_j)|Z_j\right)-E\left(w_n(Z_i,Z_j)\right)\right]\right) \\
&\to \omega^2(\gamma_0)f_\theta(\gamma_0)\int_{-\infty}^{\infty}L^2(u)\,du. \qquad (27)
\end{aligned}$$

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

$$\begin{aligned}
E\left(U_n^2\right) &= \sum_{j=1}^{n}E\left(R_{nj}^2\right) \\
&= \frac{h_n}{n}\sum_{j=1}^{n}E\left(E\left(w_n(Z_i,Z_j)|Z_j\right)-E\left(w_n(Z_i,Z_j)\right)\right)^2 \\
&= \frac{h_n}{n}\sum_{j=1}^{n}Var\left(W(Z_j)V_{1n}(\theta(X_j))\right)+o(1) \\
&= \frac{h_n}{h_n}\omega^2(\gamma_0)f_\theta(\gamma_0)\int_{-\infty}^{\infty}L^2(u)\,du+o(1) \\
&\to \omega^2(\gamma_0)f_\theta(\gamma_0)\int_{-\infty}^{\infty}L^2(u)\,du.
\end{aligned}$$

by (27). 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

$$U_n = \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) - E\left(w_n\left(Z_i, Z_j\right)\right)\right] \xrightarrow{d} N\left(0, \omega^2\left(\gamma_0\right) f_\theta\left(\gamma_0\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right). \tag{28}$$

Putting together (25), (26), (27) and (27), we get that

$$\begin{aligned}
& \sqrt{\frac{h_n}{n}} \sum_{i=1}^{n} \left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)} - \frac{\mu(X_i)}{\pi\left(X_i\right)}\right\} \left[\frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma_0}{h_n}\right)\right] \\
= & \sqrt{h_n} U_{1n} + \sqrt{h_n} U_{2n} + \sqrt{h_n} U_{3n} \\
= & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[E\left(w_n\left(Z_i, Z_j\right)|Z_j\right) - E\left(w_n\left(Z_i, Z_j\right)\right)\right] + o_p\left(1\right) \\
= & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \left[\lambda_{1n}\left(Z_j\right) - E\left\{\lambda_{1n}\left(Z_j\right)\right\}\right] \\
& \xrightarrow{d} N\left(0, \omega^2\left(\gamma_0\right) f_\theta\left(\gamma_0\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right),
\end{aligned}$$

where

$$\begin{aligned}
& \lambda_{1n}\left(Z_j\right) \\
= & \left[\frac{f\left(X_j\right)}{\pi^2\left(X_j\right)} \left\{\pi\left(X_j\right) Y_j S_j - \mu\left(X_j\right) S_j\right\}\right] \frac{1}{h_n} L\left(\frac{\theta\left(X_j\right) - \gamma_0}{h_n}\right) \\
& \text{and} \\
\omega^2\left(s\right) = & E\left\{\left\{\frac{\pi\left(X\right) Y S - \mu\left(X\right) S}{\pi^2\left(X\right)} f\left(X\right)\right\}^2 |\theta\left(X\right) = s\right\}.
\end{aligned} \tag{29}$$

Similarly, we will get that

$$\begin{aligned}
& \sqrt{\frac{h_n}{n}} \sum_{i=1}^{n} \left\{\frac{\hat{\nu}\left(X_i\right)}{\hat{\delta}(X_i)} - \frac{\nu\left(X_i\right)}{\delta\left(X_i\right)}\right\} \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma_0}{h_n}\right) \\
= & \frac{\sqrt{h_n}}{\sqrt{n}} \sum_{j=1}^{n} \lambda_{2n}\left(Z_j\right) \xrightarrow{d} N\left(0, \tau^2\left(\gamma_0\right) f_\theta\left(\gamma_0\right) \int_{-\infty}^{\infty} L^2\left(u\right) du\right),
\end{aligned}$$

where

$$\begin{aligned}
& \lambda_{2n}\left(Z_j\right) \\
= & \left[\frac{f\left(X_j\right)}{\delta^2\left(X_j\right)} \left\{\delta\left(X_j\right) Y_j \left(1 - S_j\right) - \nu\left(X_j\right)\left(1 - S_j\right)\right\}\right] \frac{1}{h_n} L\left(\frac{\theta\left(X_i\right) - \gamma_0}{h_n}\right) \\
& \text{and} \\
\tau^2\left(s\right) = & E\left\{\left\{\frac{\delta\left(X\right) Y\left(1 - S\right) - \nu\left(X\right)\left(1 - S\right)}{\delta^2\left(X\right)} f\left(X\right)\right\}^2 |\theta\left(X\right) = s\right\}.
\end{aligned} \tag{30}$$

Thus we get that

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

which establishes (21).

