

Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets

Debopam Bhattacharya*

University of Oxford

Pascaline Dupas

Stanford University

Shin Kanaya

University of Aarhus

First version: February 19, 2012; This version: January 30, 2013

Abstract

Regular use of effective health-products such as insecticide-treated mosquito nets (ITN) by a household benefits its neighbors by (a) reducing chances of infection and (b) raising awareness about product-effectiveness, thereby increasing product-use. Due to their potential social benefits and high purchase price, causing free-riding and sub-optimal private procurement, such products may be subsidized in developing countries through means-testing. Owing to associated spillover effects, cost-benefit analysis of such subsidies requires modelling behavioral responses of both the subsidized household and its neighbors. Using experimental data from Kenya where subsidies were randomized, coupled with GPS-based location information, we show how to estimate aggregate ITN use resulting from means-tested subsidies in the presence of such spatial spillovers. Accounting for spillovers introduces infinite-dimensional estimated regressors corresponding to continuously distributed location coordinates and makes the inference problem novel. We show that even if individual ITN use unambiguously increases with increasing incidence of subsidy in the neighborhood, ignoring spillovers may over- or under-predict overall ITN use resulting from a specific targeting rule, depending on the resulting aggregate incidence of subsidy. Applying our method to the Kenyan data, we find that (i) individual ITN use rises with neighborhood subsidy-rates, (ii) under means-testing, predicted ITN use is a convex increasing function of the subsidy incidence and (iii) ignoring spillovers implies a nearly-linear increasing relationship leading to over-estimation of ITN use at lower and under-estimation at higher subsidy rates.

*Bhattacharya gratefully acknowledges financial support from the Leverhulme Trusts for partly funding this project.

1 Introduction

An important goal of statistical policy analysis is to estimate the distribution of outcomes that would result from targeted policy-interventions in society. These estimates can then be fed into the formulation of effective targeting rules. The formal analysis of such "treatment assignment" problems was pioneered by Manski (2004, 2007) and subsequently extended and expanded by other researchers. This line of research has assumed that the treatment is individual-specific and does not produce any externalities for other, untreated individuals. However, in several field experiments in a variety of contexts, it has been demonstrated empirically that providing a certain treatment to an individual household has significant spillover effects on other untreated households who reside in its vicinity. These spillovers can be either behavioral, i.e., where the behavior of the untreated household is affected by the neighbors' treatment, or mechanical, i.e., where the outcome of untreated households but not its behavior is affected by the neighbors' treatment. For example, in epidemiological studies on infectious diseases, the neighborhood incidence of vaccination has been shown to be an effective disease-deterrent beyond one's own vaccination status (see Root et al., 2011, on cholera vaccines in Bangladesh). Likewise, households in sub-Saharan Africa living within the vicinity of a village with high coverage of insecticide-treated bednets (ITN) are significantly less likely to get malaria (see Hawley et al., 2004). In Kenya, students attending schools where the majority of children underwent intestinal worm infection treatment as well as children living within 3 kilometers of a school where the majority of students were dewormed experienced significant health-gains, despite not getting the treatment themselves (Miguel and Kremer, 2004). In economic contexts, Angelucci and De Giorgi (2009) have shown that means-tested cash transfers create behavioral spillovers on ineligible households. On the other hand, Bandiera and Rasul (2006) demonstrated that farmers strategically delay the time they adopt a recommended farming technology in order to free ride on the knowledge generated by those in their neighborhood who experiment before them. In the context of ITN subsidies, Dupas (2012) has recently shown that the incidence of subsidies in the neighborhood has an independent effect on a household's decision to purchase and use an ITN over and above the effect brought about by the household's own receipt or non-receipt of a subsidy. The presence of such externalities and the high private procurement costs relative to earnings for these technologies and products typically lead to socially sub-optimal levels of private investment in them and creates a case for subsidizing them, at least for the poorest and most vulnerable sections of society.

The purpose of the present paper is to develop a method for estimating the predicted aggregate effect of a given subsidy-targeting rule, taking into account the spillover effects that one household's subsidization has on neighboring households' outcomes; and to estimate the error incurred in prediction due to ignoring the spillovers. A key requirement of the method we propose is the availability of data to estimate the magnitude and shape of spillovers. In our application, we exploit data from a randomized field experiment (discussed in Dupas, 2009) in which a subsidy for antimalarial bed nets was randomly assigned across Kenyan households. Randomized program assignment, which has become common practice among applied microeconomists, yields exogenous variation in individual treatment status. More importantly, for small enough neighborhoods, it also yields exogenous variation in neighborhood treatment rates. We propose a method for estimating own-treatment and indirect-treatment (spillover) effects from such experimental variation and for using these estimates to predict the effects of means-tested subsidization on aggregate product-procurement. It is important to note that here we do not aim to identify "social interaction" type behavioral effects which capture the impact of neighborhood product-use on a household's own product use. Instead, the key ingredient of our analysis is the more direct policy-relevant impact of subsidizing neighborhood households on a household's own product-use – both when it receives and when it does not receive the subsidy itself.¹

The key methodological innovation in our cost-benefit analysis is the incorporation of neighborhood effects where neighborhoods are defined via physical distance obtained through GPS location coordinates of households. Even when individual household responses are modeled via parametric models, incorporating such neighborhood effects introduces infinite-dimensional estimated regressors corresponding to continuously distributed location coordinates. This makes the inference problem non-trivial and novel. To solve this problem, we establish an appropriate Hadamard differentiability result which, coupled with a functional central limit theorem, establishes the necessary limiting distribution result. We further establish consistency of the nonparametric bootstrap in approximating this limiting distribution, thereby giving practitioners a user-friendly method for calculating confidence intervals for the parameters of interest. Proving bootstrap validity is non-trivial in this problem, because here one needs to prove bootstrap consistency for an underlying stochastic process before applying the functional delta method.

¹In principle, one could use the experiment-led variation in neighborhood subsidy incidence as an instrument for neighborhood product use and recover such social interaction effects. But this question is somewhat tangential to our goal.

Having developed our statistical methodology, we apply it to the problem of predicting aggregate ITN use resulting from means-tested subsidies. We show that ignoring treatment externalities in the estimation of aggregate policy impacts can yield large bias, and importantly, that the sign of this bias cannot be inferred solely from the sign of the externality. For example, when individual ITN use is increasing in neighborhood subsidy rates, as in our application, intuitive reasoning might suggest that ignoring this externality would lead to under-estimation of the aggregate impact of a targeted ITN subsidy program. However, this intuition is flawed and the correct answer depends on whether the average neighborhood subsidy rate under the proposed subsidy program would be higher or lower than the average neighborhood subsidy rate observed in the data used to estimate the parameters of interest.

To get some intuition behind this finding, consider a much simpler set-up where an outcome Y is related to a scalar covariate X via the classical linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ where ε is zero-mean, independent of X and $\beta_1 > 0$. OLS estimation of this model yields estimators $\hat{\beta}_1, \hat{\beta}_0$ with probability limits (and also expected values) $\beta_1 = \text{Cov}[X, Y] / \text{Var}[X]$ and $\beta_0 = E[Y] - \beta_1 E[X]$, respectively. Corresponding to a value x of X , the predicted outcome has its probability limit of $y^* := \beta_0 + \beta_1 x = E[Y] + \beta_1 \{x - E[X]\}$. Now consider what happens if one ignores the covariate X . Then the prediction is simply the sample mean of Y which has the probability limit of $y^{miss} := E[Y]$. Therefore, $y^* < y^{miss}$ if $x < E[X]$. Thus, although the ignored covariate X has a positive effect on the outcome (since $\beta_1 > 0$), ignoring it in prediction leads to an *overestimation* of the outcome if the point x where the prediction is made is smaller than the population average of the ignored covariate. On the other hand, if $x > E[X]$, then there will be under-estimation.

For a more general model, $Y = \beta_0 + X_1 \beta_1 + X_2 \beta_2 + \varepsilon$ with X_1, X_2 independent of ε and X_2 independent of X_1 , the OLS based prediction y^* at (x_1, x_2) will have the probability limit equal to $E[Y] + \{x_1 - E[X_1]\} \beta_1 + \{x_2 - E[X_2]\} \beta_2$. Ignoring X_2 leads to OLS prediction y^{miss} with its probability limit $E[Y] + \{x_1 - E[X_1]\} \tilde{\beta}_1$, where $\tilde{\beta}_1$ is the plim of the OLS coefficient of X_1 in a regression of Y on X_1 alone. The difference between the two predictions equals

$$\begin{aligned} y^{miss} - y^* &= \{x_2 - E[X_2]\} \beta_2 + (\beta_1 - \tilde{\beta}_1) \{x_1 - E[X_1]\} \\ &= \{x_2 - E[X_2]\} \beta_2 - \{x_1 - E[X_1]\} \text{Var}[X_1]^{-1} \text{Cov}[X_1, X_2] \beta_2 \\ &= \{x_2 - E[X_2]\} \beta_2, \end{aligned}$$

since $\text{Cov}[X_1, X_2] = 0$. This is the higher dimensional analog of the previous result. If the point of

prediction is higher relative to the population mean of the omitted covariates and the coefficients on the omitted variables are positive, then ignoring the covariates leads to an over-estimation of the predicted value. Note that in this case, the omitted covariate X_2 is orthogonal to the included covariate X_1 so that there is no omitted variable bias in the *coefficient estimate*, i.e., $\beta_1 = \tilde{\beta}_1$ and yet the *prediction* is biased.²

Literature review and substantive contributions: Several papers have recently appeared in the statistics and econometrics literature on formal analysis of treatment effects under *interference* – i.e., where treatment status of peers or neighbors affects an individual’s outcome over and above his/her own treatment status. See, for example, Rosenbaum (2007), Hudgens and Halloran (2009) and Manski (2010). A different strand of the literature has investigated the impact of alternative *peer assignment* policies on population outcome distribution when peer-effects are significant determinants of individual outcomes. See, for instance, Bhattacharya (2009), Graham (2011), Carrell et al (2012) and the references therein. These papers did not require a separate formulation of neighborhood effects in terms of distance, since peer-effects were taken to work within identifiable neighborhoods (such as classes, dorm-rooms or entire villages) and hence the set of peers was "closed". In contrast, in our present context of ITN subsidies, the relevant neighborhoods are defined in terms of physical distance, which is sensible in the context of health externalities or that of social learning (c.f., Conley and Udry, 2011), especially in communities like rural Africa where the relevant households are not concentrated in isolated villages but are located continuously in space.. Therefore, the relevant neighborhoods are overlapping (not closed) in the sense that we can generically have situations where a household (hhd) i and hhd j are neighbors, hhd j and k are neighbors but hhd i and k are not neighbors. This distance-based characterization of neighborhood effects leads to new challenges for formulation and inference in our analysis.

We note that our aim here is not to devise either an outcome-maximizing (c.f., Bhattacharya, 2009; Graham, 2011) or decision-theoretically optimal (c.f., Manski, 2004; Dehejia, 2005, Hirano and Porter, 2009; Stoye, 2008; Tetenov, 2010) treatment-targeting rule. Instead, we wish to calculate marginal and average returns for what is arguably the most common subsidy allocation rule in developing (and developed) countries, viz. means-testing, where only households below a certain income level qualify for the subsidy. The methodological innovation is to conduct this analysis in

²A similar argument applies if we predict the population-average outcome corresponding to a counter-factual change in the marginal distribution of the covariates. In this case, an over-estimation will occur if the mean of the ignored covariate under the counterfactual distribution is lower than its mean under the original distribution.

the presence of infinite-dimensional neighborhood effects identified experimentally. Therefore, our problem and its analysis are, to the best of our knowledge, unique relative to the existing literature on neighborhood effects in the analysis of treatment response and treatment assignment problems.

Last, but not least, our finding that ignoring neighborhood-externalities can lead to either under or over-estimation of the aggregate impact of treatment – even when raising neighborhood treatment rates unambiguously increase individual outcomes – appears to be novel. It highlights a statistical issue concerning the impact of omitted variables on *outcome predictions* which is very different from the well-known impact of omitted variables on *regression coefficients*. As such, this finding provides a new – and to our knowledge, heretofore unnoticed – cautionary note regarding the extrapolation of experimental results involving randomized treatment assignment to the prediction of aggregate outcomes in situations where the aggregate incidence of treatment might differ from that in the experiment.

2 The context and the experiment

Malaria is a life-threatening parasitic disease transmitted from human to human through mosquitoes. In 2010, malaria is estimated to have caused up to 1,200,000 deaths, mostly among African children (Murray et al., 2012). Despite major strides in malaria eradication in the early and mid-20th century, notably in the Americas, efforts to eradicate malaria worldwide were abandoned in the 1970s. Recently, efforts to control malaria transmission have rejuvenated with the introduction of highly effective prevention tools, insecticide treated bednets (ITNs). Regular use of an ITN is believed to reduce overall child mortality by around 18 percent and to reduce morbidity for the entire population (Lengeler, 2004). However, at \$6 or more a piece, ITNs are unaffordable for many poor households, and as of 2007, ITN coverage rates in most of sub-Saharan Africa were below 25%, way short of the 100% coverage recommended by the World Health Organization (WHO, 2008). Given this, an important policy question for governments in malaria-endemic areas of Africa is how much to subsidize ITNs, and for whom. Because the returns to malaria prevention are higher among children under five and pregnant women (Lengeler, 2004), and targeting based on such easily observable characteristics is relatively cheap, initial calls were made for subsidies to be targeted at these subgroups (Webster et al., 2007). But as argued in Killeen (2007), in the presence of externalities, it might actually be more cost-effective to subsidize ITNs for a larger share of the population than just children under five and pregnant women. Evidence on the size and shape

of these externalities remain scant, however, as few experimental studies to date were designed to estimate those.

Experimental design: We exploit data from a 2007 randomized ITN subsidy experiment conducted in eight rural markets of Western Kenya, where malaria is transmitted year-round. In each market area, a list of 150 to 200 households was compiled from school registers. Households were then randomly assigned to a subsidy level. After the random assignment had been performed in office, trained enumerators visited each sampled household to administer a baseline survey. At the end of the interview, the household was given a voucher for an ITN at the randomly assigned subsidy level. Vouchers could be redeemed within three months at participating local retailers. The subsidy level varied from 40% to 100%; there were 22 corresponding final prices faced by households, ranging from 0 to 300 Kenyan shillings (US\$4.60). In what follows, we group these subsidy levels into two categories: those who received a large subsidy (i.e., faced a price of 90 Ksh or below) and those who received a lower subsidy (i.e., faced a price of 100 Ksh or above). Given the liquidity constraints faced by the study population, these two subsidy categories indeed correspond to two very different treatments: 71% of households receiving a price at or below 90 Ksh acquired the ITN, compared to only 26% of those in the low subsidy group. Furthermore, there is a level drop in take-up of around 20 percentage points (from 55 to 35%) between the price points of 90 and 100 Ksh, possibly owing to the fact that other types of bednets (less effective but better known ones) are sometimes available in retail shops for the price of 100 Ksh.³

Data: Three separate but related sources of data are combined in our analysis. First, we use administrative data on voucher redemption to obtain our main outcome of interest: whether the household purchased an ITN. Second, we use data on background household characteristics measured during a baseline survey. The three main background characteristics we consider are wealth (the combined value of all durable and animal assets owned by the household); the number of children under 10 years old; and the education level of the female head of the household. Finally, we have data on the exact geolocation (latitude and longitude, measured with a GPS device) of all households in the sample. These data enable us to compute, for each household, the number and density of study households within a given radius. We can also compute the number and density of households with a given characteristic within that radius, in particular, the share of households below a certain poverty threshold, and the share of households who received a subsidy.

³See Dupas (2012, Figure 1) for the full demand curve.

3 Methodology

Stochastic specification: There are n randomly chosen households from the population and they are indexed by i which will run from 1 to n .

Let L_i denote household i 's location (such as a latitude-longitude pair). Let X_i denote other household covariates, of which the 1st component X_{1i} is wealth-holding. Let $F_{X,L}^*(\cdot, \cdot)$ denote the (true) joint cumulative distribution function (C.D.F.) of (X_i, L_i) in the population. Let Y_i denote the binary indicator of whether the household purchased an ITN, and D_i is a binary indicator of whether it had received a subsidy-voucher entitling it to buy an ITN for under 100 Ksh (about US\$1.50). Due to simple random sampling, our experimental dataset consists of n independent and identically distributed (I.I.D.) realizations of (Y_i, X_i, D_i, L_i) .

Finally, we denote by $S_i(r) := s^*(L_i, r)$ the fraction (rate) of households residing within the r neighborhood of household i (a radius of r from L_i) who got a subsidy. In general, several values of r (say, J values, r_1, r_2, \dots, r_J) will be used in our empirical exercise; so $s^*(L_i, r)$ should be interpreted as a generic component of the vector $(s^*(L_i, r_1), s^*(L_i, r_2), \dots, s^*(L_i, r_J))$ (we regard $s^*(L_i, \cdot)$ as a random function). We will sometimes suppress the dependence of s^* on r (or that of S_i on r) when there is no possibility of confusion.

Let $Y_i(1, s^*)$ denote a potential outcome (ITN purchase) of household i if it receives the subsidy and a fraction s^* of its neighbors also receive the subsidy. Similarly, let $Y_i(0, s^*)$ denote a potential outcome when i is not receiving the subsidy itself and a fraction s^* of its neighbors receives the subsidy. Implicit in this notation is the assumption that the treatment status of a household's neighbors affects its own outcome only through the aggregate fraction of subsidized households in the neighborhood. Define

$$\mu_1(x, s) := E[Y_i(1, s) | X_i = x]; \quad \mu_0(x, s) := E[Y_i(0, s) | X_i = x],$$

where s denotes a generic incidence of subsidy in the neighborhood (while the function with an asterisk, s^* , denotes the *true* one). Here the expectation is taken with respect to the population distribution, or, equivalently, the distribution induced by random sampling from the population. These quantities can be defined without reference to any subsidy allocation *mechanism*.

Now, random assignment of subsidies in the experiment implies that $\mu_1(x, s)$ and $\mu_0(x, s)$ can be consistently estimated by the sample counterparts

$$\hat{E}[Y_i | D_i = 1, S_i = s, X_i = x]; \quad \hat{E}[Y_i | D_i = 0, S_i = s, X_i = x],$$

respectively, if one observes the realizations of (Y_i, D_i, S_i, X_i) .⁴ However, realizations of S_i are not observed, since only a subset n of the population is sampled for measurement; realizations of S_i need to be estimated off of the observed sample, and hence we have a generated regressor problem. We estimate the above quantities by

$$\hat{E}[Y_i | D_i = 1, \hat{s}(L_i) = s, X_i = x]; \quad \hat{E}[Y_i | D_i = 0, \hat{s}(L_i) = s, X_i = x],$$

respectively, where

$$\hat{s}(l) = \hat{s}(l, r) := \frac{n^{-1} \sum_{k=1}^n 1 \{ \|l - L_k\| \leq r \} D_k}{n^{-1} \sum_{u=1}^n 1 \{ \|l - L_u\| \leq r \}}.$$

Parameters of interest: Our key parameter of interest is potential ITN use resulting from income-based, i.e., means tested subsidy. Toward this end, consider the binary treatment rule: giving a subsidy if $X_{1i} < \gamma$. The expected cost of this rule is $C(\gamma) = \int 1 \{x_1 < \gamma\} dF_{X_1}^*(x_1)$. On the other hand, under this subsidy rule, the neighborhood subsidy-rate for a generic household at a location l will be given by

$$s_\gamma^*(l, r) = \Pr[\|l - L_i\| \leq r \ \& \ X_{1i} \leq \gamma] / \Pr[\|l - L_i\| \leq r], \quad (1)$$

where $\|l - \lambda\| = \|l - \lambda\|_{\mathcal{L}} = \|\lambda - l\|_{\mathcal{L}}$ is an appropriate distance measure between l and λ on \mathcal{L} . Each L_i is defined via latitude-longitude coordinates in our dataset (\mathcal{L} is a set of latitude-longitude points on the earth sphere), and therefore, we work with the distance based on the Haversine formula.⁵ This distance is used for defining each household's neighborhood.

Finally, the expected ITN usage rate corresponding to this counterfactual subsidy-rule will be given by

$$U^*(\gamma) = \int [1 \{x_1 < \gamma\} \times \mu_1(x, s_\gamma^*(l, \cdot)) + 1 \{x_1 \geq \gamma\} \times \mu_0(x, s_\gamma^*(l, \cdot))] dF_{X,L}^*(x, l). \quad (2)$$

⁴We in particular note that due to randomized subsidy allocation, the following relationship holds:

$$\begin{aligned} E[Y_i(d, s) | X_i = x] &= E[Y_i(1, s) D_i + Y_i(0, s) (1 - D_i) | D_i = d, X_i = x] \\ &= E[Y_i | D_i = 1, S_i = s, X_i = x], \quad \text{for } d = 1, 0. \end{aligned}$$

⁵For simplicity, we develop our asymptotic theory under the assumption that the earth is a sphere and $\|l - \lambda\| = \|l - \lambda\|_{\mathcal{L}}$ measures the shortest distance between two points l and λ on the sphere surface. This simplifying assumption allows us to easily compute so-called covering numbers of functional sets. Note that for the Euclidean norm, we also use the same notation $\|\cdot\|$ in the sequel, which should cause no confusion.

If we ignore the externality, we would calculate, analogous to Bhattacharya and Dupas (2011),

$\bar{\mu}_1(x)$: Probability of the ITN use if an x type household gets a subsidy;

$\bar{\mu}_0(x)$: Probability of the ITN use if an x type household gets no subsidy.

One can identify $\bar{\mu}$'s from the experiment, in which subsidy allocation was randomized. Then the naive expected ITN usage rate – ignoring the externality – is given by

$$\bar{U}(\gamma) = \int [1 \{x_1 < \gamma\} \bar{\mu}_1(x) + 1 \{x_1 \geq \gamma\} \bar{\mu}_0(x)] dF_X^*(x). \quad (3)$$

Our goals are to estimate (2) and (3) and to calculate the difference between them for various counterfactual subsidy rule (namely, various eligibility thresholds γ). We now turn to the question of how to estimate them from the data and then derive statistical properties of these estimates which are needed to conduct inference.

