SEMIPARAMETRIC ESTIMATION OF STRUCTURAL FUNCTIONS IN NONSEPARABLE TRIANGULAR MODELS

VICTOR CHERNOZHUKOV†, IVÁN FERNÁNDEZ-VAL§, WHITNEY NEWEY‡, SAMI STOULI¶, AND FRANCIS VELLA|

ABSTRACT. Triangular systems with nonadditively separable unobserved heterogeneity provide a theoretically appealing framework for the modelling of complex structural relationships. However, they are not commonly used in practice due to the need for exogenous variables with large support for identification, the curse of dimensionality in estimation, and the lack of inferential tools. This paper introduces two classes of semiparametric nonseparable triangular models that address these limitations. They are based on distribution and quantile regression modelling of the reduced form conditional distributions of the endogenous variables. We show that average, distribution and quantile structural functions are identified in these systems through a control function approach that does not require a large support condition. We propose a computationally attractive three-stage procedure to estimate the structural functions where the first two stages consist of quantile or distribution regressions. We provide asymptotic theory and uniform inference methods for each stage. In particular, we derive functional central limit theorems and bootstrap functional central limit theorems for the distribution regression estimators of the structural functions. These results establish the validity of the bootstrap for three-stage estimators of structural functions, and lead to simple inference algorithms. We illustrate the implementation and applicability of all our methods with numerical simulations and an empirical application to demand analysis.

KEYWORDS: Structural functions, nonseparable models, control function, quantile and distribution regression, semiparametric estimation, uniform inference.

Date: April 1, 2019.
† Department of Economics, MIT, vchern@mit.edu.
§ Department of Economics, Boston University, ivanf@bu.edu.
‡ Department of Economics, MIT, wnewey@mit.edu.
¶ Department of Economics, University of Bristol, s.stouli@bristol.ac.uk.
| Department of Economics, Georgetown University, Francis.Vella@georgetown.edu.

1. Introduction

Models with nonadditively separable disturbances provide an important vehicle for incorporating heterogenous effects. However, accounting for endogenous treatments in such a setting can be challenging. One methodology which has been successfully employed in a wide range of models with endogeneity is the use of control functions (see, for surveys, Imbens and Wooldridge, 2009, Wooldridge, 2015 and Blundell, Newey and Vella, 2017). The underlying logic of this approach is to account for the endogeneity by including an appropriate control function in the conditioning variables. This paper proposes some relatively simple control function procedures to estimate objects of interest in a triangular model with nonseparable disturbances. Our approach to circumventing the inherent difficulties in nonparametric estimation associated with the curse of dimensionality is to build our models upon a semiparametric specification. This also alleviates the full support requirement on the control function conditional on the treatment variable needed for nonparametric identification. Our goal is thus to provide models and methods that are essentially parametric but still allow for nonseparable disturbances in order to address strong data requirements that come with nonparametric formulations. These models can be interpreted as "baseline" models on which series approximations can be built by adding additional terms.

We consider two kinds of baseline models, quantile regression and distribution regression. These models allow the use of convenient and widely available methods to estimate objects of interest including average, distribution and quantile structural/treatment effects. A main feature of the baseline models is that interaction terms included would not usually be present as leading terms in estimation. These included terms are products of a transformation of the control function with the endogenous treatment. Their presence is meant to allow for heterogeneity in the coefficient of the endogenous variable. Such heterogenous coefficient linear models are of interest in many settings, including demand analysis and estimation of returns to education, and provide a natural starting point for more general models that allow for nonlinear effects of the endogenous treatments.

We use these baseline models to construct estimators of the average, distribution and quantile structural functions based on parametric quantile and distribution regressions. These objects fully characterize the structural relationship between the endogenous treatment and the outcome of interest, and describe the average, distribution and quantiles of the outcome across treatment values, had the treatment been exogenous. We also show how these baseline models can be expanded to include higher order terms, leading to more flexible structural function specifications. The estimation procedure consists of three stages. First, we estimate the control function via quantile regression (QR) or distribution regression (DR) of the endogenous treatment on the exogenous covariates and variables that satisfy an exclusion restriction. Second, we estimate the reduced form distribution of the outcome conditional on the treatment, covariates and estimated control function using DR or QR. Third, we construct estimators of the structural functions applying suitable functionals to the reduced form estimator from the second stage. We derive asymptotic theory for the estimators based on DR in all the stages using a trimming device that avoids tail estimation in the construction of the control function. We also establish the validity of the bootstrap for our inference on structural functions, which enables the formulation of convenient inference algorithms which we describe in detail. The modelling framework we propose thus allows us to address three key difficulties that have restricted the use of such models in empirical work – the curse of dimensionality, the full support condition for identification and the lack of easily implementable inference methods – while simultaneously retaining important features of the original nonparametric formulation. We give an empirical application based on the estimation of Engel curves which illustrates how our approach leads to flexible estimates of all structural functions and their confidence regions.

Our results for the average structural function in the linear random coefficients model are similar to Garen (1984). Florens, Heckman, Meghir and Vytlacil (2008) give identification results for a random coefficients model where the structural function is a polynomial in the endogenous treatment. Blundell and Powell (2003, 2004) introduce the average structural function, and Imbens and Newey (2009) give general models and results for a variety of objects of interest and control functions, including quantile structural functions, under a large support condition on a variable that satisfies an exclusion restriction. We formulate semiparametric specifications of their models that do not require this excluded variable to have continuous support. Our work is also related to the literature on identification and estimation in nonseparable triangular systems with as many unobservables as equations. Chesher (2003), Ma and Koenker (2006), Jun (2009), and Chernozhukov, Fernandez-Val and Kowalski (2015) consider identification and estimation of the structural function at quantiles of the unobservable in the outcome equation conditional on values of the control function; Stouli (2012) gives conditions for identification and estimation of the structural function at both marginal and conditional quantiles of the unobservable in the outcome equation given the control function, under a normalization on the distribution of the unobservable conditional on a specified value of the control function. These approaches do not apply to triangular systems with more unobservables than equations. In contrast, we consider semiparametric formulations which, if correctly specified, provide valid models for the determination of quantile structural effects irrespective of the dimensionality of unobserved heterogeneity in the outcome equation. Our work also complements the analysis and methods developed for the single-equation instrumental variable quantile regression model of Chernozhukov and Hansen (2005, 2006), which rely on monotonicity of the structural function in a scalar disturbance and therefore do not apply to the class of models we consider. In contrast, triangular systems rely on monotonicity of the first stage reduced form function in a scalar disturbance, but include more than one source of unobserved heterogeneity. In addition, control function methods can be used for any type of outcome variable, continuous, discrete or mixed continuous-discrete, but require the endogenous treatment to be continuous. In contrast, the instrumental variable quantile regression approach requires a continuous outcome but applies to any type of treatment. The two approaches therefore provide complementary modelling frameworks and estimation tools for the empirical analysis of nonseparable models with endogenous treatment[1].

This paper makes four main contributions to the existing literature. First, we establish identification of structural functions in both classes of baseline models, providing conditions that do not impose large support requirements on the excluded variable. D'Haultfoeuille and Février (2015) and Torgovitsky (2015) give identification results in the presence of an instrument with small support, but require monotonicity of the structural function in a scalar disturbance. Instead we restrict the functional form of our models. Second, we derive a functional central limit theorem and a bootstrap functional central limit theorem for the two-stage DR estimators in the second stage. These results are uniform over compact regions of values of the outcome. To the best of our knowledge, this result is new. For example, Chernozhukov, Fernandez-Val and Kowalski (2015) derived similar results for two-stage quantile regression estimators but their results are pointwise over quantile indexes and are not applicable to the problem considered here. Our analysis builds on Chernozhukov, Fernandez-Val, and Galichon (2010) and Chernozhukov, Fernandez-Val, and Melly (2013), which established the properties of the DR estimators that we use in the first stage. The theory

[1]See Section 4.2 in Chernozhukov and Hansen (2013) for a related discussion.

of the two-stage estimator, however, does not follow from these results using standard techniques due to the dimensionality and entropy properties of the first stage DR estimators. We follow the proof strategy proposed by Chernozhukov, Fernandez-Val and Kowalski (2015) to deal with these issues. Third, we derive functional central limit theorems and bootstrap functional central limit theorems for plug-in estimators of functionals of the distribution of the outcome conditional on the treatment, covariates and control function via functional delta method. These functionals include all the structural functions of interest. We also use a linear functional for the average structural function which had not been previously considered. Fourth, we show that this linear operator that relates the average of a random variable with its distribution is Hadamard differentiable. Our modelling framework and theoretical results are also of interest for the study of nonseparable triangular models in various alternative settings[2], and will allow establishing the validity of bootstrap inference for the corresponding estimators.

The rest of the paper is organized as follows. Section 2 describes the baseline models and objects of interest. Section 3 presents the estimation and inference methods. Section 4 gives asymptotic theory. Section 5 reports the results of an extensive empirical application to Engel curves, and provide implementation algorithms for all our methods. The proofs of the main result are given in the Appendix. The online Appendix Chernozhukov, Fernandez-Val, Newey, Stouli and Vella (2018) contains supplemental material, including results of numerical simulations calibrated to the application.

2. Modelling Framework

We begin with a brief review of the triangular nonseparable model and some inherent objects of interest. Let Y denote an outcome variable of interest that can be continuous, discrete or mixed continuous-discrete, X a continuous endogenous treatment, Z a vector of exogenous variables, ε a structural disturbance vector of unknown dimension, and η a scalar reduced form disturbance[3]. A general nonseparable triangular

[2]See Fernandez-Val, van Vuuren and Vella (2018) for an application to the analysis of nonseparable sample selection models with censored selection rules.

[3]In our empirical application, we use household level data to study the structural relationship between the share of expenditure on either food or leisure, Y, and the log of total expenditure, X, with gross earnings of the head of household as the excluded variable Z. Additional examples and a general economic motivation of nonseparable triangular models are given in Chesher (2003) and Imbens and Newey (2009), for instance.

model takes the form

$$\begin{aligned} Y &= g(X, \varepsilon), \\ X &= h(Z, \eta), \quad (\varepsilon, \eta) \text{ indep of } Z, \end{aligned}$$

where $\eta \mapsto h(z, \eta)$ is a one-to-one function for each z. This model implies that ε and X are independent conditional on η and that η is a one-to-one function of $V = F_X(X \mid Z)$, the cumulative distribution function (CDF) of X conditional on Z evaluated at the observed variables. Thus, V is a control function.

Objects of interest in this model include the ASF, $\mu(x)$, quantile structural function (QSF), $Q(\tau, x)$, and distribution structural function (DSF), $G(y, x)$, where

$$\mu(x) = \int g(x, \varepsilon) F_\varepsilon(d\varepsilon),$$

and

$$Q(\tau, x) = \tau^{th} \text{ quantile of } g(x, \varepsilon), \quad G(y, x) = \Pr(g(x, \varepsilon) \leq y).$$

Here $\mu(\tilde{x}) - \mu(\bar{x})$ is like an average treatment effect, $Q(\tau, \tilde{x}) - Q(\tau, \bar{x})$ is like a quantile treatment effect, and $G(y, \tilde{x}) - G(y, \bar{x})$ is like a distribution treatment effect from the treatment effects literature. If the support of V conditional on $X = x$ is the same as the marginal support of V then these objects are nonparametrically identified[4] by

$$\mu(x) = \int E[Y \mid X = x, V = v] F_V(dv),$$

and

$$Q(\tau, x) = G^{\leftarrow}(\tau, x), \quad G(y, x) = \int F_Y(y \mid X = x, V = v) F_V(dv),$$

where $G^{\leftarrow}(\tau, x)$ denotes the left-inverse of $y \mapsto G(y, x)$ for each x, i.e., $G^{\leftarrow}(\tau, x) := \inf\{y \in \mathbb{R} : G(y, x) \geq \tau\}$.

It is straightforward to extend this approach to allow for covariates in the model by further conditioning on or integrating over them. Suppose that $Z_1 \subset Z$ is included in the structural equation, which is now $g(X, Z_1, \varepsilon)$. Under the assumption that ε and V are jointly independent of Z, then ε will be independent of X and Z_1 conditional on V. Conditional on covariates and unconditional average structural functions are identified by

$$\mu(x, z_1) = \int E[Y \mid X = x, Z_1 = z_1, V = v] F_V(dv),$$

[4]Full support of V conditional on $X = x$ needed for nonparametric identification requires the excluded variable Z to have large support conditional on $X = x$; see Imbens and Newey (2009).

and

$$\mu(x) = \int E[Y \mid X = x, Z_1 = z_1, V = v] F_{Z_1}(dz_1) F_V(dv).$$

Similarly, conditional on covariates and unconditional quantile and distribution structural functions are identified by

$$Q(\tau, x, z_1) = G^{\leftarrow}(\tau, x, z_1), \quad G(y, x, z_1) = \int F_Y(y \mid X = x, Z_1 = z_1, V = v) F_V(dv),$$

and

$$Q(\tau, x) = G^{\leftarrow}(\tau, x), \quad G(y, x) = \int F_Y(y \mid X = x, Z_1 = z_1, V = v) F_{Z_1}(dz_1) F_V(dv),$$

respectively.

The structural functions can all be expressed as functionals of the control function $V = F_X(X \mid Z)$, the conditional mean function, $E[Y \mid X, Z_1, V]$, and the conditional CDF $F_Y(Y \mid X, Z_1, V)$. These reduced form functions thus constitute natural modelling targets in the context of triangular models. Without functional form restrictions, the curse of dimensionality makes them difficult to estimate, and the full support condition makes it difficult to achieve point identification of the structural functions. These difficulties motivate our specification of baseline parametric models in what follows. These baseline models provide good starting points for nonparametric estimation and may be of interest in their own right.

2.1. **Quantile Regression Baseline.** We start with a simplified specification with one endogenous treatment X, one excluded variable Z, and a continuous outcome Y. We show below how additional excluded variables and covariates can be included.

The baseline first stage is the QR model

$$X = Q_X(V \mid Z) = \pi_1(V) + \pi_2(V) Z, \quad V \mid Z \sim U(0, 1). \tag{2.1}$$

Note that $v \mapsto \pi_1(v)$ and $v \mapsto \pi_2(v)$ are infinite dimensional parameters (functions), and the mapping $v \mapsto \pi_1(v) + \pi_2(v)z$ is strictly increasing for each z.[5] We can recover

[5]The specification (2.1) restricts the support of Z, as $\partial \pi_1(v)/\partial v + z \partial \pi_2(v)/\partial v > 0$ must hold for all (z, v) for Q_X to be well-specified. In particular, if Z has support the real line, then (2.1) restricts $\partial \pi_2(v)/\partial v = 0$ for all v. When Z has positive support as in our empirical application, $\partial \pi_j(v)/\partial v > 0$, $j = 1, 2$, is sufficient for $\partial \pi_1(v)/\partial v + z \partial \pi_2(v)/\partial v > 0$. All results in the paper allow for more general specifications such as $Q_X(v \mid z) = r(z)' \pi(v)$, where $r(z)$ is a vector of transformations of z including a 1 as the first component, which alleviate the support restrictions on Z implied by (2.1).

the control function V from $V = F_X(X \mid Z) = Q_X^{-1}(X \mid Z)$ or equivalently from

$$V = F_X(X \mid Z) = \int_0^1 1\{\pi_1(v) + \pi_2(v)Z \leq X\}dv.$$

This generalized inverse representation of the CDF is convenient for estimation because it does not require the conditional quantile function to be strictly increasing to be well-defined[6].

The baseline second stage has a reduced form:

$$Y = Q_Y(U \mid X, V), \quad U \mid X, V \sim U(0,1), \tag{2.2}$$

$$Q_Y(U \mid X, V) = \beta_1(U) + \beta_2(U)X + \beta_3(U)\Phi^{-1}(V) + \beta_4(U)X\Phi^{-1}(V), \tag{2.3}$$

where Φ^{-1} is the standard normal inverse CDF. This transformation is included to expand the support of V and to encompass the normal system of equations as a special case (cf. Section 3.2.1 of the Supplementary Material for a detailed derivation), but it can be replaced by any other strictly monotonic quantile function. We can recover the conditional CDF F_Y from $F_Y(Y \mid X, V) = Q_Y^{-1}(Y \mid X, V)$ or equivalently from

$$F_Y(Y \mid X, V) = \int_0^1 1\{\beta_1(u) + \beta_2(u)X + \beta_3(u)\Phi^{-1}(V) + \beta_4(u)X\Phi^{-1}(V) \leq Y\}du.$$

An example of a structural model with reduced form (2.2)-(2.3) is the random coefficient model

$$Y = g(X, \varepsilon) = \varepsilon_1 + \varepsilon_2 X, \tag{2.4}$$

with the restrictions

$$\varepsilon_j = Q_{\varepsilon_j}(U \mid X, V) = \theta_j(U) + \gamma_j(U)\Phi^{-1}(V), \quad U \mid X, V \sim U(0,1), \quad j \in \{1, 2\}. \tag{2.5}$$

These restrictions include the control function assumption $\varepsilon_j \perp\!\!\!\perp X \mid V$ and a joint functional form restriction, where the unobservable U is the same for ε_1 and ε_2. Substituting in the second stage equation,

$$Y = \theta_1(U) + \theta_2(U)X + \gamma_1(U)\Phi^{-1}(V) + \gamma_2(U)\Phi^{-1}(V)X, \quad U \mid X, V \sim U(0,1),$$

[6]There are other valid approaches for the specification and estimation of the control function. For instance, dual regression (Spady and Stouli, 2018) provides a parametric alternative for the modelling of $F_X(X \mid Z)$, and nonparametric estimation based on locally linear and series estimators was considered by Imbens and Newey (2009). Here we focus on semiparametric approaches, QR and DR, because of their flexibility, and their well-established computational and theoretical properties.

which has the form of (2.2)-(2.3). A special case of the QR baseline is a heteroskedastic normal system of equations. We use this specification in the numerical simulations given in the Supplementary Material.