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

$$\begin{aligned}
& E\left\{\lambda_{1n}\left(Z_j\right)\lambda_{2n}\left(Z_j\right)\right\} \\
= \quad & E\left\{\frac{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j}{\pi^2\left(X_j\right)}\times\frac{\delta\left(X_j\right)Y_j\left(1-S_j\right)-\nu\left(X_j\right)\left(1-S_j\right)}{\delta^2\left(X_j\right)}\times f^2\left(X_j\right)L\left(\frac{\theta\left(X_j\right)-\gamma_0}{h_n}\right)^2\right\}
\end{aligned}$$

Now,

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

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

$$\eta^2=\left\{\tau^2\left(\gamma_0\right)+\omega^2\left(\gamma_0\right)\right\}\times f_\theta\left(\gamma_0\right)\int_{-\infty}^{\infty}L^2\left(u\right)du, \tag{31}$$

where

$$\begin{aligned}
\tau^2\left(\gamma_0\right) &= E\left\{\left\{\frac{\delta\left(X\right)Y\left(1-S\right)-\nu\left(X\right)\left(1-S\right)}{\delta^2\left(X\right)}f\left(X\right)\right\}^2|\theta\left(X\right)=\gamma_0\right\} \\
\omega^2\left(\gamma_0\right) &= E\left\{\left\{\frac{\pi\left(X\right)YS-\mu\left(X\right)S}{\pi^2\left(X\right)}f\left(X\right)\right\}^2|\theta\left(X\right)=\gamma_0\right\}.
\end{aligned}$$

Now put together (13), (14), (21) and (31) to conclude from (12) that

$$\begin{aligned}
\sqrt{nh_n}\left(\hat{\gamma}-\gamma_0\right) &= \frac{1}{f_\theta\left(\gamma_0\right)}\frac{\sqrt{h_n}}{\sqrt{n}}\sum_{j=1}^{n}\left\{\lambda_{1n}\left(Z_j\right)-\lambda_{2n}\left(Z_j\right)\right\}+o_p\left(1\right) \\
&\xrightarrow{d} N\left(0,\frac{\tau^2\left(\gamma_0\right)+\omega^2\left(\gamma_0\right)}{f_\theta\left(\gamma_0\right)}\int_{-\infty}^{\infty}L^2\left(u\right)du\right).
\end{aligned}$$

■

Theorem 3:

Proof.

$$\begin{aligned}
\hat{\zeta}-\zeta_0 &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}(X_i)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)-\zeta_0\\
&= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}(X_i)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)-\frac{1}{n}\sum_{i=1}^{n}\theta(X_i)\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)}_{T_{1n}}\\
&\quad \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\theta(X_i)\left[\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)-1\{\theta(X_i)\leq\gamma_0\}\right]}_{T_{2n}}\\
&\quad \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\{\theta(X_i)1\{\theta(X_i)\leq\gamma_0\}-\zeta_0\}}_{=o_p(1),\text{ by standard WLLN.}}.
\end{aligned}$$

Now,

$$\begin{aligned}
|T_{1n}| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i)\bar{L}\left(\frac{\hat{\gamma}-\hat{\theta}(X_i)}{h_n}\right)-\theta(X_i)\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)\right\}\right|\\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{\theta}(X_i)-\theta(X_i)}{h_n}\right|\left|h_n\bar{L}\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)-\tilde{\theta}(X_i)L\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)\right|\\
&\quad +\frac{(\hat{\gamma}-\gamma_0)}{h_n}\times\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{\theta}(X_i)L\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)\right|\\
&\leq \frac{\sup\left|\hat{\theta}(x)-\theta(x)\right|}{h_n}\frac{1}{n}\sum_{i=1}^{n}\left|h_n\bar{L}\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)-\tilde{\theta}(X_i)L\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)\right|\\
&\quad +\left(\frac{nh_n(\hat{\gamma}-\gamma_0)^2}{nh_n^3}\right)^{1/2}\times\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{\theta}(X_i)L\left(\frac{\tilde{\gamma}-\tilde{\theta}(X_i)}{h_n}\right)\right|.
\end{aligned}$$