Functional forms and estimation method: Now, we formally present our econometric framework and estimation method. To this end, we first re-formulate the fraction of subsidized households, $s^*(l, r)$, as follows. Let $F_{D,L}^*(\cdot, \cdot)$ denote the (true) joint distribution function of (D_i, L_i) induced in the population via the one-time experiment and let $F_L^*(\cdot)$ denote the marginal distribution function of L_i in the population. Then, write

$$\begin{aligned} s^*(l, r) &: = \frac{\Pr[||l - L_k|| \leq r \ \& \ D_k = 1]}{\Pr[||l - L_k|| \leq r]} \\ &= \frac{E[1 \{||l - L_k|| \leq r\} D_k]}{\kappa^*(l, r)} = \frac{\int \delta \times 1 \{||l - \lambda|| \leq r\} dF_{D,L}^*(\delta, \lambda)}{\int 1 \{||l - \lambda|| \leq r\} dF_L^*(\lambda)}, \end{aligned} \quad (4)$$

for any $r \in [\underline{r}, \infty)$ (with some constant $\underline{r} > 0$) and $l \in \mathcal{L}$, where $\kappa^*(l, r)$ is the probability that some household exists within the r neighborhood of l , i.e.,

$$\kappa^*(l, r) := \Pr[||l - L_k|| \leq r] = E[1 \{||l - L_k|| \leq r\}]. \quad (5)$$

In the above expression, probabilities and expectations are calculated with respect to the distribution corresponding to random sampling from the population, given the realizations of the subsidy experiment, and not with respect to infinite repetitions of the experiment itself. Thus, for a household located at a location l , $s^*(l, r)$ denotes the true fraction of subsidized (through the experiment) households in its r neighborhood.⁶ It is also important to note that $s^*(l, r)$ is *not* in the form of a *regression* (L_2 -projection) function.

⁶Thus although D_i is randomly assigned during the experiment, D_i and L_i are not independent with respect to the sampling distribution $F_{D,L}^*(\cdot, \cdot)$, given the realizations of the (one-time) experiment, by which the probabilities/expectations in (4) are computed.

Since the object $s^*(\cdot, \cdot)$ is required to calculate our interested parameters (as previously explained), we consider nonparametric estimation of this function (indexed by (l, r)) by

$$\hat{s}(l, r) := \frac{n^{-1} \sum_{k=1}^n 1 \{ \|l - L_k\| \leq r \} D_k}{n^{-1} \sum_{u=1}^n 1 \{ \|l - L_u\| \leq r \}}. \quad (6)$$

This is a moment based estimator corresponding to a solution of the moment equation $E[1 \{ \|l - L_k\| \leq r \} \times \{ D_k - s(l, r) \}] = 0$.⁷

Next, given the fraction of subsidized households $s^*(l, r)$, we model the probability of the ITN use for an (x, l) -type household with a treatment $d \in \{0, 1\}$ ($d = 1$ if subsidized, and $= 0$ otherwise) by

$$\Pr[Y_i = 1 | D_i = d, X_i = x, L_i = l, S_i = s] = \mu_d(x, s) = \Phi \left(x' \delta_d^* + \sum_{j=1}^J \beta_{j,d}^* s(r_j) \right), \quad (7)$$

where Φ is some cumulative distribution function of a continuously-distributed random variable (we choose a standard normal C.D.F. in our application); δ_d^* is a d_X -dimensional column vector $(\delta_{1,d}^*, \dots, \delta_{d_X,d}^*)'$ ($d = 0, 1$); and J is some finite integer and $\{r_j\}_{j=1}^J$ is a set of evaluation grid points picked by an econometrician (with $r_j \in [\underline{r}, \infty)$).

Remark 1 *Implicit in this formulation is the assumption that the location L_i itself has no impact on household decision to buy an ITN, given D_i , X_i and S_i .*

We estimate the parameters by the quasi maximum likelihood (QML) method. Here, we do not suppose that Φ is correctly specified but interpret it as a device for us to obtain a convenient and reasonable criteria for the estimation (see, e.g. discussions on the QML estimators in White, 1994), meaning that our asymptotic theory developed below is valid even under misspecification. To define our QML objective function, write

$$Z_i := (Y_i, X_i', D_i, L_i')'; \quad \boldsymbol{\eta} := (\delta_0', \delta_1', \beta_{1,0}, \dots, \beta_{J,0}, \beta_{1,1}, \dots, \beta_{J,1})'; \quad \mathbf{s}(\cdot) := (s(\cdot, r_1), \dots, s(\cdot, r_J))'.$$

In the sequel, $\boldsymbol{\eta}^*$ and $\mathbf{s}^*(\cdot)$ denote the (pseudo) true parameters (a $2(d_X + J)$ -dimensional vector of real numbers and a J -dimensional vector of functions, respectively), and $\hat{\boldsymbol{\eta}}$ and $\hat{\mathbf{s}}(\cdot)$ denote their

⁷Since we are only interested in $s^*(L_i, r)$, evaluated at $l = L_i$ (for each household i), we may use the following estimator of $s^*(L_i, r)$:

$$\hat{s}_i(r) := \frac{\sum_{1 \leq k \leq n; j \neq i} 1 \{ \|L_i - L_k\| \leq r \} D_k}{\sum_{1 \leq u \leq n; j \neq i} 1 \{ \|L_i - L_u\| \leq r \}}.$$

Since we can easily show that $\hat{s}(L_i, r) = \hat{s}_i(r) + O_{a.s.}(1/n)$ uniformly over i and r (under some weak conditions), whether we use $\hat{s}(L_i, r)$ or $\hat{s}_i(r)$ does not affect our asymptotic analyses, and therefore we subsequently work with the estimator (6).

estimators.⁸ Then, the objective function is defined as

$$l_n(\boldsymbol{\eta}, \hat{\mathbf{s}}(\cdot)) := n^{-1} \sum_{i=1}^n q(Z_i, \boldsymbol{\eta}, \hat{\mathbf{s}}(L_i)), \quad (8)$$

where

$$\begin{aligned} q(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i)) &: = D_i Y_i \ln[\Phi(X_i' \delta_0 + \sum_{j=1}^J \beta_{j,0} s(L_i, r_j))] \\ &\quad + D_i (1 - Y_i) \ln[1 - \Phi(X_i' \delta_0 + \sum_{j=1}^J \beta_{j,0} s(L_i, r_j))] \\ &\quad + (1 - D_i) Y_i \ln[\Phi(X_i' \delta_1 + \sum_{j=1}^J \beta_{j,1} s(L_i, r_j))] \\ &\quad + (1 - D_i) (1 - Y_i) \ln[1 - \Phi(X_i' \delta_1 + \sum_{j=1}^J \beta_{j,1} s(L_i, r_j))]. \end{aligned}$$

The (quasi) log-likelihood function l_n is a plug-in type, i.e., the true $\mathbf{s}^*(\cdot)$ is replaced with the preliminary (first-step) estimator $\hat{\mathbf{s}}(\cdot)$. We note that the (vector-valued) function $\mathbf{s}(\cdot)$ influences q only through its value at L_i , i.e., $\mathbf{s}(L_i) = (s(L_i, r_1), \dots, s(L_i, r_J))'$. Now, we can define the QML estimator $\hat{\boldsymbol{\eta}}$ as

$$\hat{\boldsymbol{\eta}} := \arg \sup_{\boldsymbol{\eta} \in \mathbb{R}^{2(d_X + J)}} l_n(\boldsymbol{\eta}, \hat{\mathbf{s}}(\cdot)). \quad (9)$$

To compute our object of interest, the probability of the ITN use defined in (2), we also need $s_\gamma^*(l, r)$ (defined in (1)) and the C.D.F. $F_{X,L}^*(x, l)$, which are nonparametrically estimated by

$$\begin{aligned} \hat{s}_\gamma(l, r) &: = \frac{n^{-1} \sum_{k=1}^n 1\{|l - L_k| \leq r\} \times 1\{X_{1,k} < \gamma\}}{n^{-1} \sum_{u=1}^n 1\{|l - L_u| \leq r\}} \quad \text{for each } (\gamma, r, l); \quad (10) \\ \hat{F}_{X,L}(x, l) &: = n^{-1} \sum_{i=1}^n 1\{X_i \leq x \ \& \ L_i \leq l\} \quad \text{for each } (x, l), \end{aligned}$$

respectively.

Given the estimated $\hat{\mathbf{s}}(\cdot, \cdot)$, $\hat{\boldsymbol{\eta}}$, $\hat{s}_\gamma(\cdot, \cdot)$ and $\hat{F}_{X,L}(\cdot, \cdot)$ (together with the formulation in (7)), we can now obtain an estimate of $U^*(\gamma)$ by simply replacing the true parameters with their estimates.

To formulate this and to facilitate our subsequent analyses, we define a $[0, 1]$ -valued functional Ψ :

$$\begin{aligned} \Psi(\boldsymbol{\eta}, \mathbf{s}_\gamma(\cdot), F_{X,L}(\cdot, \cdot) | \gamma) &: = \int \left\{ 1\{x_1 \geq \gamma\} \times \Phi\left(x' \delta_0 + \sum_{j=1}^J \beta_{j,0} s_\gamma(L_i, r_j)\right) \right. \\ &\quad \left. + 1\{x_1 < \gamma\} \times \Phi\left(x' \delta_1 + \sum_{j=1}^J \beta_{j,1} s_\gamma(L_i, r_j)\right) \right\} dF_{X,L}(x, l), \quad (11) \end{aligned}$$

where $\mathbf{s}_\gamma(\cdot) := (s_\gamma(\cdot, r_1), \dots, s_\gamma(\cdot, r_J))'$. In what follows, we use the notation

$$\boldsymbol{\rho}_\gamma := (\boldsymbol{\eta}, \mathbf{s}_\gamma(\cdot), F_{X,L}(\cdot, \cdot)), \quad (12)$$

⁸We often write a function or parameter with an asterisk to refer to the (pseudo) true one, that is, f^* , and \hat{f} with a hat to denote an estimator of f^* , while a function or parameter without an asterisk or hat, such as f , usually refers to a generic one.

a $(2d_X + 3J + 1)$ -dimensional vector of functions; analogously $\boldsymbol{\rho}^*$ and $\hat{\boldsymbol{\rho}}$ refer to the (pseudo) true parameter and the estimator, respectively. Ψ is a map: $\boldsymbol{\rho}_\gamma \mapsto u \in [0, 1]$, whose domain is formally defined in the Appendix before the statement of Lemma 5. Then, we can write our estimand and its estimator as

$$\begin{aligned} U^*(\gamma) &= \Psi(\boldsymbol{\rho}_\gamma^* \mid \gamma) = \Psi(\boldsymbol{\eta}^*, \mathbf{s}_\gamma^*(\cdot), F_{X,L}^*(\cdot, \cdot) \mid \gamma); \\ \hat{U}(\gamma) &= \Psi(\hat{\boldsymbol{\rho}}_\gamma \mid \gamma) = \Psi(\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma(\cdot), \hat{F}_{X,L}(\cdot, \cdot) \mid \gamma) \\ &= n^{-1} \sum_{i=1}^n \left\{ 1 \{X_{1,i} \geq \gamma\} \times \Phi \left(X_i' \hat{\delta}_0 + \sum_{j=1}^J \hat{\beta}_{j,0} \hat{s}_\gamma(L_i, r_j) \right) \right. \\ &\quad \left. + 1 \{X_{1,i} < \gamma\} \times \Phi \left(X_i' \hat{\delta}_1 + \sum_{j=1}^J \hat{\beta}_{j,1} \hat{s}_\gamma(L_i, r_j) \right) \right\}, \end{aligned} \quad (13)$$

respectively, where $\mathbf{s}_\gamma^*(\cdot) := (s_\gamma^*(\cdot, r_1), \dots, s_\gamma^*(\cdot, r_J))'$ and $\hat{\mathbf{s}}_\gamma(\cdot) := (\hat{s}_\gamma(\cdot, r_1), \dots, \hat{s}_\gamma(\cdot, r_J))'$. The integral with respect to $\hat{F}_{X,L}$ simply leads to an average as in (13) since $\hat{F}_{X,L}$ is the empirical distribution function (E.D.F.) from the observations $\{(X_i, L_i)\}_{i=1}^n$. We note that this estimator $\hat{U}(\gamma)$ can be seen as a *three-step* one: the first step is the estimation of the (nuisance) nonparametric component of $\mathbf{s}^*(\cdot)$; the second step is obtaining the estimate $\hat{\boldsymbol{\eta}}$ of the finite-dimensional parameter $\boldsymbol{\eta}^*$ whose estimation depends upon the estimate $\hat{\mathbf{s}}(\cdot)$ from the previous step; and the third step is computing an empirical average of quantities based on the second-step estimate $\hat{\boldsymbol{\eta}}$ as well as another nonparametric estimator $\hat{\mathbf{s}}_\gamma(\cdot)$ (or equivalently, computing a value of the functional Ψ with $(\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma, \hat{F}_{X,L})$ as an input).

For the case where the externality is ignored, we model $\bar{\mu}_d(x) := \Phi(x' \bar{\delta}_d)$, where $\bar{\delta}_d$ is a d_X -dimensional vector (for $d = 0, 1$), and consider the following (quasi maximum-likelihood) estimator of the parameter δ_d^{**} :

$$(\tilde{\delta}'_0, \tilde{\delta}'_1) := \arg \sup_{(\delta'_0, \delta'_1)' \in \mathbb{R}^{2d_X}} l_n((\delta'_0, \delta'_1, \mathbf{0})', \mathbf{0}),$$

where l_n is the likelihood function defined in (8) and $\mathbf{0}$ stands for a zero (column) vector (of some appropriate dimension). By using the functional defined in (11), we can also write the mis-measured usage and its estimate as

$$\bar{U}^*(\gamma) = \Psi(\bar{\boldsymbol{\eta}}^*, \mathbf{0}, F_{X,L}^*(\cdot, \cdot) \mid \gamma); \quad \tilde{U}(\gamma) = \Psi(\tilde{\boldsymbol{\eta}}, \mathbf{0}, \hat{F}_{X,L}(\cdot, \cdot) \mid \gamma),$$

respectively, where $\bar{\boldsymbol{\eta}}^* := (\bar{\delta}'_0, \bar{\delta}'_1, \mathbf{0})'$ and $\tilde{\boldsymbol{\eta}} := (\tilde{\delta}'_0, \tilde{\delta}'_1, \mathbf{0})'$. Under the null hypothesis that the externality does not matter for the ITN purchase decision, we can say that \tilde{U} is an estimator of the usage rate which incorporates the restriction due to the null hypothesis.

4 Distribution theory and methodological contributions

We are primarily interested in the asymptotic distribution of our estimator $\hat{U}(\gamma)$. Our derivation of this distribution is somewhat involved due to the complicated three-step form of $\hat{U}(\gamma)$. As discussed above, our object of interest $U^*(\gamma)$ may be regarded as a value of the functional Ψ evaluated at $\boldsymbol{\rho}_\gamma^* := (\boldsymbol{\eta}^*, \mathbf{s}_\gamma^*, F_{X,L}^*)$. And therefore, $\hat{U}(\gamma)$ is computed as a plug-in type estimator, a functional value of Ψ evaluated at estimates of parameters $\hat{\boldsymbol{\rho}}_\gamma := (\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma, \hat{F}_{X,L})$. Our strategy for deriving the asymptotic distribution of $\hat{U}(\gamma)$ relies on this formulation using the functional Ψ . That is, we first derive a joint functional central limit theorem (FCLT) for $\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma(\cdot)$ and $\hat{F}_{X,L}(\cdot, \cdot)$ to show that the (normalized) vector of the estimators weakly converges to a Gaussian process. Then we apply the functional delta method by showing that Ψ is Hadamard differentiable and computing the explicit form of the derivative, and thus obtain the \sqrt{n} asymptotic normality of $\hat{U}(\gamma) = \Psi(\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma(\cdot), \hat{F}_{X,L}(\cdot, \cdot) | \gamma)$. This approach works here because all three elements, viz., $\hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}_\gamma(\cdot)$ and $\hat{F}_{X,L}(\cdot, \cdot)$, have the same convergence rate. Combining the above steps then yields an asymptotic normality result for $\hat{U}(\gamma)$ which takes into account spatial externalities and that for $\bar{U}(\gamma)$ which does not. This result is presented as Theorem 1.

Before presenting our formal results, we state the conditions which we work with:

Assumption 1 $\{Z_i\}_{i=1}^n$ is a sequence of I.I.D. random vectors ($Z_i = (Y_i, X_i', D_i, L_i)'$), and $(X_i', \mathbf{s}^*(L_i))'$ is not linearly dependent almost surely.

Assumption 2 There exists some constant $C > 0$ such that

$$\inf_{(l,r) \in \mathcal{L} \times [\underline{r}, \infty)} \kappa^*(l, r) > C. \quad (14)$$

Assumption 3 (i) The support of X_i , \mathcal{X} , is some bounded set in \mathbb{R}^{dx} . (ii) Φ is the C.D.F. of the standard normal distribution.

Assumption 4 \mathcal{L} (the range of L_i) is a proper subset of the sphere and does not include any band/ring surrounding the sphere.

Assumption 5 For each $r \in \{r_1, \dots, r_J\}$, $s^*(l, r)$ and $\kappa^*(l, r)$ are continuous in $l (\in \mathcal{L})$. $F_{X,L}^*(x, l)$ is continuous in $(x, l) (\in \mathcal{X} \times \mathcal{L})$.

Assumption 2 is imposed to guarantee the uniform convergence of $\hat{\mathbf{s}}(l)$ and $\hat{\mathbf{s}}_\gamma(l)$ (over l). Since these are estimators of the ratio type (with the estimated $\kappa^*(l, r_j)$ in the denominator of $\hat{\mathbf{s}}(l)$ and

$\hat{\mathbf{s}}_\gamma(l)$), they are potentially unbounded if $\kappa^*(l, r)$ and its estimate are very close to zero. The assumption is useful to avoid the so-called random-denominator problem. From a practical view point, $\kappa^*(l, r)$ and its estimate would be very close to zero if r was so, but we can avoid this by setting the distance parameters $\{r_j\}$ away from zero. The boundedness of the support in (i) of Assumption 3 is a reasonable assumption in our context and greatly facilitates our formal proofs. As for (ii) of Assumption 3, we can also think of some other specification such as the logistic function. Again, we impose this condition for simplicity. Assumption 4 allows us to regard \mathcal{L} as some (bounded) subset of \mathbb{R}^2 (say, $\mathcal{L} \subset \mathbb{R}^2$), which is trivially satisfied since our data were collected in Western Kenya (only a part of the globe surface, not a "band"). By regarding $\mathcal{L} \subset \mathbb{R}^2$, we can define the continuity property of the functions in Assumption 5. This continuity condition may not necessarily be required, but it facilitates our proof by making a set of functions (indexed by (x, l)) smaller (see the condition (19.25) and the proof of Theorem 19.28 of van der Vaart, 1998).

Theorem 1 *Suppose that Assumptions 1-5 hold. Then,*

$$\sqrt{n} \begin{pmatrix} \hat{U}(\gamma) - U^*(\gamma) \\ \tilde{U}(\gamma) - \bar{U}^*(\gamma) \end{pmatrix} \xrightarrow{d} N(0, \Sigma(\gamma)),$$

for each γ , where $\Sigma(\gamma)$ is a 2-by-2 variance-covariance matrix (for the derivation of the explicit form of $\Sigma(\gamma)$, see the proof of theorem in the Appendix).

The asymptotic normality result of Theorem 1 allows us to implement econometric inference provided we can consistently estimate the asymptotic variance matrix, which takes a complicated form, as is apparent from the proof of Theorem 1 (see the Appendix), and is thus not easy to estimate. Therefore, we resort to bootstrap-based inference. We use the nonparametric (I.I.D.) bootstrap and show that it consistently estimates the asymptotic distribution of $\sqrt{n}(\hat{U}(\gamma) - U^*(\gamma), \tilde{U}(\gamma) - \bar{U}^*(\gamma))'$. This assertion is formally stated and proved via Theorem 2.

Let $\{Z_i^\#\}_{i=1}^n$ be bootstrap data drawn randomly with replacement from $\{Z_i\}_{i=1}^n$ ($Z_i = (Y_i, X_i', D_i, L_i')$), and let $(\hat{U}^\#(\gamma), \tilde{U}^\#(\gamma))$ be an estimator of $(U^*(\gamma), \bar{U}^*(\gamma))$ based on the bootstrap data (we subsequently use a superscript $\#$ to denote a bootstrapped (resampled) variable or object).

Theorem 2 *Suppose that Assumptions 1-5 hold. Then, given $\{Z_i\}_{i=1}^n$, $\sqrt{n}(\hat{U}^\#(\gamma) - \hat{U}(\gamma), \tilde{U}^\#(\gamma) - \tilde{U}(\gamma))'$ (conditionally) weakly converges to $N(0, \Sigma(\gamma))$ in probability as $n \rightarrow \infty$.⁹*

⁹For the notion of this convergence, see the definition (3.9.9) in van der Vaart and Wellner (1996, p. 378). In subsequent proofs, we use $\xrightarrow[\#]{P}$ to denote this convergence.

This theorem provides the justification of the bootstrap-based inference, i.e., we can obtain an approximation of the limit distribution by the distribution of bootstrap estimates.¹⁰

These theorems are proved in the Appendix, using the following sequence of steps (here we outline the steps for $\hat{U}(\gamma)$ only; the analogous ones for $\tilde{U}(\gamma)$ are much simpler because they do not involve the spatial externalities):

1. Lemma 1 establishes the asymptotically linear forms of the (normalized) nonparametric estimators, $\sqrt{n}[\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)]$ and $\sqrt{n}[\hat{\mathbf{s}}_\gamma(l) - \mathbf{s}_\gamma^*(l)]$.
2. By using the pathwise derivative method (Newey, 1994), we quantify effects on $\hat{\boldsymbol{\eta}}$ of preliminary estimation of \mathbf{s}^* (calculate a so-called *correction* term).
3. Lemmas 2 and 3 verify an asymptotic linear expression of the QML estimator

$$\hat{\boldsymbol{\eta}} = \arg \sup_{\boldsymbol{\eta} \in \mathbb{R}^{2(d_X+J)}} l_n(\boldsymbol{\eta}, \hat{\mathbf{s}}(\cdot))$$

defined in (9) (this expression reflects the correction due to estimating \mathbf{s}^* obtained in the previous step).