The specification (2.2)-(2.3) is a baseline, or starting point, for a more general series approximation to the quantiles of Y conditional on X and V based on including additional functions of X and $\Phi^{-1}(V)$. The baseline is unusual as it includes the interaction term $\Phi^{-1}(V)X$; it is more usual to take the starting point to be $(1, \Phi^{-1}(V), X)$, which is linear in the regressors X and $\Phi^{-1}(V)$. The inclusion of the interaction term is motivated by allowing the coefficient of X to vary with individuals, so that $\Phi^{-1}(V)$ then interacts X in the conditional distribution of ε_2 given the control functions. Augmenting the baseline with splines or power transformations of X and $\Phi^{-1}(V)$ and their interactions gives rise to more flexible semiparametric specifications. In this more general case, the inclusion of interaction terms is motivated by increasing the flexibility with which the coefficients of X and its transformations can vary across individuals – as would arise, for instance, when including power transformations of X and $\Phi^{-1}(V)$ in the random coefficient models (2.4) and (2.5), respectively. All the parameters of both the baseline and augmented specifications can be identified when the support of Z is discrete, and estimated using the QR estimator (Koenker and Bassett, 1978). Thus the baseline (2.2)-(2.3) incorporates a flexible heterogeneous coefficients structure that can easily be augmented for the purpose of modelling more complex relationships, while only relying on weak identification conditions and preserving ease of estimation.[7]

The ASF of the baseline specification is:

$$\mu(x) = \int_0^1 E[Y \mid X = x, V = v]dv = \beta_1 + \beta_2 x,$$

where the second equality follows by $\int_0^1 \Phi^{-1}(v)dv = 0$ and

$$E[Y \mid X, V] = \int_0^1 Q_Y(u \mid X, V)du = \beta_1 + \beta_2 X + \beta_3 \Phi^{-1}(V) + \beta_4 X \Phi^{-1}(V), \tag{2.6}$$

[7]Our formal definition of the model in Section 2.3 explicitly allows for augmented semiparametric baseline specifications, and Section 5 and the Supplementary Material illustrate our methods when incorporating spline and power transformations of X, $\Phi^{-1}(V)$ and their interactions.

with $\beta_j := \int_0^1 \beta_j(u)du$, $j \in \{1, \dots, 4\}$. The QSF does not appear to have a closed form expression. It is the solution to

$$
\begin{aligned}
Q(\tau, x) &= G^{\leftarrow}(\tau, x), \\
G(y, x) &= \int_0^1 \int_0^1 1\{\beta_1(u) + \beta_2(u)x + \beta_3(u)\Phi^{-1}(v) + \beta_4(u)\Phi^{-1}(v)x \leq y\} du dv.
\end{aligned}
$$

Remark 1. Another way to arrive at that conditional mean specification (2.6) is to start with the random coefficients model $Y = \varepsilon_1 + \varepsilon_2 X$ and assume that the conditional mean of ε_1 and ε_2 given V are linear in $(1, \Phi^{-1}(V))$. Then

$$E[Y \mid X, V] = E[\varepsilon_1 \mid V] + E[\varepsilon_2 \mid V]X = \bar{\theta}_1 + \bar{\gamma}_1\Phi^{-1}(V) + \bar{\theta}_2 X + \bar{\gamma}_2\Phi^{-1}(V)X.$$

This is the same conditional mean specification implied by the reduced form (2.3). The ASF can thus be equivalently specified as

$$\mu(x) = \int_0^1 E[Y \mid X = x, V = v]dv = \bar{\theta}_1 + \bar{\theta}_2 x,$$

where $(\bar{\theta}_1, \bar{\theta}_2)$, the regression coefficients for the intercept and X from the linear projection of Y on the full set of regressors, satisfy $\bar{\theta}_j = \int_0^1 \theta_j(u)du$, $j = 1, 2$.

2.2. **Distribution Regression Baseline.** We start again with a simplified specification with one endogenous treatment X and one excluded Z, but now the outcome Y can be continuous, discrete or mixed.

Let Γ denote a strictly increasing continuous CDF such as the standard normal or logistic CDF. The first stage equation is the distribution regression model

$$\eta = \pi_1(X) + \pi_2(X)Z, \quad \eta \mid Z \sim \Gamma,$$

where $x \mapsto \pi_1(x)$ and $x \mapsto \pi_2(x)$ are infinite dimensional parameters, and the mapping $x \mapsto \pi_1(x) + \pi_2(x)z$ is strictly increasing for each z.[8] This model corresponds to the specification of the control variable V as

$$V = F_X(X \mid Z) = \Gamma(\pi_1(X) + \pi_2(X)Z). \tag{2.7}$$

While the first stage QR model specifies the conditional quantile function of X given Z to be linear in Z, the DR model (2.7) specifies the conditional distribution of X given Z to be generalized linear in Z, i.e., linear after applying the link function Γ.

[8]Analogously to the QR specification (2.1), the DR specification also restricts the support of Z, and the remarks of footnote 5 also apply to the DR baseline.

The second stage baseline has a reduced form:

$$F_Y(Y \mid X, V) = \Gamma(\beta_1(Y) + \beta_2(Y)X + \beta_3(Y)\Phi^{-1}(V) + \beta_4(Y)\Phi^{-1}(V)X). \tag{2.8}$$

When Y is continuous, an example of a structural model that has reduced form (2.8) is the latent random coefficient model

$$\xi = \varepsilon_1 + \varepsilon_2\Phi^{-1}(V), \quad \xi \mid X, V \sim \Gamma, \tag{2.9}$$

with the restrictions

$$\varepsilon_j = \theta_j(Y) + \gamma_j(Y)X, \quad j \in \{1, 2\},$$

such that the mapping $y \mapsto \theta_j(y) + \gamma_j(y)x$ is strictly increasing for each x, and the following conditional independence property is satisfied:

$$F_{\varepsilon_j}(\varepsilon_j \mid V) = F_{\varepsilon_j}(\varepsilon_j \mid X, V), \quad j \in \{1, 2\}. \tag{2.10}$$

Substituting the expression for ε_1 and ε_2 in (2.9) yields

$$\xi = \theta_1(Y) + \gamma_1(Y)X + \theta_2(Y)\Phi^{-1}(V) + \gamma_2(Y)\Phi^{-1}(V)X,$$

which has a reduced form for the distribution of Y conditional on (X, V) as in (2.8). The numerical simulations in the Supplementary Material provide an example of a special case of the DR model.

As in the quantile baseline, the specification (2.8) can be used as starting point for a more general series approximation to the distribution of Y conditional on X and V based on including additional functions of X and $\Phi^{-1}(V)$. All the parameters of model (2.7)-(2.8) are identified when the support of Z is discrete and can be estimated by DR.

For the DR baseline, the QSF is the solution to

$$Q(\tau, x) = G^{\leftarrow}(\tau, x), \;\; G(y, x) = \int_0^1 \Gamma(\beta_1(y) + \beta_2(y)x + \beta_3(y)\Phi^{-1}(v) + \beta_4(y)\Phi^{-1}(v)x)dv.$$

Compared to the QR baseline model, the ASF cannot be obtained as a linear projection but it can be conveniently expressed as a linear functional of $G(y, x)$. Let $\mathcal{Y}$ denote the support of Y assumed to be bounded, $\mathcal{Y}^+ = \mathcal{Y} \cap [0, \infty)$ and $\mathcal{Y}^- = \mathcal{Y} \cap (-\infty, 0)$. The ASF can be characterized as

$$\mu(x) = \int_0^1 E[Y \mid X = x, V = v]dv = \int_{\mathcal{Y}^+} [1 - G(y, x)]\nu(dy) - \int_{\mathcal{Y}^-} G(y, x)\nu(dy), \tag{2.11}$$

where ν is either the counting measure when $\mathcal{Y}$ is countable or the Lebesgue measure otherwise, and we exploit the linear relationship between the expected value and the distribution of a random variable. This characterization simplifies both the computation and theoretical treatment of the DR-based estimator for the ASF. It also applies to the QR specification upon using the corresponding expression for $G(y,x)$.

2.3. **Identification.** The most general specifications that we consider include several excluded variables, covariates and transformations of the regressors in both stages. Denote the sets of regressors in the first and second stages by

$$R := r(Z) \text{ and } W := w(X, Z_1, V) := p(X) \otimes r_1(Z_1) \otimes q(V),$$

where r, r_1, p and q are vectors of transformations such as powers, b-splines and interactions, and $\otimes$ denotes the Kronecker product. The simplest case is when $r(Z) = (1, Z)'$, $r_1(Z_1) = (1, Z_1)'$, $p(X) = (1, X)'$ and $q(V) = (1, \Phi^{-1}(V))'$, so that $w(X, Z_1, V) = (1, \Phi^{-1}(V), X, X\Phi^{-1}(V), Z_1, Z_1\Phi^{-1}(V), XZ_1, XZ_1\Phi^{-1}(V))'$. The following assumption gathers the baseline specifications for the first and second stages.

Assumption 1. *[Baseline Models] The outcome Y has a conditional density function $y \mapsto f_Y(y \mid X, Z_1, V)$ with respect to the measure ν that is a.s. bounded away from zero uniformly in $\mathcal{Y}$; and (a) X conditional on Z follows the QR model*

$$X = Q_X(V \mid Z) = R'\pi(V), \quad V \mid Z \sim U(0,1),$$

and Y conditional on (X, Z_1, V) follows the QR model

$$Y = Q_Y(U \mid X, Z_1, V) = W'\beta(U), \quad V = F_X(X \mid Z), \quad U \mid X, Z_1, V \sim U(0,1);$$

or (b) X conditional on Z follows the DR model

$$V = \Lambda(R'\pi(X)), \quad V \mid Z \sim U(0,1),$$

and Y conditional on (X, Z_1, V) follows the DR model,

$$U = \Gamma(W'\beta(Y)), \quad V = F_X(X \mid Z), \quad U \mid X, Z_1, V \sim U(0,1),$$

where Γ is either the standard normal or logistic CDF.

The structural functions of the baseline models involve quantile and distribution regressions on the same set of regressors. A sufficient condition for identification of the coefficients of these regressions is that the second moment matrix of those regressors is nonsingular. The regressors have a Kronecker product form $p(X) \otimes r_1(Z_1) \otimes q(V)$.

The second moment matrix for these regressors will be nonsingular if the joint distribution dominates a distribution where X, Z_1 and V are independent and the second moment matrices of X, Z_1 and V are positive definite[9].

Assumption 2. *The joint probability distribution of* X, Z_1 *and* V *dominates a product probability measure* $\mu(x) \times \varsigma(z_1) \times \rho(v)$ *such that* $E_\mu[p(X)p(X)']$, $E_\varsigma[r_1(Z_1)r_1(Z_1)']$, *and* $E_\rho[q(V)q(V)']$ *are positive definite.*

When $p(X) = (1, X)'$, $r_1(Z_1) = (1, Z_1)'$ with $\dim(Z_1) = 1$, and $q(V) = (1, \Phi^{-1}(V))'$, Assumption 2 simplifies to the requirement that the joint distribution of X, Z_1 and V be dominating one such that $\text{Var}_\mu(X) > 0$, $\text{Var}_\varsigma(Z_1) > 0$, and $\text{Var}_\rho(\Phi^{-1}(V)) > 0$. For general specifications where the regressors are higher order power series, it is sufficient for Assumption 2 that the joint distribution of X, Z_1 and V be dominating one that has density bounded away from zero on a hypercube. That will mean that the joint distribution dominates a uniform distribution on that hypercube, and for a uniform distribution on a hypercube $E[w(X, Z_1, V)w(X, Z_1, V)']$ is nonsingular.

Lemma 1. *If Assumption 2 holds, then* $E[w(X, Z_1, V)w(X, Z_1, V)']$ *is nonsingular.*

For the QR and DR specifications, Assumptions 1 and 2 are then sufficient for identification of the reduced form functions $Q_Y(U \mid X, Z_1, V)$ and $F_Y(Y \mid X, Z_1, V)$, respectively, and therefore for identification of the structural functions.

Theorem 1. *If Assumptions 1 and 2 hold, then the DSF, QSF and ASF are identified.*

For the specifications in Assumption 1, identification of structural functions does not require the support of Z to be continuous. This result also holds if the vector of regressors $w(X, Z_1, V)$ is specified as a subset of $p(X) \otimes r_1(Z_1) \otimes q(V)$. When $q(V) = (1, \Phi^{-1}(V))'$, for the second moment matrix of regressors to be nonsingular only requires the control function to have positive variance across the support of X, which can be satisfied if the support of Z is binary or discrete (Newey and Stouli, 2018). This is in sharp contrast with nonparametric identification which requires Z to have a large support (Imbens and Newey, 2009). Theorem 1 thus illustrates the identifying power of semiparametric restrictions and the trade-off between these restrictions and the full support condition for identification of structural functions.

[9]This condition is sufficient for identification and is in principle testable. However, in practice it will be easier to check directly if the sample second moment matrix for the regressors is of full rank.

3. Estimation and Inference Methods

The QR and DR baselines of the previous section lead to three-stage analog estimation and inference methods for the DSF, QSF and ASF. The first stage estimates the control function $V = F_X(X \mid Z)$. The second stage estimates the conditional distribution function $F_Y(Y \mid X, Z_1, V)$, replacing V by the estimator from the first stage. The third stage obtains estimators of the structural functions, which are functionals of the first and second stages building blocks. We provide a detailed description of the implementation of each step for both QR and DR methods. We also describe a weighted bootstrap procedure to perform uniform inference on all structural functions considered. Detailed implementation algorithms are given in Section 5.1

We assume that we observe a sample of n independent and identically distributed realizations $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ of the random vector (Y, X, Z), and that $\dim(X) = 1$. Calligraphic letters such as $\mathcal{Y}$ and $\mathcal{X}$ denote the supports of Y and X; and $\mathcal{YX}$ denotes the joint support of (Y, X). The description of all the stages includes individual weights e_i which are set to 1 for the estimators, or drawn from a distribution that satisfies Assumption 3 in Section 4 for the weighted bootstrap version of the estimators.

3.1. **First Stage: Estimation of Control Function.**

The first stage estimates the n target values of the control function, $V_i = F_X(X_i \mid Z_i)$, $i = 1, \dots, n$. We estimate the conditional distribution of X in a trimmed support $\overline{\mathcal{X}}$ that excludes extreme values. The purpose of the trimming is to avoid the far tails. We consider a fixed trimming rule, which greatly simplifies the derivation of the asymptotic properties. In our numerical and empirical examples we find that the results are not sensitive to the trimming rule and the choice of $\overline{\mathcal{X}}$ as the observed support of X, i.e., no trimming, works well. We use bars to denote trimmed supports with respect to X, e.g., $\overline{\mathcal{XZ}} = \{(x, z) \in \mathcal{XZ} : x \in \overline{\mathcal{X}}\}$. A subscript in a set denotes a finite grid covering the set, where the subscript is the number of grid points. Unless otherwise specified, the points of the grid are sample quantiles of the corresponding variable at equidistant probabilities in $[0, 1]$. For example, $\mathcal{X}_5$ denotes a grid of 5 points covering $\mathcal{X}$ located at the 0, 1/4, 1/2, 3/4 and 1 sample quantiles of X.

Denoting the usual check function by $\rho_v(z) = (v - 1(z < 0))z$, the first stage in the QR baseline is

$$\text{(3.1)} \quad \widehat{F}^e_X(x \mid z) \;=\; \epsilon + \int_{\epsilon}^{1-\epsilon} 1\{R'\widehat{\pi}^e(v) \leq x\} dv, \quad R = r(z), \quad (x,z) \in \overline{\mathcal{X}\mathcal{Z}},$$

$$\text{(3.2)} \quad \widehat{\pi}^e(v) \;\in\; \arg \min_{\pi \in \mathbb{R}^{\dim(R)}} \sum_{i=1}^{n} e_i \rho_v (X_i - R_i'\pi),$$

for some small constant $\epsilon > 0$. The adjustment in the limits of the integral in (3.1) avoids tail estimation of quantiles[10]. The first stage in the DR baseline is,

$$\text{(3.3)} \quad \widehat{F}^e_X(x \mid z) = \Gamma(R'\widehat{\pi}^e(x)), \quad R = r(z), \quad (x,z) \in \overline{\mathcal{X}\mathcal{Z}},$$

$$\text{(3.4)} \quad \widehat{\pi}^e(x) \in \arg \min_{\pi \in \mathbb{R}^{\dim(R)}} -\sum_{i=1}^{n} e_i \left[1\left(X_i \leq x\right) \log \Gamma(R_i'\pi) + 1\left(X_i > x\right) \log\left(1 - \Gamma(R_i'\pi)\right)\right].$$

When $e_i = 1$ for all $i = 1, \ldots, n$, expressions (3.1)-(3.2) and (3.3)-(3.4) define $\widehat{F}_X$, the QR and DR estimators of F_X. For $(X_i, Z_i) \in \overline{\mathcal{X}\mathcal{Z}}$, the estimator and weighted bootstrap version of the control function are then $\widehat{V}_i = \widehat{F}_X(X_i \mid Z_i)$ and $\widehat{V}^e_i = \widehat{F}^e_X(X_i \mid Z_i)$, respectively, and we set $\widehat{V}_i = \widehat{V}^e_i = 0$ otherwise.

3.2. **Second Stage: Estimation of $F_Y(Y \mid X, Z_1, V)$.** With the estimated control function in hand, the second building block required for the estimation of structural functions is an estimate of the reduced form CDF of Y given (X, Z_1, V). The baseline models provide direct estimation procedures based on QR and DR.

Let $T := 1(X \in \overline{\mathcal{X}})$ be a trimming indicator, which is formally defined in Assumption 4 of Section 4. The estimator of F_Y in the QR baseline is

$$\text{(3.5)} \quad \widehat{F}^e_Y(y \mid x, z_1, v) = \epsilon + \int_{\epsilon}^{1-\epsilon} 1\{w(x,z_1,v)'\widehat{\beta}^e(u) \leq y\} du, \quad (y,x,z_1,v) \in \mathcal{Y}\overline{\mathcal{X}\mathcal{Z}_1\mathcal{V}},$$

$$\text{(3.6)} \quad \widehat{\beta}^e(u) \in \arg \min_{\beta \in \mathbb{R}^{\dim(W)}} \sum_{i=1}^{n} e_i T_i \rho_u (Y_i - \widehat{W}^{e\prime}_i \beta), \quad \widehat{W}^e_i = w(X_i, Z_{1i}, \widehat{V}^e_i),$$

[10] Chernozhukov, Fernandez-Val and Melly (2013) provide conditions under which this adjustment does not introduce bias.