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

$$\leq\frac{\sup_{x\in\mathcal{X}}\left|\hat{\theta}(x)-\theta(x)\right|}{h_n}\times O_p(1)+\left(\frac{nh_n(\hat{\gamma}-\gamma_0)^2}{nh_n^3}\right)^{1/2}\times O_p(1).$$

Now, theorem 3 implies that $nh_n(\hat{\gamma}-\gamma_0)^2=O_p(1)$, Assumptions B1 and B4 (i) imply that $\frac{\sup\left|\hat{\theta}(x)-\theta(x)\right|}{h_n}=o_p(1)$ and that $nh_n^3\to\infty$. Thus we have that $T_{1n}=o_p(1)$.

As for T_{2n}, observe that since $\theta(\cdot)$ is uniformly bounded, by using steps exactly analogous to step 2 in the proof of theorem 2 (leading to (14)), we will get by the DCT that $T_{2n}=o_p(1)$. ■

Theorem 4:

Proof. We will work with the following expansion

$$
\begin{aligned}
&\hat{\zeta}-\zeta_0 \\
= \; & \underbrace{\frac{1}{n}\sum_{i=1}^{n}\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\zeta_0}_{T_{1n}} \\
& \underbrace{+\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\frac{1}{h_n}\theta\left(X_i\right)L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\}}_{T_{2n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma_0\right)\frac{1}{nh_n}\sum_{i=1}^{n}\theta\left(X_i\right)L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)}_{T_{3n}} \\
& \underbrace{-\frac{1}{4nh_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}^2\left\{-2h_nL\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)+\tilde{\theta}\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right\}}_{T_{4n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma_0\right)^2\times\frac{1}{4nh_n^2}\sum_{i=1}^{n}\theta\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)}_{T_{5n}} \\
& \underbrace{+\left(\hat{\gamma}-\gamma_0\right)\frac{1}{2nh_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left[h_nL\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)-\tilde{\theta}\left(X_i\right)L'\left(\frac{\tilde{\gamma}-\tilde{\theta}\left(X_i\right)}{h_n}\right)\right]}_{T_{6n}} \qquad (32)
\end{aligned}
$$

Step 4: Under assumptions B1 and B8, the fourth term in (32) 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 $\frac{(\hat{\gamma}-\gamma_0)^2}{h_n^2}=O_p\left(\frac{1}{nh_n^3}\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$ by assumption B12. So the fifth term in (32) 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 (32) equals

$$
\begin{aligned}
\frac{1}{nh_n}\sum_{i=1}^{n}\theta\left(X_i\right)L\left(\frac{\theta\left(X_i\right)-\gamma_0}{h_n}\right) &\rightarrow E\left(\theta\left(X_i\right)\frac{1}{h_n}\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right) \\
&= \gamma_0 f_\theta\left(\gamma_0\right)+O\left(h_n^r\right) \\
&\rightarrow \gamma_0 f_\theta\left(\gamma_0\right),
\end{aligned}
$$

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

Combining steps 4 and 5, we get that

$$
\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\} \\
= \quad & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\zeta_0\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\frac{1}{h_n}\theta\left(X_i\right)L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+\sqrt{n}\left(\hat{\gamma}-\gamma_0\right)\times\gamma_0 f_\theta\left(\gamma_0\right)+o_p\left(1\right). \qquad (33)
\end{aligned}
$$

Replacing in the previous display the asymptotic expansion of $(\hat{\gamma}-\gamma_0)$ from (12), we have that

$$
\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\} \\
= \quad & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\zeta_0\right\}-\frac{\gamma_0}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma_0\right)-F_\theta\left(\gamma_0\right)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+\gamma_0\frac{1}{2\sqrt{n}h_n^2}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}^2 L'\left(\frac{\tilde{\theta}\left(X_i\right)-\gamma_0}{h_n}\right) \\
&+o_p\left(1\right). \qquad (34)
\end{aligned}
$$