4. Combining the previous steps, one then obtains the asymptotically linear uniform expression of the (stacked) vector of the normalized estimators:

$$\sqrt{n}[\hat{\boldsymbol{\rho}}_\gamma(x, l) - \boldsymbol{\rho}_\gamma^*(x, l)] \equiv \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \hat{\mathbf{s}}_\gamma(l) - \mathbf{s}_\gamma^*(l) \\ \hat{F}_{X,L}(x, l) - F_{X,L}^*(x, l) \end{pmatrix} = \hat{\mathbf{G}}_\gamma(x, l) + o_p(1), \quad (15)$$

for each γ , in terms of an empirical process $\hat{\mathbf{G}}_\gamma$ (defined in (55) in the Appendix).

Lemma 4 then proves a functional central limit theorem for the process $\{\hat{\mathbf{G}}_\gamma(x, l)\}_{(x,l) \in \mathcal{X} \times \mathcal{L}}$, which forms the basis of our inference and makes our analysis different from previous analyses of multi-step estimation methods owing to the non-standard form of $\hat{\mathbf{s}}$.

5. Lemma 5 then establishes that $U^*(\gamma)$ is a Hadamard differentiable functional Ψ of $(\boldsymbol{\eta}^*, \mathbf{s}_\gamma^*, F_{X,L}^*)$.
6. Steps 4 and 5 are then combined to establish Theorem 1.

¹⁰We estimate the limit distribution by a simulation. Simulation errors may be made arbitrarily small by letting the number of simulations large enough, and therefore we do not evaluate such errors, following the standard manner in the literature.

7. Finally, the above steps are combined with two auxiliary results in Lemmas 6 and 7 to obtain Theorem 2. The key step for establishing the bootstrap validity is to show the asymptotic uniform negligibility of

$$\sqrt{n}[\hat{\rho}_\gamma^\#(x, l) - \hat{\rho}_\gamma(x, l)] - \hat{\mathbf{G}}_\gamma^\#(x, l), \quad (16)$$

where $\hat{\rho}_\gamma^\#(x, l)$ and $\hat{\mathbf{G}}_\gamma^\#(x, l)$ are bootstrap counterparts of $\hat{\rho}_\gamma(x, l)$ and $\hat{\mathbf{G}}_\gamma(x, l)$, respectively ($\hat{\mathbf{G}}_\gamma^\#(x, l)$ is formally defined in (69) in the proof of Theorem 2). Note that $\sqrt{n}[\hat{\rho}_\gamma^\#(x, l) - \hat{\rho}_\gamma(x, l)]$ is a bootstrap counterpart of the normalized (actual) estimator $\sqrt{n}[\hat{\rho}_\gamma(x, l) - \rho_\gamma^*(x, l)]$, and thus the negligibility of (16) states that the expression as (15) holds true also in the bootstrap world. To verify this negligibility of (16), we exploit U-statistic-like expressions of estimators, as well as explicit expressions of bootstrapped estimators using multinomial random variables (see the proof of Lemma 6), where the multinomial distribution arises from the nonparametric bootstrap procedure.

Relation of our derivations with the existing literature: Our derivations, outlined above and proved rigorously in the Appendix, may be contrasted with distribution theory for (i) two-step semiparametric estimators (e.g., Newey, 1994; Chen, Linton and van Keilegom, 2003; Mammen, Rothe and Schienle, 2012) and (ii) three-step semiparametric estimators (e.g., Hahn and Ridder, 2012).

The two key points of difference in our set-up, relative to other papers in the relevant literature, are that (i) our first step-estimators $\hat{\mathbf{s}}(\cdot)$ and $\hat{\mathbf{s}}_\gamma(\cdot)$, as is clear from the definitions (6) and (10), do not involve kernel (local) or series-based estimators, and consequently (ii) every component of our second-step stochastic process, viz., $[\hat{\rho}_\gamma(\cdot, \cdot) - \rho_\gamma^*(\cdot, \cdot)]$, has a \sqrt{n} -rate of convergence. Hence our proof of the final step can be based on the functional CLT coupled with Hadamard differentiability arguments. This makes the derivation different from other two or three-step estimators in the literature.

The two key technical contributions in our derivations are (a) calculating the correction term, which quantifies the impact of the estimation of $\hat{\mathbf{s}}(\cdot)$ on the second-step parametric estimator $\hat{\eta}$ (see equations in (33) and (34), and Lemma 3), and (b) establishing bootstrap consistency for our final, third step estimator (Theorem 2). As is well-known from e.g., Newey (1994), the effect of first-step estimators of infinite-dimensional parameters may be calculated via pathwise derivatives leading to a correction term in the asymptotic linearization of the second-step estimator. But in general, one needs to find these correction terms on a case by case basis, as in the calculations

for semiparametric efficiency bounds. The calculation of the corrections term here is non-trivial and involves a clever use of the law of repeated conditional expectations, which is purportedly of independent interest (see Remark 2 in the Appendix).

As regards bootstrap validity, to our knowledge, ours is the first result of bootstrap consistency for three-step estimators. For *two-step* ones, Chen, Linton and Keilegom (2003) have shown bootstrap consistency based on an entropy-based technique. However, for our proof for the three-step estimators, a corresponding "entropy condition" is difficult to verify since the nonparametric estimators $\hat{\mathbf{s}}(\cdot)$ and $\hat{\mathbf{s}}_\gamma(\cdot)$ are constructed with indicator functions and thus are discontinuous and non-smooth.¹¹ Therefore, we develop independent proofs based on properties of the multinomial distribution induced by the bootstrap resampling, as mentioned above. Thus it may be fair to say that our results and proofs complement existing results and proofs on multi-step estimators in the semiparametric literature.

5 The application and empirical results

We now apply our methods to data collected via the experiment described above. See Section 2 for descriptions of the context and details on the design of the experiment generating the data.

The outcome we consider is whether the household acquires the ITN. As shown in Dupas (2012), this is a very good proxy for ITN coverage (what, ultimately, the social planner that subsidizes ITNs cares about) since ITN usage conditional on ownership is high, reaching 90% within 12 months.

As regressors X_i , we include household wealth (represented by the total value of the durable and animal assets reported by the household at a baseline survey), the number of children under the age of ten years and years of education of the oldest female member of the household. This choice of regressors is motivated by previous studies on determinants of ITN use in Kenya (c.f., Bhattacharya and Dupas, 2011 and Dupas, 2012).

In order to take into account the neighborhood effects, we include as additional regressors the fraction of households who received a large subsidy and reside within 250 meters, 500 meters and 1000 meters respectively of a given household. In terms of the notation of section 3, we included (estimated counterparts of) the three additional regressors $s^*(L_i, 250)$, $s^*(L_i, 500)$ and $s^*(L_i, 1000)$, in addition to X_i . Table 1 reports summary statistics on these regressors, as well as

¹¹We could work with some smoothed versions of $\hat{\mathbf{s}}(\cdot)$ and $\hat{\mathbf{s}}_\gamma(\cdot)$, but such smoothing complicates our estimation procedure, requiring some smoothing parameter choice.

on the randomized treatment dummy (= 1 if the household received a large subsidy voucher) and the outcome of interest (a dummy equal to 1 if the household purchased the ITN).

Tables 2A and 2B report the determinants of buying an ITN as a function of household covariates, estimated using the experimental data where subsidy allocation was randomized. Table 2A reports the probit coefficient estimates corresponding to subsidized households, i.e., $\hat{\delta}_1, \hat{\beta}_1$, and Table 2B reports probit coefficients for the unsubsidized ones, i.e., $\hat{\delta}_0, \hat{\beta}_0$ in (9). The standard errors, which adjust for the first-step estimation of the \hat{s} 's, are calculated using the bootstrap. The justification of the bootstrap follows from lemma 6 and 7 proved in the appendix (c.f., step 7 in section 4, above).

It can be seen from these tables that the covariates have somewhat different impacts on the decision to purchase a net depending on whether the household itself received the subsidy. The wealth variable is a significant predictor of ITN purchase among households not receiving the large subsidy but an insignificant one among heavily subsidized households. Similarly, the log-likelihood test statistic for testing the joint significance of the three neighborhood variables yields a chi-square test statistic of 31.97 (p-value below 0.0001) when the households received the subsidy themselves and a test statistic value of 5.40 (p-value = 0.145) when they did not.¹²

These findings are intuitive because the ITN is expensive relative to average earnings in our population of interest. Consequently, it is rare for an unsubsidized household to buy an ITN, no matter what the neighborhood incidence of subsidy is, unless it is relatively wealthy. However, upon receiving the subsidy, a household has the option of deciding whether to buy an ITN and it

¹²The QML objective function in (8) is constructed for obtaining the estimates of two sets of parameters, $(\hat{\delta}_1, \hat{\beta}_1)$ and $(\hat{\delta}_0, \hat{\beta}_0)$ *jointly*, i.e. $\hat{\boldsymbol{\eta}} = (\hat{\delta}'_0, \hat{\delta}'_1, \hat{\beta}'_{1,0}, \dots, \hat{\beta}'_{J,0}, \hat{\beta}'_{1,1}, \dots, \hat{\beta}'_{J,1})' = (\hat{\delta}'_0, \hat{\delta}'_1, \hat{\beta}'_0, \hat{\beta}'_1)'$. While this definition of the objective function facilitates our asymptotic analysis, we here calculate these two sets *separately*. That is, we use $n^{-1} \sum_{i=1}^n q_0(Z_i, \boldsymbol{\eta}, \hat{\mathbf{s}}(L_i))$ as an objective for $(\hat{\delta}_0, \hat{\beta}_0)$ (q_0 is the sum of the first two terms of q), and $n^{-1} \sum_{i=1}^n q_1(Z_i, \boldsymbol{\eta}, \hat{\mathbf{s}}(L_i))$ for $(\hat{\delta}_1, \hat{\beta}_1)$ (q_0 is the sum of the last two terms of q). We note that these joint and separate estimation procedures yield numerically the same result. This is because the objective function q consists of two parts, one only relevant for (δ_0, β_0) and the other for (δ_1, β_1) , and there is no interaction between two sets of the parameters.

We also note the two chi-square test statistic values reported here are based on two separate QML procedures. It is known that in a standard setting with no generated regressor, a likelihood ratio statistic is asymptotically chi-square distributed. This also holds true in our case with generated regressors $\hat{\mathbf{s}}(L_i)$. For the asymptotic chi-square distribution result of a likelihood ratio statistic, what is important is asymptotic normality of the estimator. Since we have verified the asymptotic normality of the second step estimator (Lemma 4), we can as well prove the asymptotic result for the test statistic (proof arguments are quite standard as in Section 9 of Newey and McFadden, 1994, and we omit details).

is only then that neighborhood effects impact this decision.

Table 3 reports the estimated impact of means-tested subsidies, first including and then excluding the neighborhood effects. The estimates are reported for each of 20 different choices of wealth thresholds γ for means-testing. Columns 6 and 7 report bootstrapped pointwise 95% equal-tailed confidence intervals for estimates of the difference $U^*(\gamma) - \bar{U}^*(\gamma)$, as we vary the poverty line γ such that households with income below γ would receive the subsidy. The theoretical justification of the bootstrap comes from our Theorem 2.

The entries in Table 3 may be interpreted as follows. For example, row 10 in Table 3 (highlighted in bold font) shows the expected impact if households with wealth no more than 10,000 Kenyan shillings are made eligible for the subsidy. The estimated fraction of such households is about 36% of the total population of households. An estimate of the average ITN ownership rate resulting from this rule is 0.396 if externalities are taken into account and 0.431 if externalities are ignored. This extent of overestimation is seen to be associated with a 95% confidence interval $(-0.049, -0.019)$ which does not include zero and thus suggests statistically significant over-estimation. The reason why we see such overestimation is that the average neighborhood rate in the experiment used for the estimation was significantly higher (43%) than under the means-tested subsidy considered in this case (36%), c.f., our discussion in the Introduction. This means that when the positive externality is not properly accounted for, the aggregate effect of the subsidy on subsidy-recipients is *over-*, rather than *under-*, estimated.

Figure 1 summarizes the overall findings from Table 3. The horizontal axis represents column 2 of Table 3, showing the fraction subsidized under each threshold used for means-testing. The light grey dashed line in the background is the 45-degree line, the solid black line represents the predicted fraction – including the neighborhood effects – of households who will purchase an ITN. Finally, the thick dashed line represents the predicted take-up rate if the neighborhood effects are ignored. It can be seen from Figure 1 that ignoring externalities leads to overestimation of ITN coverage at low subsidy incidence rates and to underestimation at higher rates. When the incidence rate equals that in the experiment (about 0.43), the lines corresponding to including or ignoring externalities coincide and there is no bias. This is the point discussed earlier. This feature is illustrated in Figure 2 where we plot the difference between the predicted usage including externality and the one ignoring the externality together with the corresponding (pointwise) 95% confidence intervals.

Some policy-relevant conclusions from Table 3 may be stated in terms of average and marginal returns on the subsidy. We show these calculations in Table 4 where we list the average and marginal

returns to the incidence of subsidy. The average return is the ratio of the predicted take-up rate (column 3 of Table 3) to the share of the population eligible for the subsidy (column 2 of Table 3). The (approximate) marginal return is defined to be the ratio of the difference in (predicted) ITN ownership rates to the difference in subsidy incidence rate. For example, row 10 in Table 4 suggests that when the 36% poorest households of the population are eligible for the large subsidy, the average return is 110%; while the marginal return is 81% when the spatial externality is taken into account, and 51% when it is ignored. For the most part, ignoring the positive externalities via neighborhood effects leads to under-estimation of marginal returns. The overall pattern of average and marginal returns are plotted in Figure 3. Because a fixed share of households acquires the ITN even if not eligible for the large subsidy, average returns appear relatively high at low levels of subsidy, and steadily decline as the subsidy rate increases. For example, subsidizing the poorest ten percent households leads to expected ITN ownership by about 24 percent of all households; subsidizing the poorest 55 percent leads to average use of just around 55 percent. However, the marginal returns to subsidy, i.e., the increase in ITN ownership corresponding to an increase in subsidy-incidence rates, is increasing in the subsidy rate. For example, when subsidizing the poorest ten percent of households via means-testing, increasing the subsidy incidence rate by one percentage point leads to increase in expected ITN ownership by about 0.5 percentage points; when subsidizing the poorest 55 percent, increasing the subsidy-rate by one percentage point leads to increase in expected ITN ownership by about 0.8 percentage points.

The third feature, shown graphically in Figure 4 is that the marginal returns, taking into account the externalities, tend to be higher than those obtained ignoring externalities, except at very low (less than 15%) incidence-rates. The latter curve is almost flat in contrast to the former, which is generally increasing due to the positive feedback effects coming through neighborhood effects. This feature corresponds to the convexity of the solid black line and the linearity of the thick dashed grey line in Figure 1. The difference between these two curves shown by the dashed line in Figure 4 can be interpreted as the impact of the externality.

6 Conclusion

In this paper we have addressed the problem of predicting the aggregate effect of treatment-targeting policies in a population where spatial externalities are present, i.e., where treating some households (e.g., whose income is lower than a threshold) affects the behavior of households living in close vicin-

ity of the treated households. Our set-up is quite general in that we do not assume that all of the spillover takes place within a closed and identifiable locality (e.g., classroom/village/school/hospital ward) but not across localities – as is often assumed in empirical studies of spillover effects.¹³ In other words, our set-up allows for households i, j, k where i, j are neighbors, j, k are neighbors but i, k are not neighbors. This distance-based notion of neighborhood is the relevant one in contexts like ours, involving the spatial dissipation of infection or of social learning when the relevant households are not concentrated in isolated villages/hamlets but are located continuously in space. However, allowing for such distance-based, overlapping neighborhoods in the statistical analysis gives rise to an infinite dimensional estimated regressor – viz., the fraction of treated households in a neighborhood around a (continuously varying) location coordinate. We develop the relevant theory of inference which (i) accommodates such predicted, infinite dimensional regressors, and (ii) establishes bootstrap consistency for the resulting three-step semiparametric estimator. On a more substantive note, we point out that ignoring spatial spillovers may, statistically, over or underestimate the aggregate effect of targeted subsidy regimes even when the sign of the spillover effects is unambiguously positive. This, somewhat counter-intuitive, conclusion obtains from discrepancies between aggregate incidence of subsidies under the proposed regime and that under the experiment used to obtain estimates of the spillover effects.

The methods developed above are applicable to other estimation situations where there are spatial externalities and where the set of units generating and experiencing the spillovers is not closed or, equivalently, “neighborhoods” are overlapping. The availability and use of GPS based location data, an important ingredient of our analysis, are becoming increasingly common in quantitative social sciences. In some contexts, such as epidemiological studies, actuarial projections and social network analysis, such data are indispensable. Our methods, therefore, are likely to have applicability to many other empirical contexts, beyond the important policy-question of predicting welfare effects of health-product subsidies.

Turning to the specific application in this paper, our results contribute to the ongoing debate regarding ITN subsidization. Our findings above suggest that the marginal impact of subsidies is generally increasing and remains fairly high even when nearly half the population receives a subsidy. Most of this increasing marginal effect is brought about by behavioral externalities whereby high

¹³To be clear, we also define the neighborhood of a household as the collection of all households which reside within a fixed distance from it. However, this definition makes the neighborhoods overlapping and does not partition the population of households into disjoint and collectively exhaustive "equivalent classes" of neighborhoods.

subsidy levels in the neighborhood lead to higher ITN take-up, conditional on a household's own subsidy-status. The marginal effect stops increasing beyond the point at which 50% of households become eligible for the subsidy but remains positive throughout. In the light of these findings and (a) given that the World Health Organization recommends that every community members sleeps under an ITN,¹⁴ and (b) assuming means-testing remains politically feasible and practically implementable, there is a strong case, based on behavioral externalities alone, for extending the subsidy to at least the poorest 50% of the rural Kenyan population. If there are in addition some epidemiological externalities (e.g. reduced malaria transmission probabilities when coverage is high enough, as suggested in Hawley et al., 2003), the marginal returns in terms of public health may continue increasing well beyond a subsidy incidence rate of 50%.

¹⁴See <http://www.who.int/malaria/publications/atoz/itnspospaperfinal.pdf>

Figure 1

**Estimated aggregate outcome by share of the population subsidized,
with and without taking into account the externality**

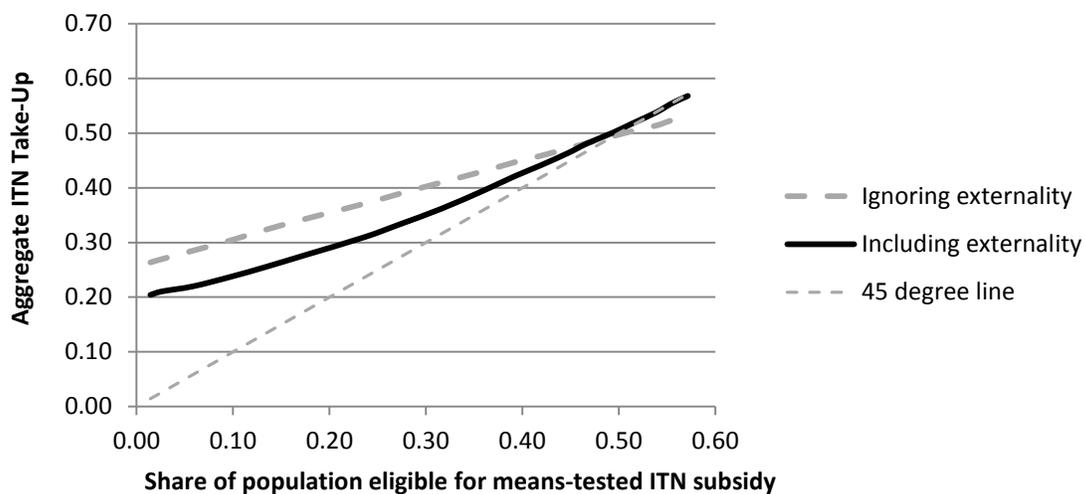
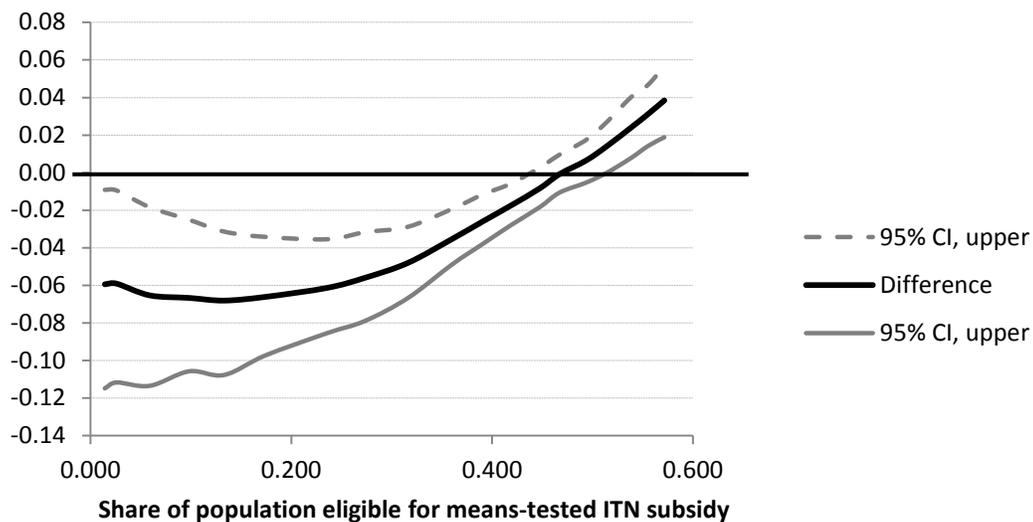


Figure 2

**95% Confidence interval for the difference between the two predictions
(with and without externality)**



Note: The share subsidized in the experiment that generated the data used for estimation was 43% (see Table 1).