As for the first stage, the adjustment in the limits of the integral in (3.5) avoids tail estimation of quantiles. The estimator of F_Y in the DR baseline is

$$\widehat{F}^e_Y(y \mid x, z_1, v) = \Gamma(w(x,z_1,v)'\widehat{\beta}^e(y)), \quad (y,x,z_1,v) \in \mathcal{Y}\overline{\mathcal{X}\mathcal{Z}_1\mathcal{V}}, \tag{3.7}$$

$$\widehat{\beta}^e(y) \in \arg\min_{\beta\in\mathbb{R}^{\dim(W)}} -\sum_{i=1}^n e_i T_i \Big[1\,(Y_i \le y)\log\Gamma(\widehat{W}^{e\prime}_i\beta) + 1\,(Y_i > y)\log(1-\Gamma(\widehat{W}^{e\prime}_i\beta))\Big]. \tag{3.8}$$

When $e_i = 1$ for all $i = 1,\dots,n$, expressions (3.5)-(3.6) and (3.7)-(3.8) define $\widehat{F}_Y$, the quantile and distribution regression estimators of F_Y, respectively.

3.3. **Third Stage: Estimation of Structural Functions.** Given the estimators $(\{\widehat{V}_i\}_{i=1}^n, \widehat{F}_Y)$ and their bootstrap draws $(\{\widehat{V}^e_i\}_{i=1}^n, \widehat{F}^e_Y)$, we can form estimators of the structural functions as functionals of these building blocks.

The estimator and bootstrap draw of the DSF are

$$\widehat{G}(y,x) = \frac{1}{n_T}\sum_{i=1}^n \widehat{F}_Y(y \mid x, Z_{1i}, \widehat{V}_i)T_i, \tag{3.9}$$

where $n_T = \sum_{i=1}^n T_i$, and

$$\widehat{G}^e(y,x) = \frac{1}{n^e_T}\sum_{i=1}^n e_i\widehat{F}^e_Y(y \mid x, Z_{1i}, \widehat{V}^e_i)T_i, \tag{3.10}$$

where $n^e_T = \sum_{i=1}^n e_i T_i$. For the DR estimator, $y \mapsto \widehat{G}(y,x)$ may not be monotonic. This can be addressed by applying the rearrangement method of Chernozhukov, Fernandez-Val and Galichon (2010).

Given the DSF estimate and bootstrap draw, $\widehat{G}(y,x)$ and $\widehat{G}^e(y,x)$, the estimator and bootstrap draw of the QSF are

$$\widehat{Q}(\tau,x) = \int_{\mathcal{Y}^+} 1\{\widehat{G}(y,x) \le \tau\}\nu(dy) - \int_{\mathcal{Y}^-} 1\{\widehat{G}(y,x) \ge \tau\}\nu(dy), \tag{3.11}$$

and

$$\widehat{Q}^e(\tau,x) = \int_{\mathcal{Y}^+} 1\{\widehat{G}^e(y,x) \le \tau\}\nu(dy) - \int_{\mathcal{Y}^-} 1\{\widehat{G}^e(y,x) \ge \tau\}\nu(dy), \tag{3.12}$$

respectively. Finally, the estimator and bootstrap draw of the ASF are

$$\widehat{\mu}(x) = \int_{\mathcal{Y}^+} [1 - \widehat{G}(y,x)]\nu(dy) - \int_{\mathcal{Y}^-} \widehat{G}(y,x)\nu(dy), \tag{3.13}$$

and

$$\widehat{\mu}^e(x) = \int_{\mathcal{Y}^+}[1 - \widehat{G}^e(y,x)]\nu(dy) - \int_{\mathcal{Y}^-}\widehat{G}^e(y,x)\nu(dy), \tag{3.14}$$

respectively. When the set $\mathcal{Y}$ is uncountable and bounded, we approximate the previous integrals by sums over a fine mesh of equidistant points $\mathcal{Y}_S := \{\inf[y \in \mathcal{Y}] = y_1 < \cdots < y_S = \sup[y \in \mathcal{Y}]\}$ with mesh width δ such that $\delta\sqrt{n} \to 0$. For example, (3.12) and (3.14) are approximated by

$$\widehat{Q}^e_S(\tau,x) = \delta \sum_{s=1}^{S}\left[1(y_s \geq 0) - 1\{\widehat{G}^e(y_s,x) \geq \tau\}\right], \tag{3.15}$$

and

$$\widehat{\mu}^e_S(x) = \delta \sum_{s=1}^{S}\left[1(y_s \geq 0) - \widehat{G}^e(y_s,x)\right]. \tag{3.16}$$

3.4. **Weighted Bootstrap Inference on Structural Functions.** We consider inference uniform over regions of values of (y, x, τ). We denote the region of interest as $\mathcal{I}_G$ for the DSF, $\mathcal{I}_Q$ for the QSF, and $\mathcal{I}_\mu$ for the ASF. Examples include:

(1) The DSF, $y \mapsto \widehat{G}^e(y,x)$, for fixed x and over $y \in \widetilde{\mathcal{Y}} \subset \mathcal{Y}$, by setting $\mathcal{I}_G = \widetilde{\mathcal{Y}} \times \{x\}$.
(2) The QSF, $\tau \mapsto \widehat{Q}^e(\tau,x)$ for fixed x and over $\tau \in \widetilde{\mathcal{T}} \subset (0,1)$, by setting $\mathcal{I}_Q = \widetilde{\mathcal{T}} \times \{x\}$,
(3) The ASF, $\widehat{\mu}^e(x)$, over $x \in \widetilde{\mathcal{X}} \subset \overline{\mathcal{X}}$, by setting $\mathcal{I}_\mu = \widetilde{\mathcal{X}}$.

When the region of interest is not a finite set, we approximate it by a finite grid. All the details of the procedure we implement are summarized in Section 5.1.

The weighted bootstrap versions of the DSF, QSF and ASF estimators are obtained by rerunning the estimation procedure introduced in Section 3.3 with sampling weights drawn from a distribution that satisfies Assumption 3 in Section 4; see Algorithm 2 in Section 5.1 for details. They can then be used to perform uniform inference over the region of interest.

For instance, a $(1-\alpha)$-confidence band for the DSF over the region $\mathcal{I}_G$ can be constructed as

$$\left[\widehat{G}(y,x) \pm \widehat{k}_G(1-\alpha)\widehat{\sigma}_G(y,x), (y,x) \in \mathcal{I}_G\right], \tag{3.17}$$

where $\widehat{\sigma}_G(y,x)$ is an estimator of $\sigma_G(y,x)$, the asymptotic standard deviation of $\widehat{G}(y,x)$, such as the rescaled weighted bootstrap interquartile range[11]

$$\widehat{\sigma}_G(y,x) = \text{IQR}\left[\widehat{G}^e(y,x)\right]/1.349, \tag{3.18}$$

and $\widehat{k}_G(1-\alpha)$ denote a consistent estimator of the $(1-\alpha)$-quantile of the maximal t-statistic

$$\|t_G(y,x)\|_{\mathcal{I}_G} = \sup_{(y,x)\in\mathcal{I}_G} \left| \frac{\widehat{G}(y,x) - G(y,x)}{\sigma_G(y,x)} \right|,$$

such as the $(1-\alpha)$-quantile of the bootstrap draw of the maximal t-statistic

$$\|t^e_G(y,x)\|_{\mathcal{I}_G} = \sup_{(y,x)\in\mathcal{I}_G} \left| \frac{\widehat{G}^e(y,x) - \widehat{G}(y,x)}{\widehat{\sigma}_G(y,x)} \right|. \tag{3.19}$$

Confidence bands for the ASF can be constructed by a similar procedure, using the bootstrap draws of the ASF estimator. For the QSF, we can either use the same procedure based on the bootstrap draws of the QSF, or invert the confidence bands for the DSF following the generic method of Chernozhukov, Fernandez-Val, Melly and Wuthrich (2016). The first possibility works only when Y is continuous, whereas the second method is more generally applicable. We provide algorithms for the construction of the bands in Section 5.1.

4. Asymptotic Theory

We derive asymptotic theory for the estimators of the ASF, DSF and QSF where both the first and second stages are based on DR. The theory for the estimators based on QR can be derived using similar arguments.

In what follows, we shall use the following notation. We let the random vector $A = (Y, X, Z, W, V)$ live on some probability space $(\Omega_0, \mathcal{F}_0, P)$. Thus, the probability measure P determines the law of A or any of its elements. We also let $A_1, ..., A_n$, i.i.d. copies of A, live on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which contains the infinite product of $(\Omega_0, \mathcal{F}_0, P)$. Moreover, this probability space can be suitably enriched to carry also the random weights that appear in the weighted bootstrap. The distinction between the two laws P and $\mathbb{P}$ is helpful to simplify the notation in the

[11]An alternative is to use the bootstrap standard deviation, but its validity requires convergence of bootstrap moments in addition to convergence of the bootstrap distribution; cf. Remark 3.2 in Chernozhukov, Fernandez-Val and Melly (2013) .

proofs and in the analysis. Unless explicitly mentioned, all functions appearing in the statements are assumed to be measurable.

We now state formally the assumptions. The first assumption is about sampling and the bootstrap weights.

Assumption 3. *[Sampling and Bootstrap Weights] (a) Sampling: the data* $\{Y_i, X_i, Z_i\}_{i=1}^n$ *are a sample of size n of independent and identically distributed observations from the random vector* (Y, X, Z). *(b) Bootstrap weights:* $(e_1, ..., e_n)$ *are i.i.d. draws from a random variable* $e \geq 0$, *with* $\mathrm{E}_P[e] = 1$, $\mathrm{Var}_P[e] = 1$, *and* $\mathrm{E}_P|e|^{2+\delta} < \infty$ *for some* $\delta > 0$; *live on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$; *and are independent of the data* $\{Y_i, X_i, Z_i\}_{i=1}^n$ *for all n.*

The second assumption is about the first stage where we estimate the control function $(x, z) \mapsto \vartheta_0(x, z)$ defined as

$$\vartheta_0(x, z) := F_X(x \mid z),$$

with trimmed support $\overline{\mathcal{V}} = \{\vartheta_0(x, z) : (x, z) \in \overline{\mathcal{X}\mathcal{Z}}\}$. We assume a logistic DR model for the conditional distribution of X in the trimmed support $\overline{\mathcal{X}}$.

Assumption 4. *[First Stage] (a) Trimming: we consider a trimming rule defined by the tail indicator*

$$T = 1(X \in \overline{\mathcal{X}}),$$

where $\overline{\mathcal{X}} = [\underline{x}, \overline{x}]$ *for some* $-\infty < \underline{x} < \overline{x} < \infty$, *such that* $P(T = 1) > 0$. *(b) Model: the distribution of X conditional on Z follows Assumption 1(b) with* $\Gamma = \Lambda$, *where* Λ *is the logit link function; the coefficients* $x \mapsto \pi_0(x)$ *are three times continuously differentiable with uniformly bounded derivatives;* $\overline{\mathcal{R}}$ *is compact; and the minimum eigenvalue of* $\mathrm{E}_P\left[\Lambda(R'\pi_0(x))[1 - \Lambda(R'\pi_0(x))]RR'\right]$ *is bounded away from zero uniformly over* $x \in \overline{\mathcal{X}}$.

For $x \in \overline{\mathcal{X}}$, let

$$\widehat{\pi}^e(x) \in \arg \min_{\pi \in \mathbb{R}^{\dim(R)}} -\frac{1}{n} \sum_{i=1}^n e_i \{1(X_i \leq x) \log \Lambda(R_i'\pi) + 1(X_i > x) \log[1 - \Lambda(R_i'\pi)]\},$$

and set

$$\vartheta_0(x, r) = \Lambda(r'\pi_0(x)); \ \widehat{\vartheta}^e(x, r) = \Lambda(r'\widehat{\pi}^e(x)),$$

if $(x, r) \in \overline{\mathcal{X}\mathcal{R}}$, and $\vartheta_0(x, r) = \widehat{\vartheta}^e(x, r) = 0$ otherwise.

Theorem 4 of Chernozhukov, Fernandez-Val and Kowalski (2015) established the asymptotic properties of the DR estimator of the control function. We repeat the result here as a lemma for completeness and to introduce notation that will be used in the results below. Let $T(x) := 1(x \in \overline{\mathcal{X}})$, $\|f\|_{T,\infty} := \sup_{a \in \mathcal{A}} |T(x) f(a)|$ for any function $f : \mathcal{A} \mapsto \mathbb{R}$, $\lambda := \Lambda(1 - \Lambda)$, the density of the logistic distribution.

Lemma 2. *[First Stage] Suppose that Assumptions 3 and 4 hold. Then, (1)*

$$\begin{aligned}
\sqrt{n}(\widehat{\vartheta}^e(x,r) - \vartheta_0(x,r)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \ell(A_i, x, r) + o_{\mathbb{P}}(1) \rightsquigarrow \Delta^e(x,r) \text{ in } \ell^\infty(\overline{\mathcal{X}\mathcal{R}}), \\
\ell(A, x, r) &:= \lambda(r'\pi_0(x))[1\{X \leq x\} - \Lambda(R'\pi_0(x))] \times \\
&\quad \times r' \mathrm{E}_P \left\{ \Lambda(R'\pi_0(x))[1 - \Lambda(R'\pi_0(x))] R R' \right\}^{-1} R, \\
\mathrm{E}_P[\ell(A, x, r)] &= 0, \mathrm{E}_P[T \ell(A, X, R)^2] < \infty,
\end{aligned}$$

where $(x,r) \mapsto \Delta^e(x,r)$ is a Gaussian process with uniformly continuous sample paths and covariance function given by $\mathrm{E}_P[\ell(A,x,r)\ell(A,\tilde{x},\tilde{r})']$. (2) There exists $\widetilde{\vartheta}^e : \overline{\mathcal{X}\mathcal{R}} \mapsto [0,1]$ that obeys the same first order representation uniformly over $\overline{\mathcal{X}\mathcal{R}}$, is close to $\widehat{\vartheta}^e$ in the sense that $\|\widetilde{\vartheta}^e - \widehat{\vartheta}^e\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$ and, with probability approaching one, belongs to a bounded function class such that the covering entropy satisfies[12]

$$\log N(\epsilon, \Upsilon, \|\cdot\|_{T,\infty}) \lesssim \epsilon^{-1/2}, \quad 0 < \epsilon < 1.$$

The next assumptions are about the second stage. We assume a logistic DR model for the conditional distribution of Y given (X, Z_1, V), impose compactness and smoothness conditions, and provide sufficient conditions for identification of the parameters. Compactness is imposed over the trimmed supports and can be relaxed at the cost of more complicated and cumbersome proofs. The smoothness conditions are fairly tight. The assumptions on $\mathcal{Y}$ cover continuous, discrete and mixed outcomes in the second stage. We denote partial derivatives as $\partial_x f(x,y) := \partial f(x,y)/\partial x$.

Assumption 5. *[Second Stage] (a) Model: the distribution of Y conditional on (X, Z_1, V) follows Assumption 1(b) with $\Gamma = \Lambda$. (b) Compactness and smoothness: the set $\overline{\mathcal{X}\mathcal{Z}\mathcal{W}}$ is compact; the set $\mathcal{Y}$ is either a compact interval in $\mathbb{R}$ or a finite subset of $\mathbb{R}$; X has a continuous conditional density function $x \mapsto f_X(x \mid z)$ that is bounded above by a constant uniformly in $z \in \overline{\mathcal{Z}}$; if $\mathcal{Y}$ is an interval, then Y has a conditional density function $y \mapsto f_Y(y \mid x, z)$ that is uniformly continuous in $y \in \mathcal{Y}$ uniformly in $(x,z) \in \overline{\mathcal{X}\mathcal{Z}}$, and bounded above by a constant uniformly in $(x,z) \in \overline{\mathcal{X}\mathcal{Z}}$; the derivative vector $\partial_v w(x, z_1, v)$ exists and its components are uniformly continuous in $v \in \overline{\mathcal{V}}$ uniformly in $(x, z_1) \in \overline{\mathcal{X}\mathcal{Z}_1}$, and are bounded in absolute value by a constant, uniformly*

[12] See Appendix B for a definition of the covering entropy.

in $(x,w,v) \in \overline{\mathcal{X}\mathcal{Z}_1\mathcal{V}}$*; and for all* $y \in \mathcal{Y}$*,* $\beta_0(y) \in \mathcal{B}$*, where* $\mathcal{B}$ *is a compact subset of* $\mathbb{R}^{\dim(W)}$*. (c) Nondegeneracy: the matrix* $C(y,v) := \mathrm{Cov}_P[f_y(A)+g_y(A), f_v(A)+g_v(A)]$ *is finite and is of full rank for all* $y, v \in \mathcal{Y}$*, where*

$$f_y(A) := \{\Lambda(W'\beta_0(y)) - 1(Y \leq y)\}WT,$$

and, for $\dot{W} = \partial_v w(X, Z_1, v)|_{v=V}$*,*

$$g_y(A) := \mathrm{E}_P[\{[\Lambda(W'\beta_0(y)) - 1(Y \leq y)]\dot{W} + \lambda(W'\beta_0(y))\dot{W}'\beta_0(y)W\}T\ell(a,X,R)]\big|_{a=A}.$$

For $y \in \mathcal{Y}$, let

$$\widehat{\beta}(y) = \arg\min_{\beta \in \mathbb{R}^{\dim(W)}} \frac{1}{n}\sum_{i=1}^{n} T_i \rho_y(Y_i, \beta'\widehat{W}_i), \quad \widehat{W}_i = w(X_i, Z_{1i}, \widehat{V}_i), \quad \widehat{V}_i = \widehat{\vartheta}(X_i, R_i),$$

where

$$\rho_y(Y,B) := -\{1(Y \leq y)\log\Lambda(B) + 1(Y > y)\log[1-\Lambda(B)]\},$$

and $\widehat{\vartheta}$ is the estimator of the control function in the unweighted sample; and

$$\widehat{\beta}^e(y) = \arg\min_{\beta \in \mathbb{R}^{\dim(W)}} \frac{1}{n}\sum_{i=1}^{n} e_i T_i \rho_y(Y_i, \beta'\widehat{W}_i^e), \quad \widehat{W}_i^e = w(X_i, Z_{1i}, \widehat{V}_i^e), \quad \widehat{V}_i^e = \widehat{\vartheta}^e(X_i, R_i),$$

where $\widehat{\vartheta}^e$ is the estimator of the control function in the weighted sample.