The third term in (34) in absolute value is dominated by

$$
\begin{aligned}
& \frac{\gamma_0}{2}\times\sup_u\left\{\hat{\theta}\left(u\right)-\theta\left(u\right)\right\}^2\times\frac{\sqrt{n}}{h_n^2}\frac{1}{n}\sum_{i=1}^{n}\left|L'\left(\frac{\tilde{\theta}\left(X_i\right)-\gamma_0}{h_n}\right)\right| \\
= \quad & \frac{\gamma_0}{2}\times\left\{\left(\frac{\ln n}{n\sigma_n^p}\right)^{1/2}+\sigma_n^q\right\}^2\times\frac{\sqrt{n}}{h_n^2}\times O_p\left(1\right) \\
= \quad & o_p\left(1\right), \text{ by assumption B4(iii).}
\end{aligned}
$$

Thus from (34), we have that

$$
\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\} \\
= \quad & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\zeta_0\right\}-\frac{\gamma_0}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma_0\right)-F_\theta\left(\gamma_0\right)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
&+o_p\left(1\right). \qquad (35)
\end{aligned}
$$

Step 6A: Consider the first term in (35)

$$
\begin{aligned}
&= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right]}_{T_{4n}} \\
&\quad \underbrace{+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}-\zeta_0\right]}_{T_{5n}=O_p(1),\text{ by CLT.}} \qquad (36)
\end{aligned}
$$

We will show that $T_{4n}=o_p(1)$ using the arguments similar to the ones used for showing (17).

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

$$
\begin{aligned}
E\left(T_{4n}\right) &= \sqrt{n}E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-G_\theta\left(\gamma_0\right)\right\} \\
&= \sqrt{n}\left\{\int_{-A}^{A}\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)g_\theta(s)\,ds-G_\theta\left(\gamma_0\right)\right\} \\
&= \sqrt{n}\left[\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)G_\theta(s)\right\}|_{-A}^{A}-\int_{-A}^{A}\frac{1}{h_n}L\left(\frac{\gamma_0-s}{h_n}\right)G_\theta(s)\,ds-G_\theta\left(\gamma_0\right)\right] \\
&= \sqrt{n}\left[-\int_{\frac{\gamma_0-A}{h_n}}^{\frac{\gamma_0+A}{h_n}}L(u)\,G_\theta\left(\gamma_0-uh_n\right)du-G_\theta\left(\gamma_0\right)\right] \\
&= O\left(\sqrt{n}h_n^r\right)\to 0,\text{ by B12.} \qquad (37)
\end{aligned}
$$

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

$$
\begin{aligned}
&E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right\}^2 \\
&= \int_{-A}^{A}\left\{\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)-1\left\{s<\gamma_0\right\}\right\}^2 g_\theta(s)\,ds \\
&= \int_{-A}^{A}\bar{L}^2\left(\frac{\gamma_0-s}{h_n}\right)g_\theta(s)\,ds \\
&\quad +\int_{-A}^{\gamma_0}1\left\{s<\gamma_0\right\}g_\theta(s)\,ds \\
&\quad -2\int_{-A}^{A}\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)\times 1\left\{s<\gamma_0\right\}\times g_\theta(s)\,ds
\end{aligned}
$$

Using the DCT repeatedly (c.f. the steps leading to (15)), we get that

$$\begin{aligned}
& E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right\}^2 \\
= & \int_{-A}^{A}\bar{L}^2\left(\frac{\gamma_0-s}{h_n}\right)g_\theta\left(s\right)ds+\int_{-A}^{\gamma_0}1\left\{s<\gamma_0\right\}g_\theta\left(s\right)ds \\
& -2\int_{-A}^{A}\bar{L}\left(\frac{\gamma_0-s}{h_n}\right)\times 1\left\{s<\gamma_0\right\}\times g_\theta\left(s\right)ds \\
\to & G_\theta\left(\gamma_0\right)+G_\theta\left(\gamma_0\right)-2G_\theta\left(\gamma_0\right)=0.
\end{aligned}$$

This implies that for T_{4n} defined in (36),

$$\begin{aligned}
Var\left(T_{4n}\right) &= Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right]\right) \\
&= Var\left(\theta\left(X_i\right)\times\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right) \\
&\leq E\left\{\theta\left(X_i\right)\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)-\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}\right\}^2 \\
&\to 0. \qquad (38)
\end{aligned}$$

From (37) and (38), we get that $E\left(T_{4n}\right)^2\to 0$ and thus $T_{4n}=o_p\left(1\right)$.

