I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/18

Semiparametric transformation model with endogeneity: a control function approach

VANHEMS, A. and I. VAN KEILEGOM

Semiparametric transformation model with endogeneity: a control function approach

Anne Vanhems * Ingrid Van Keilegom §

May 8, 2013

Abstract

We consider a semiparametric transformation model, in which the regression function has an additive nonparametric structure and the transformation of the response is assumed to belong to some parametric family. We suppose that endogeneity is present in the explanatory variables. Using a control function approach, we show that the proposed model is identified under suitable assumptions, and propose a profile likelihood estimation method for the transformation. The proposed estimator is shown to be asymptotically normal under certain regularity conditions. A simulation study shows that the estimator behaves well in practice. Finally, we give an empirical example using the U.K. Family Expenditure Survey.

Key Words: Additive models; Control function; Endogeneity; Instrumental variable; Nonseparability; Profile likelihood; Semiparametric regression; Transformation models.

---

*Université de Toulouse, Toulouse Business School and Toulouse School of Economics, a.vanhems@esc-toulouse.fr

§Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium. E-mail address: ingrid.vankeilegom@uclouvain.be.
This research was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy) and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

1 Introduction

Consider the following semiparametric transformation model:

$$\Lambda_\theta(Y) = \phi(X, Z) + \epsilon, \tag{1.1}$$

where $\{\Lambda_\theta : \theta \in \Theta\}$ is a parametric family of strictly increasing functions, and the function $\phi(\cdot, \cdot)$ is of unknown form. The response Y is a real valued continuously distributed random scalar and the vector of regressors (X, Z) consists of real valued continuously distributed variables, X takes values in $\mathbb{R}^{d_x}$, and Z in $\mathbb{R}^{d_z}$, with $d_x \geq 1$ and $d_z \geq 0$. We assume moreover that X is endogenous and correlated with the error term ϵ, while Z represents a vector of exogenous random variables. Our objective is to identify the structure $(\Lambda_\theta, \phi, F_\epsilon)$ using a control function approach (where $F_\epsilon(\cdot) = \Pr(\epsilon \leq \cdot)$), to estimate θ and ϕ given a sample of observations and to do inference on these estimators.

Our work is mainly related to the literature on semiparametric transformation models and semiparametric modeling with endogeneity. Transformation models lie at the heart of many problems in structural econometrics[1] and various papers have studied them under different sets of assumptions. In a fully exogenous setting, some papers have considered nonparametric forms for Λ and ϕ, like Horowitz (2001), Jacho-Chavez, Lewbel and Linton (2010) or Ekeland, Heckman and Nesheim (2004). In the same setting, other papers have analyzed a semiparametric model with a parametric form for ϕ like in Horowitz (1996) or a parametric form for Λ as in Linton, Sperlich and Van Keilegom (2008). As in this last paper, our work focusses on a parametric transformation but we extend this framework by considering a vector X of endogenous variables.

The issue of endogeneity is very crucial in econometrics (see Wooldridge 2008 for a general overview). It can arise as a result of e.g. omitted variables, autoregression with autocorrelated errors or sample selection errors and it has also been investigated in the specific setting of transformation models. Chiappori, Komunjer and Kristensen (2010) consider a fully nonparametric setting and, with a little stronger assumption of conditional independence between ϵ and one coordinate of X, are able to identify the model and recover a parametric rate of convergence for the estimated transformation operator. On the other hand, Florens and Sokullu (2012) and Fève and Florens (2010) consider a semiparametric form for the function ϕ and identify and estimate the model using an instrument W and by

[1] Important applications of transformation models are given by duration models, with the classical mixed proportional hazards model (see Heckman and Singer 1984, Nielsen et al. 1992), and various applications in labor economics (see for example Keifer 1988) or industrial organization (Sokullu 2011). Another class of applications is given by the hedonic model with the paper by Ekeland, Heckman and Nesheim (2004) or Heckman, Matzkin and Nesheim (2005).

imposing very few technical assumptions (like conditional mean independence) in the line of ill-posed inverse problems theory (see Carrasco, Florens and Renault 2007 for an overview of inverse problem theory in econometrics). In our case, the parametric assumption concerns the operator Λ and we identify the model using a control function approach.