The following lemma establishes a functional central limit theorem and a functional central limit theorem for the bootstrap for the estimator of the DR coefficients in the second stage. Let $d_w := \dim(W)$, and $\ell^\infty(\mathcal{Y})$ be the set of all uniformly bounded real functions on $\mathcal{Y}$, and define the matrix $J(y) := \mathrm{E}_P[\lambda(W'\beta_0(y))WW'T]$ for $y \in \mathcal{Y}$. We use $\rightsquigarrow_{\mathbb{P}}$ to denote bootstrap consistency, i.e., weak convergence conditional on the data in probability, which is formally defined in Appendix B.1.

Lemma 3. *[FCLT and Bootstrap FCLT for* $\widehat{\beta}(y)$*] Under Assumptions 1–5, in* $\ell^\infty(\mathcal{Y})^{d_w}$*,*

$$\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) \rightsquigarrow J(y)^{-1}G(y), \quad \textit{and} \quad \sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y)) \rightsquigarrow_{\mathbb{P}} J(y)^{-1}G(y),$$

where $y \mapsto G(y)$ *is a* d_w*-dimensional zero-mean Gaussian process with uniformly continuous sample paths and covariance function*

$$\mathrm{E}_P[G(y)G(v)'] = C(y,v), \quad y, v \in \mathcal{Y}.$$

We consider now the estimators of the main quantities of interest – the structural functions. Let $W_x := w(x, Z_1, V)$, $\widehat{W}_x := w(x, Z_1, \widehat{V})$, and $\widehat{W}_x^e := w(x, Z_1, \widehat{V}^e)$. The

DR estimator and bootstrap draw of the DSF in the trimmed support, $G_T(y,x) = \mathrm{E}_P\{\Lambda[\beta_0(y)'W_x] \mid T = 1\}$, are $\widehat{G}(y,x) = \sum_{i=1}^n \Lambda[\widehat{\beta}(y)'\widehat{W}_{xi}]T_i/n_T$, and $\widehat{G}^e(y,x) = \sum_{i=1}^n e_i\Lambda[\widehat{\beta}^e(y)'\widehat{W}^e_{xi}]T_i/n^e_T$. Let $p_T := P(T = 1)$. The next result gives large sample theory for these estimators.

Theorem 2 (FCLT and Bootstrap FCLT for DSF)**.** *Under Assumptions 1–5, in* $\ell^\infty(\overline{\mathcal{YX}})$,

$$\sqrt{np_T}(\widehat{G}(y,x) - G_T(y,x)) \rightsquigarrow Z(y,x) \text{ and } \sqrt{np_T}(\widehat{G}^e(y,x) - \widehat{G}(y,x)) \rightsquigarrow_{\mathbb{P}} Z(y,x),$$

where $(y,x) \mapsto Z(y,x)$ *is a zero-mean Gaussian process with covariance function*

$$\mathrm{Cov}_P[\Lambda[W_x'\beta_0(y)] + h_{y,x}(A), \Lambda[W_u'\beta_0(v)] + h_{v,u}(A) \mid T = 1],$$

with

$$h_{y,x}(A) = \mathrm{E}_P\{\lambda[W_x'\beta_0(y)]W_xT\}'^{-1}[f_y(A) + g_y(A)] + \\ \mathrm{E}_P\{\lambda[W_x'\beta_0(y)]\dot{W}_x'\beta_0(y)T\ell(a,X,R)\}\big|_{a=A}.$$

When Y is continuous and $y \mapsto G_T(y,x)$ is strictly increasing, we can also characterize the asymptotic distribution of $\widehat{Q}(\tau,x)$, the estimator of the QSF in the trimmed support. Let $g_T(y,x)$ be the density of $y \mapsto G_T(y,x)$, $\overline{\mathcal{T}} := \{\tau \in (0,1) : Q(\tau,x) \in \mathcal{Y}, g_T(Q(\tau,x),x) > \epsilon, x \in \overline{\mathcal{X}}\}$ for fixed $\epsilon > 0$, and $Q_T(\tau,x)$ the QSF in the trimmed support $\overline{\mathcal{TX}}$ defined as

$$Q_T(\tau,x) = \int_{\mathcal{Y}^+} 1\{G_T(y,x) \leq \tau\}dy - \int_{\mathcal{Y}^-} 1\{G_T(y,x) \geq \tau\}dy.$$

The estimator and its bootstrap draw given in (3.11)-(3.12) follow the functional central limit theorem:

Theorem 3 (FCLT and Bootstrap FCLT for QSF)**.** *Assume that* $y \mapsto G_T(y,x)$ *is strictly increasing in* $\overline{\mathcal{Y}}$ *and* $(y,x) \mapsto G_T(y,x)$ *is continuously differentiable in* $\overline{\mathcal{YX}}$*. Under Assumptions 1–5, in* $\ell^\infty(\overline{\mathcal{TX}})$,

$$\sqrt{np_T}(\widehat{Q}(\tau,x) - Q_T(\tau,x)) \rightsquigarrow -\frac{Z(Q(\tau,x),x)}{g_T(Q(\tau,x),x)} \text{ and} \\ \sqrt{np_T}(\widehat{Q}^e(\tau,x) - \widehat{Q}(\tau,x)) \rightsquigarrow_{\mathbb{P}} -\frac{Z(Q(\tau,x),x)}{g_T(Q(\tau,x),x)},$$

where $(y,x) \mapsto Z(y,x)$ *is the same Gaussian process as in Theorem 2.*

Finally, we consider the ASF in the trimmed support

$$\mu_T(x) = \int_{\mathcal{Y}^+} [1 - G_T(y,x)]\nu(dy) - \int_{\mathcal{Y}^-} G_T(y,x)\nu(dy).$$

The estimator and its bootstrap draw given in (3.13)-(3.14) follow the functional central limit theorem:

Theorem 4 (FCLT and Bootstrap FCLT for ASF)**.** *Under Assumptions 1–5, in* $\ell^\infty(\overline{\mathcal{X}})$*,*

$$\sqrt{np_T}(\widehat{\mu}(x) - \mu_T(x)) \rightsquigarrow - \int_{\mathcal{Y}} Z(y,x)\nu(dy) \text{ and}$$
$$\sqrt{np_T}(\widehat{\mu}^e(x) - \widehat{\mu}(x)) \rightsquigarrow_{\mathbb{P}} - \int_{\mathcal{Y}} Z(y,x)\nu(dy),$$

where $(y,x) \mapsto Z(y,x)$ *is the same Gaussian process as in Theorem 2.*

5. Implementation and Application to Estimation of Engel Curves

In this section we provide algorithms for the implementation of our methods, and apply them to the estimation of a semiparametric nonseparable triangular model for Engel curves. We focus on the structural relationship between household's total expenditure and household's demand for two goods: food and leisure. We take the outcome Y to be the expenditure share on either food or leisure, and X the logarithm of total expenditure. Endogeneity in the estimation of Engel curves arises because the decision to consume a particular good may occur simultaneously with the allocation of income between consumption and savings. Following Blundell, Chen and Kristensen (2007) we use the logarithm of gross earnings of the head of household as the variable that satisfies an exclusion restriction. We also include an additional binary covariate Z_1 accounting for the presence of children in the household.

There is an extensive literature on Engel curve estimation (e.g., see Lewbel, 2006, for a review), and the use of nonseparable triangular models for the identification and estimation of Engel curves has been considered in the recent literature. Blundell, Chen and Kristensen (2007) estimate semi-nonparametrically Engel curves for several categories of expenditure, Imbens and Newey (2009) estimate the QSF nonparametrically for food and leisure, and Chernozhukov, Fernandez-Val and Kowalski (2015) estimate Engel curves for alcohol accounting for censoring. For comparison purposes we use the same dataset as these papers, the 1995 U.K. Family Expenditure Survey.

We restrict the sample to 1,655 married or cohabiting couples with two or fewer children, in which the head of the household is employed and between the ages of 20 and 55 years. For this sample we estimate the DSF, QSF and ASF for both goods. Unlike Imbens and Newey (2009) we also account for the presence of children in the household and we impose semiparametric restrictions through our baseline models. In contrast to Chernozhukov, Fernandez-Val and Kowalski (2015), we do not impose separability between the control function and other regressors, and we estimate the structural functions.

5.1. **Implementation of Estimation and Inference Methods.** In order to guide implementation of our methods, we provide step-by-step implementation algorithms for the three-stage estimation procedure, weighted bootstrap, and the construction of uniform bands for the structural functions. All structural functions are estimated by both QR and DR methods, following exactly the description of the implementation presented in Section 3 with the specifications $r(Z) = (1, Z)'$, $r_1(Z_1) = (1, Z_1)'$, $p(X) = (1, X)'$, and $q(V) = (1, \Phi^{-1}(V))'$. We implement our methods in the software R (R Development Core Team, 2019), using the open source `quantreg` R package (Koenker, 2018) for QR, and the `glm` function for DR.

In the empirical application the regions of interest for the structural functions are $\widetilde{\mathcal{X}} = [\widehat{Q}_X(0.1), \widehat{Q}_X(0.9)]$ and $\widetilde{\mathcal{Y}} = [\widehat{Q}_Y(0.1), \widehat{Q}_Y(0.9)]$, where $\widehat{Q}_X(u)$ and $\widehat{Q}_Y(u)$ are the sample u-quantiles of X and Y, respectively. We approximate $\widetilde{\mathcal{X}}$ by a grid $\widetilde{\mathcal{X}}_K$ with $K = 3, 5$, and $\widetilde{\mathcal{Y}}$ by a grid $\widetilde{\mathcal{Y}}_{15}$. We estimate the structural functions and perform uniform inference on the structural functions over the following regions:

(1) For the QSF, $\widehat{Q}(\tau, x)$, we take $\widetilde{\mathcal{T}} = \{0.25, 0.5, 0.75\}$, and then set: $\mathcal{I}_Q = \widetilde{\mathcal{T}}\widetilde{\mathcal{X}}_5$.
(2) For the DSF, $\widehat{G}(y, x)$, we set: $\mathcal{I}_G = \widetilde{\mathcal{Y}}_{15}\widetilde{\mathcal{X}}_3$.
(3) For the ASF, $\widehat{\mu}(x)$, we set: $\mathcal{I}_\mu = \widetilde{\mathcal{X}}_5$.

5.1.1. *Estimation.* Algorithm 1 is implemented for estimation of structural functions. In Algorithm 1, the choice of the link function for DR, of ϵ for QR, and the size of the grids M can differ across stages and methods. In the empirical application we implement the DR estimator using the logit link function, and we set $\epsilon = 0.01$ for QR, and $M = 599$ throughout. For the third stage, we approximate the integrals (3.12) and (3.14) using $S = 599$ points. For DR, since the estimated DSF may be non-monotonic in y, we apply rearrangement to $y \mapsto \widehat{G}(y, x)$ at each value of x in the region of interest, using the `Rearrangement` R package (Graybill, Chen, Chernozhukov, Fernandez-Val,

Algorithm 1 ***Three-Stage Estimation Procedure.***

For $i = 1, \dots, n$, set $e_i = 1$.

First Stage. *[Control function estimation]*

(1) *(QR) For ϵ in $(0, 0.5)$ (e.g., $\epsilon = .01$) and a fine mesh of M values $\{\epsilon = v_1 < \cdots < v_M = 1 - \epsilon\}$, estimate $\{\widehat{\pi}^e(v_m)\}_{m=1}^M$ by solving (3.2). Then set $\widehat{V}_i^e = \widehat{F}_X^e(X_i \mid Z_i)$, $i = 1, \dots, n$, as in (3.1).*

(2) *(DR) Estimate $\{\widehat{\pi}(X_i)\}_{i=1}^n$ by solving (3.4). Then set $\widehat{V}_i^e = \widehat{F}_X^e(X_i \mid Z_i)$, $i = 1, \dots, n$, as in (3.3).*

Second Stage. *[Reduced-form CDF estimation]*

(1) *(QR) (a) For ϵ in $(0, 0.5)$ (e.g., $\epsilon = .01$) and a fine mesh of M values $\{\epsilon = u_1, \dots, u_M = 1 - \epsilon\}$, estimate $\{\widehat{\beta}^e(u_m)\}_{m=1}^M$ by solving (3.6).*
(b) Obtain $\widehat{F}_Y^e(y \mid x, Z_{1i}, \widehat{V}_i^e)$ as in (3.5)

(2) *(DR) (a) For each $y_m \in \mathcal{Y}_M$, estimate $\{\widehat{\beta}(y_m)\}_{m=1}^M$* by solving (3.8).
(b) Obtain $\widehat{F}_Y^e(y \mid x, Z_{1i}, \widehat{V}_i^e)$ as in (3.7).

Third Stage. *[Structural functions estimation] For $n_T^e = \sum_{i=1}^n e_i T_i$ and a fine mesh of S values $\{\inf[y \in \mathcal{Y}] = y_1 < \cdots < y_S = \sup[y \in \mathcal{Y}]\}$, compute*

$$\widehat{G}^e(y, x) = \frac{1}{n_T^e} \sum_{i=1}^n e_i \widehat{F}_Y^e(y \mid x, Z_{1i}, \widehat{V}_i^e) T_i,$$

$$\widehat{Q}_S^e(\tau, x) = \delta \sum_{s=1}^S \left[1(y_s \geq 0) - 1\{\widehat{G}^e(y_s, x) \geq \tau\} \right], \widehat{\mu}_S^e(x) = \delta \sum_{s=1}^S \left[1(y_s \geq 0) - \widehat{G}^e(y_s, x) \right].$$

and Galichon, 2016). Overall, for the empirical application we have found that the estimates are not very sensitive to M and the choice of link function, and are also robust to varying values of ϵ and S. None of the methods uses trimming, that is we set $T = 1$ a.s.

Remark 2. For DR, the estimation of $\pi(x)$ at each $x = X_i$ can be computationally expensive. Substantial gains in computational speed is achieved by first estimating $\pi(x)$ in a grid $\overline{\mathcal{X}}_M$, and then obtaining $\widehat{\pi}(x)$ at each $x = X_i$ by interpolation.

Remark 3. All the estimation steps can also be implemented keeping Z_1, or some component of Z_1, fixed as a conditioning variable. The estimated structural functions are then evaluated at values of the conditioning variable(s) of interest. Denoting the DSF estimator and bootstrap draw by $\widehat{G}(y, x, z_1) = \sum_{i=1}^n \widehat{F}_Y(y \mid x, z_1, \widehat{V}_i) T_i / n_T$ and $\widehat{G}^e(y, x, z_1) = \sum_{i=1}^n e_i \widehat{F}_Y^e(y \mid x, z_1, \widehat{V}_i^e) T_i / n_T^e$, the corresponding QSF and ASF

Algorithm 2 ***Weighted Bootstrap.***

For $b = 1, \ldots, B$, repeat the following steps:

Step 0. *Draw $e_b := \{e_{ib}\}_{i=1}^n$ i.i.d. from a random variable that satisfies Assumption 3 (e.g., the standard exponential distribution).*

Step 1. *Reestimate the control function $\widehat{V}_{ib}^e = \widehat{F}_{X,b}^e(X_i \mid Z_i)$ in the weighted sample, according to (3.1)-(3.2) or (3.3)-(3.4).*

Step 2. *Reestimate the reduced form CDF $\widehat{F}_{Y,b}^e$ in the weighted sample according to (3.5)-(3.6) or (3.7)-(3.8).*

Step 3. *For $n_{Tb}^e = \sum_{i=1}^n e_{ib} T_i$ and the fine mesh of S values specified in the Third Stage of Algorithm 1, compute*

$$\widehat{G}_b^e(y, x) = \frac{1}{n_{Tb}^e} \sum_{i=1}^n e_{ib} \widehat{F}_{Y,b}^e(y \mid x, Z_{1i}, \widehat{V}_{ib}^e) T_i,$$

$$\widehat{Q}_b^e(\tau, x) = \delta \sum_{s=1}^S \left[1(y_s \geq 0) - 1\{\widehat{G}_b^e(y_s, x) \geq \tau\} \right], \widehat{\mu}_b^e(x) = \delta \sum_{s=1}^S \left[1(y_s \geq 0) - \widehat{G}_b^e(y_s, x) \right].$$

estimators and bootstrap draws obtain upon substituting $\widehat{G}(y, x, z_1)$ and $\widehat{G}^e(y, x, z_1)$ for $\widehat{G}(y, x)$ and $\widehat{G}^e(y, x)$ in (3.9)-(3.10).

Remark 4. For the QR specification, the estimator of the ASF in the second and third stages can be replaced by $\widehat{\mu}(x) = w(x, \bar{Z}_1, 0)'\widehat{\beta}$, where $\bar{Z}_1 = \sum_{i=1}^n Z_{1i}/n$ and $\widehat{\beta}$ the least squares estimator of the linear regression of Y on $\widehat{W}_i^e$. Our numerical implementation in the Supplementary Material shows that estimates thus obtained are very similar to those formed according to (3.16).

5.1.2. *Inference.* Algorithm 2 is implemented in order to obtain B weighted bootstrap versions of our estimators. The B weighted bootstrap estimates are used for the construction of confidence bands for the structural functions, according to Algorithm 3. The resulting confidence bands are then valid uniformly over specified regions of interest.