Replacing in (35), we get that

$$\begin{aligned}
& \sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\} \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}-\zeta_0\right]-\frac{\gamma_0}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\left(\theta\left(X_i\right)\leq\gamma_0\right)-F_\theta\left(\gamma_0\right)\right\} \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\} \\
& +o_p\left(\frac{1}{\sqrt{n}}\right). \qquad (39)
\end{aligned}$$

The final step is to analyze the third term in (39), using U-statistic type decompositions.

First notice that analogous to (24) above, we have here that up to $o_p(1)$ terms:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}(X_i)-\theta(X_i)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)+\{\gamma_0-\theta(X_i)\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)\right\}\\
= &\sqrt{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left[\begin{array}{c}\frac{1}{\pi^2(X_i)}\{\pi(X_i)Y_jS_j-\mu(X_i)S_j\}\frac{1}{\sigma_n^p}K\left(\frac{X_j-X_i}{\sigma_n}\right)\\ \times\left[\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)+\{\gamma_0-\theta(X_i)\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)\right]\end{array}\right]\\
\equiv &\sqrt{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}w_n(Z_i,Z_j)\\
= &\underbrace{\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left[w_n(Z_i,Z_j)-E(w_n(Z_i,Z_j)|Z_i)-E(w_n(Z_i,Z_j)|Z_j)+E(w_n(Z_i,Z_j))\right]}_{U_{1n}}\\
&+\underbrace{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[E(w_n(Z_i,Z_j)|Z_j)-E(w_n(Z_i,Z_j))\right]}_{U_{2n}}\\
&+\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E(w_n(Z_i,Z_j)|Z_i).}_{U_{3n}}
\end{aligned}
$$

It is straightforward (replace the kernel involving terms) to verify that we will get the same conclusion as (26) and (25) 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_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) \\ \times\left[\bar{L}\left(\frac{\gamma_0-\theta(X_i)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\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_j S_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_0-\theta(X_i)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\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_j S_j-\mu\left(X_j+u\sigma_n\right) S_j\right\} K\left(u\right) \\ \times\left[\bar{L}\left(\frac{\gamma_0-\theta(X_j+u\sigma_n)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_j+u\sigma_n\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\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_j S_j-\mu\left(X_j\right) S_j\right\} f\left(X_j\right) \\ \times\left[\bar{L}\left(\frac{\gamma_0-\theta(X_j)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_j\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\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_j S_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_0-\theta\left(X_j\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_j\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\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,

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{j=1}^n\left[E\left(w_n\left(Z_i, Z_j\right) | Z_j\right)-E\left(w_n\left(Z_i, Z_j\right)\right)\right] \\
= \; & \frac{1}{\sqrt{n}} \sum_{j=1}^n\left[E\left(w_n\left(Z_i, Z_j\right) | Z_j\right)-W\left(Z_j\right) \times 1\left(\theta\left(X_j\right) \leq \gamma_0\right)\right] \\
& +\frac{1}{\sqrt{n}} \sum_{j=1}^n\left\{W\left(Z_j\right) \times 1\left(\theta\left(X_j\right) \leq \gamma_0\right)\right\} \\
= \; & \frac{1}{\sqrt{n}} \sum_{j=1}^n\left\{W\left(Z_j\right) \times 1\left(\theta\left(X_j\right) \leq \gamma_0\right)\right\} \\
& +\frac{1}{\sqrt{n}} \sum_{j=1}^n \underbrace{W\left(Z_j\right)\left[\begin{array}{c} \left\{\bar{L}\left(\frac{\gamma_0-\theta(X_j)}{h_n}\right)-1\left(\theta\left(X_j\right) \leq \gamma_0\right)\right\} \\ +\left\{\gamma_0-\theta\left(X_j\right)\right\} \frac{1}{h_n} L\left(\frac{\gamma_0-\theta(X_j)}{h_n}\right) \end{array}\right]}_{T_{nj}}. \qquad (40)
\end{aligned}
$$