Figure 3
Average and Marginal Return, by share subsidized

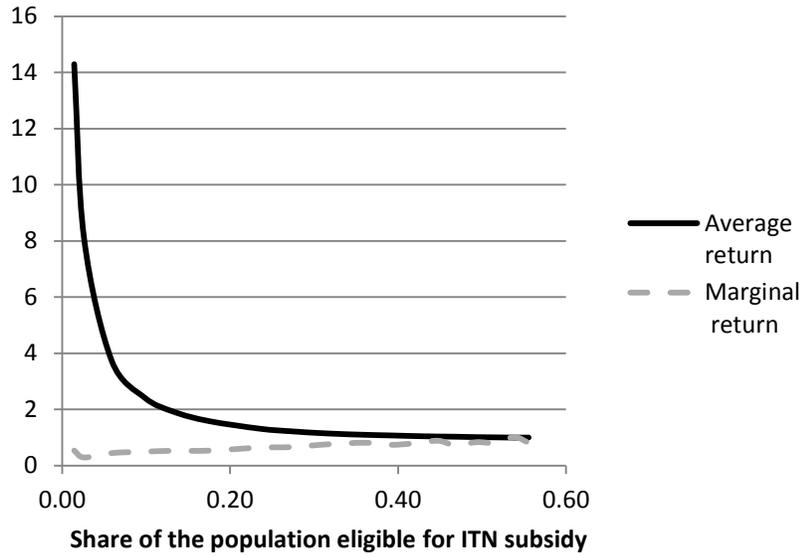
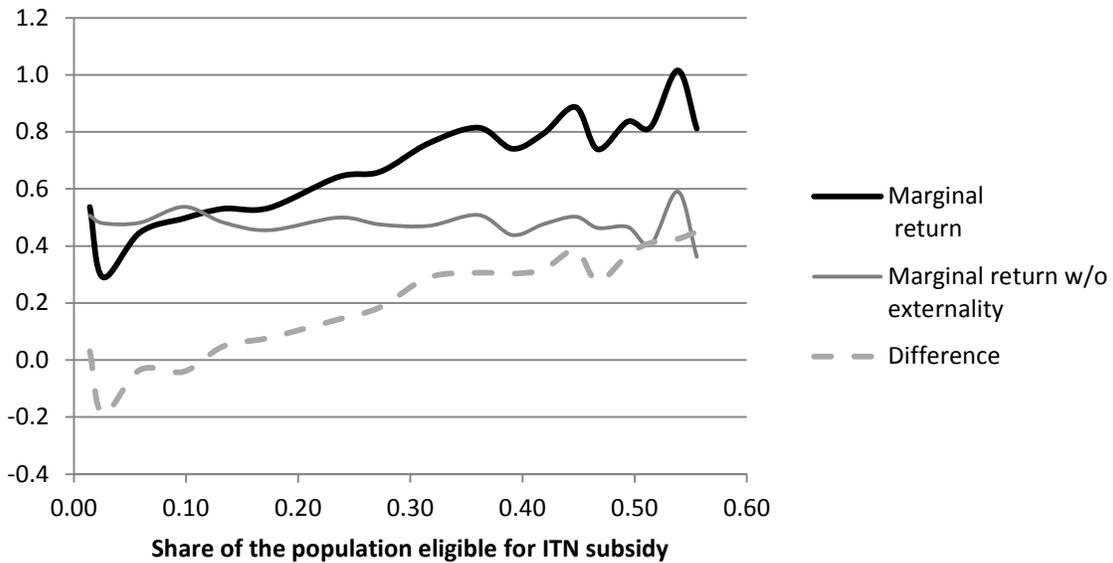


Figure 4
Difference in estimated marginal returns, with and without externality



Notes: The average return is the ratio of the predicted take-up rate (col. 3 of Table 3) to the share of the population eligible for the subsidy. The (approximate) marginal return is the ratio of the change (between two rows) in predicted take-up rate to the change (between two rows) in the share eligible for the subsidy.

Table 1: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
<i>Regressors</i>					
Household wealth ('000 Ksh)	1094	21.98	21.12	0.00	1727.28
# of children 10 or younger	1094	5.46	2.82	0.00	35.00
Years of education of female head	1060	4.94	3.84	0.00	15.00
Share of subsidy recipients within 1000m	1094	0.43	0.22	0.00	1.00
Share of subsidy recipients within 500m	1094	0.41	0.25	0.00	1.00
Share of subsidy recipients within 250 m	1094	0.38	0.32	0.00	1.00
<i>Randomized Treatment status</i>					
Received large subsidy (price <100 Ksh)	1094	0.43	0.49	0.00	1.00
<i>Outcome of Interest</i>					
Household purchased ITN	1094	0.46	0.50	0.00	1.00

Note: Sample of Kenyan households surveyed in 2007.

Table 2A: Determinants of ITN Take-up among households that received large subsidy

	Coef.	Std. Err.	t-stat
Household wealth ('000 Ksh)	0.0001	0.0003	0.30
# of children 10 or younger	-0.0010	0.0229	-0.04
Years of education of female head	0.0018	0.0160	0.11
Share of subsidy recipients within 1000m	1.8720	0.5170	3.62
Share of subsidy recipients within 500m	-0.3500	0.4770	-0.73
Share of subsidy recipients within 250 m	0.3630	0.2620	1.39

Notes: Probit estimates. Dependent Variable: Household purchased ITN. N=486, Pseudo-Rsq= 0.06. p-value for joint-significance of three neighborhood covariates<0.0001

Table 2B: Determinants of ITN Take-up among households that did not receive large subsidy

	Coef.	Std. Err.	t-stat
HH wealth	0.0050	0.0020	2.50
# of children 10 or younger	0.0033	0.0215	0.15
Years of education of female head	0.0115	0.0150	0.77
Share of subsidy recipients within 1000m	0.0180	0.4600	0.04
Share of subsidy recipients within 500m	0.3260	0.4620	0.71
Share of subsidy recipients within 250 m	0.1810	0.2470	0.73

Notes: Probit estimates. Dependent Variable: Household purchased ITN. N=634, Pseudo-Rsq= 0.02. p-value for joint-significance of three neighborhood covariates=0.145

Table 3: Impact Estimates

(1)	(2)	(3)		(4)	(5)	(6)		(7)
Poverty Line (Υ) (in '000s Ksh)	Share of the population eligible for ITN subsidy	Estimate of Aggregate ITN Take-up...		... ignoring externality	Difference	Bootstrapped Confidence Interval for Difference Estimate		
		...including externality				95% CI, lower		95% CI, upper
1	0.014	0.2042		0.2636	-0.059	-0.115		-0.009
2	0.026	0.2104		0.2695	-0.059	-0.112		-0.009
3	0.059	0.2200		0.2853	-0.065	-0.113		-0.018
4	0.098	0.2376		0.3042	-0.067	-0.106		-0.025
5	0.133	0.2549		0.3230	-0.068	-0.108		-0.031
6	0.174	0.2767		0.3428	-0.066	-0.097		-0.034
7	0.236	0.3095		0.3708	-0.061	-0.085		-0.035
8	0.273	0.3336		0.3896	-0.056	-0.079		-0.032
9	0.316	0.3619		0.4099	-0.048	-0.067		-0.029
10	0.361	0.3958		0.4309	-0.035	-0.049		-0.019
11	0.391	0.4205		0.4464	-0.026	-0.038		-0.012
12	0.419	0.4410		0.4585	-0.017	-0.028		-0.006
13	0.447	0.4637		0.4721	-0.008	-0.018		0.003
14	0.467	0.4811		0.4820	-0.001	-0.011		0.010
15	0.494	0.5009		0.4944	0.007	-0.005		0.018
16	0.514	0.5181		0.5039	0.014	0.000		0.027
17	0.538	0.5378		0.5137	0.024	0.008		0.040
18	0.555	0.5550		0.5237	0.031	0.014		0.047
19	0.571	0.5680		0.5296	0.038	0.019		0.057
20	0.598	0.5913		0.5435	0.048	0.026		0.068

Note: The poverty line is the threshold household wealth below which the household is eligible for the subsidy.

Table 4: Average and (approximate) marginal returns

(1)	(2)	(3)	(4)	(5)	(6)
Poverty Line (Y) (in '000s Ksh)	Share of the population eligible for ITN subsidy	Average return	Marginal return	Marginal return w/o externality	Difference
1	0.014	14.294	0.536	0.506	0.0305
2	0.026	8.127	0.290	0.480	-0.1894
3	0.059	3.734	0.447	0.481	-0.0345
4	0.098	2.419	0.497	0.537	-0.0406
5	0.133	1.916	0.530	0.483	0.0477
6	0.174	1.589	0.533	0.455	0.0780
7	0.236	1.313	0.642	0.499	0.1423
8	0.273	1.221	0.660	0.475	0.1853
9	0.316	1.145	0.760	0.470	0.2891
10	0.361	1.097	0.815	0.508	0.3060
11	0.391	1.075	0.741	0.438	0.3026
12	0.419	1.053	0.794	0.476	0.3179
13	0.447	1.037	0.887	0.503	0.3843
14	0.467	1.030	0.739	0.463	0.2756
15	0.494	1.015	0.836	0.467	0.3691
16	0.514	1.007	0.816	0.406	0.4107
17	0.538	0.999	1.015	0.591	0.4247
18	0.555	0.999	0.811	0.362	0.4496

Notes: The average return is the ratio of the predicted take-up rate (col. 3 of Table 3) to the share of the population eligible for the subsidy. The (approximate) marginal return is the ratio of the change (between two rows) in predicted take-up rate to the change (between two rows) in the share eligible for the subsidy.

References

- [1] Andrews, D. (1994a) Asymptotics for semiparametric econometrics models via stochastic equicontinuity, *Econometrica*, 62, 43-72.
- [2] Andrews, D.W.K. (1994b) Empirical Process Methods in econometrics, *Handbook of Econometrics*, Chapter 37, Volume IV (Ed. R.F. Engle and D.L. McFadden), Elsevier Science B.V.
- [3] Angelucci, M. and G. De Giorgi (2008) Indirect effects of an aid program: How do cash transfers affect non-eligibles' consumption? *American Economic Review*, 99, 486-508.
- [4] Bandiera, O. and I. Rasul (2006) Social networks and technology adoption in northern Mozambique, *Economic Journal*, 116-514, 869-902.
- [5] Bhattacharya, D. (2009) Inferring optimal peer allocation using experimental data, *Journal of the American Statistical Association*, 104-486, 486-500.
- [6] Bhattacharya, D. and P. Dupas (2012) Inferring welfare maximizing treatment assignment under budget constraints, *Journal of Econometrics*, 167-1, 168-196.
- [7] Carrell, Scott, B. Sacerdote and James E. West (2012). "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects," mimeo. Dartmouth College.
- [8] Chen, X., O. Linton and I. van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth, *Econometrica*, 71, 1591-1608.
- [9] Conley, T. and C.R. Udry (2010) Learning About a New Technology: Pineapple in Ghana, *American Economic Review*, 100, 35-69.
- [10] Dehejia, R.H. (2005) Program evaluation as a decision problem, *Journal of Econometrics*, 125, 141-173.
- [11] Dupas, P. (2009) What matters (and what does not) in households' decision to invest in malaria prevention? *American Economic Review*, 99-2, 224-230.
- [12] Dupas, P. (2012) Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a field experiment, Working Paper.

- [13] Kong, E., O. Linton and Y. Xia (2010) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model, *Econometric Theory*, 26, 1529-1564.
- [14] Graham, B.S. (2011) Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers, *Handbook of Social Economics*, 1B (Ed. J. Benhabib, A. Bisin and M. Jackson), 965-1052, North-Holland.
- [15] Hahn, J. and G. Ridder (2012) The asymptotic variance of semi-parametric estimators with generated regressors, forthcoming in *Econometrica*.
- [16] Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory*, 24, 726-748.
- [17] Hawley, W.A., P.A. Phillips-Howard, F.O. ter Kuile, D.J. Terlouw, J.M. Vulule, M. Ombok, B.L. Nahlen, J.E. Gimnig, S.K. Kariuki, M.S. Kolczak, A.W. Hightower (2003) Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya, *American Journal of Tropical Medicine and Hygiene*, 68 (4 Suppl), 121-127.
- [18] Hirano, K. and J. Porter (2009) Asymptotics for Statistical Treatment Rules, *Econometrica*, 77-5, 1683-1701.
- [19] Hudgens, M.G. and M.E. Halloran (2008) Toward Causal Inference With Interference, *Journal of the American Statistical Association*, 103-482, 832-842.
- [20] Kanaya, S. (2012) Uniform convergence rates of kernel-based nonparametric estimators for diffusion processes: A damping function approach, Working Paper, University of Oxford.
- [21] Killeen, G.F., T.A. Smith, H. M. Ferguson, H. Mshinda, S. Abdulla, C. Lengeler and S.P. Kachur (2007) Preventing childhood malaria in Africa by protecting adults from mosquitoes with insecticide-treated nets, *PLoS Medicine*, 4-7, e229.
- [22] Lengeler, C. (2004) Insecticide-treated bed nets and curtains for preventing malaria, *Cochrane Database of Systematic Reviews*, 2, Art. No. CD000363.
- [23] Mammen, E., C. Rothe and M. Schienle (2012) Semiparametric estimation with generated covariates, Working Paper, Toulouse School of Economics.

- [24] Manski, C. (2004) Statistical Treatment Rules for Heterogeneous Populations, *Econometrica*, 72-4, 1221-1246.
- [25] Manski, Charles F. (2007) Identification for Prediction and Decision, Harvard University Press, Cambridge, MA.
- [26] Manski, C. (2010) Identification of treatment response with social interactions, CeMMAP working paper CWP01/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- [27] Miguel, E., and M. Kremer (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities, *Econometrica*, 72, 159-217.
- [28] Mosimann, J.E. (1962) On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions, *Biometrika*, 49, 65-82.
- [29] Murray, C.J.L, L.C. Rosenfeld, S.S. Lim, K.G. Andrews, K.J. Foreman, D. Haring, N. Fullman, M. Naghavi, R. Lozano and A.D. Lopez (2012) Global malaria mortality between 1980 and 2010: a systematic analysis, *The Lancet*, 379-9814, 413-431.
- [30] Newey, W. (1994) The asymptotic variance of semiparametric estimators, *Econometrica*, 62, 1349-1382.
- [31] Newey, W. and D.L. McFadden (1994) Large sample estimation and hypothesis testing, *Handbook of Econometrics*, Chapter 36, Volume IV (Ed. R.F. Engle and D.L. McFadden), Elsevier Science B.V.
- [32] Pagan, A. and A. Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press.
- [33] Root, E.D., S. Giebultowicz, M. Ali, M. Yunus and M. Emch (2011) The Role of Vaccine Coverage within Social Networks in Cholera Vaccine Efficacy, *PLoS ONE*, 6-7: e22971.
- [34] Rosenbaum, P. (2007) Interference Between Units in Randomized Experiments, *Journal of the American Statistical Association*, 102-477, 191-200.
- [35] Stoye, J (2009) Minimax Regret Treatment Choice with Finite Samples, *Journal of Econometrics*, 151, 70-81.

- [36] Tetenov, A. (2008) Statistical treatment choice based on asymmetric minmax regrest criterion, forthcoming in *Journal of Econometrics*.
- [37] Van der Vaart, A.W. and J. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.
- [38] Van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press.
- [39] Webster, J., J. Lines and L. Smith (2007) Protecting all pregnant women and children under five years living in malaria endemic areas in Africa with insecticide treated mosquito nets, Working Paper, World Health Organization.
- [40] White, H. (1994) *Estimation, Inference and Specification Analysis*, Cambridge University Press.
- [41] WHO (2008) World Malaria Report 2008, World Health Organization, Geneva.

A Appendix

This appendix is divided into two subsections – the first subsection dealing with Theorem 1 and the second one with Theorem 2. In this Appendix, terminologies related to empirical process theory follow those of van der Vaart and Wellner (1996), which is hereafter referred to as VW (1996).

A.1 Lemmas and proofs for Theorem 1

Step 1: Linearization of the nonparametric estimators $\hat{s}(l)$ and $\hat{s}_\gamma(l)$.

Lemma 1 *Suppose that Assumptions 1-2 hold. Then,*

$$\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} |\hat{s}(l, r) - s^*(l, r)| = n^{-1} \sum_{k=1}^n \xi(Z_k, l, r) + O_{a.s.}((\ln n)/n); \quad (17)$$

$$\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} |\hat{s}_\gamma(l, r) - s_\gamma^*(l, r)| = n^{-1} \sum_{k=1}^n \xi_\gamma(Z_k, l, r) + O_{a.s.}((\ln n)/n). \quad (18)$$

where

$$\xi(Z_k, l, r) : = 1\{|l - L_k| \leq r\} [D_k - s^*(l, r)] / \kappa^*(l, r); \quad (19)$$

$$\xi_\gamma(Z_k, l, r) : = 1\{|l - L_k| \leq r\} [1\{X_{1,k} < \gamma\} - s_\gamma^*(l, r)] / \kappa^*(l, r); \quad (20)$$

$\kappa^*(l, r)$ is defined in (5); and we note that L_k and D_k are subcomponents of $Z_k (= (Y_k, X_k', D_k, L_k)')$.

Furthermore, it also holds that

$$\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} |\hat{s}(l, r) - s^*(l, r)| = O_{a.s.}(\sqrt{(\ln n)/n});$$

$$\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} |\hat{s}_\gamma(l, r) - s_\gamma^*(l, r)| = O_{a.s.}(\sqrt{(\ln n)/n}).$$

These linearized expressions have some resemblance to the Bahadur expression for kernel-based estimation (see, e.g., Sections 3.3-3.4 of Pagan and Ullah, 1999; Kong, Linton and Xia, 2010). However, we do *not* consider the *localization*, and therefore the convergence rates in (17) and (18) are much faster. By this lemma, we have

$$\sqrt{n}[\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)] = (1/\sqrt{n}) \sum_{k=1}^n \boldsymbol{\xi}(Z_k, l) + o_{a.s.}(1); \quad (21)$$

$$\sqrt{n}[\hat{\mathbf{s}}_\gamma(l) - \mathbf{s}_\gamma^*(l)] = (1/\sqrt{n}) \sum_{k=1}^n \boldsymbol{\xi}_\gamma(Z_k, l) + o_{a.s.}(1), \quad (22)$$

uniformly over $l \in \mathcal{L}$, where

$$\boldsymbol{\xi}(Z_k, l) := (\xi(Z_k, l, r_1), \dots, \xi(Z_k, l, r_J))'; \quad \boldsymbol{\xi}_\gamma(Z_k, l) := (\xi_\gamma(Z_k, l, r_1), \dots, \xi_\gamma(Z_k, l, r_J))'.$$

The former result (21) allows us to derive the linearized expression of $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$, while the latter (22) is used to derive that of the nonparametric component of $\hat{\boldsymbol{\rho}}_\gamma(x, l) - \boldsymbol{\rho}_\gamma^*(x, l)$ in (12).

Proof of Lemma 1. Since we can show (17) and (18) exactly in the same manner, we only consider the former. Let

$$\begin{aligned}\hat{\kappa}(l, r) &: = n^{-1} \sum_{k=1}^n \mathbf{1}\{\|l - L_k\| \leq r\}; \\ \hat{t}(l, r) &: = n^{-1} \sum_{k=1}^n \mathbf{1}\{\|l - L_k\| \leq r\} D_k; \quad \text{and} \quad t^*(l, r) := E[\mathbf{1}\{\|l - L_k\| \leq r\} D_k].\end{aligned}$$

Then, we can write

$$\hat{s}(l, r) - s^*(l, r) = \frac{\hat{t}(l, r) - \hat{\kappa}(l, r) s^*(l, r)}{\kappa^*(l, r)} \left[1 + \frac{\kappa^*(l, r) - \hat{\kappa}(l, r)}{\hat{\kappa}(l, r)} \right],$$

where $\kappa^*(l, r)$ is defined in (5). By noting the fact that

$$\hat{t}(l, r) - \hat{\kappa}(l, r) s^*(l, r) = n^{-1} \sum_{k=1}^n \mathbf{1}\{\|l - L_k\| \leq r\} [D_k - s^*(r|l)]$$

and using the following results:

$$\sup_{(l,r) \in \mathcal{L} \times (0, \infty)} |\hat{t}(l, r) - \hat{\kappa}(l, r) s^*(l, r)| = O_{a.s.}(\sqrt{\ln n/n}); \quad (23)$$

$$\sup_{(l,r) \in \mathcal{L} \times (0, \infty)} |\hat{t}(l, r) - t^*(l, r)| = O_{a.s.}(\sqrt{\ln n/n}); \quad (24)$$

$$\sup_{(l,r) \in \mathcal{L} \times (0, \infty)} |\hat{\kappa}(l, r) - \kappa^*(l, r)| = O_{a.s.}(\sqrt{\ln n/n}); \quad (25)$$

$$\sup_{(l,r) \in \mathcal{L} \times [\underline{r}, \infty)} 1/\hat{\kappa}(l, r) = O_{a.s.}(1), \quad (26)$$

whose proofs are provided below, we can obtain the desired result (17).

We can easily check that (26) holds by the boundedness condition of (14) and the result (25).

We now show the three convergence results (23)-(25). Since these can be proved in the same manner, we only consider the last one (25). Denote by \mathcal{F} the following set of functions indexed by $(l, r) \in \mathcal{L} \times (0, \infty)$: $\mathcal{F} := \{f(p|l, r) \mid l \in \mathcal{L} \text{ and } r > 0\}$, where \mathcal{L} is a set of latitude-longitude points on the earth surface; each $f(\cdot|l, r)$ is a function with the form $f(p|l, r) := \mathbf{1}\{\|l - p\| \leq r\}$; $\|l - p\| = \|l - p\|_{\mathcal{L}}$ measures a distance between two points l and p on \mathcal{L} . Recall that we have assumed that the earth is a sphere and $\|l - p\|$ measures the shortest distance between l and p on the sphere (Footnote 5), and therefore, the number of all closed balls on \mathcal{L} is bounded by that of all closed balls in \mathbb{R}^3 (a ball in \mathbb{R}^3 is defined with respect to the standard Euclidean norm). By (ii) of Lemma 9.12 of Kosorok (2008), the set of all closed balls in \mathbb{R}^3 is a VC-class of index 6.

Let \mathcal{A} denote the set of all closed balls on \mathcal{L} . Then, it obviously holds that $\mathcal{F} \subset \{1_A(\cdot) \mid A \in \mathcal{A}\}$ where $1_A(\cdot)$ is an indicator function ($1_A(p) = 1$ if $p \in A$ and $= 0$ otherwise). Then, by using the

fact that the VC index of \mathcal{A} is bounded by 6, together with Theorem 9.2 of Kosorok (2008), the covering number of the set of functions \mathcal{F} satisfies

$$\sup_Q N(\epsilon, \mathcal{F}, L_r(Q)) \leq \Lambda \epsilon^{-5r} \quad \text{for } \epsilon \in (0, 1), \quad (27)$$

where $N(\epsilon, \mathcal{F}, L_r(Q))$ stands for the L_r covering number of the set of function \mathcal{F} (for the definition, see Section 2.2.2 of Kosorok, 2008 or Section 2.2.1 of VW, 1996); the supremum is over all probability measures on \mathcal{L} ; and $\Lambda (> 0)$ is some constant independent of Q . Now, using (27), the Bernstein exponential inequality (see, e.g., page 102 of VW, 1996) and the Borel-Cantelli lemma as in (Hansen, 2008, Proofs of Theorems 2 and 3) and Kanaya (2012, Proof of Theorem 4), we can easily show the desired result (25), where we note that the number of $L_r(Q)$ ϵ -balls needed to cover \mathcal{F} is finite with the polynomial order (at most) while the convergence of relevant probabilities has an exponential decay. Now, the proof of Lemma 1 is completed. ■

Step 2: Quantifying effects on $\hat{\boldsymbol{\eta}}$ of preliminary estimation of \mathbf{s}^* .