In the empirical application, we run $B = 199$ bootstrap replications in Algorithm 2 for both methods. We then implement Algorithm 3 in order to perform uniform inference on the structural functions over the specified regions $\mathcal{I}_Q$, $\mathcal{I}_G$ and $\mathcal{I}_\mu$. While we found confidence bands to be robust to the choice of B in the empirical application, the choice of the regions of interest over which to construct the uniform bands has a noticeable effect on the width and shape of the bands. This is especially the case for DR, while

Algorithm 3 ***Uniform Inference for Structural Functions.***

Step 1. *Given B bootstrap draws $\{(\widehat{G}_b^e(y,x), \widehat{Q}_b^e(\tau,x), \widehat{\mu}_b^e(x))\}_{b=1}^B$, compute the standard errors of $\widehat{G}(y,x)$, $\widehat{Q}(\tau,x)$, and $\widehat{\mu}(x)$ as*

$$\widehat{\sigma}_G(y,x) = IQR\left[\{\widehat{G}_b^e(y,x)\}_{b=1}^B\right]/1.349,$$

$$\widehat{\sigma}_Q(\tau,x) = IQR\left[\{\widehat{Q}_b^e(\tau,x)\}_{b=1}^B\right]/1.349, \qquad \widehat{\sigma}_\mu(x) = IQR\left[\{\widehat{\mu}_b^e(x)\}_{b=1}^B\right]/1.349.$$

Step 2. *For $b = 1, \dots, B$, compute the bootstrap draws of the maximal t-statistics for the DSF, QSF, and ASF as*

$$\left\|t_{G,b}^e(y,x)\right\|_{\mathcal{I}_G} = \sup_{(y,x)\in\mathcal{I}_G}\left|\frac{\widehat{G}_b^e(y,x) - \widehat{G}(y,x)}{\widehat{\sigma}_G(y,x)}\right|,$$

$$\left\|t_{Q,b}^e(\tau,x)\right\|_{\mathcal{I}_Q} = \sup_{(\tau,x)\in\mathcal{I}_Q}\left|\frac{\widehat{Q}_b^e(\tau,x) - \widehat{Q}(\tau,x)}{\widehat{\sigma}_Q(\tau,x)}\right|, \quad \left\|t_{\mu,b}^e(x)\right\|_{\mathcal{I}_\mu} = \sup_{x\in\mathcal{I}_\mu}\left|\frac{\widehat{\mu}_b^e(x) - \widehat{\mu}(x)}{\widehat{\sigma}_\mu(x)}\right|.$$

Step 3. *(i) (DSF and ASF) Form $(1-\alpha)$-confidence bands for the DSF and ASF as*

$$\left\{\widehat{G}(y,x) \pm \widehat{k}_G(1-\alpha)\widehat{\sigma}_G(y,x) : (y,x) \in \mathcal{I}_G\right\}, \quad \left\{\widehat{\mu}(x) \pm \widehat{k}_\mu(1-\alpha)\widehat{\sigma}_\mu(x) : x \in \mathcal{I}_\mu\right\},$$

where $\widehat{k}_G(1-\alpha)$ is the sample $(1-\alpha)$-quantile of $\{||t_{G,b}^e(y,x)||_{\mathcal{I}_G} : 1 \le b \le B\}$, and $\widehat{k}_\mu(1-\alpha)$ is the sample $(1-\alpha)$-quantile of $\{||t_{\mu,b}^e(x)||_{\mathcal{I}_\mu} : 1 \le b \le B\}$.
(ii) (QSF) If Y is continuous, form a $(1-\alpha)$-confidence band for the QSF as

$$\left\{\widehat{Q}(\tau,x) \pm \widehat{k}_Q(1-\alpha)\widehat{\sigma}_Q(\tau,x) : (\tau,x) \in \mathcal{I}_Q\right\},$$

where $\widehat{k}_Q(1-\alpha)$ is the sample $(1-\alpha)$-quantile of$\{||t_{Q,b}^e(\tau,x)||_{\mathcal{I}_Q} : 1 \le b \le B\}$. Otherwise, form a $(1-\alpha)$-confidence band for the QSF as

$$\left\{\left[\widehat{G}_U^{\leftarrow}(\tau,x), \widehat{G}_L^{\leftarrow}(\tau,x)\right] : (\tau,x) \in \mathcal{I}_G^{\leftarrow}\right\},$$

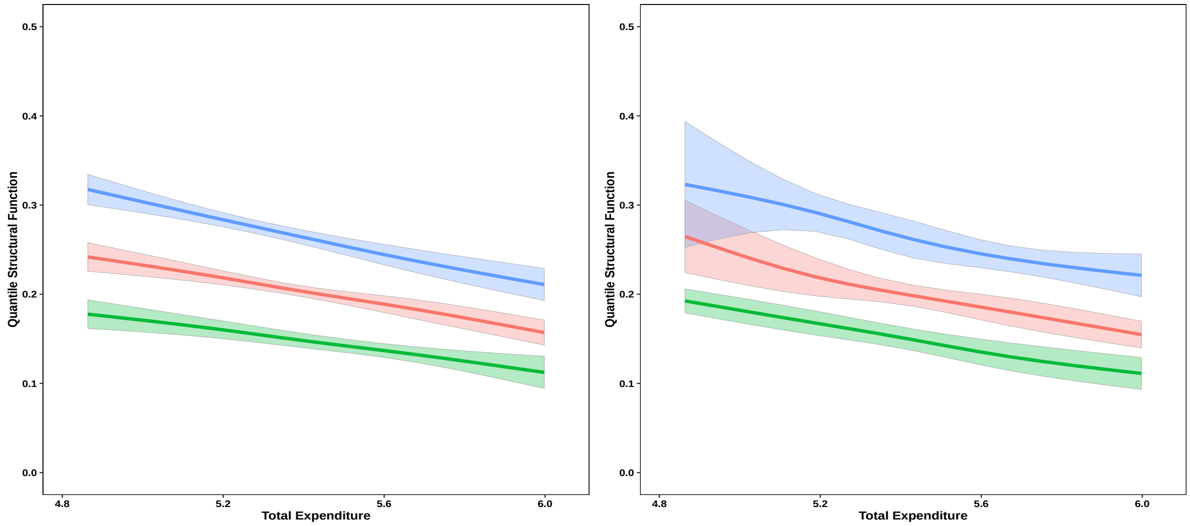
where

$$\mathcal{I}_G^{\leftarrow} = \{(\tau,x) : \widehat{G}_L(y,x) = \tau, (y,x) \in \mathcal{I}_G\} \cap \{(\tau,x) : \widehat{G}_U(y,x) = \tau, (y,x) \in \mathcal{I}_G\},$$

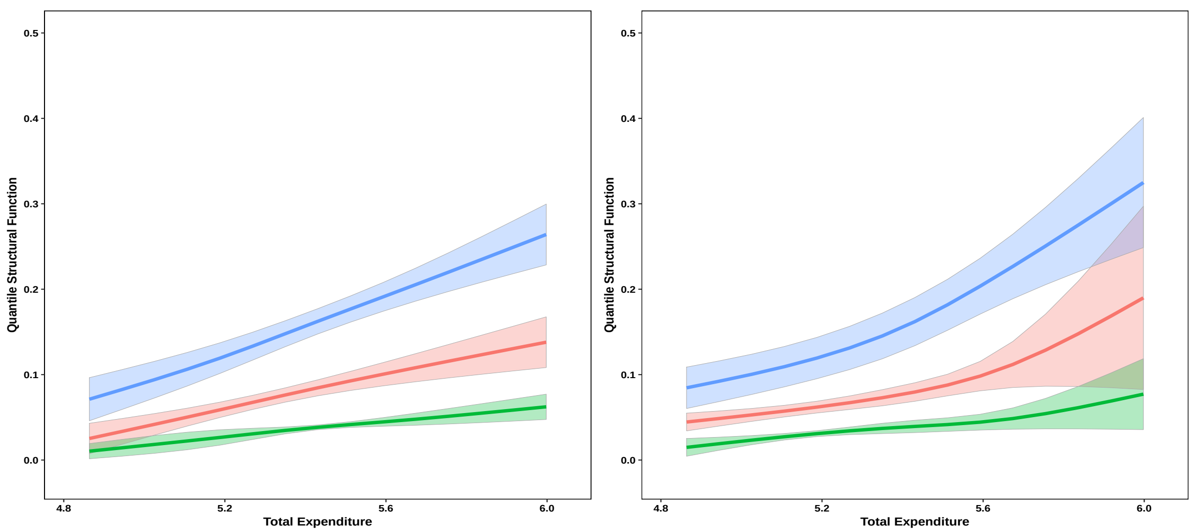
with

$$\widehat{G}_L(y,x) = \widehat{G}(y,x) - \widehat{k}_G(1-\alpha)\widehat{\sigma}_G(y,x), \quad \widehat{G}_U(y,x) = \widehat{G}(y,x) + \widehat{k}_G(1-\alpha)\widehat{\sigma}_G(y,x).$$

QR inference appears to be less sensitive to the choice of regions of interest. In the Supplementary Material we illustrate the effect of varying the definition of regions of interest on weighted bootstrap confidence bands.



(A) Food.



(B) Leisure.

FIGURE 5.1. QSF. 0.25-QSF (green), 0.5-QSF (red), and 0.75-QSF (blue). Quantile (left) and distribution regression (right).

5.2. **Empirical Results.** Figures 5.1-5.3 show the QSF, ASF and DSF for both goods[13]. For each structural function, we report weighted bootstrap 90%-confidence bands that are uniform over the corresponding region specified above. Our empirical results illustrate that QR and DR specifications are able to capture different features

[13]For graphical representation the QSF and ASF are interpolated by splines over $\overline{\mathcal{X}}$ and the DSF over $\overline{\mathcal{Y}}$.

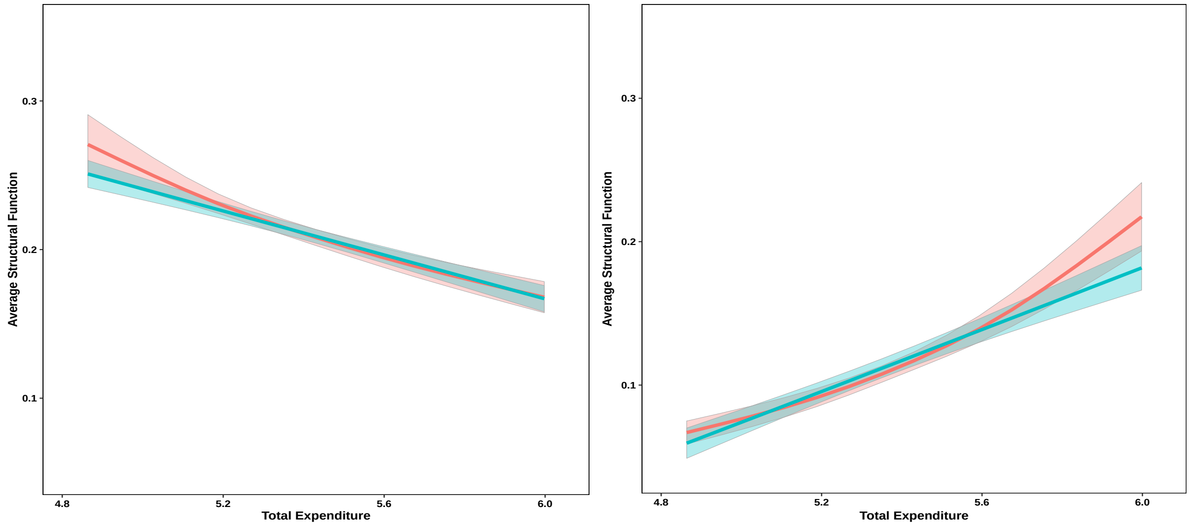
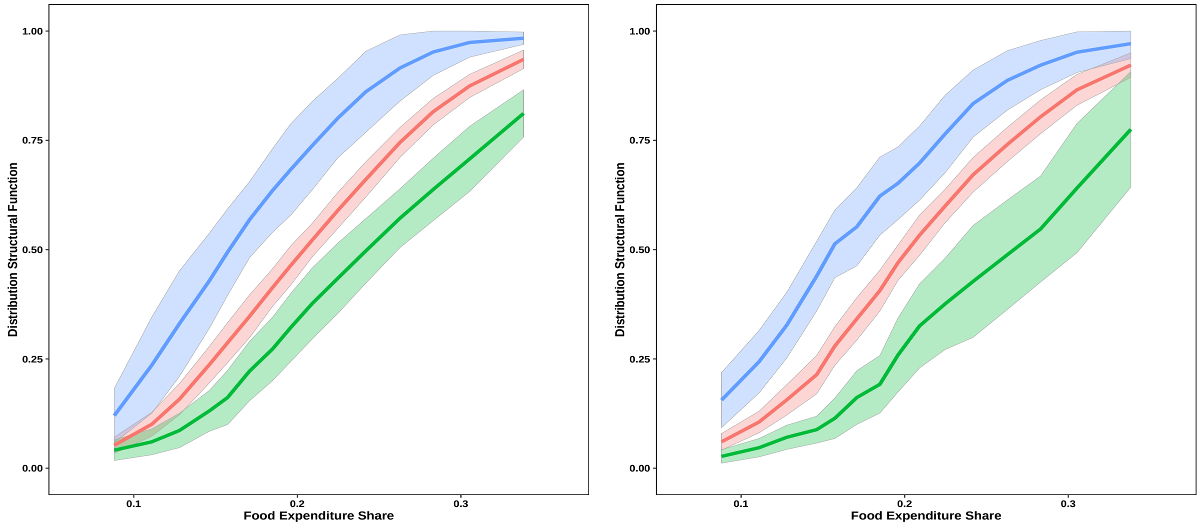


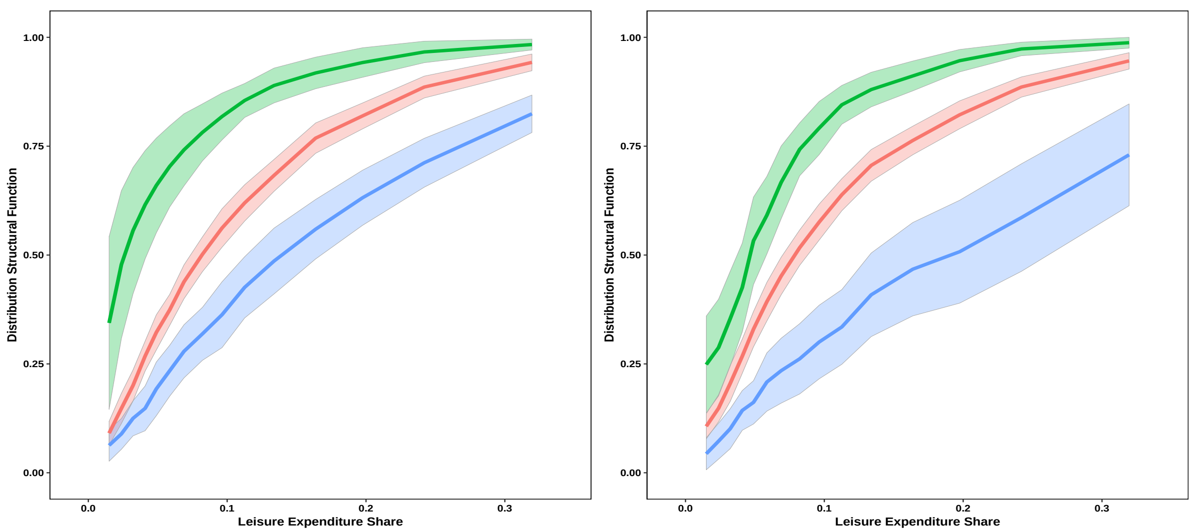
FIGURE 5.2. ASF for food (left) and leisure (right). Quantile (blue) and distribution regression (red).

of structural functions, and are therefore complementary. For food, both estimation methods deliver very similar QSF estimates, close to being linear, although linearity is not imposed in the estimation procedure. For leisure, the QSF and ASF estimated by DR are able to capture some nonlinearity which is absent from those obtained by QR. For QR, this reflects the specified linear structure of the ASF which also constrains the shape of the QSF. In addition, some degree of heteroskedasticity appears to be a feature of the structural model for both goods, although much more markedly for leisure, so our methods are well-suited for this problem. Increased dispersion across quantile levels in Figure 5.1 is reflected by the increasing spread across probability levels between the two extreme DSF estimates in Figure 5.3, an important feature of the data highlighted in Imbens and Newey (2009).

Our baseline models naturally allow for the inclusion of transformations of covariates - for instance spline transformations - in order to account for potential nonlinearities in data. In practice, these augmented specifications are useful to verify the robustness of the empirical findings based on the baseline specifications. In order to illustrate nonlinear implementations of our approach and robustness of our baseline estimates, the QSF for food and leisure obtained by taking cubic B-splines transformations with 4 knots of log-total expenditure are shown in Figure 5.4, for both DR and QR methods. A complete description of the structural stochastic relationship between total



(A) Food.



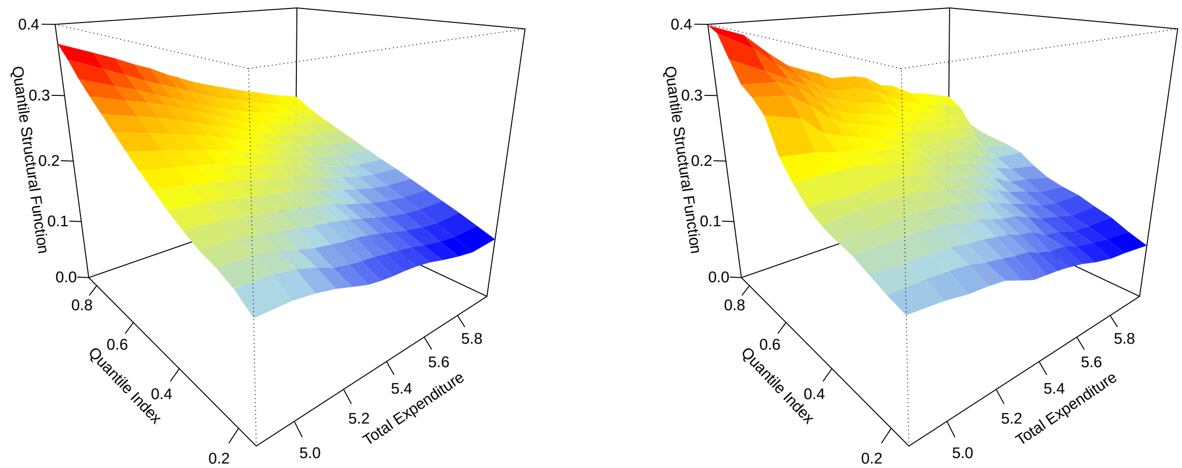
(B) Leisure.