The control function approach has been detailed in several papers, see e.g. Newey, Powell and Vella (1999), or Blundell and Powell (2003) for a separable setting, and Imbens and Newey (2009) for a nonseparable setting. As stressed in Matzkin (2003), transformation models can be written as a nonseparable model with $Y = \Lambda^{-1}(\phi(X, Z) + \epsilon)$ and there is indeed a connection between both frameworks. However, in addition to the fact that we consider a semiparametric model, one major difference is that we need to identify and estimate both functions Λ and ϕ whereas only one is involved in classical nonseparable models.

At last, one could also relate our work to the semiparametric analysis with generated covariates developed in Mammen, Rothe and Schienle (2012b) since the control function needs to be estimated in a first step. However, we also need to take into account the estimation of the density of the error term ϵ in the global estimation process, and our estimation procedure is therefore, from a structural point of view, quite different from theirs. We will detail this main difference in the next section.

The paper is organized as follows. In the next section we discuss the contribution of our model and present some challenges with the estimation and inference. Section 3 is devoted to the identification and estimation of the model. In Section 4 we state the asymptotic normality of the estimators of θ and ϕ, we propose a bootstrap procedure to estimate the distribution of $\widehat{\theta}$ in practice, and we also illustrate the asymptotic results for some particular specifications of the control variable. A finite sample study is presented in Section 5, including some simulations and an application to real data. Some general conclusions are given in Section 6, and finally the proofs of the asymptotic results are collected in the Appendix.

2 Contribution

2.1 Semiparametric transformation models with generated covariates

In this section, our objective is to present the contribution of our paper with respect to the existing literature, and in particular to link our work to the general setting of semiparametric transformation models with generated covariates. Consider an absolute continuous random vector (X, Z, W, Y) with density $f_{X,Z,W,Y}$, whose support is $R_{X,Z,W,Y} \subset \mathbb{R}^{d_x+d_z+d_w+1}$, and suppose that (X, Z, Y) satisfies model (1.1) and that W is an appropriate instrument. [2]

As explained earlier, the variable X is endogenous and we use a control function approach to treat the endogeneity. Consider the variable $V = r(X, Z, W)$ where r is an unknown infinite dimensional parameter defined on $R_{X,Z,W}$, the support of (X, Z, W). The variable V is a control variable if (X, Z) and ϵ are independent conditional on V (see Imbens and Newey 2009 for a general definition in the nonseparable case). Examples of control variables V will be given in Subsection 4.4.

A classical issue in nonparametric estimation is the evaluation at the boundary of the support $R_{X,Z,V}$ where the functions are imprecisely estimated. For simplicity, in what follows, we will restrict the evaluation of the functions of interest over some strict compact subset, which is useful to accommodate fixed trimming schemes in the estimation procedure, and we will simply denote this compact subset by $R_{X,Z,V}$.

Our model, defined by equation (1.1) and the control variable V, belongs to the general setting studied in Chen, Linton and Van Keilegom (2003) and we will use their framework to establish our paper's contribution. Basically, they denote Θ for a finite dimensional parameter set (of dimension p) and $\mathcal{H}$ for an infinite dimensional parameter space. They also assume that there exists a non-random measurable vector-valued function G defined on $\Theta \times \mathcal{H}$ such that $G(\theta, s_\theta) = 0$ at $\theta = \theta_0 \in \Theta$. Note that $\theta_0 \in \Theta$ and $s_0 \equiv s_{\theta_0}$ are the true unknown finite and infinite dimensional parameters.