To investigate asymptotic behavior of the parametric component $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$, we define our QML estimator $\hat{\boldsymbol{\eta}}$ as the solution to the following equation (the first-order condition of the maximization problem in (9)):

$$\hat{\boldsymbol{\eta}} := \left\{ \boldsymbol{\eta} \in \mathbb{R}^{d_\eta} \mid n^{-1} \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}, \hat{\mathbf{s}}(L_i)) = 0 \right\}, \quad (28)$$

where \mathbf{m} is a d_η -dimensional (column) vector of functions:

$$\mathbf{m}(z, \boldsymbol{\eta}, \mathbf{s}(l)) = (\partial/\partial\boldsymbol{\eta}) q(z, \boldsymbol{\eta}, \mathbf{s}(l)).$$

By the convexity property of the function q , we can easily check the existence of $\hat{\boldsymbol{\eta}}$ and its uniqueness (for given $\hat{\mathbf{s}}(\cdot)$) and we here define the (pseudo) true parameter $(\boldsymbol{\eta}^*, \mathbf{s}^*(\cdot))$ as the one satisfying the following population moment equation:

$$E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))] = 0. \quad (29)$$

The interpretation of $\hat{\boldsymbol{\eta}}$ as a moment-based estimator in (28) helps us to find a linearized (asymptotic) expression of $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$. One can usually obtain linearized expressions for standard parametric estimators by using Taylor expression arguments. This is also the case here. However, our estimator $\hat{\boldsymbol{\eta}}$ depends up on the preliminary estimation of $\hat{\mathbf{s}}$, and we need to quantify its effect. This can be done by Newey' (1994) pathwise derivative method, which allows us to obtain a so-called *correction* term (i.e., a term capturing the effect of preliminary estimation in the linearized expression of $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$).

Pathwise derivative method: Let \mathfrak{F} be the set of all distribution functions of Z_i , and $\{F_\theta : F_\theta \in \mathfrak{F}\}$ be a one-dimensional subfamily of \mathfrak{F} , i.e., a path in \mathfrak{F} (indexed by $\theta \in \mathbb{R}$), and also let $E_\theta[\cdot]$ denote the expectation at the distribution F_θ . We let $F = F_0 = F_\theta|_{\theta=0}$ the true distribution, and $E[\cdot] (= E_0[\cdot])$ denote the expectation at the true F_0 . Suppose that F_θ has a density dF_θ (with respect to some dominating measure which is common for all F_θ). Let $S_\theta(z) := \partial \ln(dF_\theta) / \partial \theta$. Then, $S_0(z) = \partial \ln(dF_\theta) / \partial \theta|_{\theta=0}$ is called a score corresponding to dF_θ , where we also suppose that $E_\theta[|S_\theta(Z_k)|^2] < \infty$ for each θ .

Our nonparametric estimator $\hat{s}(l, r)$ for each (l, r) is a moment estimator which solves the moment condition: $E[1\{|L_k - l| \leq r\} [D_k - s(l, r)]] = 0$. Therefore, we can see $\hat{s}(l, r)$ as the sample counterpart of the following (population) quantity:

$$s(l, r|F_\theta) := \frac{E_\theta[1\{|L_k - l| \leq r\} D_k]}{E_\theta[1\{|L_k - l| \leq r\}]} \quad (30)$$

when $Z_k \sim F_\theta$.

When $Z_k \sim F_\theta$, our (finite-dimensional) parameter of interest, $\boldsymbol{\eta}^* = \boldsymbol{\eta}^*(\theta)$, solves the following population moment equation:

$$E_\theta[\mathbf{m}(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i|F_\theta))] = 0,$$

which corresponds to equation (3.8) in Newey (1994). If we can find some function $\boldsymbol{\alpha}$ such that

$$(\partial/\partial\theta) E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}(L_i|F_\theta))]|_{\theta=0} = E[\boldsymbol{\alpha}(Z_i) S_0(Z_i)] \quad \text{and} \quad E[\boldsymbol{\alpha}(Z_i)] = 0, \quad (31)$$

then we can obtain the linear expression (or influence function) of the estimator $\hat{\boldsymbol{\eta}}$ (see equation (3.10) in Newey, 1994). $\boldsymbol{\alpha}(Z_i)$ satisfying (31) is called a correction term (in Newey's terminology).

We below present a way for deriving this correction term.

Given the expression (30) under $Z_k \sim F_\theta$, we are able to calculate

$$\begin{aligned} (\partial/\partial\theta) s(l, r|F_\theta) &= \{E_\theta[1\{|L_k - l| \leq r\}]\}^{-2} \left\{ E_\theta[D_k 1\{|L_k - l| \leq r\} S_\theta(Z_i)] E_\theta[1\{|L_k - l| \leq r\}] \right. \\ &\quad \left. - E_\theta[D_k 1\{|L_k - l| \leq r\}] E_\theta[1\{|L_k - l| \leq r\} S_\theta(Z_k)] \right\}, \end{aligned}$$

which quantifies the influence of the infinitesimal change of F_θ on $s(l, r|F_\theta)$ at F_θ . Let

$$\xi(Z_k, l, r) = 1\{|L_k - l| \leq r\} [D_k - s^*(l, r)] / \kappa^*(l, r),$$

as defined in (19), then, we can write

$$(\partial/\partial\theta) s(l, r|F_0) = (\partial/\partial\theta) s(l, r|F_\theta)|_{\theta=0} = E[\xi(Z_k, l, r) S_0(Z_k)]. \quad (32)$$

For notational convenience, define the following vectors of functions:

$$\mathbf{s}(l|F_\theta) := (s(l, r_1|F_\theta), \dots, s(l, r_J|F_\theta))' \quad \text{and} \quad \boldsymbol{\xi}(Z_k, l) := (\xi(Z_k, l, r_1), \dots, \xi(Z_k, l, r_J))'.$$

We also define a $d_{\boldsymbol{\eta}}$ -by- J matrix of functions as

$$\bar{\mathbf{D}}^*(z) := (\partial/\partial \mathbf{s}) \mathbf{m}(z, \boldsymbol{\eta}^*, \mathbf{s})|_{\mathbf{s}=\mathbf{s}^*(l)},$$

where we note that $\mathbf{s}^*(l) = \mathbf{s}(l|F_0)$.

Given these, we have

$$\begin{aligned} (\partial/\partial \theta) E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}(L_i|F_\theta))]_{\theta=0} &= E[\bar{\mathbf{D}}^*(Z_i) (\partial/\partial \theta) \mathbf{s}(L_i|F_\theta)]_{\theta=0} \\ &= E[\bar{\mathbf{D}}^*(Z_i) E[\boldsymbol{\xi}(Z_k, L_i) S_0(Z_k) | L_i]] \\ &= E[\bar{\mathbf{D}}^*(Z_i) E[\boldsymbol{\xi}(Z_k, L_i) S_0(Z_k) | Z_i]] \\ &= E[E[\bar{\mathbf{D}}^*(Z_i) \boldsymbol{\xi}(Z_k, L_i) | Z_k] S_0(Z_k)] \\ &= E[\boldsymbol{\alpha}(Z_k) S_0(Z_k)], \end{aligned} \tag{33}$$

where the second equality follows from (32) and the definition of the column vector $\boldsymbol{\xi}(Z_k, l)$; and the third holds since L_i is a subcomponent of Z_i ; and the fourth holds due to the law of repeated conditional expectations. The last equality in (33) holds with

$$\begin{aligned} \boldsymbol{\alpha}(z) &: = E[\bar{\mathbf{D}}^*(Z_i) \boldsymbol{\xi}(z, L_i)] = E\left[\frac{\partial}{\partial \mathbf{s}} \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*(L_i)} \boldsymbol{\xi}(z, L_i)\right] \\ &= E\left[\frac{\partial}{\partial \mathbf{s}} \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*(L_i)} \right. \\ &\quad \left. \times \left(\frac{1\{\|l - L_i\| \leq r_1\} [d - s^*(L_i, r_1)]}{\kappa^*(L_i, r_1)}, \dots, \frac{1\{\|l - L_i\| \leq r_J\} [d - s^*(L_i, r_J)]}{\kappa^*(L_i, r_J)} \right) \right] \end{aligned} \tag{34}$$

which is a $d_{\boldsymbol{\eta}}$ -dimensional column vector of functions, where we note that l and d are subcomponents of z (recall $Z_k = (Y_k, X'_k, D_k, L'_k)'$). By the I.I.D. condition, we can check

$$(\partial/\partial \theta) E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}(L_i|F_\theta))]_{\theta=0} = E[\boldsymbol{\alpha}(Z_k) S_0(Z_k)] = E[\boldsymbol{\alpha}(Z_i) S_0(Z_i)],$$

and $E[\boldsymbol{\alpha}(Z_i)] (= E[\boldsymbol{\alpha}(Z_k)]) = 0$.

Remark 2 *The calculation of the pathwise derivative $(\partial/\partial \theta) E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}(L_i|F_\theta))]_{\theta=0}$ in (33) is basically a chain rule. However, unlike a standard simple case for composite functions, we need to consider an involved object, i.e., the expectation of a composite function. In this case, a derivative calculation rule is still similar to a standard one, but we use the use of the law of conditional*

expectations. This calculation method seems to have general applicability to several other cases, as long as a preliminary estimator affects the next step only through its values at observations ($\hat{\mathbf{s}}(L_i)$ in our case). If one can obtain the pathwise derivative of the preliminary estimator, which is not difficult in many cases, we can almost immediately derive the form of the correction term by the law of repeated conditional expectations.

Remark 3 Given the form of the correction term as in (34), we need to verify that it indeed constitutes the asymptotic linear expression of $[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*]$. This can be done by direct calculations as in our proof of Lemma 3 below.

Step 3: Linearization of the parametric component $\hat{\boldsymbol{\eta}}$.

By the boundedness of relevant functions and their derivatives as well as the uniform convergence of the nonparametric first step estimator $\hat{\mathbf{s}}(\cdot)$, we can easily prove the consistency of $\hat{\boldsymbol{\eta}}$ for $\boldsymbol{\eta}^*$ (note that by the convexity, we can restrict the domain of $\boldsymbol{\eta}$ as some compact neighborhood of $\boldsymbol{\eta}^*$, which we denote by $\mathcal{C}_{\boldsymbol{\eta}}(\subset \mathbb{R}^{d_{\boldsymbol{\eta}}})$). By using the Taylor-expansion arguments, we can write

$$\begin{aligned} \sqrt{n}[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*] &= -[\mathbf{M}^*]^{-1} (1/\sqrt{n}) \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \hat{\mathbf{s}}(L_i)) + o_P(1/\sqrt{n}) \\ &= -[\mathbf{M}^*]^{-1} (1/\sqrt{n}) \left\{ \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] \right\} + o_P(1) \end{aligned} \quad (35)$$

where \mathbf{M}^* is the limit Hessian matrix of the objective function defined as

$$\mathbf{M}^* := E \left[(\partial^2 / \partial \boldsymbol{\eta}' \partial \boldsymbol{\eta}) q(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) \right] = E \left[(\partial / \partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) \right], \quad (36)$$

and $\bar{\mathbf{D}}^*(\cdot)$ is a matrix-valued function defined as

$$\bar{\mathbf{D}}^*(z) := (\partial / \partial \mathbf{s}') \mathbf{m}(z, \boldsymbol{\eta}^*, \mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*(l)}. \quad (37)$$

This result (35) is formally stated below as Lemma 2.

We note that while the first term inside the curly brackets on the right-hand side (RHS) of (35) is the score function for the fully parametric case with $\mathbf{s}^*(\cdot)$ known, the second term captures the effect due to the nonparametrically estimated component. We can verify that this second term can be actually replaced with the correction term derived in Step 2, and as a result, the main component on the RHS of (35) can be written as an empirical process. That is, we have

$$\sqrt{n}[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*] = (1/\sqrt{n}) \sum_{i=1}^n \boldsymbol{\psi}(Z_i) + o_P(1), \quad (38)$$

where $\boldsymbol{\psi}$ is the influence function defined as

$$\boldsymbol{\psi}(Z_i) := -[\mathbf{M}^*]^{-1} \{ \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \boldsymbol{\alpha}(Z_i) \}. \quad (39)$$

Lemma 3 (presented below), together with Lemma 2, formally proves the validity of this linearized expression (38).

As stated previously, the term $\boldsymbol{\alpha}$ in (39) is called a "correction term", since it *corrects* sampling errors stemming from the preliminary nonparametric estimation, which are *additional* to usual sampling errors in standard parametric estimation. The sum of \mathbf{m} and $\boldsymbol{\alpha}$ left-multiplied by $-[\mathbf{M}^*]^{-1}$ is called the *influence* function of $\hat{\boldsymbol{\eta}}$, which fully characterizes the (first-order) asymptotic properties of the estimator $\hat{\boldsymbol{\eta}}$.

Lemma 2 *Suppose that Assumptions 1-3 hold. Then,*

$$\begin{aligned} & \sqrt{n} [\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*] \\ = & -[\mathbf{M}^*]^{-1} (1/\sqrt{n}) \left\{ \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] \right\} + o_P(1) \end{aligned} \quad (40)$$

where \mathbf{M}^* and $\bar{\mathbf{D}}^*$ are defined in (36) and (37), respectively.

Proof. First, we claim the consistency of $\hat{\boldsymbol{\eta}}$, that is, $|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*| = o_P(1)$ ($\boldsymbol{\eta}^*$ is defined in (29)). The proof of this result is omitted for the sake of brevity (it can be easily shown by using arguments similar to those in this proof and in the proof of Lemma 3).

By the definition of our semiparametric estimator $\hat{\boldsymbol{\eta}}$ and the Taylor expansion, we have

$$\begin{aligned} 0 &= (1/\sqrt{n}) \sum_{i=1}^n \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) \\ &= (1/\sqrt{n}) \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \hat{\mathbf{s}}(L_i)) + n^{-1} \sum_{i=1}^n (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \tilde{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) \sqrt{n} [\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*], \end{aligned} \quad (41)$$

where $\tilde{\boldsymbol{\eta}}$ is on the linesegment connecting to $\hat{\boldsymbol{\eta}}$ to $\boldsymbol{\eta}^*$ (it may be different for a different element of $(\partial/\partial \boldsymbol{\eta}') \mathbf{m}$).

We now consider the limit of the component $n^{-1} \sum_{i=1}^n (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \tilde{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i))$. To this end, note that by the convexity property of the function q , which is induced by the functional form of Φ (the C.D.F. of the standard normal), we may restrict the parameter space of $\boldsymbol{\eta}$ to some compact neighborhood of $\boldsymbol{\eta}^*$ in \mathbb{R}^{d_η} . We denote by $C_{\boldsymbol{\eta}^*}$ this compact neighborhood. For notational simplicity, we write

$$\mathbf{h}(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i)) = (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i)) = (\partial^2/\partial \boldsymbol{\eta}' \partial \boldsymbol{\eta}) q(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i)),$$

and denote by $h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}(L_i))$ each element of the matrix \mathbf{h} with $1 \leq a, b \leq d_\eta$. Then, by the Taylor expansion again, we have

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \hat{\mathbf{s}}(L_i)) \\
&= n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i)) + n^{-1} \sum_{i=1}^n (\partial/\partial \mathbf{s}') h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}) \Big|_{\mathbf{s}=\tilde{\mathbf{s}}(L_i)} [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] \\
&= n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i)) + O_{a.s.}(\sqrt{(\log n)/n}), \quad \text{uniformly over } \boldsymbol{\eta} \in C_{\boldsymbol{\eta}^*}, \tag{42}
\end{aligned}$$

for each (a, b) element, where $\tilde{\mathbf{s}}(L_i)$ is on the linesegment connecting $\hat{\mathbf{s}}(L_i)$ to $\mathbf{s}^*(L_i)$, and the last equality holds since $(\partial/\partial \mathbf{s}')h_{a,b}(z, \boldsymbol{\eta}, \mathbf{s})$ is uniformly bounded over $z, \boldsymbol{\eta}$ and \mathbf{s} (recall that the functional form of Φ and the linear specification in (7) and (8)) and

$$\hat{\mathbf{s}}(l) - \mathbf{s}^*(l) = O_{a.s.}(\sqrt{(\log n)/n}) \quad \text{uniformly over } l, \tag{43}$$

(this can be easily shown by using arguments similar to those in the proof of Lemma 1).

Given the result (42), we next consider the uniform law of large numbers for $n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i))$. By the specification of q with the C.D.F. of the standard normal Φ (in (7) and (8)), as well as the boundedness condition of the support of X_i (in Assumption 3), we can see that $h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i))$ is Lipschitz continuous with respect to $\boldsymbol{\eta}$ (with a Lipschitz constant uniform over $\boldsymbol{\eta} \in C_{\boldsymbol{\eta}^*}$). Therefore, together with the I.I.D. condition, $n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i))$ is stochastically equicontinuous (over $\boldsymbol{\eta} \in C_{\boldsymbol{\eta}^*}$). Note also that $E[h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i))]$ is continuous uniformly over $\boldsymbol{\eta} \in C_{\boldsymbol{\eta}^*}$. Then, by the law of large numbers (recall the I.I.D. condition) and the compactness of the parameter space $C_{\boldsymbol{\eta}^*}$, we have

$$\sup_{\boldsymbol{\eta} \in C_{\boldsymbol{\eta}^*}} \left| n^{-1} \sum_{i=1}^n h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i)) - E[h_{a,b}(Z_i, \boldsymbol{\eta}, \mathbf{s}^*(L_i))] \right| = o_P(1), \tag{44}$$

for each (a, b) . By the consistency of $\hat{\boldsymbol{\eta}}$, (42) and (44), together with recalling the definition of $\mathbf{M}^*(Z_i)$ (given in (36)) and that of $h_{a,b}$ (given earlier in this proof), we now obtain

$$n^{-1} \sum_{i=1}^n (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) \xrightarrow{P} \mathbf{M}^*. \tag{45}$$

Therefore, by this and (41), together with the Taylor expansion, we have

$$\begin{aligned}
& [\mathbf{M}^* + o_P(1)] \sqrt{n} [\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*] = -(1/\sqrt{n}) \sum_{i=1}^n \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \hat{\mathbf{s}}(L_i)) \\
&= -(1/\sqrt{n}) \sum_{i=1}^n \left\{ \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \frac{\partial}{\partial \mathbf{s}'} \mathbf{m}(z, \boldsymbol{\eta}^*, \mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*(L_i)} [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] \right\}, \\
&+ O_P\left(\sqrt{n} \max_{1 \leq a, b \leq d_\eta} \max_{1 \leq i \leq n} \sup_{\mathbf{s}(\cdot)} |h_{a,b}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}(L_i))| \times [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)]^2\right), \tag{46}
\end{aligned}$$

where the supremum in the last line is taken over the set of $[0, 1]$ -valued functions on $l \in \mathcal{L}$. By the specification of q and the boundedness condition (in Assumption 3), $h_{a,b}(z, \boldsymbol{\eta}, \mathbf{s})$ is uniformly bounded (over $(z, \boldsymbol{\eta}, \mathbf{s})$). Also, by (43), $\sup_l [\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)]^2 = O_{a.s.}((\log n)/n)$. Therefore, the second component on the RHS of (46) is $O_{a.s.}((\log n)/\sqrt{n}) = o_{a.s.}(1)$. Finally, recalling the definition of $\bar{\mathbf{D}}^*(z)$ in (37) and noting the invertibility of \mathbf{M}^* , which follows from the convexity property of q and the non linear-dependence condition in Assumption 1, we obtain the desired result (40). This completes the proof of Lemma 2. ■

Lemma 3 *Suppose that Assumptions 1-3. Then, it holds that*

$$(1/\sqrt{n}) \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] = (1/\sqrt{n}) \sum_{i=1}^n \boldsymbol{\alpha}(Z_i) + o_P(1), \quad (47)$$

with $E[\boldsymbol{\alpha}(Z_i)] = 0$ and $E[||\boldsymbol{\alpha}(Z_i)||^2] < \infty$, where $\boldsymbol{\alpha}(\cdot)$ is a vector-valued function defined as

$$\boldsymbol{\alpha}(z) : = E[\bar{\mathbf{D}}^*(Z_i) \boldsymbol{\xi}(z, L_i)] = E\left[\frac{\partial}{\partial \mathbf{s}'} \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*(L_i)} \boldsymbol{\xi}(z, L_i)\right]; \quad (48)$$

$$\boldsymbol{\xi}(z, L_i) : = (\xi(z, L_i, r_1), \dots, \xi(z, L_i, r_J))'; \quad (49)$$

$$\xi(z, L_i, r) : = 1 \{||L_i - l|| \leq r_j\} [d - s^*(L_i, r)] / \kappa^*(L_i, r);$$

and we note that d and l are subcomponents of $z = (y, x', d, l)'$ and the functional form of $\xi(\cdot, \cdot, r)$ defined here is the same as that defined in (19).

The proof of the lemma follows subsequently. Note that the left-hand side (LHS) of (47) includes an infinite-dimensional component, $\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)$. To handle this sort of object (and verify a result like (47)), we use some U-statistic-like arguments.

Proof of Lemma 3. First, we check the moment conditions of $\boldsymbol{\alpha}(Z_i)$. By the boundedness of the support of X_i and the specification of Φ (in Assumption 3), $\bar{\mathbf{D}}^*(z)$ is uniformly bounded over the support of Z_i . Also, by Assumption 2, $|\xi(z, \lambda, r)| \leq 2/C$ uniformly over (z, λ, r) . Therefore, $\boldsymbol{\alpha}(z)$ is uniformly bounded over the support of Z_i and $E[||\boldsymbol{\alpha}(Z_i)||^2] < \infty$. By the definition of $\boldsymbol{\alpha}(z)$ and $\boldsymbol{\xi}(z, L_u)$ (given in (48) and (49)), we can easily see that $E[\boldsymbol{\alpha}(Z_i)] = 0$ (recall also the way we have derived the form of $\boldsymbol{\alpha}(\cdot)$ in (33)).