FIGURE 5.3. DSF. DSF at $X = \widehat{Q}_X(0.1)$ (green), $X = \widehat{Q}_X(0.5)$ (red), $X = \widehat{Q}_X(0.9)$ (blue). Quantile (left) and distribution regression (right).

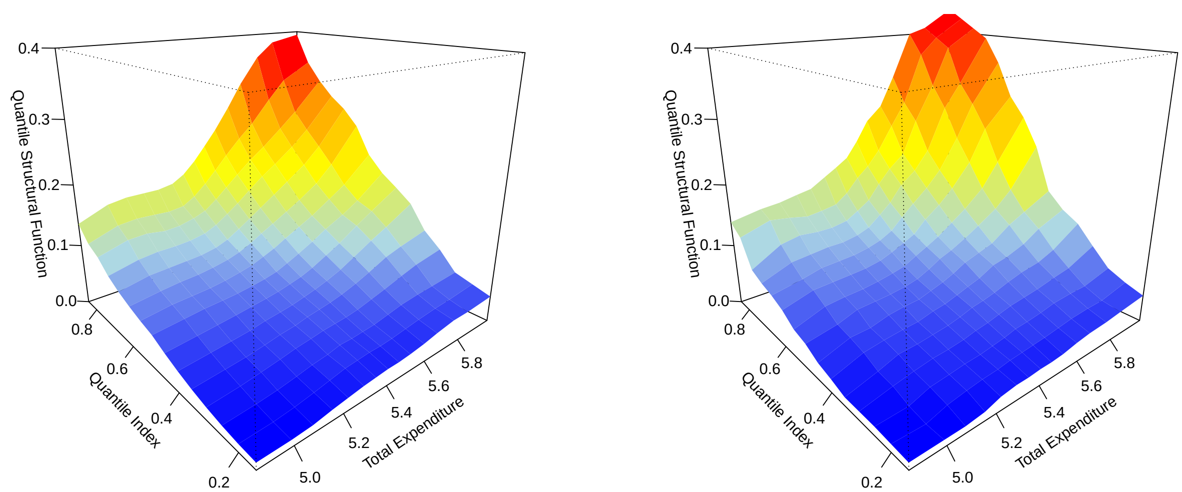
expenditure and food and leisure shares is then obtained, and confirms the essentially linear form of the QSF for food, as well as the nonlinearity already detected by DR for leisure in the empirical application - without the inclusion of nonlinear transformations of log-total expenditure.

Compared to existing studies of this dataset, the empirical results presented for the DSF are new. Our semiparametric estimates of the ASF and QSF capture the main features displayed by the nonparametric estimates of Imbens and Newey (2009), or those we obtain with more flexible specifications in Figure 5.4. Our ASF estimates are also similar to those obtained by nonparametric instrumental variable regression in Blundell, Chen and Kristensen (2007), who do not consider estimation of quantile and distributional structural relationships. In particular, our DR estimates for leisure capture the nonlinearities at higher total expenditure levels. In contrast to these previous studies, our results and methods further make it possible to construct uniform confidence regions for all structural functions, thereby providing applied researchers with useful inferential tools.

Overall the empirical results we present illustrate that our parsimonious models are able to capture complex features of the data, such as asymmetric distributions and nonlinear structural relationships, while leading to relatively easy-to-implement estimators and inferential methods that can be augmented straightforwardly for robustness checks and additional flexibility. This is demonstrated further in the Supplementary Material where we perform a thorough sensitivity analysis which further shows that our empirical results are robust to the modelling, estimation and integration choices.



(A) Food.



(B) Leisure.

FIGURE 5.4. Flexible QSF specification. QR (left) and DR (right) .

Appendix A. Identification

A.1. **Proof of Lemma 1.** By Assumption 2, $E_\mu[p(X)p(X)']$, $E_\varsigma[r_1(Z_1)r_1(Z_1)']$, and $E_\rho[q(V)q(V)']$ are positive definite. Also, a measure μ_1 dominates a measure μ_2 if $\mu_1(A) \geq \mu_2(A)$ for any measurable set A. This implies that for any nonnegative function $f(x, z_1, v)$:

$$\int f(x, z_1, v)\mu_1(dx, dz_1, dv) \geq \int f(x, z_1, v)\mu_2(dx, dz_1, dv).$$

Therefore, taking $f(X, Z_1, V) = \{b'W\}^2$ with $W = w(X, Z_1, V)$ and $b \neq 0$, Assumption 2 implies

$$\begin{aligned} E[\{b'W\}^2] &\geq \int \{b'W\}^2[\mu(dx) \times \varsigma(dz_1) \times \rho(dv)] \\ &= b'\left\{\int \{WW'\}[\mu(dx) \times \varsigma(dz_1) \times \rho(dv)]\right\} b \\ &= b'\left\{\int \{p(x)p(x)'\} \otimes \{r_1(z_1)r_1(z_1)'\} \otimes \{q(v)q(v)'\}[\mu(dx) \times \varsigma(dz_1) \times \rho(dv)]\right\} b \\ &= b'\{E_\mu[p(X)p(X)'] \otimes E_{\varsigma_1}[r_1(Z_1)r_1(Z_1)'] \otimes E_\rho[q(V)q(V)']\}b > 0, \end{aligned}$$

where the result follows by positive definiteness of the kronecker product of positive definite matrices. □

A.2. **Proof of Theorem 1.** We show identification of $Q_Y(U \mid X, Z_1, V)$ and $F_Y(Y \mid X, Z_1, V)$ for the QR specification, and of $F_Y(Y \mid X, Z_1, V)$ for the DR specification. Identification of structural functions then follows from their definitions in Section 2.

For the QR specification, suppose there exists $\tilde\beta(u) \neq \beta(u)$ such that $\beta(u)'w(X, Z_1, V) = \tilde\beta(u)'w(X, Z_1, V)$, $u \in (0, 1)$. Then $\{\beta(u) - \tilde\beta(u)\}'w(X, Z_1, V) = 0$, and, for $u \in (0, 1)$,

$$\begin{aligned} 0 &= E\left[(\beta(u) - \tilde\beta(u))'\{w(X, Z_1, V)w(X, Z_1, V)'\}(\beta(u) - \tilde\beta(u))\right] \\ &= (\beta(u) - \tilde\beta(u))'E[w(X, Z_1, V)w(X, Z_1, V)'](\beta(u) - \tilde\beta(u)), \end{aligned}$$

which yields a contradiction since $E[w(X, Z_1, V)w(X, Z_1, V)']$ is positive definite under Assumption 2, by Lemma 1. Thus $\beta(u)$ is identified for all $u \in (0, 1)$. Identification of the QR coefficients implies that $Q_Y(U \mid X, Z_1, V)$ is identified under Assumption 1(a). Identification of $F_Y(Y \mid X, Z_1, V)$ then follows by $y \mapsto F_Y(y \mid X, Z_1, V)$ being the inverse of $u \mapsto Q_Y(u \mid X, Z_1, V)$.

For the DR specification, positive definiteness of $E[w(X, Z_1, V)w(X, Z_1, V)']$ under Assumption 2 by Lemma 1 is also sufficient for identification of the DR coefficients by standard identification results for Logit and Probit models, e.g., see Example 1.2 in Newey and McFadden (1994). The reduced form CDF $F_Y(Y \mid X, Z_1, V)$ is then identified under Assumption 1(b) by Γ and $w(X, Z_1, V)$ being known functions. $\square$

Appendix B. Asymptotic Theory

B.1. **Notation.** In what follows ϑ denotes a generic value for the control function. It is convenient also to introduce some additional notation, which will be extensively used in the proofs. Let $V_i(\vartheta) := \vartheta(X_i, Z_i)$, $W_i(\vartheta) := w(X_i, Z_{1i}, V_i(\vartheta))$, and $\dot{W}_i(\vartheta) := \partial_v w(X_i, Z_{1i}, v)|_{v=V_i(\vartheta)}$. When the previous functions are evaluated at the true values we use $V_i = V_i(\vartheta_0)$, $W_i = W_i(\vartheta_0)$, and $\dot{W}_i = \dot{W}_i(\vartheta_0)$. Also, let $\rho_y(u, v) := -1(u \leq y) \log \Lambda(v) - 1(u > y) \log \Lambda(-v)$. Recall that $A := (Y, X, Z, W, V)$, $T(x) = 1(x \in \overline{\mathcal{X}})$, and $T = T(X)$. For a function $f : \mathcal{A} \mapsto \mathbb{R}$, we use $\|f\|_{T,\infty} = \sup_{a \in \mathcal{A}} |T(x)f(a)|$; for a K-vector of functions $f : \mathcal{A} \mapsto \mathbb{R}^K$, we use $\|f\|_{T,\infty} = \sup_{a \in \mathcal{A}} \|T(x)f(a)\|_2$. We make functions in Υ as well as estimators $\widehat{\vartheta}$ to take values in $[0, 1]$, the support of the control function V. This allows us to simplify notation in what follows.

We adopt the standard notation in the empirical process literature (see, e.g., van der Vaart, 1998),

$$\mathbb{E}_n[f] = \mathbb{E}_n[f(A)] = n^{-1} \sum_{i=1}^{n} f(A_i),$$

and

$$\mathbb{G}_n[f] = \mathbb{G}_n[f(A)] = n^{-1/2} \sum_{i=1}^{n} (f(A_i) - \mathrm{E}_P[f(A)]).$$

When the function $\widehat{f}$ is estimated, the notation should interpreted as:

$$\mathbb{G}_n[\widehat{f}\,] = \mathbb{G}_n[f] \mid_{f=\widehat{f}} \text{ and } \mathrm{E}_P[\widehat{f}\,] = \mathrm{E}_P[f] \mid_{f=\widehat{f}}.$$

We also use the concepts of covering entropy and bracketing entropy in the proofs. The covering entropy $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of $\|\cdot\|$-balls of radius ϵ needed to cover the set of functions $\mathcal{F}$. The bracketing entropy $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of ϵ-brackets in $\|\cdot\|$ needed to cover the set of functions $\mathcal{F}$. An ϵ-bracket $[\ell, u]$ in $\|\cdot\|$ is the set of functions f with $\ell \leq f \leq u$ and $\|u - \ell\| < \epsilon$.

For a sequence of random functions $y \mapsto f_n(y)$ and a deterministic sequence a_n, we use $f_n(y) = \bar{o}_{\mathbb{P}}(a_n)$ and $f_n(y) = \bar{O}_{\mathbb{P}}(a_n)$ to denote uniform in $y \in \mathcal{Y}$ orders in probability, i.e., $\sup_{y\in\mathcal{Y}} f_n(y) = o_{\mathbb{P}}(a_n)$ and $\sup_{y\in\mathcal{Y}} f_n(y) = O_{\mathbb{P}}(a_n)$, respectively. The uniform in $y \in \mathcal{Y}$ deterministic orders $\bar{o}(a_n)$ and $\bar{O}(a_n)$ are defined analogously suppressing the $\mathbb{P}$ subscripts.

We follow the notation and definitions in van der Vaart and Wellner (1996) of bootstrap consistency. Let D_n denote the data vector and E_n be the vector of bootstrap weights. Consider the random element $Z_n^e = Z_n(D_n, E_n)$ in a normed space $\mathbb{Z}$. We say that the bootstrap law of Z_n^e consistently estimates the law of some tight random element Z and write $Z_n^e \rightsquigarrow_{\mathbb{P}} Z$ in $\mathbb{Z}$ if

(B.1) $$\sup_{h\in \mathrm{BL}_1(\mathbb{Z})} \left|\mathrm{E}_P^e h\left(Z_n^e\right) - \mathrm{E}_P h(Z)\right| \to_{\mathbb{P}^*} 0,$$

where $\mathrm{BL}_1(\mathbb{Z})$ denotes the space of functions with Lipschitz norm at most 1, E_P^e denotes the conditional expectation with respect to E_n given the data D_n, and $\to_{\mathbb{P}^*}$ denotes convergence in (outer) probability.

B.2. **Proof of Lemma 3.** We only consider the case where $\mathcal{Y}$ is a compact interval of $\mathbb{R}$. The case where $\mathcal{Y}$ is finite is simpler and follows similarly.

B.2.1. *Auxiliary Lemmas.* We start with two results on stochastic equicontinuity and a local expansion for the second stage estimators that will be used in the proof of Lemma 3.

Lemma 4. *[Stochastic equicontinuity] Let $e \geq 0$ be a positive random variable with $\mathrm{E}_P[e] = 1$, $\mathrm{Var}_P[e] = 1$, and $\mathrm{E}_P|e|^{2+\delta} < \infty$ for some $\delta > 0$, that is independent of (Y, X, Z, W, V), including as a special case $e = 1$, and set, for $A = (e, Y, X, Z, W, V)$,*

$$f_y(A, \vartheta, \beta) := e \cdot [\Lambda(W(\vartheta)'\beta) - 1(Y \leq y)] \cdot W(\vartheta) \cdot T.$$

Under Assumptions 3–5 the following relations are true.

(a) *Consider the set of functions*

$$\mathcal{F} = \{f_y(A, \vartheta, \beta)'\alpha : (\vartheta, \beta, y) \in \Upsilon_0 \times \mathcal{B} \times \mathcal{Y}, \alpha \in \mathbb{R}^{\dim(W)}, \|\alpha\|_2 \leq 1\},$$

where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$, $\mathcal{B}$ is a compact set under the $\|\cdot\|_2$ metric containing $\beta_0(y)$ for all $y \in \mathcal{Y}$, Υ_0 is the intersection of Υ, defined in Lemma 2, with a neighborhood of ϑ_0 under the $\|\cdot\|_{T,\infty}$ metric. This class is P-Donsker with a square integrable envelope of the form e times a constant.

(b) *Moreover, if* $(\vartheta, \beta(y)) \to (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$*, then*

$$\sup_{y \in \mathcal{Y}} \|f_y(A, \vartheta, \beta(y)) - f_y(A, \vartheta_0, \beta_0(y))\|_{P,2} \to 0.$$

(c) *Hence for any* $(\widetilde{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$ *such that* $\widetilde{\vartheta} \in \Upsilon_0$*,*

$$\sup_{y \in \mathcal{Y}} \|\mathbb{G}_n f_y(A, \widetilde{\vartheta}, \widetilde{\beta}(y)) - \mathbb{G}_n f_y(A, \vartheta_0, \beta_0(y))\|_2 \to_{\mathbb{P}} 0.$$

(d) *For any* $(\widehat{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$*, so that*

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n}), \text{ where } \widetilde{\vartheta} \in \Upsilon_0,$$

we have that

$$\sup_{y \in \mathcal{Y}} \|\mathbb{G}_n f_y(A, \widehat{\vartheta}, \widetilde{\beta}(y)) - \mathbb{G}_n f_y(A, \vartheta_0, \beta_0(y))\|_2 \to_{\mathbb{P}} 0.$$

Proof of Lemma 4. The proof is divided in subproofs of each of the claims.

Proof of Claim (a). The proof proceeds in several steps.

Step 1. Here we bound the bracketing entropy for

$$\mathcal{I}_1 = \{[\Lambda(W(\vartheta)'\beta) - 1(Y \leq y)]T : \beta \in \mathcal{B}, \vartheta \in \Upsilon_0, y \in \mathcal{Y}\}.$$

For this purpose consider a mesh $\{\vartheta_k\}$ over Υ_0 of $\|\cdot\|_{T,\infty}$ width δ, a mesh $\{\beta_l\}$ over $\mathcal{B}$ of $\|\cdot\|_2$ width δ, and a mesh $\{y_j\}$ over $\mathcal{Y}$ of $\|\cdot\|_2$ width δ. A generic bracket over $\mathcal{I}_1$ takes the form

$$[i_1^0, i_1^1] = [\{\Lambda(W(\vartheta_k)'\beta_l - \kappa\delta) - 1(Y \leq y_j - \delta)\}T, \{\Lambda(W(\vartheta_k)'\beta_l + \kappa\delta) - 1(Y \leq y_j + \delta)\}T],$$

where $\kappa = L_W \max_{\beta \in \mathcal{B}} \|\beta\|_2 + L_W$, and $L_W := \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$.

Note that this is a valid bracket for all elements of $\mathcal{I}_1$ because for any ϑ located within δ from ϑ_k and any β located within δ from β_l,

$$\begin{aligned} |W(\vartheta)'\beta - W(\vartheta_k)'\beta_l|T &\leq |(W(\vartheta) - W(\vartheta_k))'\beta|T + |W(\vartheta_k)'(\beta - \beta_l)|T \\ &\leq L_W \delta \max_{\beta \in \mathcal{B}} \|\beta\|_2 + L_W \delta \leq \kappa\delta, \end{aligned} \quad \text{(B.2)}$$

and the $\|\cdot\|_{P,2}$-size of this bracket is given by

$$
\begin{aligned}
\|i_1^0 - i_1^1\|_{P,2} &\leq \sqrt{\mathrm{E}_P[P\{Y \in [y \pm \delta] \mid X, Z\}T]} \\
&+ \sqrt{\mathrm{E}_P[\{\Lambda(W(\vartheta_k)'\beta_l + \kappa\delta) - \Lambda(W(\vartheta_k)'\beta_l - \kappa\delta)\}^2 T]} \\
&\leq \sqrt{\|f_Y(\cdot \mid \cdot)\|_{T,\infty} 2\delta} + \kappa\delta/2,
\end{aligned}
$$

because $\|\lambda(\cdot)\|_{T,\infty} \leq 1/4$, where $\lambda = \Lambda(1-\Lambda)$ is the derivative of Λ.

Hence, counting the number of brackets induced by the mesh created above, we arrive at the following relationship between the bracketing entropy of $\mathcal{I}_1$ and the covering entropies of Υ_0, $\mathcal{B}$, and $\mathcal{Y}$,

$$
\begin{aligned}
\log N_{[]}(\epsilon, \mathcal{I}_1, \|\cdot\|_{P,2}) &\lesssim \log N(\epsilon^2, \Upsilon_0, \|\cdot\|_{T,\infty}) + \log N(\epsilon^2, \mathcal{B}, \|\cdot\|_2) + \log N(\epsilon^2, \mathcal{Y}, \|\cdot\|_2) \\
&\lesssim 1/(\epsilon^2 \log^4 \epsilon) + \log(1/\epsilon) + \log(1/\epsilon),
\end{aligned}
$$

and so $\mathcal{I}_1$ is P-Donsker with a constant envelope.