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

$$
\begin{aligned}
& E\left(T_{nj}^2\right) \\
= & \int_{-A}^{A} \omega^2(s) \left[\left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1(s \leq \gamma_0) \right\} + \{\gamma_0 - s\} \frac{1}{h_n} L\left(\frac{\gamma_0 - s}{h_n}\right) \right]^2 f_\theta(s)\, ds \\
= & \int_{-A}^{A} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1(s \leq \gamma_0) \right\}^2 f_\theta(s)\, ds \\
& + \int_{-A}^{A} \omega^2(s) \left(\frac{\gamma_0 - s}{h_n}\right)^2 L^2\left(\frac{\gamma_0 - s}{h_n}\right) f_\theta(s)\, ds \\
& + 2 \int_{-A}^{A} \omega^2(s) \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) \left(\frac{\gamma_0 - s}{h_n}\right) L\left(\frac{\gamma_0 - s}{h_n}\right) f_\theta(s)\, ds. \qquad (41)
\end{aligned}
$$

The first term in (41) equals

$$
\begin{aligned}
& \int_{-A}^{A} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1(s \leq \gamma_0) \right\}^2 f_\theta(s)\, ds \\
= & \int_{-A}^{\gamma_0} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1(s \leq \gamma_0) \right\}^2 f_\theta(s)\, ds \\
& + \int_{\gamma_0}^{A} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1(s \leq \gamma_0) \right\}^2 f_\theta(s)\, ds \\
= & \int_{-A}^{\gamma_0} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) - 1 \right\}^2 f_\theta(s)\, ds + \int_{\gamma_0}^{A} \omega^2(s) \left\{ \bar{L}\left(\frac{\gamma_0 - s}{h_n}\right) \right\}^2 f_\theta(s)\, ds
\end{aligned}
$$

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 (41) converges to zero by the DCT since $\lim_{u\to\pm\infty} u^2 L^2(u) = 0$. The third integral in (41) 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

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

Next,

$$
\begin{aligned}
& \sqrt{n} E \left\{ W(Z_j) \left[\left\{ \bar{L}\left(\frac{\gamma_0 - \theta(X_j)}{h_n}\right) - 1(\theta(X_j) \leq \gamma_0) \right\} + \{\gamma_0 - \theta(X_j)\} \frac{1}{h_n} L\left(\frac{\gamma_0 - \theta(X_j)}{h_n}\right) \right] \right\} \\
= & \sqrt{n} E_{X_j} \left\{ E\{W\underbrace{(Z_j)}_{=0}|X_j\} \times \left[\begin{array}{c} \left\{ \bar{L}\left(\frac{\gamma_0 - \theta(X_j)}{h_n}\right) - 1(\theta(X_j) \leq \gamma_0) \right\} \\ + \{\gamma_0 - \theta(X_j)\} \frac{1}{h_n} L\left(\frac{\gamma_0 - \theta(X_j)}{h_n}\right) \end{array} \right] \right\} \\
= & \ 0.
\end{aligned}
$$

So it follows that the second term in (40) is $o_p(1)$.

Thus we have that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_i\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_i\right)}{h_n}\right)\right\}\\
=\;&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\left\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\right\}}{\pi^2\left(X_j\right)}f_X\left(X_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma_0\right)\right\}+o_p\left(1\right)\\
&\xrightarrow{d} N\left(0,\int_{-A}^{\gamma_0}\omega^2\left(s\right)f_\theta\left(s\right)ds\right),
\end{aligned}$$

where

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

Using exactly analogous steps, we will also get that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\nu}\left(X_i\right)}{\hat{\delta}(X_i)}-\frac{\nu\left(X_i\right)}{\delta\left(X_i\right)}\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_j\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_j\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_j\right)}{h_n}\right)\right\}\\
=\;&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\delta\left(X_j\right)Y_j\left(1-S_j\right)-\nu\left(X_j\right)\left(1-S_j\right)}{\delta^2\left(X\right)}f_X\left(X_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma_0\right)\right\}+o_p\left(1\right).\\
&\xrightarrow{d} N\left(0,\int_{-A}^{\gamma_0}\tau^2\left(s\right)f_\theta\left(s\right)ds\right),
\end{aligned}$$

where

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

Finally, we get that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\theta}\left(X_i\right)-\theta\left(X_i\right)\right\}\left\{\bar{L}\left(\frac{\gamma_0-\theta\left(X_j\right)}{h_n}\right)+\left\{\gamma_0-\theta\left(X_j\right)\right\}\frac{1}{h_n}L\left(\frac{\gamma_0-\theta\left(X_j\right)}{h_n}\right)\right\}\\
&\xrightarrow{d} N\left(0,\int_{-A}^{\gamma_0}\left\{\omega^2\left(s\right)+\tau^2\left(s\right)\right\}f_\theta\left(s\right)ds\right), \qquad (42)
\end{aligned}$$