Next, we verify the equality in (47). By the uniform convergence result (17) in Lemma 1, we have

$$\begin{aligned} & (1/\sqrt{n}) \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] \\ = & (1/\sqrt{n}) \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [n^{-1} \sum_{k=1}^n \boldsymbol{\xi}(Z_k, L_i)] + O_P((\ln n)/\sqrt{n}). \end{aligned} \quad (50)$$

By the definition of $\boldsymbol{\alpha}(\cdot)$ given in (48), we can write

$$\sum_{i=1}^n \boldsymbol{\alpha}(Z_i) = \int \bar{\mathbf{D}}^*(z) [n^{-1} \sum_{k=1}^n \boldsymbol{\xi}(Z_k, l)] dF_Z^*(z).$$

By using this and letting

$$\mathbf{a}(i, k) := \bar{\mathbf{D}}^*(Z_i) \boldsymbol{\xi}(Z_k, L_i) - \int \bar{\mathbf{D}}^*(z) \boldsymbol{\xi}(Z_k, l) dF_Z^*(z),$$

we can also write

$$\begin{aligned} & (1/\sqrt{n}) \left\{ \sum_{i=1}^n \bar{\mathbf{D}}^*(Z_i) [n^{-1} \sum_{k=1}^n \boldsymbol{\xi}(Z_k, L_i)] - \sum_{i=1}^n \boldsymbol{\alpha}(Z_i) \right\} \\ &= (1/\sqrt{n}) n^{-1} \sum_{i=1}^n \sum_{k=1}^n \left\{ \bar{\mathbf{D}}^*(Z_i) \boldsymbol{\xi}(Z_k, L_i) - \int \bar{\mathbf{D}}^*(z) \boldsymbol{\xi}(Z_k, l) dF_Z^*(z) \right\} \\ &= (1/\sqrt{n}) n^{-1} \left\{ \sum_{i=1}^n \mathbf{a}(i, i) + \sum \sum_{1 \leq i, j \leq n; i \neq j} \mathbf{a}(i, k) \right\} = O_P(1/\sqrt{n}) = o_P(1), \end{aligned} \quad (51)$$

where the last equality holds since $E[\|\mathbf{a}(i, i)\|^2] = O(1)$ (uniformly over any i , which follows from the boundedness of relevant functions) and

$$E[\|\sum \sum_{1 \leq i, j \leq n; i \neq j} \mathbf{a}(i, j)\|^2] = O(n^2). \quad (52)$$

We can see that this result (52) holds true by noting the following expansion:

$$\begin{aligned} & E[\|\sum \sum_{1 \leq i, j \leq n; i \neq j} \mathbf{a}(i, j)\|^2] = n(n-1) \{E[\mathbf{a}(1, 2)' \mathbf{a}(1, 2)] + E[\mathbf{a}(1, 2)' \mathbf{a}(2, 1)]\} \\ & + n(n-1)(n-2) \{E[\mathbf{a}(1, 2)' \mathbf{a}(1, 3)] + E[\mathbf{a}(1, 2)' \mathbf{a}(3, 1)] + E[\mathbf{a}(1, 2)' \mathbf{a}(2, 3)] + E[\mathbf{a}(1, 2)' \mathbf{a}(3, 2)]\} \\ & + n(n-1)(n-2)(n-3) E[\mathbf{a}(1, 2)' \mathbf{a}(3, 4)], \end{aligned} \quad (53)$$

where $\|\cdot\|$ stands for the usual Euclidean norm; the equality holds by the I.I.D. condition; the second and third terms on the right-hand side are zero since $E[\mathbf{a}(i, k) | Z_k] = E[\mathbf{a}(i, k) | Z_i] = 0$ for any $i \neq k$; and it holds that $E[\mathbf{a}(1, 2)' \mathbf{a}(1, 2)] = O(1)$ and $E[\mathbf{a}(1, 2)' \mathbf{a}(2, 1)] = O(1)$ by the boundedness of the relevant functions. Now, by (50) and (51), we obtain the desired result (47). ■

Step 4: Functional CLT of $\hat{\mathbf{G}}_\gamma$ and weak convergence of $\sqrt{n}[\hat{\boldsymbol{\rho}}_\gamma - \boldsymbol{\rho}_\gamma^*]$.

Recall the definitions of $\hat{\boldsymbol{\rho}}_\gamma$ and $\boldsymbol{\rho}_\gamma^*$ stated in defining (12). Then, by (22) and (38) together with the definition of the E.D.F., we have uniformly over $(x, l) \in \mathcal{X} \times \mathcal{L}$,

$$\sqrt{n}[\hat{\boldsymbol{\rho}}_\gamma(x, l) - \boldsymbol{\rho}_\gamma^*(x, l)] = \hat{\mathbf{G}}_\gamma(x, l) + o_P(1), \quad (54)$$

where

$$\begin{aligned} \hat{\mathbf{G}}_\gamma(x, l) & : = (1/\sqrt{n}) \sum_{i=1}^n \begin{pmatrix} \boldsymbol{\psi}(Z_i) \\ \boldsymbol{\xi}_\gamma(Z_i, l) \\ g(Z_i, x, l) \end{pmatrix}; \\ g(Z_i, x, l) & : = 1\{X_i \leq x \ \& \ L_i \leq l\} - F_{X, L}^*(x, l); \end{aligned} \quad (55)$$

From this expression, we can see that the asymptotic properties of $\sqrt{n}[\hat{\rho}_\gamma - \rho_\gamma^*]$ are determined by those of $\hat{\mathbf{G}}_\gamma$. We below prove a functional CLT of this $\hat{\mathbf{G}}_\gamma$, i.e., the weak convergence of $\hat{\mathbf{G}}_\gamma$ to a tight Gaussian process. To formally prove the FCLT result, we hereafter work as if \mathcal{L} (the range of L_i) is a bounded subset of \mathbb{R}^2 , while \mathcal{L} is in fact a set of latitude-longitude points on the sphere. Accordingly, we regard $s^*(\cdot, r)$, $\kappa^*(\cdot, r)$, $F_{X,L}^*(x, \cdot)$ and all relevant functions as functions on (the subset of) \mathbb{R}^2 . This way of treating \mathcal{L} (and the functions on \mathcal{L}) is possible by decimalizing a pair of latitude-longitude figures appropriately, if \mathcal{L} does not include any *band* or *ring* surrounding the sphere.¹⁵ Also, note that this way does not at all restrict a way of calculating/defining a distance on \mathcal{L} . We can still use a distance measured along a path on the sphere's surface (say, one based on the Haversine formula), but by (re-)defining the relevant function as functions on \mathbb{R}^2 , we can easily define some continuity properties of functions.¹⁶

We now derive the weak convergence of $\hat{\mathbf{G}}_\gamma$ (as elements of $l^\infty(\mathcal{X} \times \mathcal{L})$, which denotes the set of all bounded $(2d_X + 3J + 1)$ -dimensional vector-valued functions, $\mathcal{X} \times \mathcal{L}(\subset \mathbb{R}^{d_X} \times \mathbb{R}^2) \rightarrow \mathbb{R}^{2d_X + 3J + 1}$).¹⁷ While our main purpose is an inference on $\hat{U}(\gamma)$, we are also interested in examining whether it differs from $\tilde{U}(\gamma)$, i.e., the effect of taking into account the externality. To implement an econometric inference jointly on $\hat{U}(\gamma)$ and $\tilde{U}(\gamma)$, we also consider the joint result of $\hat{\rho}_\gamma(x, l)$ and $(\tilde{\delta}'_0, \tilde{\delta}'_1)'$, which is useful to derive the joint asymptotic distribution of $(\hat{U}(\gamma), \tilde{U}(\gamma))'$.

Lemma 4 *Suppose that Assumptions 1-5.*

Then, (i) it holds that

$$\hat{\mathbf{G}}_\gamma \Longrightarrow \mathbf{G}_\gamma^*, \quad (56)$$

for each γ , where \Longrightarrow denotes the weak convergence; the weak convergence takes place in $l^\infty(\mathcal{X} \times \mathcal{L})$; $l^\infty(\mathcal{X} \times \mathcal{L})$ is the set of all vector-valued bounded maps, $\mathcal{X} \times \mathcal{L}(\subset \mathbb{R}^{d_X} \times \mathbb{R}^2) \rightarrow \mathbb{R}^{2d_X + 3J + 1}$;

¹⁵For example, it is possible if \mathcal{L} is the set of all points satisfying $45^\circ\text{N} \leq \text{latitude} \leq 55^\circ\text{N}$ and $5^\circ\text{S} \leq \text{longitude} \leq 15^\circ\text{E}$; but not if \mathcal{L} is the set of all points satisfying $45^\circ\text{N} \leq \text{latitude} \leq 55^\circ\text{N}$. For the latter, \mathcal{L} is a latitude "band" surrounding the sphere, and it is hard to think of its projection on \mathbb{R}^2 since the points of 180°E (+180) are also those of 180°S (-180).

¹⁶For example, in defining the right continuity of $F_{X,L}(\cdot, \cdot)$, we should consider that $F_{X,L}(\cdot, \cdot)$ is a function on (a subset of) $\mathbb{R}^{d_X} \times \mathbb{R}^2$.

¹⁷Note that in empirical process theory, it is often standard to consider a set of functions as an index set (rather than $\mathcal{X} \times \mathcal{L}$, a subset of $\mathbb{R}^{d_X} \times \mathbb{R}^2$, as here). To verify the desired weak convergence result in our subsequent proof, we actually consider such a set of functions \mathcal{F} (defined in (57)). However, the weak convergence in $l^\infty(\mathcal{F})$ (the space of all bounded $(2d_X + 3J + 1)$ -dimensional vectors of maps, $\mathcal{F} \rightarrow \mathbb{R}^{2d_X + 3J + 1}$) implies that in $l^\infty(\mathcal{X} \times \mathcal{L})$ under standard conditions (see Section 19.5 of van der Vaart, 1998). Therefore, we can work here without explicitly defining some set of functions as an index set.

$\hat{\mathbf{G}}_\gamma = \{\hat{\mathbf{G}}_\gamma(x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$; and $\mathbf{G}_\gamma^* = \{\mathbf{G}_\gamma^*(x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$ is a tight mean-zero Gaussian process (indexed by $(x, l) \in \mathcal{X} \times \mathcal{L}$, for each γ) whose covariance kernel $\mathbf{\Lambda}_\gamma^*$ is given as

$$\mathbf{\Lambda}_\gamma^*((x_1, l_1), (x_2, l_2)) := E \left[\begin{pmatrix} \psi(Z_i) \\ \boldsymbol{\xi}_\gamma(Z_i, l_1) \\ g(Z_i, x_1, l_1) \end{pmatrix} \begin{pmatrix} \psi(Z_i) \\ \boldsymbol{\xi}_\gamma(Z_i, l_2) \\ g(Z_i, x_2, l_2) \end{pmatrix}' \right].$$

(ii) It holds that

$$\sqrt{n}[\hat{\boldsymbol{\rho}}_\gamma - \boldsymbol{\rho}_\gamma^*] \implies \mathbf{G}_\gamma^*,$$

for each γ , where $[\hat{\boldsymbol{\rho}}_\gamma - \boldsymbol{\rho}_\gamma^*] = \{[\hat{\boldsymbol{\rho}}_\gamma(x, l) - \boldsymbol{\rho}_\gamma^*(x, l)]\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$; and the weak convergence takes place in the space $l^\infty(\mathcal{X} \times \mathcal{L})$.

(iii) It also holds that

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\rho}}_\gamma - \boldsymbol{\rho}_\gamma^* \\ (\tilde{\delta}'_0, \tilde{\delta}'_1)' - (\delta^{**'}_0, \delta^{**'}_1)' \end{pmatrix} \implies \mathbf{H}_\gamma^*,$$

where the weak convergence takes place in the space of $\bar{l}^\infty(\mathcal{X} \times \mathcal{L})$, the set of all vector-valued bounded maps, $\mathcal{X} \times \mathcal{L} (\subset \mathbb{R}^{d_X} \times \mathbb{R}^2) \rightarrow \mathbb{R}^{4d_X + 3J + 1}$; and $\mathbf{H}_\gamma^* = \{\mathbf{H}_\gamma^*(x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$ is a tight mean-zero Gaussian process (indexed by $(x, l) \in \mathcal{X} \times \mathcal{L}$, for each γ ; the first $2d_X + 3J + 1$ components of $\mathbf{H}_\gamma^*(x, l)$ coincide with $\mathbf{G}_\gamma^*(x, l)$) whose covariance kernel is given as

$$\mathbf{\Pi}_\gamma^*((x_1, l_1), (x_2, l_2)) := E \left[\begin{pmatrix} \psi(Z_i) \\ \boldsymbol{\xi}_\gamma(Z_i, l_1) \\ g(Z_i, x_1, l_1) \\ \bar{\boldsymbol{\psi}}(Z_i) \end{pmatrix} \begin{pmatrix} \psi(Z_i) \\ \boldsymbol{\xi}_\gamma(Z_i, l_2) \\ g(Z_i, x_2, l_2) \\ \bar{\boldsymbol{\psi}}(Z_i) \end{pmatrix}' \right];$$

$$\bar{\boldsymbol{\psi}}(Z_i) := -[\bar{\mathbf{M}}^*]^{-1} \mathbf{m}(Z_i, \bar{\boldsymbol{\eta}}^*, \mathbf{0}); \quad \bar{\mathbf{M}}^* := E[(\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \bar{\boldsymbol{\eta}}^*, \mathbf{0})].$$

Proof. We can check the (uniform) validity of the linear representation in (15) by the results proved in Lemmas 1, 2 and 3. Therefore, given the part (i) of the lemma, the result (ii) follows immediately by using standard results on stochastic convergence (see, e.g., (iv) of Theorem 18.10 in van der Vaart, 1998) and we omit details. We also omit the proof of the part (iii) for brevity, which is quite analogous to the proof of the parts (i) and (ii).

To prove the result (i), we consider the following set of functions:

$$\mathcal{F} = \{(\boldsymbol{\psi}(\cdot), \boldsymbol{\xi}_\gamma(\cdot, l)', g(\cdot, x, l))'\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}. \quad (57)$$

Given this, in order to obtain the weak convergence of $\{\mathbf{G}_\gamma^*(x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$ in the space $l^\infty(\mathcal{X} \times \mathcal{L})$, we follow the same strategy as in the proof of Theorem 19.28 of van der Vaart (1998) (or that as in Theorem 2.11.22 of VW, 1996). While van der Vaart's theorem deals with a set of univariate (real-valued) functions, it can be easily extended to the multivariate (vector-valued) function case. We omit the proof for the multivariate case (since it is a trivial modification of van der Vaart's proof), but we can verify the weak convergence result for the multivariate case by checking the following four conditions (in addition to the I.I.D. condition in Assumption 1): (I) There exists some (pseudo) metric ϱ with which the space $\mathcal{T}(= \mathcal{X} \times \mathcal{L})$ is a totally bounded. (II) For any $\mathbf{f}_s, \mathbf{f}_t \in \mathcal{F}$, it holds that

$$\sup_{\varrho(s, t) < \delta_n; s, t \in \mathcal{T}} E[\|\mathbf{f}_s(Z_i) - \mathbf{f}_t(Z_i)\|^2] \rightarrow 0 \text{ as } \delta_n \rightarrow 0, \quad (58)$$

where $\|\cdot\|$ stands for the usual Euclid norm. (III) $\sup_{\mathbf{f} \in \mathcal{F}} \sup_{z \in \mathcal{Z}} \|\mathbf{f}(z)\| < \infty$, where \mathcal{Z} denotes the support of a random vector Z_i . (IV) Let \mathcal{F}_q be each q -th element of \mathcal{F} ($q \in \{1, \dots, (d_\eta + J + 1)\}$); and then for each q , there exists some positive constants c_1, c_2 such that

$$\max_{1 \leq q \leq d_\eta + J + 1} \sup_Q N(\epsilon, \mathcal{F}_q, L_2(Q)) \leq c_1 \epsilon^{-c_2}, \quad (59)$$

where the supremum is taken over all probability measures on $\mathcal{T}(= \mathcal{X} \times \mathcal{L})$.

Now, we check the four conditions (I)-(IV). The condition (I) is trivially satisfied by setting ϱ as the usual Euclidean metric, since \mathcal{T} is a bonded subset of the $d_X + 2$ dimensional Euclidean space. To verify (II), recall that by the continuity of Assumption 5, each element of $E[\hat{\mathbf{G}}_\gamma(x, l)\hat{\mathbf{G}}_\gamma(x, l)']$ is continuous. Since the domain $\mathcal{X} \times \mathcal{L}$ is totally bounded, this continuity actually holds uniformly over $(x, l) \in \mathcal{X} \times \mathcal{L}$, leading to (58). The condition III) can be also easily checked by noting that each component of $\hat{\mathbf{G}}_\gamma$ is uniformly bounded over $(x, l) \in \mathcal{X} \times \mathcal{L}$, i.e., $\boldsymbol{\psi}(z)$, which is independent of (x, l) , is uniformly bounded over the support of Z_i (see the proof of Lemma 3); $|\boldsymbol{\xi}_\gamma(z, l)| \leq 2/C$ by Assumption 2; and the uniform boundedness of $g(X_i, L_i, x, l) := 1\{X_i \leq x \ \& \ L_i \leq l\} - F_{X, L}^*(x, l)$ is also obvious by the definition of the C.D.F.

Finally, to verify the condition (IV), recall that for $q \in \{1, \dots, d_\eta\}$, \mathcal{F}_q consists of a singleton (the q -th element of the vector-valued function $\boldsymbol{\psi}(\cdot)$); for $q \in \{d_\eta + 1, \dots, d_\eta + J\}$, $\mathcal{F}_q := \{\boldsymbol{\xi}_\gamma(\cdot, l, r_{q-d_\eta})\}_{l \in \mathcal{L}}$; and for $q = d_\eta + J + 1$, $\mathcal{F}_q := \{g(\cdot, x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$. Here, we verify (59) by explicitly computing the covering number bound of $\{\boldsymbol{\xi}_\gamma(\cdot, l, r)\}_{l \in \mathcal{L}}$ for each $r \in \{r_1, \dots, r_J\}$ and that of $\{g(\cdot, x, l)\}_{(x, l) \in \mathcal{X} \times \mathcal{L}}$. Now, fix any $r \in \{r_1, \dots, r_J\}$ (we here suppress the dependence of variables/functions on r for notational simplicity), and recall that the form of the function defined

in (20): $\xi_\gamma(Z_i, l, r) = 1\{|l - L_i| \leq r\} [1\{X_{1,i} < \gamma\} - s_\gamma^*(l, r)] / \kappa^*(l, r)$. This may be regarded as a composite of several functions, and we consider each subcomponent of $\xi_\gamma(\cdot, l, r)$. First, we let \mathcal{H}_1 be the set of functions: $\mathcal{H}_1 := \{h_1(\cdot|l) \mid l \in \mathcal{L}\}$ with $h_1(\tilde{z}|l, r) = 1\{|\tilde{z} - \tilde{l}| \leq r\}$. Then, by the arguments as in the proof of Lemma 1,

$$\sup_Q N(\epsilon, \mathcal{H}_1, L_2(Q)) \leq \Lambda_1 \epsilon^{-10} \quad \text{for } \epsilon \in (0, 1),$$

where the supremum is over all probability measures on $\mathcal{X} \times \mathcal{L}$; and $\Lambda_1 (> 0)$ is some constant independent of Q (note that $|h_1| \leq 1$ for any $h_1 \in \mathcal{H}_1$). Second, let $\mathcal{H}_2 := \{s_\gamma^*(l, r)\}_{l \in \mathcal{L}}$ and $\mathcal{H}_3 := \{1/\kappa^*(l, r)\}_{l \in \mathcal{L}}$. \mathcal{H}_2 is a set of constant functions (of Z_i) since $s_\gamma^*(l, r)$ is independent of Z_i . By the same token, \mathcal{F}_3 is also a set of constant functions. Therefore, by the definition of the covering number, we can easily derive the following bounds:

$$\sup_Q N(\epsilon, \mathcal{H}_2, L_2(Q)) \leq \Lambda_2 \epsilon^{-1} \quad \text{and} \quad \sup_Q N(C^{-1}\epsilon, \mathcal{H}_3, L_2(Q)) \leq \Lambda_3 \epsilon^{-1},$$

for $\epsilon \in (0, 1)$, where each supremum is over all probability measures on $\mathcal{X} \times \mathcal{L}$; Λ_2 and Λ_3 are some positive constants independent of Q (note that $\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} s^*(l, r) \leq 1$ by the definition of s^* ; and $\sup_{(l,r) \in \mathcal{L} \times [\underline{x}, \infty)} \{1/\kappa^*(l, r)\} \leq 1/C$ by Assumption 2). Third, look at

$$\{\xi_\gamma(\cdot, l, r)\}_{l \in \mathcal{L}} \subset \{h_1 \times [1\{\cdot < \gamma\} - h_2] \times h_3 \mid h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2, h_3 \in \mathcal{H}_3\}.$$

Using this fact with (v) of Lemma 9.9, Lemmas 9.18 and 9.25 of Kosorok (2008), we can see that the covering number of $\{\xi_\gamma(\cdot, l, r)\}_{l \in \mathcal{L}}$ satisfies

$$\sup_Q N(\bar{c}\epsilon, \{\xi_\gamma(\cdot, l, r)\}_{l \in \mathcal{L}}, L_2(Q)) \leq \Lambda_1 \Lambda_3 \Lambda_3 \epsilon^{-12} \quad \text{for } \epsilon \in (0, 1), \quad (60)$$

where the supremum is over all probability measures on $\mathcal{X} \times \mathcal{L}$, and \bar{c} is some constant independent of Q .