Step 2. Similarly to Step 1, it follows that

$$
\mathcal{I}_2 = \{W(\vartheta)'\alpha T : \vartheta \in \Upsilon_0, \alpha \in \mathbb{R}^{\dim(W)}, \|\alpha\|_2 \leq 1\}
$$

also obeys a similar bracketing entropy bound

$$
\log N_{[]}(\epsilon, \mathcal{I}_2, \|\cdot\|_{P,2}) \lesssim 1/(\epsilon^2 \log^4 \epsilon) + \log(1/\epsilon)
$$

with a generic bracket taking the form $[i_2^0, i_2^1] = [\{W(\vartheta_k)'\beta_l - \kappa\delta\}T, \{W(\vartheta_k)'\beta_l + \kappa\delta\}T]$. Hence, this class is also P-Donsker with a constant envelope.

Step 3. In this step we verify the claim (a). Note that $\mathcal{F} = e \cdot \mathcal{I}_1 \cdot \mathcal{I}_2$. This class has a square-integrable envelope under P. The class $\mathcal{F}$ is P-Donsker by the following argument. Note that the product $\mathcal{I}_1 \cdot \mathcal{I}_2$ of uniformly bounded classes is P-Donsker, e.g., by Theorem 2.10.6 of van der Vaart and Wellner (1996). Under the stated assumption the final product of the random variable e with the P-Donsker class remains to be P-Donsker by the Multiplier Donsker Theorem, namely Theorem 2.9.2 in van der Vaart and Wellner (1996).

Proof of Claim (b). The claim follows by the Dominated Convergence Theorem, since any $f \in \mathcal{F}$ is dominated by a square-integrable envelope under P, and, uniformly in $y \in \mathcal{Y}$, $\Lambda[W(\vartheta)'\beta(y)]T \to \Lambda[W'\beta_0(y)]T$ and $|W(\vartheta)'\beta(y)T - W'\beta_0(y)T| \to 0$ in view of the relation such as (B.2).

Proof of Claim (c). This claim follows from the asymptotic equicontinuity of the empirical process $(\mathbb{G}_n[f_y], f_y \in \mathcal{F})$ under the $L_2(P)$ metric, and hence also with respect to the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$ in view of Claim (b).

Proof of Claim (d). It is convenient to set $\widehat{f}_y := f_y(A, \widehat{\vartheta}, \widetilde{\beta}(y))$ and $\widetilde{f}_y := f_y(A, \widetilde{\vartheta}, \widetilde{\beta}(y))$. Note that

$$\begin{aligned} \max_{1\le j\le \dim W} |\mathbb{G}_n[\widehat{f}_y - \widetilde{f}_y]|_j &\le \max_{1\le j\le \dim W} |\sqrt{n}\mathbb{E}_n[\widehat{f}_y - \widetilde{f}_y]|_j + \max_{1\le j\le \dim W} |\sqrt{n}\mathrm{E}_P(\widehat{f}_y - \widetilde{f}_y)|_j \\ &\lesssim \sqrt{n}\mathbb{E}_n[\widehat{\zeta}\,] + \sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,] \lesssim \mathbb{G}_n[\widehat{\zeta}\,] + 2\sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,], \end{aligned}$$

where $|f_y|_j$ denotes the jth element of an application of absolute value to each element of the vector f_y, and $\widehat{\zeta}$ is defined by the following relationship, which holds with probability approaching one uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned} \max_{1\le j\le \dim W} |\widehat{f}_y - \widetilde{f}_y|_j &\lesssim |e| \cdot \{\|W(\widehat{\vartheta}) - W(\widetilde{\vartheta})\|_2 + |\Lambda[W(\widehat{\vartheta})'\widetilde{\beta}(y)] - \Lambda[W(\widetilde{\vartheta})'\widetilde{\beta}(y)]|\} \cdot T \\ &\lesssim \widehat{\zeta} := e \cdot \kappa\Delta_n, \end{aligned}$$

where $\kappa = L_W \max_{\beta\in\mathcal{B}} \|\beta\|_2 + L_W$, $L_W = \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$, and $\Delta_n = o(1/\sqrt{n})$ is a deterministic sequence such that

$$\Delta_n \ge \|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty}.$$

By part (c) the result follows from

$$\mathbb{G}_n[\widehat{\zeta}\,] = \bar{o}_{\mathbb{P}}(1), \qquad \sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,] = \bar{o}_{\mathbb{P}}(1).$$

Indeed,

$$\|e \cdot \kappa\Delta_n\|_{P,2} = \bar{o}(1) \Rightarrow \mathbb{G}_n[\widehat{\zeta}\,] = \bar{o}_{\mathbb{P}}(1),$$

and

$$\|e \cdot \kappa\Delta_n\|_{P,1} \le \mathrm{E}_P|e| \cdot \kappa\Delta_n = \bar{o}(1/\sqrt{n}) \Rightarrow \mathrm{E}_P|\widehat{\zeta}| = \bar{o}_{\mathbb{P}}(1/\sqrt{n}),$$

since $\Delta_n = o(1/\sqrt{n})$.

Lemma 5. *[Local expansion] Under Assumptions 3–5, for*

$$\begin{aligned} &\widehat{\delta}(y) = \sqrt{n}(\widetilde{\beta}(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1); \\ &\widehat{\Delta}(x, r) = \sqrt{n}(\widehat{\vartheta}(x, r) - \vartheta_0(x, r)) = \sqrt{n}\ \mathbb{E}_n[\ell(A, x, r)] + o_{\mathbb{P}}(1) \text{ in } \ell^\infty(\overline{\mathcal{XR}}), \\ &\|\sqrt{n}\ \mathbb{E}_n[\ell(A, \cdot)]\|_{T,\infty} = O_{\mathbb{P}}(1), \end{aligned}$$

we have that

$$\sqrt{n}\ \mathrm{E}_P[\{\Lambda[W(\widehat{\vartheta})'\widetilde{\beta}(y)] - 1(Y \leq y)\}W(\widehat{\vartheta})T] = J(y)\widehat{\delta}(y) + \sqrt{n}\ \mathbb{E}_n\left[g_y(A)\right] + \bar{o}_{\mathbb{P}}(1),$$

where

$$g_y(a) = \mathrm{E}_P\{[\Lambda(W'\beta_0(y)) - 1(Y \leq y)]\dot{W} + \lambda(W'\beta_0(y))W\dot{W}'\beta_0(y)\}T\ell(a, X, R).$$

Proof of Lemma 5.

Uniformly in $\xi := (X, Z) \in \overline{\mathcal{X}\mathcal{Z}}$ and $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sqrt{n}\mathrm{E}_P\{\Lambda[W(\widehat{\vartheta})'\widetilde{\beta}(y)] - 1(Y \leq y) \mid X, Z\}T \\
&= \sqrt{n}\mathrm{E}_P\{\Lambda[W'\beta_0(y)] - 1(Y \leq y) \mid X, Z\}T \\
&+\lambda[W(\bar{\vartheta}_\xi)'\bar{\beta}_\xi(y)]\{W(\bar{\vartheta}_\xi)'\widehat{\delta}(y) + \dot{W}(\bar{\vartheta}_\xi)'\bar{\beta}_\xi\widehat{\Delta}(X, R)\}T \\
&= \sqrt{n}\mathrm{E}_P\{\Lambda[W'\beta_0(y)] - 1(Y \leq y) \mid X, Z\}T \\
&+\lambda[W'\beta_0(y)]\{W'\widehat{\delta}(y) + \dot{W}'\beta_0(y)\widehat{\Delta}(X, R)\}T + R_\xi(y),
\end{aligned}$$

and

$$\bar{R}(y) = \sup_{\{\xi \in \overline{\mathcal{X}\mathcal{Z}}\}} |R_\xi(y)| = \bar{o}_{\mathbb{P}}(1)$$

where $\bar{\vartheta}_\xi$ is on the line connecting ϑ_0 and $\widehat{\vartheta}$ and $\bar{\beta}_\xi(y)$ is on the line connecting $\beta_0(y)$ and $\widetilde{\beta}(y)$. The first equality follows by the mean value expansion. The second equality follows by uniform continuity of $\lambda(\cdot)$, uniform continuity of $W(\cdot)$ and $\dot{W}(\cdot)$, and by $\|\widehat{\vartheta} - \vartheta_0\|_{T,\infty} \to_{\mathbb{P}} 0$ and $\sup_{y \in \mathcal{Y}} \|\widetilde{\beta}(y) - \beta_0(y)\|_2 \to_{\mathbb{P}} 0$.

Since $\lambda(\cdot)$ and the entries of W and $\dot{W}$ are bounded, $\widehat{\delta}(y) = \bar{O}_{\mathbb{P}}(1)$, and $\|\widehat{\Delta}\|_{T,\infty} = O_{\mathbb{P}}(1)$, with probability approaching one uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sqrt{n}\mathrm{E}_P\{\Lambda[W(\widehat{\vartheta})'\widetilde{\beta}(y)] - 1(Y \leq y)\}W(\widehat{\vartheta})T = \mathrm{E}_P\{\Lambda(W'\beta_0(y)) - 1(Y \leq y)\}\dot{W}T\widehat{\Delta}(X, R) \\
&+ \mathrm{E}_P\{\lambda[W'\beta_0(y)]WW'T\}\widehat{\delta}(y) + \mathrm{E}_P\{\lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)T\widehat{\Delta}(X, R)\} + O_{\mathbb{P}}(\bar{R}(y)) \\
&= J(y)\widehat{\delta}(y) + \mathrm{E}_P[\{\Lambda(W'\beta_0(y)) - 1(Y \leq y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T\widehat{\Delta}(X, R) + o_{\mathbb{P}}(1).
\end{aligned}$$

Substituting in $\widehat{\Delta}(x, r) = \sqrt{n}\ \mathbb{E}_n[\ell(A, x, r)] + o_{\mathbb{P}}(1)$ and interchanging E_P and $\mathbb{E}_n$, we obtain

$$\mathrm{E}_P[\{\Lambda(W'\beta_0(y)) - 1(Y \leq y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T\widehat{\Delta}(X, R) = \sqrt{n}\ \mathbb{E}_n[g_y(A)] + \bar{o}_{\mathbb{P}}(1),$$

since $[\{\Lambda(W'\beta_0(y)) - 1(Y \leq y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T$ is bounded uniformly in $y \in \mathcal{Y}$. The claim of the lemma follows. □

B.2.2. *Proof of Lemma 3.* The proof is divided in two parts corresponding to the FCLT and bootstrap FCLT.

Part 1: FCLT

In this part we show $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) \rightsquigarrow J(y)^{-1}G(y)$ in $\ell^\infty(\mathcal{Y})^{d_w}$.

Step 1. This step shows that $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1)$.

Recall that

$$\widehat{\beta}(y) = \arg\min_{\beta \in \mathbb{R}^{\dim(W)}} \mathbb{E}_n[\rho_y(Y, W(\widehat{\vartheta})'\beta)T].$$

Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that uniformly in $y \in \mathcal{Y}$,

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\|\eta\|_2=1} \sqrt{n}\eta'\mathbb{E}_n\left[\widehat{f}_{\eta,B_\epsilon,y}\right] > 0 \right) \geq 1 - \epsilon, \tag{B.3}$$

where

$$\widehat{f}_{\eta,B_\epsilon,y}(A) := \left\{ \Lambda[W(\widehat{\vartheta})'(\beta_0(y) + B_\epsilon\eta/\sqrt{n})] - 1(Y \leq y) \right\} W(\widehat{\vartheta})T.$$

Let

$$f_y(A) := \{\Lambda[W'\beta_0(y)] - 1(Y \leq y)\} WT.$$

Then uniformly in $\|\eta\|_2 = 1$,

$$\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n[\widehat{f}_{\eta,B_\epsilon,y}] &= \eta'\mathbb{G}_n[\widehat{f}_{\eta,B_\epsilon,y}] + \sqrt{n}\eta'\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon,y}] \\
&=_{(1)} \eta'\mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon,y}] \\
&=_{(2)} \eta'\mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \eta'\mathbb{G}_n[g_y] + \bar{o}_{\mathbb{P}}(1) \\
&=_{(3)} \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 4 and Lemma 5 with $\widetilde{\beta}(y) = \beta_0(y) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, $\|\widetilde{\vartheta} - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0(y) + B_\epsilon\eta/\sqrt{n} - \beta_0(y)\|_2 = \bar{O}(1/\sqrt{n})$; relation (3) holds because f_y and g_y are P-Donsker by step-2 below. Since uniformly in $y \in \mathcal{Y}$, $J(y)$ is positive definite, with minimal eigenvalue bounded away from zero, the inequality (B.3) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step we show the main result. Let

$$\widehat{f}_y(A) := \left\{ \Lambda[W(\widehat{\vartheta})'\widehat{\beta}(y)] - 1(Y \leq y) \right\} W(\widehat{\vartheta})T.$$

From the first order conditions of the distribution regression problem,

$$
\begin{aligned}
0 = \sqrt{n}\mathbb{E}_n\left[\widehat{f}_y\right] &= \mathbb{G}_n\left[\widehat{f}_y\right] + \sqrt{n}\mathrm{E}_P\left[\widehat{f}_y\right] \\
&=_{(1)} \mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\left[\widehat{f}_y\right] \\
&=_{(2)} \mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + J(y)\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) + \mathbb{G}_n[g_y] + \bar{o}_{\mathbb{P}}(1),
\end{aligned}
$$

where relations (1) and (2) follow by Lemma 4 and Lemma 5 with $\widetilde{\beta}(y) = \widehat{\beta}(y)$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, and $\|\widetilde{\vartheta} - \vartheta\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\widehat{\beta}(y) - \beta_0(y)\|_2 = \bar{O}_{\mathbb{P}}(1/\sqrt{n})$.

Therefore by uniform invertibility of $J(y)$ in $y \in \mathcal{Y}$,

$$
\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1).
$$

The function f_y is P-Donsker by standard argument for distribution regression (e.g., step 3 in the proof of Theorem 5.2 of Chernozhukov, Fernandez-Val and Melly, 2013). Similarly, g_y is P-Donsker by Example 19.7 in van der Vaart (1998) because $g_y \in \{h_y(A) : |h_y(A) - h_v(A)| \leq M(A)|y - v|; \mathrm{E}_P M(A)^2 < \infty; y, v \in \mathcal{Y}\}$, since

$$
|g_y - g_v| \leq L\mathrm{E}_P[T|\ell(a, X, R)|]\big|_{a=A}|y - v|,
$$

with $L = 2L_W + L_W^2 \max_{\beta \in \mathcal{B}} \|\beta\|_2/4$, $L_W := \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$, and $\mathrm{E}_P[T\ell(A, X, R)^2] < \infty$ by Lemma 2. Hence, by the Functional Central Limit Theorem

$$
\mathbb{G}_n(f_y + g_y) \rightsquigarrow G(y) \text{ in } \ell^\infty(\mathcal{Y})^{d_w},
$$

where $y \mapsto G(y)$ is a zero mean Gaussian process with uniformly continuous sample paths and the covariance function $C(y, v)$ specified in the lemma. Conclude that

$$
\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) \rightsquigarrow J(y)^{-1}G(y) \text{ in } \ell^\infty(\mathcal{Y})^{d_w}.
$$

□

Part 2: Bootstrap FCLT

In this part we show $\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y)) \rightsquigarrow_{\mathbb{P}} J(y)^{-1}G(y)$ in $\ell^\infty(\mathcal{Y})^{d_w}$.

Step 1. This step shows that $\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

Recall that

$$
\widehat{\beta}^e(y) = \arg \min_{\beta \in \mathbb{R}^{\dim(W)}} \mathbb{E}_n[e\rho_y(Y, W(\widehat{\vartheta}^e)'\beta)T],
$$

where e is the random variable used in the weighted bootstrap. Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that uniformly in $y \in \mathcal{Y}$,

$$\text{(B.4)} \qquad \liminf_{n\to\infty} \mathbb{P}\left(\inf_{\|\eta\|_2=1} \sqrt{n}\eta'\mathbb{E}_n\left[\widehat{f}^e_{\eta,B_\epsilon,y}\right] > 0 \right) \geq 1-\epsilon,$$

where

$$\widehat{f}^e_{\eta,B_\epsilon,y}(A) := e \cdot \left\{ \Lambda[W(\widehat{\vartheta}^e)'(\beta_0(y) + B_\epsilon\eta/\sqrt{n})] - 1(Y \leq y) \right\} W(\widehat{\vartheta}^e)T.$$

Let

$$f^e_y(A) := e \cdot \{\Lambda[W'\beta_0(y)] - 1(Y \leq y)\} WT.$$

Then uniformly in $\|\eta\|_2 = 1$,

$$\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n[\widehat{f}^e_{\eta,B_\epsilon,y}] &= \eta'\mathbb{G}_n[\widehat{f}^e_{\eta,B_\epsilon,y}] + \sqrt{n}\eta'\mathrm{E}_P[\widehat{f}^e_{\eta,B_\epsilon,y}] \\
&=_{(1)} \eta'\mathbb{G}_n[f^e_y] + \bar{o}_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P[\widehat{f}^e_{\eta,B_\epsilon,y}] \\
&=_{(2)} \eta'\mathbb{G}_n[f^e_y] + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \eta'\mathbb{G}_n[g^e_y] + \bar{o}_{\mathbb{P}}(1) \\
&=_{(3)} \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 4 and Lemma 5 with $\widetilde{\beta}(y) = \beta_0(y) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$ and $\|\widetilde{\vartheta}^e - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\beta_0(y) + B_\epsilon\eta/\sqrt{n} - \beta_0(y)\|_2 = \bar{O}(1/\sqrt{n})$; relation (3) holds because $f^e_y = e \cdot f_y$ and $g^e_y = e \cdot g_y$, where f_y and g_y are P-Donsker by step-2 of the proof of Theorem 3 and $\mathrm{E}_P e^2 < \infty$. Since uniformly in $y \in \mathcal{Y}$, $J(y)$ is positive definite, with minimal eigenvalue bounded away from zero, the inequality (B.4) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step we show that $\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f^e_y + g^e_y) + \bar{o}_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

Let

$$\widehat{f}^e_y(A) := e \cdot \{\Lambda[W(\widehat{\vartheta}^e)'\widehat{\beta}^e(y)] - 1(Y \leq y)\} W(\widehat{\vartheta}^e)T.$$

From the first order conditions of the distribution regression problem in the weighted sample, uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned}
0 = \sqrt{n}\mathbb{E}_n\left[\widehat{f}^e_y\right] &= \mathbb{G}_n\left[\widehat{f}^e_y\right] + \sqrt{n}\mathrm{E}_P\left[\widehat{f}^e_y\right] \\
&=_{(1)} \mathbb{G}_n[f^e_y] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\left[\widehat{f}^e_y\right] \\
&=_{(2)} \mathbb{G}_n[f^e_y] + \bar{o}_{\mathbb{P}}(1) + J(y)\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) + \mathbb{G}_n[g^e_y] + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 4 and Lemma 5 with $\widetilde{\beta}(y) = \widehat{\beta}^e(y)$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$ and $\|\widetilde{\vartheta}^e - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\widehat{\beta}^e(y) - \beta_0(y)\|_2 = \bar{O}_{\mathbb{P}}(1/\sqrt{n})$.