In our specific setting, the function G is defined via a moment condition: $G(\theta, s_\theta) = \mathrm{E}\{M(\theta, s_\theta, X, Z, W, Y)\}$ for a certain function M which we will define in Section 4 and which depends on the observable random variables X, Z, W and Y, on the parameter θ, and on a vector s_θ of nuisance functions, which is defined as follows :

$$s_\theta = (\phi_\theta, \dot{\phi}_\theta, f_{\epsilon(\theta)}, f'_{\epsilon(\theta)}, \dot{f}_{\epsilon(\theta)}, r)^t.$$

Here, the function

$$\phi_\theta(.) = \int \mathrm{E}(\Lambda_\theta(Y)|X = ., Z = ., V = v)dF_V(v), \tag{2.1}$$

corresponds to the function defined in (1.1) for a fixed value of θ, $\dot{\phi}_\theta$ denotes the vector of partial derivatives of ϕ_θ with respect to the components of θ, $f_{\epsilon(\theta)}$ is the density function of the residual term $\epsilon(\theta) = \Lambda_\theta(Y) - \phi_\theta(X, Z)$ for a fixed value of θ, $f'_{\epsilon(\theta)}$ represents the derivative of $f_{\epsilon(\theta)}$ with respect to the principal argument, and $\dot{f}_{\epsilon(\theta)}$ is the vector of partial derivatives with respect to the components of θ. Note that each of these functions depends

[2]In what follows, the support of any random variable T will be denoted by R_T and its density function will be denoted by f_T.

on the true control function r, which will play a major role in the estimation procedure in Section 3 and in the asymptotic theory developed in Section 4.

2.2 Challenges with the estimation and inference on the model

In this subsection, we would like to highlight the main challenges with the estimation and the inference on our model and what are the differences between our paper and other results in the literature.

As it has been stressed in the previous subsection, our model lies in the general setting of semiparametric transformation models with a generated covariate defined by $V = r(X, Z, W)$ which is involved in the function ϕ_θ as well as in the density function $f_{\epsilon(\theta)}$, and belongs to the general framework studied in Chen, Linton and Van Keilegom (2003). However, from a structural point of view, it differs from recent papers also related to Chen, Linton and Van Keilegom (2003), in the estimation process as well as the asymptotic properties.

First of all, the paper by Mammen, Rothe and Schienle (2012b) considers also a general class of semiparametric optimization estimators with infinite-dimensional nuisance parameters that include a conditional expectation function estimated nonparametrically using generated covariates. In our model, as presented in the previous subsection, the generated covariate V affects the function ϕ_θ, defined as the marginal integration of a conditional expectation (see equation (2.1)), its derivative with respect to θ, the residual density function $f_{\epsilon(\theta)}$ as well as its derivatives with respect to the principal argument and to θ. This structural difference between both models of course impacts the estimation step as well as the inference.

Second, our model extends the Linton, Sperlich and Van Keilegom (2008) setting which does not take into account the endogenous variable X and then the generated covariate V. As it has just been stressed, the estimation of V appears in each step and thus affects all the nuisance functions. In addition, the assumption of endogeneity implies that (X, Z) and ϵ are not independent anymore, which complicates a lot the derivation of the asymptotic variance in Theorem 4.1. This second main difference is stressed in the second comment after Theorem 4.1 as well as in the proof where more lemmas are required to derive the asymptotic normality (Lemmas 7.1 and 7.2).

Third, our framework is also very different from Imbens and Newey (2009) although the identification proof is partly based on their arguments. From a structural point of view, we need to identify two functions namely Λ and ϕ whereas they only consider the identification of ϕ. Moreover, we consider a semiparametric model and our estimation procedure includes the estimation of the parameter θ (whereas they consider a fully nonparametric setting).

As we have stressed above, the estimation of θ in an endogenous setting complicates a lot the estimation step, since our model also requires the estimation of the function ϕ and the density f_ϵ of the error, as well as the derivatives of ϕ and f_ϵ.