We now turn to $\mathcal{F}_{d_\eta+J+1} = \{g(\cdot, x, l)\}_{(x,l) \in \mathcal{X} \times \mathcal{L}}$. To bound its covering number, we use the following two results:

$$\begin{aligned} \sup_Q N(\epsilon, \{1\{\cdot \leq x \ \& \ \cdot \leq l\}\}_{(x,l) \in \mathcal{X} \times \mathcal{L}}, L_2(Q)) &\leq \Lambda_4 \epsilon^{-2[(d_X+4)-1]} = \Lambda_4 \epsilon^{-2(d_X+3)}, \\ \sup_Q N(\epsilon, \{F_{X,L}^*(x, l)\}_{(x,l) \in \mathcal{X} \times \mathcal{L}}, L_2(Q)) &\leq \Lambda_5 \epsilon^{-1}, \end{aligned}$$

where each supremum is over all probability measures on $\mathcal{X} \times \mathcal{L}$; Λ_4 and Λ_5 are some positive constants independent of Q ; the former result follows from (i) of Lemma 9.12 and Lemma 9.2, Kosorok (2008) (recall that we can regard \mathcal{L} as a bounded subset of \mathbb{R}^2 by Assumption 4); and the

latter follows from the fact that the functional set in question consists of only constant functions. From these, we can obtain

$$\sup_Q N(\epsilon, \{g(\cdot, x, l)\}_{(x,l) \in \mathcal{X} \times \mathcal{L}}, L_2(Q)) \leq \Lambda_4 \Lambda_5 \epsilon^{-2d_X - 7} \quad \text{for } \epsilon \in (0, 1), \quad (61)$$

where the supremum is over all probability measures on $\mathcal{X} \times \mathcal{L}$, and \bar{c} is some constant independent of Q . By (60) and (61), we can obtain (59) and check the condition (IV). Thus, part (i) of the theorem is verified, completing the proof of Lemma 4. ■

Steps 5 & 6: Functional delta method.

Next, we use the functional delta method to find the asymptotic distribution of $\hat{U}(\gamma)$ (for each γ). For our formal presentation, we need to define the domain of the functional Ψ precisely. For this purpose, let $C[\mathcal{L}]$ be the set of all J -dimensional vector-valued functions, $\mathbf{s} : \mathcal{L} \rightarrow (-\infty, \infty)^J$ with $\mathbf{s} = (s_1, \dots, s_J)'$ and each element s_j being a càdlàg function: $\mathcal{L} \rightarrow (-\infty, \infty)$, and $\text{BV}[\mathcal{X} \times \mathcal{L}]$ be the set of all càdlàg functions, $\mathcal{X} \times \mathcal{L} \rightarrow (-\infty, \infty)$, of variation bounded by some $M(< \infty)$.¹⁸ Given these, we can re-define Ψ as a functional $\mathbb{D} := \mathbb{R}^{d_n} \times C[\mathcal{L}] \times \text{BV}[\mathcal{X} \times \mathcal{L}] \rightarrow (-\infty, \infty)$. This re-formulation allows us to verify the following differentiability condition required for the functional delta method:

Lemma 5 *Suppose that Assumptions 3-5 hold. Then, the functional $\Psi(\cdot : \mathbb{D} \rightarrow \mathbb{R})$ whose form is given in (11) is Hadamard differentiable at $\boldsymbol{\rho}_\gamma^*$ tangentially to the domain \mathbb{D} , and its Hadamard derivative $\Gamma_\gamma^*(\boldsymbol{\sigma})(\cdot : \mathbb{D} \rightarrow \mathbb{R})$ is given by*

$$\begin{aligned} \Gamma_\gamma^*(\boldsymbol{\sigma}) &= \int \left\{ 1\{x_1 \geq \gamma\} \times \phi \left(x' \delta_0^* + \sum_{j=1}^J \beta_{j,0}^* s_\gamma^*(l, r_j) \right) \right. \\ &\times \left[x'(\delta_0 - \delta_0^*) + \sum_{j=1}^J (\beta_{j,0} - \beta_{j,0}^*) s_\gamma^*(l, r_j) + \sum_{j=1}^J \beta_{j,0}^* (s_j(l) - s_\gamma^*(l, r_j)) \right] \\ &+ 1\{x_1 < \gamma\} \times \phi \left(x' \delta_1^* + \sum_{j=1}^J \beta_{j,1}^* s_\gamma^*(l, r_j) \right) \\ &\times \left[x'(\delta_1 - \delta_1^*) + \sum_{j=1}^J (\beta_{j,1} - \beta_{j,1}^*) s_\gamma^*(l, r_j) + \sum_{j=1}^J \beta_{j,1}^* (s_j(l) - s_\gamma^*(l, r_j)) \right] \left. \right\} dF_{X,L}^*(x, l) \\ &+ \int \left\{ 1\{x_1 \geq \gamma\} \times \Phi \left(x' \delta_0^* + \sum_{j=1}^J \beta_{j,0}^* s_\gamma^*(l, r_j) \right) \right. \\ &+ 1\{x_1 < \gamma\} \times \Phi \left(x' \delta_1^* + \sum_{j=1}^J \beta_{j,1}^* s_\gamma^*(l, r_j) \right) \left. \right\} d(F_{X,L}(x, l) - F_{X,L}^*(x, l)), \end{aligned} \quad (62)$$

for each γ , where ϕ is the probability density function of the standard normal; $\boldsymbol{\sigma} := (\boldsymbol{\rho} - \boldsymbol{\rho}_\gamma^*) \in \mathbb{D}$ with $\boldsymbol{\rho} = (\delta_0', \delta_1', \beta_{1,0}, \dots, \beta_{J,0}, \beta_{1,1}, \dots, \beta_{J,1}, s_1(\cdot), \dots, s_J(\cdot), F(\cdot, \cdot))'$, a generic element of \mathbb{D} .

¹⁸A càdlàg function is a function which is right-continuous and whose left limits exist at any points.

Proof. Recall the functional form of Φ , the linear specification in (7), and the form of the likelihood function (8). Then, we can easily verify the existence of the Hadamard derivative and its form (62) by using the standard chain rule for Hadamard differentiable functionals and arguments analogous to those in the proof of Lemma 20.10 of van der Vaart (1998). We omit details of the proof for brevity. ■

Given these preparations, we can now prove Theorem 1.

Proof of Theorem 1. We here use the functional delta method (see, e.g., Theorem 20.8 of van der Vaart, 1998). To do so, let $\bar{\Gamma}_\gamma^*$ be the Hadamard derivative of the functional $\Psi(\cdot: \mathbb{D} \rightarrow \mathbb{R})$ at $\bar{\rho}_\gamma^* := (\bar{\delta}_0^*, \bar{\delta}_1^*, \mathbf{0}'_{3J}, F^*(\cdot, \cdot))'$ with $\mathbf{0}_{3J}$ being the $3J$ -dimensional vector of zeros. By the same arguments as for Lemma 5, we can verify the existence of $\bar{\Gamma}_\gamma^*$ and derive its form:

$$\begin{aligned} \bar{\Gamma}_\gamma^*(\bar{\sigma}) &= \int \left\{ 1\{x_1 \geq \gamma\} \phi(x'\bar{\delta}_0^*) x'(\delta_0 - \bar{\delta}_0^*) + 1\{x_1 < \gamma\} \phi(x'\bar{\delta}_1^*) x'(\delta_1 - \bar{\delta}_1^*) \right\} dF_{X,L}^*(x, l) \\ &+ \int \left\{ 1\{x_1 \geq \gamma\} \Phi(x'\bar{\delta}_0^*) + 1\{x_1 < \gamma\} \Phi(x'\bar{\delta}_1^*) \right\} d(F_{X,L}(x, l) - F_{X,L}^*(x, l)), \end{aligned} \quad (63)$$

for each γ , where a generic form of $\bar{\sigma}$ can be written as $\bar{\sigma} = (\boldsymbol{\rho} - \bar{\rho}_\gamma^*) \in \mathbb{D}$.

Let $\bar{\mathbb{D}} := \mathbb{D} \times \mathbb{R}^{2d_X}$, and J_γ^* be a map from $\bar{\mathbb{D}}$ to \mathbb{R}^2 :

$$J_\gamma^*(\mathbf{h}) := (\Gamma_\gamma^*(\boldsymbol{\sigma}), \bar{\Gamma}_\gamma^*(\bar{\sigma}))',$$

where $\boldsymbol{\sigma}(x, l)$ and $\bar{\boldsymbol{\sigma}}(x, l)$ are $2d_X + 3J + 1$ -dimensional vector (for each $(x, l) \in \mathcal{X} \times \mathcal{L}$); $\boldsymbol{\sigma}(x, l)$ is the first $2d_X + 3J + 1$ components of $\mathbf{h}(x, l)$; $\bar{\boldsymbol{\sigma}}(x, l)$ consists of the last $2d_X$ components of $\mathbf{h}(x, l)$, $\mathbf{0}'_{3J}$ and the $(2d_X + 3J + 1)$ -th component of $\mathbf{h}(x, l)$ (with the stated order).

Now, applying the functional delta method, together with the results of Lemmas 4, we can obtain

$$\sqrt{n}(\hat{U}(\gamma) - U^*(\gamma), \tilde{U}(\gamma) - \bar{U}^*(\gamma))' \xrightarrow{d} J_\gamma^*(\mathbf{H}_\gamma^*).$$

Since \mathbf{H}_γ^* is a mean-zero Gaussian process (as defined in Lemma 4) and J_γ^* is a continuous linear map (recall the forms of Γ_γ^* and $\bar{\Gamma}_\gamma^*$ given in (62) and (63)), $\Gamma_\gamma^*(\mathbf{G}_\gamma^*)$ is normally distributed with mean zero, and the form of the variance matrix $\boldsymbol{\Sigma}(\gamma)$ can be computed as in Section 3.9.2 of VW (1996) (with the aid of the Riesz representation and Fubini theorems). This completes the proof. ■

A.2 Bootstrap validity

Step 7 Bootstrap validity: We here provide the proof the validity of our bootstrap inference (Theorem 2) with two auxiliary lemmas which follow the proof of the theorem.

Proof of Theorem 2. We below consider the bootstrap consistency result only for $\sqrt{n}(\hat{U}^\#(\gamma) - \hat{U}(\gamma))$. The joint result for $\sqrt{n}(\hat{U}^\#(\gamma) - \hat{U}(\gamma), \tilde{U}^\#(\gamma) - \tilde{U}(\gamma))'$ can be shown quite analogously, and we omit details.

We first consider a convergence result for $\sqrt{n}[\hat{\rho}_\gamma^\#(x, l) - \hat{\rho}_\gamma(x, l)]$. To this end, we define

$$\tilde{\mathbf{G}}_\gamma^\#(x, l) := (1/\sqrt{n}) \sum_{i=1}^n \begin{pmatrix} \hat{\psi}(Z_i^\#) \\ \hat{\xi}_\gamma(Z_i^\#, l) \\ \hat{g}(Z_i^\#, x, l) \end{pmatrix}, \quad (64)$$

where

$$\begin{aligned} \hat{\psi}(Z_i^\#) &: = -[\hat{\mathbf{M}}]^{-1} \left\{ \mathbf{m}(Z_i^\#, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i^\#)) + \hat{\boldsymbol{\alpha}}(Z_i^\#) \right\}; \\ \hat{\mathbf{M}} &: = n^{-1} \sum_{i=1}^n E \left[(\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) \right]; \\ \hat{\boldsymbol{\alpha}}(z) &: = n^{-1} \sum_{u=1}^n \hat{\mathbf{D}}(Z_u) \hat{\boldsymbol{\xi}}(z, L_u); \quad \hat{\mathbf{D}}(z) := (\partial/\partial \mathbf{s}') \mathbf{m}(z, \hat{\boldsymbol{\eta}}, \mathbf{s}) \Big|_{\mathbf{s}=\hat{\mathbf{s}}(l)}; \\ \hat{\boldsymbol{\xi}}(z, L_u) &: = (\hat{\xi}(z, L_u, r_1), \dots, \hat{\xi}(z, L_u, r_J))'; \end{aligned} \quad (65)$$

$$\begin{aligned} \hat{\xi}(z, L_u, r) &: = 1\{|L_u - l| \leq r\} [d - \hat{\mathbf{s}}(L_u, r)] / \hat{\kappa}(L_u, r); \\ \hat{\xi}_\gamma(Z_i^\#, l) &: = (\hat{\xi}_\gamma(Z_i^\#, l, r_1), \dots, \hat{\xi}_\gamma(Z_i^\#, l, r_J))'; \\ \hat{\xi}_\gamma(Z_i^\#, l, r) &: = 1\{|l - L_i^\#| \leq r\} \left[1\{X_{1,i}^\# < \gamma\} - \hat{\mathbf{s}}_\gamma(l, r) \right] / \hat{\kappa}(l, r); \\ \hat{g}(Z_i^\#, x, l) &: = 1\{X_i^\# \leq x \ \& \ L_i^\# \leq l\} - \hat{F}_{X,L}(x, l). \end{aligned} \quad (66)$$

We note that $\tilde{\mathbf{G}}_\gamma^\#(x, l)$ is a bootstrap counterpart of $\hat{\mathbf{G}}_\gamma(x, l)$ (i.e., it is constructed by replacing all the true parameters in $\hat{\mathbf{G}}_\gamma(x, l)$ with their empirical estimates, and that by the definitions,

$$E^\#[\hat{\psi}(Z_i^\#)] = 0; \quad E^\#[\hat{\xi}_\gamma(Z_i^\#, l)] = 0; \quad \text{and} \quad E^\#[\hat{g}(Z_i^\#, x, l)] = 0, \quad (67)$$

where $E^\#$ denotes the mathematical expectation with respect to the E.D.F., e.g., $E^\#[\hat{\psi}(Z_i^\#)] = (1/n) \sum_{i=1}^n \hat{\psi}(Z_i)$.

Now, we can easily show that

$$\sup_{(x,l) \in \mathcal{X} \times \mathcal{L}} \left| \sqrt{n}[\hat{\rho}_\gamma^\#(x, l) - \hat{\rho}_\gamma(x, l)] - \tilde{\mathbf{G}}_\gamma^\#(x, l) \right| \xrightarrow[\#]{P} 0, \quad (68)$$

whose proof proceeds in the same way as the proof of (15) (but with replacing the true C.D.F. and estimators replaced by the E.D.F. and their bootstrap counterparts, respectively), and we omit details. We further consider the following object:

$$\hat{\mathbf{G}}_\gamma^\#(x, l) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \psi(Z_i^\#) \\ \xi_\gamma(Z_i^\#, l) \\ g(Z_i^\#, x, l) \end{pmatrix} - \hat{\mathbf{G}}_\gamma(x, l), \quad (69)$$

which is a bootstrap counterpart of $\hat{\mathbf{G}}_\gamma(x, l)$ (note that $\hat{\mathbf{G}}_\gamma^\#(x, l)$ is re-centered and $E[\hat{\mathbf{G}}_\gamma(x, l)] = 0$). Then, we show the following convergence result:

$$\sup_{(x, l) \in \mathcal{X} \times \mathcal{L}} \left| \tilde{\mathbf{G}}_\gamma^\#(x, l) - \hat{\mathbf{G}}_\gamma^\#(x, l) \right| \xrightarrow[\#]{P} 0. \quad (70)$$

To prove this, recall the zero-mean conditions in (67), as well as the definitions of $\tilde{\mathbf{G}}_\gamma^\#$ and $\hat{\mathbf{G}}_\gamma^\#$ in (64) and (69), respectively. Then, we can write

$$\begin{aligned} & \tilde{\mathbf{G}}_\gamma^\#(x, l) - \hat{\mathbf{G}}_\gamma^\#(x, l) \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \hat{\psi}(Z_i^\#) - \psi(Z_i^\#) \\ \hat{\xi}_\gamma(Z_i^\#, l) - \xi_\gamma(Z_i^\#, l) \\ \hat{g}(Z_i^\#, x, l) - g(Z_i^\#, x, l) \end{pmatrix} - \frac{1}{\sqrt{n}} \sum_{k=1}^n \begin{pmatrix} \hat{\psi}(Z_k) - \psi(Z_k) \\ \hat{\xi}_\gamma(Z_k, l) - \xi_\gamma(Z_k, l) \\ \hat{g}(Z_k, x, l) - g(Z_k, x, l) \end{pmatrix}, \end{aligned} \quad (71)$$

where the convergence results of the components involving $\hat{\psi}(Z_i^\#)$, $\psi(Z_i^\#)$, $\hat{\xi}_\gamma$ and ξ_γ are presented in Lemmas 7; and the convergence of the last row component can be shown analogously and its proof is omitted (recall the fact that the (usual) convergence in probability implies the conditional weak convergence in probability).

Note that $\hat{\mathbf{G}}_\gamma^\#$ is simply in the form of a bootstrapped empirical process, to which we can directly apply the bootstrap CLT (e.g., Theorem 3.6.1 of VW, 1996; or Theorem 2.6 of Kosorok, 2008). That is, by the weak-convergence/Donsker result of $\hat{\mathbf{G}}_\gamma$ ((i) of Lemma 4), it holds that

$$\hat{\mathbf{G}}_\gamma^\# \xrightarrow[\#]{P} \mathbf{G}_{\gamma}^*,$$

where the conditional weak convergence in probability take places in the space $l^\infty(\mathcal{X} \times \mathcal{L})$ (given the original observations $\{Z_i\}_{i=1}^n$). This, together with (68) and (70), implies that

$$\sqrt{n}[\hat{\rho}_\gamma^\# - \hat{\rho}_\gamma] \xrightarrow[\#]{P} \mathbf{G}_\gamma^*. \quad (72)$$

Now, we have obtained the weak convergence of $\sqrt{n}[\hat{\rho}_\gamma - \rho_\gamma^*]$ ((ii) of Lemma 4), the conditional weak convergence of $\sqrt{n}[\hat{\rho}_\gamma^\# - \hat{\rho}_\gamma]$ (the result (72)), and the Hadamard differentiability of Ψ (Lemma 5). By these three results, we can apply the functional delta method for bootstrap (Theorem 3.9.11 of VW, 1996), which verifies that $\sqrt{n}(\hat{U}^\#(\gamma) - \hat{U}(\gamma)) \xrightarrow[\#]{P} N(0, \Sigma(\gamma))$ as $n \rightarrow \infty$, as desired (note that the separability requirement of Theorem 3.9.11 of VW (1996) is trivially satisfied since \mathbf{G}_γ^* is a Gaussian process and $\mathcal{X} \times \mathcal{L}$ is a subset of $\mathbb{R}^{d_x} \times \mathbb{R}^2$). This completes the proof of the theorem.

■

Lemma 6 *Suppose that the same conditions hold as in Theorem 2. Then,*

$$\sup_{(x,l) \in \mathcal{X} \times \mathcal{L}} (1/\sqrt{n}) \left\{ \sum_{i=1}^n [\hat{\psi}(Z_i^\#) - \psi(Z_i^\#)] - \sum_{k=1}^n [\hat{\psi}(Z_k) - \psi(Z_k)] \right\} = o_P(1).$$

Proof. Let $(M_{n,1}, \dots, M_{n,n})$ be a vector of multinomial-distributed random variables with the parameter n , and the probability vector $(1/n, \dots, 1/n)$ which is independent of $\{Z_i\}_{i=1}^n$. Then, by the definition of our bootstrap procedure, we can write

$$\sum_{i=1}^n [\hat{\psi}(Z_i^\#) - \psi(Z_i^\#)] = \sum_{i=1}^n M_{n,i} [\hat{\psi}(Z_i) - \psi(Z_i)]. \quad (73)$$

By the (uniform) law of large numbers, which follows from the same arguments for deriving (45) in Lemma 2, we can write $\hat{\mathbf{M}}^{-1} = [\mathbf{M}^*]^{-1} + \mathbf{e}_n$ with $\mathbf{e}_n = o_P(1)$. Therefore, recalling the definition of $\hat{\psi}$, we can also write

$$\hat{\psi}(Z_i) = -[\mathbf{M}^*]^{-1} \{ \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(Z_i)) + \hat{\boldsymbol{\alpha}}(Z_i) \} - \mathbf{e}_n \{ \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) + \hat{\boldsymbol{\alpha}}(Z_i) \}. \quad (74)$$

Now, by these expressions (73)-(74), we have

$$\begin{aligned} & (1/\sqrt{n}) \left\{ \sum_{i=1}^n [\hat{\psi}(Z_i^\#) - \psi(Z_i^\#)] - \sum_{k=1}^n [\hat{\psi}(Z_k) - \psi(Z_k)] \right\} \\ &= (1/\sqrt{n}) \sum_{i=1}^n (M_{n,i} - 1) [\hat{\psi}(Z_i) - \psi(Z_i)] \\ &= -(1/\sqrt{n}) [\mathbf{M}^*]^{-1} \sum_{i=1}^n (M_{n,i} - 1) [\mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) - \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))]; \\ & \quad + (1/\sqrt{n}) [\mathbf{M}^*]^{-1} \sum_{i=1}^n (M_{n,i} - 1) [\hat{\boldsymbol{\alpha}}(Z_i) - \boldsymbol{\alpha}(Z_i)]; \\ & \quad - (1/\sqrt{n}) \mathbf{e}_n \sum_{i=1}^n M_{n,i} [\mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) + \hat{\boldsymbol{\alpha}}(Z_i)]. \\ &= [\mathbf{M}^*]^{-1} [-\mathbf{J}_{n,1} + \mathbf{J}_{n,2}] - \mathbf{J}_{n,3}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{J}_{n,1} & : = (1/\sqrt{n}) \sum_{i=1}^n (M_{n,i} - 1) [\mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) - \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))]; \\ \mathbf{J}_{n,2} & : = (1/\sqrt{n}) \sum_{i=1}^n (M_{n,i} - 1) [\hat{\boldsymbol{\alpha}}(Z_i) - \boldsymbol{\alpha}(Z_i)]; \\ \mathbf{J}_{n,3} & : = (1/\sqrt{n}) \mathbf{e}_n \sum_{i=1}^n M_{n,i} [\mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) + \hat{\boldsymbol{\alpha}}(Z_i)]. \end{aligned}$$

We can show that

$$\mathbf{J}_{n,1} = O_P((\ln n)/\sqrt{n}); \quad (75)$$

$$\mathbf{J}_{n,2} = O_P((\ln n)/\sqrt{n}); \quad (76)$$

$$\mathbf{J}_{n,3} = o_P(1), \quad (77)$$

whose proofs are below provided. These results imply the conclusion of the lemma.