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

Replacing in (33), we finally arrive at

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\zeta}-\zeta_0\right\} \\
= \;& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\{\pi\left(X_j\right)Y_jS_j-\mu\left(X_j\right)S_j\}}{\pi^2\left(X_j\right)}f_X\left(X_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma_0\right)\right\} \\
&-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\delta\left(X_j\right)Y_j\left(1-S_j\right)-\nu\left(X_j\right)\left(1-S_j\right)}{\delta^2\left(X\right)}f_X\left(X_j\right)\times 1\left(\theta\left(X_j\right)\leq\gamma_0\right)\right\} \\
&+\gamma_0\times\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{F_\theta\left(\gamma_0\right)-1\left(\theta\left(X_j\right)\leq\gamma_0\right)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\theta\left(X_i\right)\times 1\left\{\theta\left(X_i\right)<\gamma_0\right\}-\zeta_0\right\}+o_p\left(1\right). \qquad (43)
\end{aligned}$$

■

Lemma 3:

Proof. Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\hat{\phi}\left(X_j,1\right)-\phi\left(X_j,1\right)\right] \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)}{\hat{\pi}\left(X_i\right)}-\frac{\mu(X_i)}{\pi\left(X_i\right)}\right\} \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\hat{\mu}\left(X_i\right)-\mu(X_i)}{\pi\left(X_i\right)}\right\}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\mu\left(X_i\right)}{\pi\left(X_i\right)}\frac{\hat{\pi}\left(X_i\right)-\pi(X_i)}{\pi\left(X_i\right)}\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\{\hat{\mu}\left(X_i\right)-\mu(X_i)\}\{\hat{\pi}\left(X_i\right)-\pi(X_i)\}}{\pi\left(X_i\right)\hat{\pi}\left(X_i\right)} \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu(X_i)\{\hat{\pi}\left(X_i\right)-\pi(X_i)\}^2}{\pi^2\left(X_i\right)\hat{\pi}\left(X_i\right)}. \qquad (44)
\end{aligned}$$

The last two terms are bounded above by a positive scalar times $\sqrt{n}\sup_x\|\{\hat{\mu}\left(x\right)-\mu(x)\}\{\hat{\pi}\left(x\right)-\pi(x)\}\|$ and $\sqrt{n}\sup_x\|\{\hat{\pi}\left(x\right)-\pi(x)\}\|^2$, 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 (44) is asymptotically equivalent to ($\sqrt{n}$ times) a centered sample average.

Now, the first two terms in (44) add up to

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

We will show that

$$
\begin{aligned}
E\left(U_{1n}\right)^2 &= o\left(1\right), && (45) \\
U_{2n} &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ E\left(S | X_j\right) \times Y_j S_j - E\left(SY | X_j\right) \times S_j \right\} + o_p\left(1\right), && (46) \\
U_{3n} &= o_p\left(1\right). && (47)
\end{aligned}
$$

Observe that

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

for some uniformly bounded function H by assumption. Therefore, $U_{3n} = O_p\left(\sqrt{n}\sigma_n^q\right) = o_p\left(1\right)$ by assumption A3 and this establishes (47).

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 (46).

One can establish (45) by essentially repeating the proof of Powell, Stoker and Stock (1989) lemma 3.1.

Combining (45), (46) and (47), we get that

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

■

Corollary to Theorem 4

Proof. Observe that

$$
\begin{aligned}
& \hat{\rho}-\rho_0 \\
= & \underbrace{\frac{1}{n} \sum_{j=1}^n\left[\hat{\phi}\left(X_j, 1\right)-\phi\left(X_j, 1\right)\right]}_{S_{1n}}+\underbrace{\frac{1}{n} \sum_{j=1}^n\left[\phi\left(X_j, 1\right)-E\left\{\phi\left(X_j, 1\right)\right\}\right]}_{S_{2n}}-\left\{\hat{\zeta}-\zeta_0\right\}
\end{aligned}
$$

By lemma 1, $S_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$, then S_{2n} is a standard empirical process and so $O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\left\{\hat{\zeta} - \zeta_0\right\}$ is $O_p\left(\frac{1}{\sqrt{n}}\right)$, by theorem 2. ■