At last, we would like to stress a very nice feature of our estimation procedure since the estimator $\widehat{\theta}$, which we will define in Section 3.2 below, behaves asymptotically as if the control variable V would be observed for each observation in the sample. This property is surprising, since the estimation of the control variable is needed for estimating each of the nuisance functions $\phi_\theta, \dot{\phi}_\theta, f_{\epsilon(\theta)}, f'_{\epsilon(\theta)}$ and $\dot{f}_{\epsilon(\theta)}$, and hence it is an intrinsic part of the estimation procedure. We have investigated this property via finite sample simulations in Subsection 5.1. The simulations confirm our theoretical findings. Note that the impact of of the estimation of a generated covariate on the asymptotic variance has been investigated in several papers, as in Mammen, Rothe and Schienle (2012b) or Hahn and Ridder (2013) who prove that indeed in some cases the influence of the generated covariate may disappear in the asymptotic variance.

3 Identification and estimation

3.1 Identification

The result we present below allows to identify the fully nonparametric structure $(\Lambda, \phi, F_\epsilon)$ and therefore, in this subsection, we omit the index θ for the operator Λ and the functions ϕ and F_ϵ.

Recall that we assume that V can be written as an unknown function r of (X, Z, W), i.e. $V = r(X, Z, W)$, and identification of model (1.1) will be studied under the general assumption that V acts as a control function, that is:

(A.1) (X, Z) and ϵ are independent conditional on V

(A.2) The support of V conditional on (X, Z) equals the support of V.

These assumptions are standard in the literature on nonseparable models (see Imbens and Newey 2009) and will allow to identify the functions ϕ and F_ϵ.

We also need to identify Λ and based on Chiappori, Komunjer and Kristensen (2010) and Linton, Sperlich and Van Keilegom (2008), we impose the following additional assumptions:

(A.3) Λ is a continuously differentiable and strictly increasing function defined on the support R_Y of Y.

(A.4) The support $R_{X,Z}$ of (X,Z) is compact, and for almost all $(x,z) \in R_{X,Z}$, the density $f_{\epsilon|X,Z}(\cdot|x,z)$ exists, is strictly positive and continuously differentiable.

(A.5) The derivative of ϕ with respect to x_1 (the first coordinate of x) exists and the set $\{(x,z) \in R_{X,Z} : \frac{\partial}{\partial x_1}\phi(x,z) \neq 0\}$ has a nonempty interior.

(A.6) $\mathrm{E}(\Lambda(Y)) = 1$, $\Lambda(0) = 0$, and $\mathrm{E}(\epsilon) = 0$.

Our result is based on the equality:

$$\begin{aligned} F_{Y|X,Z,V}(y|x,z,v) &= \Pr[\Lambda(Y) \leq \Lambda(y)|X=x,\ Z=z,\ V=v] \\ &= \Pr[\epsilon \leq \Lambda(y) - \phi(X,Z)|X=x,\ Z=z,\ V=v] \\ &= \Pr[\epsilon \leq \Lambda(y) - \phi(x,z)|V=v], \end{aligned}$$

where the first equality comes from the monotonicity Assumption $(A.3)$, and the third one follows from Assumption $(A.1)$. Then, following Imbens and Newey (2009) we have:

$$\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv) = F_\epsilon(\Lambda(y) - \phi(x,z)). \tag{3.1}$$

Proposition 3.1. *Under Assumptions $(A.1)-(A.6)$, the structure $(\Lambda, \phi, F_\epsilon)$ is identified.*

The proof is given in the Appendix.