To derive the convergence result for $\mathbf{J}_{n,1}$, observe that

$$\begin{aligned}
& \mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) - \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) \\
&= (\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) [\hat{\mathbf{s}}(L_i) - \mathbf{s}^*(L_i)] + (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \\
&\quad + O(\sup_{l \in \mathcal{L}} \|\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)\|^2) + O(\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|^2) \\
&= (\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) [n^{-1} \sum_{u=1}^n \boldsymbol{\xi}(Z_u, L_i)] + (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) [\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*] \\
&\quad + O_P((\ln n)/n), \tag{78}
\end{aligned}$$

uniformly over i , where the equalities follow from the Taylor expansion, the (uniform) boundedness of the derivatives of \mathbf{m} , the \sqrt{n} -consistency of $\hat{\boldsymbol{\eta}}$, and the result (18) in Lemma 1. This implies that

$$\begin{aligned}
\mathbf{J}_{n,1} &= (1/\sqrt{n}) \sum_{i=1}^n (M_{n,i} - 1) (\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) [n^{-1} \sum_{u=1}^n \boldsymbol{\xi}(Z_u, L_i)] \\
&\quad + (1/\sqrt{n}) \sum_{i=1}^n (M_{n,i} - 1) (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + O_P((\ln n)/\sqrt{n}) \\
&= n^{-3/2} \left\{ \sum_{i=1}^n \mathbf{b}(i, i) + \sum \sum_{1 \leq i, u \leq n; i \neq u} \mathbf{b}(i, u) \right\} + O_P(1/\sqrt{n}) + O_P((\ln n)/\sqrt{n}) \\
&= O_P((\ln n)/\sqrt{n}), \tag{79}
\end{aligned}$$

where

$$\mathbf{b}(i, u) := [M_{n,i} - 1] (\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) \boldsymbol{\xi}(Z_u, L_i).$$

The second equality in (79) holds since

$$\begin{aligned}
& \left\| \sum_{i=1}^n (M_{n,i} - 1) (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \right\|^2 \\
&= O_P(\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|^2) \times \sum_{i=1}^n (M_{n,i} - 1)^2 + \sum \sum_{1 \leq i, k \leq n; i \neq k} (M_{n,i} - 1) (M_{n,k} - 1) \\
&\quad \times (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*)' [(\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))] (\partial/\partial \boldsymbol{\eta}') \mathbf{m}(Z_k, \boldsymbol{\eta}^*, \mathbf{s}^*(L_k)) (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \\
&= O_P(1),
\end{aligned}$$

which follows from the uniform boundedness of $(\partial/\partial \boldsymbol{\eta}') \mathbf{m}$, the \sqrt{n} -consistency of $\hat{\boldsymbol{\eta}}$, the independence between $(M_{n,1}, \dots, M_{n,n})$ and $\{Z_i\}_{i=1}^n$, and the facts that

$$\begin{aligned}
E[(M_{n,i} - 1)^2] &= \text{Var}[M_{n,i}] = n(1/n)[1 - (1/n)] = 1 - 1/n, \\
E[(M_{n,i} - 1)(M_{n,k} - 1)] &= \text{Cov}[M_{n,i}, M_{n,k}] = -1/n \quad (\text{for } i \neq k).
\end{aligned}$$

To see the validity of third equality in (79), we observe that $\sum_{i=1}^n \mathbf{b}(i, i) = O_P(n)$ and

$$\begin{aligned}
& E[\|\sum \sum_{1 \leq i, u \leq n; i \neq u} \mathbf{b}(i, u)\|^2] = \sum \sum_{1 \leq i, u \leq n; i \neq u} \{E[\mathbf{b}(i, u)' \mathbf{b}(i, u)] + E[\mathbf{b}(i, u)' \mathbf{b}(u, i)]\} \\
& + \sum \sum \sum_{1 \leq i, u, v \leq n; i \neq u \neq v} \{E[\mathbf{b}(i, u)' \mathbf{b}(u, v)] + E[\mathbf{b}(i, u)' \mathbf{b}(v, i)] \\
& + E[\mathbf{b}(i, u)' \mathbf{b}(u, v)] + E[\mathbf{b}(i, u)' \mathbf{b}(v, u)]\} \\
& + \sum \sum \sum \sum_{1 \leq i, u, v, w \leq n; i \neq u \neq v \neq w} E[\mathbf{b}(i, u)' \mathbf{b}(v, w)] \\
& = \sum \sum_{1 \leq i, u \leq n; i \neq u} \{E[\mathbf{b}(i, u)' \mathbf{b}(i, u)] + E[\mathbf{b}(i, u)' \mathbf{b}(u, i)]\} \\
& + \sum \sum \sum_{1 \leq i, u, v \leq n; i \neq u \neq v} E[\mathbf{b}(i, u)' \mathbf{b}(v, u)] \\
& = n^2 \times O(1) + n^3 \times O(1/n) = O(n^2), \tag{80}
\end{aligned}$$

where the equalities use the following facts: $E[\mathbf{b}(i, u) | Z_i] = 0$ for $i \neq u$;

$$E[\mathbf{b}(i, u)' \mathbf{b}(i, u)] = O(1) \quad \text{and} \quad E[\mathbf{b}(i, u)' \mathbf{b}(u, i)] = O(1),$$

uniformly over i and u ; and

$$\begin{aligned}
& E[\mathbf{b}(i, u)' \mathbf{b}(v, u)] \\
& = \text{Cov}[M_{n,i}, M_{n,v}] \times E[\boldsymbol{\xi}(Z_u, L_i)' [(\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))]' (\partial/\partial \mathbf{s}') \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_v)) \boldsymbol{\xi}(Z_u, L_v)] \\
& = -(1/n) \times O(1), \quad \text{uniformly over } i, u \text{ and } v \ (i \neq u \neq v),
\end{aligned}$$

which follows from the independence between $(M_{n,1}, \dots, M_{n,n})$ and $\{Z_i\}_{i=1}^n$, and the uniform boundedness of $(\partial/\partial \mathbf{s}') \mathbf{m}$ and $\boldsymbol{\xi}$. Now, we have proved the desired result (75)

We now look at $\mathbf{J}_{n,2}$. Analogously to (83) and (84), we can write

$$\hat{\boldsymbol{\xi}}(Z_i, l, r) = \boldsymbol{\xi}(Z_i, l, r) + \boldsymbol{\nu}_n(Z_i, L_u, r) + O_P((\ln n)/n),$$

uniformly over Z_i, L_u, l and r , where

$$\boldsymbol{\nu}_n(\tilde{z}, l, r) := [\varepsilon_n(l, r) / \kappa^*(l, r)] \mathbf{1}\{\|l - \tilde{l}\| \leq r\} + [\epsilon_n(l, r) / \kappa^*(l, r)] \mathbf{1}\{\|l - \tilde{l}\| \leq r\} [d - s_\gamma^*(l, r)];$$

$\varepsilon_n(l, r)$ and $\epsilon_n(l, r)$ are the same components used in (82), and $\boldsymbol{\nu}_n(\tilde{z}, l, r) = O_P(\sqrt{(\ln n)/n})$ uniformly over \tilde{z}, l and r . Let $\boldsymbol{\nu}_n(\tilde{z}, l) := (\boldsymbol{\nu}_n(\tilde{z}, l, r_1), \dots, \boldsymbol{\nu}_n(\tilde{z}, l, r_J))'$. Then, we can write

$$\hat{\boldsymbol{\xi}}(Z_i, L_u) = \boldsymbol{\xi}(Z_i, L_u) + \boldsymbol{\nu}_n(Z_i, L_u) + O_P((\ln n)/n). \tag{81}$$

For notational simplicity, we denote by $\mathbf{w}_\iota(z, \boldsymbol{\eta}, \mathbf{s})$ the ι -th row vector of the by the d_η -by- J matrix $(\partial/\partial \mathbf{s}') \mathbf{m}(z, \boldsymbol{\eta}, \mathbf{s})$ ($1 \leq \iota \leq d_\eta$), i.e., each $\mathbf{w}_\iota(z, \boldsymbol{\eta}, \mathbf{s})$ is a 1-by- J vector with

$$(\partial/\partial \mathbf{s}') \mathbf{m}(z, \boldsymbol{\eta}, \mathbf{s}) = \begin{pmatrix} \mathbf{w}_1(z, \boldsymbol{\eta}, \mathbf{s}) \\ \vdots \\ \mathbf{w}_{d_\eta}(z, \boldsymbol{\eta}, \mathbf{s}) \end{pmatrix}.$$

Then, by using the Taylor expansion (as in 78)), the boundedness of the derivatives of \mathbf{w}_ι , and the expression (81),

$$\begin{aligned} & \mathbf{w}_\iota(Z_u, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_u)) \hat{\boldsymbol{\xi}}(Z_i, L_u) = \mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \hat{\boldsymbol{\xi}}(Z_i, L_u) \\ &= (\partial/\partial \mathbf{s}') \left[\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \hat{\boldsymbol{\xi}}(Z_i, L_u) \right] [\hat{\mathbf{s}}(L_u) - \mathbf{s}^*(L_u)] + O(\sup_{l \in \mathcal{L}} \|\hat{\mathbf{s}}(l) - \mathbf{s}^*(l)\|^2) \\ & \quad + (\partial/\partial \boldsymbol{\eta}') \left[\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \hat{\boldsymbol{\xi}}(Z_i, L_u) \right] (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + O(\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*\|^2) \\ &= \mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) [\boldsymbol{\xi}(Z_i, L_u) + \boldsymbol{\nu}_n(Z_i, L_u)] \\ & \quad + (\partial/\partial \mathbf{s}') [\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u)] [\hat{\mathbf{s}}(L_u) - \mathbf{s}^*(L_u)] \\ & \quad + (\partial/\partial \boldsymbol{\eta}') [\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u)] (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + O_P((\log n)/n), \end{aligned}$$

uniformly over u and i , for $\iota = 1, \dots, d_\eta$. Therefore, recalling the following expression:

$$\begin{aligned} & \hat{\boldsymbol{\alpha}}(Z_i) - \boldsymbol{\alpha}(Z_i) \\ &= n^{-1} \sum_{u=1}^n (\partial/\partial \mathbf{s}') \mathbf{m}(Z_u, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_u)) \hat{\boldsymbol{\xi}}(Z_i, L_u) - \int (\partial/\partial \mathbf{s}') \mathbf{m}(z, \boldsymbol{\eta}^*, \mathbf{s}^*(l)) \boldsymbol{\xi}(Z_i, l) dF_Z(z), \end{aligned}$$

we can write

$$\begin{aligned} & \text{the } \iota\text{-th element of the column vector } \mathbf{J}_{n,2} \\ &= K_{n,1}(\iota) + K_{n,2}(\iota) + K_{n,3}(\iota) + K_{n,4}(\iota) + O_P((\log n)/\sqrt{n}), \end{aligned}$$

where

$$\begin{aligned} K_{n,1}(\iota) & : = (1/n^{3/2}) \sum_{i=1}^n \sum_{u=1}^n (M_{n,i} - 1) \\ & \quad \times \left[\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u) - \int \mathbf{w}_\iota(z, \boldsymbol{\eta}^*, \mathbf{s}^*(l)) \boldsymbol{\xi}(Z_i, l) dF_Z(z) \right]; \\ K_{n,2}(\iota) & : = (1/n^{3/2}) \sum_{i=1}^n \sum_{u=1}^n (M_{n,i} - 1) \mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\nu}_n(Z_i, L_u); \\ K_{n,3}(\iota) & : = (1/n^{3/2}) \sum_{i=1}^n \sum_{u=1}^n (M_{n,i} - 1) \frac{\partial}{\partial \mathbf{s}'} [\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u)] [\hat{\mathbf{s}}(L_u) - \mathbf{s}^*(L_u)]; \\ K_{n,4}(\iota) & : = (1/n^{3/2}) \sum_{i=1}^n \sum_{u=1}^n (M_{n,i} - 1) \frac{\partial}{\partial \boldsymbol{\eta}'} [\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u)] (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*). \end{aligned}$$

We can show that $E[|K_{n,1}(\iota)|^2] = O(1/n)$ by considering its expansion as in (80) and using the facts that

$$\begin{aligned} E[\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u) - \int \mathbf{w}_\iota(z, \boldsymbol{\eta}^*, \mathbf{s}^*(l)) \boldsymbol{\xi}(Z_i, l) dF_Z(z) | Z_i] &= 0; \\ E[\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u) - \int \mathbf{w}_\iota(z, \boldsymbol{\eta}^*, \mathbf{s}^*(l)) \boldsymbol{\xi}(Z_i, l) dF_Z(z) | Z_u] &= 0, \end{aligned}$$

where the latter follows from the definition of $\boldsymbol{\xi}(Z_i, l)$ (given in (19)). We can also show that $K_{n,2}(\iota) = O_P(\sqrt{(\ln n)/n})$ and $K_{n,3}(\iota) = O_P(\sqrt{(\ln n)/n})$ for each ι . To prove these two, we expand each of $E[|K_{n,2}(\iota)|^2]$ and $E[|K_{n,3}(\iota)|^2]$ as in (80) (both exist since \mathbf{w}_ι and its derivatives, $\boldsymbol{\nu}_n(Z_i, L_u)$, $\boldsymbol{\xi}(Z_i, L_u)$ and $[\hat{\mathbf{s}}(L_u) - \mathbf{s}^*(L_u)]$ are uniformly bounded), and then use the results: $\boldsymbol{\xi}(Z_i, L_u) = O_P(\sqrt{(\ln n)/n})$ and $[\hat{\mathbf{s}}(L_u) - \mathbf{s}^*(L_u)] = O_P(\sqrt{(\ln n)/n})$ are $O_P(\sqrt{(\ln n)/n})$ uniformly over i and u ; the independence condition between $(M_{n,1}, \dots, M_{n,n})$ and $\{Z_i\}_{i=1}^n$; and the following fact:

$$E[(M_{n,i} - 1)(M_{n,u} - 1)(M_{n,v} - 1)(M_{n,w} - 1)] = 3(n-2)/n^3 = O(1/n^2),$$

for $1 \leq i \neq u \neq v \neq w \leq n$, which follows from the generalized factorial moment formula for multinomial distributions (page 67, Mosimann, 1962). By using analogous arguments with the \sqrt{n} -consistency of $\hat{\boldsymbol{\eta}}$ and the fact that

$$E\left[\frac{\partial}{\partial \boldsymbol{\eta}'} [\mathbf{w}_\iota(Z_u, \boldsymbol{\eta}^*, \mathbf{s}^*(L_u)) \boldsymbol{\xi}(Z_i, L_u)] \Big| Z_u\right] = 0,$$

we can also show $|K_{n,4}(\iota)|^2 = O_P(1/n)$. From these, we can obtain the desired result (76).

Finally, we look at $\mathbf{J}_{n,3}$. Using the fact that $\mathbf{e}_n = O_P(1/\sqrt{n})$, we can write

$$\begin{aligned} \mathbf{J}_{n,3} &= O_P(1/n) \times \left\{ \sum_{i=1}^n M_{n,i} [\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \boldsymbol{\alpha}(Z_i)] \right. \\ &\quad \left. + \sum_{i=1}^n M_{n,i} [\mathbf{m}(Z_i, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(L_i)) - \mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i))] + \sum_{i=1}^n M_{n,i} [\hat{\boldsymbol{\alpha}}(Z_i) - \boldsymbol{\alpha}(Z_i)] \right\}. \end{aligned}$$

The first term within the curly braces is $O_P(1/\sqrt{n})$ since $E[\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \boldsymbol{\alpha}(Z_i)] = 0$ and thus

$$E[|\sum_{i=1}^n M_{n,i} [\mathbf{m}(Z_i, \boldsymbol{\eta}^*, \mathbf{s}^*(L_i)) + \boldsymbol{\alpha}(Z_i)]|^2] = O(n).$$

The second and third terms within the curly braces are $o_P(n)$ since we have the uniform convergence of $[\mathbf{m}(z, \hat{\boldsymbol{\eta}}, \hat{\mathbf{s}}(l)) - \mathbf{m}(z, \boldsymbol{\eta}^*, \mathbf{s}^*(l))]$ and $[\hat{\boldsymbol{\alpha}}(z) - \boldsymbol{\alpha}(z)]$ (over z ; in probability), which can be easily shown by using the Taylor expansion arguments with (uniform) convergence results of $\hat{\boldsymbol{\eta}}$ and $\hat{\mathbf{s}}(l)$.

Therefore, we have obtained $\mathbf{J}_{n,3} = O_P(1/n) \times \{O_P(\sqrt{n}) + o_P(n)\} = o_P(1)$, as desired. Now, the proof of Lemma 6 is completed. ■

Lemma 7 *Suppose that the same conditions hold as in Theorem 2. Then,*

$$\sup_{(x,l) \in \mathcal{X} \times \mathcal{L}} (1/\sqrt{n}) \left\{ \sum_{i=1}^n [\hat{\xi}_\gamma(Z_i^\#, l) - \xi_\gamma(Z_i^\#, l)] - \sum_{k=1}^n [\hat{\xi}_\gamma(Z_k, l) - \xi_\gamma(Z_k, l)] \right\} \xrightarrow{P} 0.$$

Proof. Noting the uniform convergence results of $\hat{\kappa}(l, r)$ and $\hat{s}_\gamma(l, r)$ in Lemma 1, and the boundedness condition of $\kappa^*(l, r)$ in Assumption 2, we can write

$$\hat{s}_\gamma(l, r) = s^*(l, r) + \varepsilon_n(l, r) \quad \text{and} \quad 1/\hat{\kappa}(l, r) = [1/\kappa^*(l, r)] + \epsilon_n(l, r), \quad (82)$$

where $\sup_{(l,r) \in \mathcal{X} \times [r, \infty)} \max\{|\varepsilon_n(l, r)|, |\epsilon_n(l, r)|\} = O_P(\sqrt{(\ln n)/n})$, and $\max\{|\varepsilon_n(l, r)|, |\epsilon_n(l, r)|\} < C$ for some $C \in (0, \infty)$ (the existence of such C follows from the boundedness of the relevant functions and estimators). Given these (with recalling the definition of ξ_γ and $\hat{\xi}_\gamma$ in (49) and (66)), we can write each component of the $J \times 1$ vector $[\hat{\xi}_\gamma(Z_i^\#, l) - \xi_\gamma(Z_i^\#, l)]$ as

$$\begin{aligned} & \hat{\xi}_\gamma(Z_i^\#, l, r) - \xi_\gamma(Z_i^\#, l, r) = 1\{\|l - L_i^\#\| \leq r\} \left[1\{X_{1,i}^\# < \gamma\} - \hat{s}_\gamma(l, r) \right] / \hat{\kappa}(l, r) \\ & - 1\{\|l - L_i^\#\| \leq r\} \left[1\{X_{1,i}^\# < \gamma\} - s_\gamma^*(l, r) \right] / \kappa^*(l, r) \\ = & -[1/\kappa^*(l, r)] \left\{ \varepsilon_n(l, r) 1\{\|l - L_i^\#\| \leq r\} + \epsilon_n(l, r) 1\{\|l - L_i^\#\| \leq r\} \left[1\{X_{1,i}^\# < \gamma\} - s_\gamma^*(l, r) \right] \right. \\ & \left. + \varepsilon_n(l, r) \epsilon_n(l, r) 1\{\|l - L_i^\#\| \leq r\} \right\}, \end{aligned}$$

for each $r \in \{r_1, \dots, r_J\}$. This leads to the following expression:

$$\begin{aligned} & \hat{\xi}_\gamma(Z_i^\#, l, r) - \xi_\gamma(Z_i^\#, l, r) = -[\varepsilon_n(l, r) / \kappa^*(l, r)] 1\{\|l - L_i^\#\| \leq r\} \\ & + [\epsilon_n(l, r) / \kappa^*(l, r)] 1\{\|l - L_i^\#\| \leq r\} \left[1\{X_{1,i}^\# < \gamma\} - s_\gamma^*(l, r) \right] + O_P((\ln n)/n), \quad (83) \end{aligned}$$

uniformly over $i \in \{1, \dots, n\}$ and $l \in \mathcal{L}$. In the same way, we can also write

$$\begin{aligned} & \hat{\xi}_\gamma(Z_k, l, r) - \xi_\gamma(Z_k, l, r) = -[\varepsilon_n(l, r) / \kappa^*(l, r)] 1\{\|l - L_k\| \leq r\} \\ & + [\epsilon_n(l, r) / \kappa^*(l, r)] 1\{\|l - L_k\| \leq r\} \left[1\{X_{1,k} < \gamma\} - s_\gamma^*(l, r) \right] + O_P((\ln n)/n), \quad (84) \end{aligned}$$

uniformly over $i \in \{1, \dots, n\}$ and $l \in \mathcal{L}$. Therefore, for each r ,

$$\begin{aligned} & (1/\sqrt{n}) \sum_{i=1}^n \{\hat{\xi}_\gamma(Z_i^\#, l, r) - \xi_\gamma(Z_i^\#, l, r)\} - (1/\sqrt{n}) \sum_{k=1}^n \{\hat{\xi}_\gamma(Z_k, l, r) - \xi_\gamma(Z_k, l, r)\} \\ = & -[\varepsilon_n(l, r) / \kappa^*(l, r)] (1/\sqrt{n}) \left\{ \sum_{i=1}^n 1\{\|l - L_i^\#\| \leq r\} - \sum_{k=1}^n 1\{\|l - L_k\| \leq r\} \right\} \\ & + [\epsilon_n(l, r) / \kappa^*(l, r)] (1/\sqrt{n}) \left\{ \sum_{i=1}^n 1\{\|l - L_i^\#\| \leq r\} [1\{X_{1,i}^\# < \gamma\} - s_\gamma^*(l, r)] \right. \\ & \left. - \sum_{k=1}^n 1\{\|l - L_k\| \leq r\} [1\{X_{1,k} < \gamma\} - s_\gamma^*(l, r)] \right\} + O_P((\ln n)/\sqrt{n}). \quad (85) \end{aligned}$$

Recall the Donsker property of the two sets of functions, $\{1\{\|l - \cdot\| \leq r\} \mid l \in \mathcal{L} \ \& \ r > 0\}$ and $\{1\{\|l - \cdot\| \leq r\} [1\{\cdot < \gamma\} - s_\gamma^*(l, r)] \mid l \in \mathcal{L} \ \& \ r > 0\}$ (argued in the proof of Lemma 1). By applying the functional CLT for bootstrap empirical processes (e.g., Theorem 3.6.1 of VW, 1996; or Theorem 2.6 of Kosorok, 2008), each of two parts within the curly braces on the RHS of (85), normalized by $(1/\sqrt{n})$, conditionally weakly converges to some tight random function in probability. Therefore, we have

$$\text{the LHS of (85)} \xrightarrow[\#]{P} 0 \quad \text{uniformly over } (l, r) \in \mathcal{L} \times \{r_1, \dots, r_J\},$$

where we also have used the boundedness of $1/\kappa^*(l, r)$ (Assumption 2), the convergence property of $\varepsilon_n(l, r)$ and $\epsilon_n(l, r)$, and the fact that the (usual) convergence in probability implies the conditional weak convergence in probability. This completes the proof of Lemma 7. ■