Therefore by uniform invertibility of $J(y)$ in $y \in \mathcal{Y}$,

$$\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y^e + g_y^e) + \bar{o}_{\mathbb{P}}(1).$$

Step 3. In this final step we establish the behavior of $\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y))$ under $\mathbb{P}^e$. Note that $\mathbb{P}^e$ denotes the conditional probability measure, namely the probability measure induced by draws of $e_1, ..., e_n$ conditional on the data $A_1, ..., A_n$. By Step 2 of the proof of Theorem 1 and Step 2 of this proof, we have that under $\mathbb{P}$:

$$\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y^e + g_y^e) + \bar{o}_{\mathbb{P}}(1),$$
$$\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1).$$

Hence, under $\mathbb{P}$

$$\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y)) = -J(y)^{-1}\mathbb{G}_n(f_y^e - f_y + g_y^e - g_y) + r_n(y)$$
$$= -J(y)^{-1}\mathbb{G}_n((e-1)(f_y + g_y)) + r_n(y),$$

where $r_n(y) = \bar{o}_{\mathbb{P}}(1)$. Note that it is also true that

$$r_n(y) = \bar{o}_{\mathbb{P}^e}(1) \text{ in } \mathbb{P}\text{-probability},$$

where the latter statement means that for every $\epsilon > 0$, $\mathbb{P}^e(\|r_n(y)\|_2 > \epsilon) = \bar{o}_{\mathbb{P}}(1)$. Indeed, this follows from Markov inequality and by

$$\mathrm{E}_{\mathbb{P}}[\mathbb{P}^e(\|r_n(y)\|_2 > \epsilon)] = \mathbb{P}(\|r_n(y)\|_2 > \epsilon) = \bar{o}(1),$$

where the latter holds by the Law of Iterated Expectations and $r_n(y) = \bar{o}_{\mathbb{P}}(1)$.

Note that $f_y^e = e \cdot f_y$ and $g_y^e = e \cdot g_y$, where f_y and g_y are P-Donsker by step-2 of the proof of the first part and $\mathrm{E}_P e^2 < \infty$. Then, by the Conditional Multiplier Functional Central Limit Theorem, e.g., Theorem 2.9.6 in van der Vaart and Wellner (1996),

$$G_n^e(y) := \mathbb{G}_n((e-1)(f_y + g_y)) \rightsquigarrow_{\mathbb{P}} G(y) \text{ in } \ell^\infty(\mathcal{Y})^{d_w}.$$

Conclude that

$$\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y)) \rightsquigarrow_{\mathbb{P}} J(y)^{-1}G(y) \text{ in } \ell^\infty(\mathcal{Y})^{d_w}.$$

□

B.3. **Proof of Theorems 2–4.** In this section we use the notation $W_x(\vartheta) = w(x, Z_1, V(\vartheta))$ such that $W_x = w(x, Z_1, V(\vartheta_0))$. Again we focus on the case where $\mathcal{Y}$ is a compact interval of $\mathbb{R}$.

B.3.1. *Proof of Theorem 2.* The result follows by a similar argument to the proof of Lemma 3 using Lemmas 6 and 7 in place of Lemmas 4 and 5, and the delta method. For the sake of brevity, here we just outline the proof of the FCLT.

Let $\psi_x(A, \vartheta, \beta) := \Lambda(W_x(\vartheta)'\beta)T$ such that $G_T(y, x) = \mathrm{E}_P\psi_x(A, \vartheta_0, \beta_0(y))/\mathrm{E}_P T$ and $\widehat{G}(y, x) = \mathbb{E}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))/\mathbb{E}_n T$. Then, for $\widehat{\psi}_{y,x} := \psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))$ and $\psi_{y,x} := \psi_x(A, \vartheta_0, \beta_0(y))$,

$$\begin{aligned}\sqrt{n}\left[\mathbb{E}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y)) - \mathrm{E}_P\psi_x(A, \vartheta_0, \beta_0(y))\right] &= \mathbb{G}_n\left[\widehat{\psi}_{y,x}\right] + \sqrt{n}\mathrm{E}_P\left[\widehat{\psi}_{y,x} - \psi_{y,x}\right] \\ &=_{(1)} \mathbb{G}_n[\psi_{y,x}] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\left[\widehat{\psi}_{y,x} - \psi_{y,x}\right] \\ &=_{(2)} \mathbb{G}_n[\psi_{y,x}] + \bar{o}_{\mathbb{P}}(1) + \mathbb{G}_n[h_{y,x}] + \bar{o}_{\mathbb{P}}(1),\end{aligned}$$

where relations (1) and (2) follow by Lemma 6 and Lemma 7 with $\widetilde{\beta}(y) = \widehat{\beta}(y)$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, and $\|\widetilde{\vartheta} - \vartheta\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1)$ from step 2 of the proof of Lemma 3.

The functions $(y, x) \mapsto \psi_{y,x}$ and $(y, x) \mapsto h_{y,x}$ are P-Donsker by Example 19.7 in van der Vaart (1998) because they are Lipschitz continuous on $\mathcal{Y}\overline{\mathcal{X}}$. Hence, by the Functional Central Limit Theorem

$$\mathbb{G}_n(\psi_{y,x} + h_{y,x}) \rightsquigarrow Z(y, x) \text{ in } \ell^\infty(\mathcal{Y}\overline{\mathcal{X}}),$$

where $(y, x) \mapsto Z(y, x)$ is a zero mean Gaussian process with uniformly continuous sample paths and covariance function

$$\mathrm{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u}], \quad (y, x), (v, u) \in \mathcal{Y}\overline{\mathcal{X}}.$$

The result follows by the functional delta method applied to the ratio of $\mathbb{E}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))$ and $\mathbb{E}_n T$ using that

$$\begin{pmatrix} \mathbb{G}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y)) \\ \mathbb{G}_n T \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z(y, x) \\ Z_T \end{pmatrix},$$

where $Z_T \sim N(0, p_T(1 - p_T))$,

$$\mathrm{Cov}_P(Z(y, x), Z_T) = G_T(y, x)p_T(1 - p_T),$$

and

$$\operatorname{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u} \mid T = 1]$$
$$= \frac{\operatorname{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u}] - G_T(y,x)G_T(v,u)p_T(1-p_T)}{p_T}.$$

□

Lemma 6. *[Stochastic equicontinuity] Let $e \geq 0$ be a positive random variable with $\mathrm{E}_P[e] = 1$, $\mathrm{Var}_P[e] = 1$, and $\mathrm{E}_P|e|^{2+\delta} < \infty$ for some $\delta > 0$, that is independent of (Y, X, Z, W, V), including as a special case $e = 1$, and set, for $A = (e, Y, X, Z, W, V)$,*

$$\psi_x(A, \vartheta, \beta) := e \cdot \Lambda(W_x(\vartheta)'\beta) \cdot T.$$

Under Assumptions 3–5, the following relations are true.

(a) *Consider the set of functions*

$$\mathcal{F} := \{\psi_x(A, \vartheta, \beta) : (\vartheta, \beta, x) \in \Upsilon_0 \times \mathcal{B} \times \overline{\mathcal{X}}\},$$

where $\overline{\mathcal{X}}$ is a compact subset of $\mathbb{R}$, $\mathcal{B}$ is a compact set under the $\|\cdot\|_2$ metric containing $\beta_0(y)$ for all $y \in \mathcal{Y}$, Υ_0 is the intersection of Υ, defined in Lemma 2, with a neighborhood of ϑ_0 under the $\|\cdot\|_{T,\infty}$ metric. This class is P-Donsker with a square integrable envelope of the form e times a constant.

(b) *Moreover, if $(\vartheta, \beta(y)) \to (\vartheta_0, \beta_0(y))$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$, then*

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \|\psi_x(A, \vartheta, \beta(y)) - \psi_x(A, \vartheta_0, \beta_0(y))\|_{P,2} \to 0.$$

(c) *Hence for any $(\widetilde{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$ such that $\widetilde{\vartheta} \in \Upsilon_0$,*

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \|\mathbb{G}_n\psi_x(A, \widetilde{\vartheta}, \widetilde{\beta}(y)) - \mathbb{G}_n\psi_x(A, \vartheta_0, \beta_0(y))\|_2 \to_{\mathbb{P}} 0.$$

(d) *For any $(\widehat{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$, so that*

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n}), \text{ where } \widetilde{\vartheta} \in \Upsilon_0,$$

we have that

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \|\mathbb{G}_n\psi_x(A, \widehat{\vartheta}, \widetilde{\beta}(y)) - \mathbb{G}_n\psi_x(A, \vartheta_0, \beta_0(y))\|_2 \to_{\mathbb{P}} 0.$$

Proof of Lemma 6. The proof is omitted because is similar to the proof of Lemma 4. □

Lemma 7. *[Local expansion] Under Assumptions 3–5, for*

$$\widehat{\delta}(y) = \sqrt{n}(\widetilde{\beta}(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1);$$
$$\widehat{\Delta}(x,r) = \sqrt{n}(\widehat{\vartheta}(x,r) - \vartheta_0(x,r)) = \sqrt{n}\ \mathbb{E}_n[\ell(A,x,r)] + o_{\mathbb{P}}(1) \text{ in } \ell^{\infty}(\overline{\mathcal{X}\mathcal{R}}),$$
$$\|\sqrt{n}\ \mathbb{E}_n[\ell(A,\cdot)]\|_{T,\infty} = O_{\mathbb{P}}(1),$$

we have that

$$\sqrt{n}\ \left\{ \mathrm{E}_P \Lambda[W_x(\widehat{\vartheta})'\widetilde{\beta}(y)]T - \mathrm{E}_P \Lambda[W_x'\beta_0(y)]T \right\} = \mathrm{E}_P\{\lambda[W_x'\beta_0(y)]W_x T\}'\widehat{\delta}(y) + \mathrm{E}_P\{\lambda[W_x'\beta_0(y)]\dot{W}_x'\beta_0(y)T\ell(a,X,R)\}\big|_{a=A} + \bar{o}_{\mathbb{P}}(1),$$

where $\bar{o}_{\mathbb{P}}(1)$ denotes order in probability uniform in $(y,x) \in \mathcal{Y}\overline{\mathcal{X}}$.

Proof of Lemma 7. The proof is omitted because is similar to the proof of Lemma 5. □

B.3.2. *Proof of Theorem 3.* The result follows from Theorem 2 and the functional delta method, because the map $\phi : H \mapsto \int_{\mathcal{Y}^+} 1(H(y,x) \leq \tau)dy - \int_{\mathcal{Y}^-} 1(H(y,x) \geq \tau)dy$ is Hadamard differentiable at $H = G_T$ under the conditions of the theorem by Proposition 2 of Chernozhukov, Fernandez-Val and Galichon (2010) with derivative

$$\phi'_{G_T}(h) = -\frac{h(\phi(\cdot,x),x)}{g_T(\phi(\cdot,x),x)}.$$

B.3.3. *Proof of Theorem 4.* The result follows from Theorem 2 and the functional delta method, because the map $\varphi : H \mapsto \int_{\mathcal{Y}}[1(y \geq 0) - H(y,x)]dy$ is Hadamard differentiable at $H = G_T$ by Lemma 8 with derivative

$$\varphi'_{G_T}(h) = -\int_{\mathcal{Y}} h(y,x)\nu(dy).$$

Lemma 8. *[Hadamard Differentiability of ASF Map] Let $\mathcal{Y}$ be bounded. The ASF map $\varphi : \ell^{\infty}(\mathcal{Y}\overline{\mathcal{X}}) \to \ell^{\infty}(\overline{\mathcal{X}})$ defined by*

$$H \mapsto \varphi(H) := \int_{\mathcal{Y}} [1(y \geq 0) - H(y,x)]\nu(dy),$$

is Hadamard-differentiable at $H = G$, tangentially to the set of uniformly continuous functions on $\mathcal{Y}\overline{\mathcal{X}}$, with derivative map $h \mapsto \varphi'_G(h)$ defined by

$$\varphi'_G(h) := -\int_{\mathcal{Y}} h(y,x)\nu(dy),$$

where the derivative is defined and is continuous on $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$.

Proof of Lemma 8. Consider any sequence $H^t \in \ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ such that for $h^t := (H^t - G)/t$, $h^t \to h$ in $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ as $t \searrow 0$, where h is a uniformly continuous function on $\mathcal{Y}\overline{\mathcal{X}}$. We want to show that as $t \searrow 0$,

$$\frac{\varphi(H^t) - \varphi(G)}{t} - \varphi'_G(h) \to 0 \text{ in } \ell^\infty(\mathcal{Y}\overline{\mathcal{X}}).$$

The result follows because by linearity of the map φ

$$\frac{\varphi(H^t) - \varphi(G)}{t} = -\int_{\mathcal{Y}} h^t(y,x)\nu(dy) \to -\int_{\mathcal{Y}} h(y,x)\nu(dy) = \varphi'_G(h).$$

The derivative is well-defined over $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ and continuous with respect to the sup-norm on $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$.

References

[1] Blundell, R., X. Chen, and D. Kristensen (2007), "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves." *Econometrica*, 75, 1613–1669.

[2] Blundell, R., W. K. Newey and F. Vella. "Control Functions." *Unpublished manuscript.*

[3] Blundell, R., and J. L. Powell (2003), "Endogeneity in nonparametric and semiparametric regression models." *Econometric society monographs*, 36, 312–357.

[4] Blundell, R., and J. L. Powell (2004) "Endogeneity in semiparametric binary response models." *The Review of Economic Studies*, 71, 655–679.

[5] Chernozhukov, V. I. Fernandez-Val, and A. Galichon (2010) "Quantile and Probability Curves without Crossing." *Econometrica*, 78, 1093–1125.

[6] Chernozhukov, V., I. Fernandez-Val, and A. Kowalski (2015), "Quantile Regression with Censoring and Endogeneity." *Journal of Econometrics*, 186, 201–221.

[7] Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013), "Inference on Counterfactual Distributions." *Econometrica*, 81, 2205–2268.

[8] Chernozhukov, V., I. Fernandez-Val, B. Melly, and K. Wuthrich (2016) "Generic Inference on Quantile and Quantile Effect Functions for Discrete Outcomes." *eprint arXiv:1608.05142*.

[9] Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73, 245–261.

[10] Chernozhukov, V. and C. Hansen (2006), "Instrumental quantile regression inference for structural and treatment effect models." *Journal of Econometrics*, 132, 491–525.

[11] Chernozhukov, V. and C. Hansen (2013), "Quantile models with endogeneity." *Annu. Rev. Econ.*, 5, 57–81.

[12] Chesher, A. (2003), "Identification in nonseparable models." *Econometrica*, 71, 1405–1441.

[13] D'Haultfoeuille, X., and P. Février (2015), "Identification of nonseparable triangular models with discrete instruments". *Econometrica*, 83, 1199–1210.

[14] Fernandez-Val, I., van Vuuren, A., and F. Vella (2018), "Nonseparable Sample Selection Models with Censored Selection Rules." *eprint arXiv:1801.08961.*

[15] Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008), "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects." *Econometrica*, 76, 1191–1206.

[16] Garen, J. (1984) "The Returns to Schooling: A Selectivity Approach with a Continuous Choice Variable." *Econometrica*, 52, 1199–1218.

[17] Graybill, W., M. Chen, V. Chernozhukov, I., Fernandez-Val, and A. Galichon (2016), Rearrangement. R package version 2.1.

[18] Imbens, G. W., and W. K. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77, 1481–1512.

[19] Imbens, G. W., and J. Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47, 5–86.

[20] Jun, S. J. (2009), "Local structural quantile effects in a model with a nonseparable control variable." *Journal of Econometrics*, 151, 82–97.

[21] Koenker, R. (2018), quantreg: Quantile Regression. R package version 5.38.

[22] Koenker, R., and G. Basset Jr (1978) "Regression Quantiles." *Econometrica*, 46, 33–50.

[23] Koenker, R., and Z. Xiao (2002), "Inference on the quantile regression process." *Econometrica*, 70, 4, 1583–1612.

[24] Lewbel, A. (2006), Engel curves: Entry for the New Palgrave Dictionary of Economics.

[25] Ma, L., and R. Koenker (2006), "Quantile regression methods for recursive structural equation models." *Journal of Econometrics*, 134, 471–506.

[26] Newey, W., and S. Stouli (2018), "Control variables, discrete instruments, and identification of structural functions.*eprint arXiv:1809.05706.*

[27] R Development Core Team (2019), *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

[28] Spady, R., and S. Stouli (2018), "Dual regression." *Biometrika*, 105, 1–18.

[29] Stouli, S. (2012), "Construction of structural functions in conditional independence models." Working paper, University College London.

[30] Torgovitsky, A. (2015), "Identification of nonseparable models using instruments with small support." *Econometrica*, 83, 1185–1197.

[31] van der Vaart, A.W. (2000), *Asymptotic Statistics.* Cambridge university press.

[32] van der Vaart, A.W. and Wellner, J. A. (1996), *Weak convergence and empirical processes.* Springer.

[33] Wooldridge, J. (2015), "Control function methods in applied econometrics." *Journal of Human Resources*, 50, 420–445.