Remark 3.1. *1. Note that Chiappori, Komunjer and Kristensen (2010) suggest a slightly different independence assumption, instead of $(A.1)$: ϵ is independent of X_1 conditional on (X_{-1}, Z, V) (where $X = (X_1, X_{-1})$). Although an equivalent identification result could be derived with their set of assumptions, the estimation of the parameter θ would become more tricky since the distribution of ϵ would remain conditional on (X_{-1}, Z).*

2. Note also that Proposition 3.1 only gives sufficient conditions to identify the structure $(\Lambda, \phi, F_\epsilon)$. In particular, Assumption $(A.2)$ could be weakened using a separability assumption as proposed in Newey, Powell and Vella (1999). Indeed, once Λ is identified using Assumptions $(A.1)$, $(A.3)-(A.6)$, we get:

$$\mathrm{E}\left(\Lambda(Y)|X=x, Z=z, V=v\right) = \phi(x,z) + \lambda(v),$$

where $\lambda(v) = \mathrm{E}\left[\epsilon|V=v\right]$. Then, using Theorem 2.2 in Newey, Powell and Vella (1999) and the normalization assumption $(A.6)$, we conclude that if there is no functional relationship between (X,Z) and V, then ϕ is identified.

3.2 Estimation

Although a fully nonparametric approach is possible, we return now (and for the rest of the paper) to the case where the transformation of the response variable is parametric, i.e. $\Lambda(\cdot) \equiv \Lambda_\theta(\cdot)$, for some parametric family $\{\Lambda_\theta(\cdot) : \theta \in \Theta\}$, where we suppose that Θ is compact. Indeed, considering a parametric transformation can lead to easier interpretation, like for the family of power transformations proposed by Box and Cox (1964), and the Bickel and Doksum (1981) class of transformations.

From equation (3.1) we obtain:

$$\int f_{Y|X,Z,V}(y|x,z,v)dF_V(v) = f_{\epsilon(\theta_0)}\left(\Lambda_0(y) - \phi_0(x,z)\right) . \Lambda_0'(y), \tag{3.2}$$

where $f_{\epsilon(\theta_0)}$ and $f_{Y|X,Z,V}$ are the probability density functions of ϵ and of Y given (X, Z, V), respectively, and where $\phi_0 \equiv \phi_{\theta_0}$ is defined in (2.1) and $\Lambda_0 \equiv \Lambda_{\theta_0}$.

Consider now a randomly drawn i.i.d. sample $(X_i, Z_i, W_i, Y_i),\ i = 1, ..., n$ from the random vector (X, Z, W, Y). Then, we can define $V_i = r(X_i, Z_i, W_i)$ and the log-likelihood function is derived from equation (3.2) by:

$$\sum_{i=1}^{n} \Big\{ \log[f_{\epsilon(\theta_0)}\left(\Lambda_0(Y_i) - \phi_0(X_i, Z_i)\right)] + \log[\Lambda_0'(Y_i)] \Big\}. \tag{3.3}$$

This log-likelihood function depends on the unknown functions $f_{\epsilon(\theta_0)}$, r and ϕ_0. The idea is now to estimate θ be replacing all unknown quantities in the above log-likelihood by nonparametric estimators for a fixed value of θ, and to maximize the so-obtained expression with respect to the unknown parameter θ.

Let us first of all consider the estimation of the function ϕ_θ defined in (2.1) for a fixed arbitrary value of θ. Define

$$m_\theta(x,z,v) = \mathrm{E}\left(\Lambda_\theta(Y)|X = x, Z = z, V = v\right)$$

for $\theta \in \Theta$, and note that $\phi_\theta(x,z) = \mathrm{E}[m_\theta(x,z,V)]$. Hence, for estimating $\phi_\theta(x,z)$, we first need to estimate $m_\theta(x,z,v)$.

Denoting $m_0 \equiv m_{\theta_0}$, we have that

$$m_0(x,z,v) \;=\; \phi_0(x,z) + \lambda(v), \tag{3.4}$$

where $\lambda(v) = \mathrm{E}\left[\epsilon|V = v\right]$ using assumption $(A.1)$. Note that, under Assumption $(A.6)$ we have:

$$\mathrm{E}\left[\lambda(V)\right] = \mathrm{E}\left[\mathrm{E}\left(\epsilon|\ V\right)\right] = \mathrm{E}\epsilon \;=\; 0.$$

We assume in what follows that we dispose of a nonparametric estimator of V_i, denoted by $\widehat{V}_i = \widehat{r}(X_i, Z_i, W_i)$ $(i = 1, \ldots, n)$. In Subsection 4.1 we will develop conditions on $\widehat{V}_i - V_i$ that are needed for the asymptotic theory, while illustrations of nonparametric estimators of V_i will be provided in Subsection 4.4. We first estimate the function $m_\theta(x, z, v)$ by using a nonparametric kernel estimator based on $(X_i, Z_i, \widehat{V}_i, Y_i)$ $(i = 1, \ldots, n)$:

$$\begin{aligned} \widehat{m}_\theta(x, z, v) &= \widehat{\mathrm{E}}\left[\Lambda_\theta(Y) | X = x, Z = z, \widehat{V} = v\right] \\ &= \frac{\sum_{i=1}^n \Lambda_\theta(Y_i) K_h(x - X_i) K_h(z - Z_i) K_h(v - \widehat{V}_i)}{\sum_{i=1}^n K_h(x - X_i) K_h(z - Z_i) K_h(v - \widehat{V}_i)}, \end{aligned}$$

where K is a d-dimensional product kernel of the form $K(u_1, \ldots, u_d) = \prod_{j=1}^d k_1(u_j)$, with $d = d_x, d_z$ or d_v (where d_v represents the dimension of V) and k_1 is a univariate kernel function. As usual, h is a bandwidth converging to zero when n tends to infinity, $k_{1h}(.) = \frac{1}{h} k_1(\frac{\cdot}{h})$ and $K_h(u_1, \ldots, u_d) = \prod_{j=1}^d k_{1h}(u_j)$. For simplifying the presentation, we work with the same bandwidth for all variables, and we use a fixed trimming of the covariate space $R_{X,Z,V}$ as explained already in Section 2.1. Note that we could also work with a shrinking trimming procedure, but for reasons of clarity of presentation we prefer to work here with a fixed trimming scheme. Also note that instead of trimming the covariate space, we could also use boundary corrected kernels or local polynomial estimators, which would however make the notations unnecessarily heavy.

The kernel estimator of ϕ_θ is now defined as:

$$\widehat{\phi}_\theta(x, z) = \frac{1}{n} \sum_{i=1}^n \widehat{m}_\theta(x, z, \widehat{V}_i). \tag{3.5}$$

Using the estimator of $\phi_\theta(x, z)$ we can now estimate the error density $f_{\epsilon(\theta)}$ of the variable $\epsilon(\theta) = \Lambda_\theta(Y) - \phi_\theta(X, Z)$ for a fixed value of θ:

$$\widehat{f}_{\epsilon(\theta)}(e) = \frac{1}{n} \sum_{i=1}^n k_{2g}\left(e - \widehat{\epsilon}_i(\theta)\right), \tag{3.6}$$

where $\widehat{\epsilon}_i(\theta) = \Lambda_\theta(Y_i) - \widehat{\phi}_\theta(X_i, Z_i)$, k_2 is a univariate kernel, and g is a bandwidth parameter.

Finally, we are in position to estimate the transformation parameter θ, by plugging-in all unknown quantities in the log-likelihood given in (3.3):

$$\widehat{\theta} = \arg\max_\theta \sum_{i=1}^n \Big\{ \log[\widehat{f}_{\epsilon(\theta)}(\Lambda_\theta(Y_i) - \widehat{\phi}_\theta(X_i, Z_i))] + \log[\Lambda'_\theta(Y_i)] \Big\}. \tag{3.7}$$

Once θ is estimated we can re-estimate the regression function $\phi(x, y)$, this time using $\widehat{\theta}$ instead of an arbitrary value of θ. This gives

$$\widehat{\phi}(x, z) = \widehat{\phi}_{\widehat{\theta}}(x, z)$$

for any x and z.

4 Large sample properties

4.1 Assumptions and notations

As explained in Subsection 2.1, our model fits in the general framework considered in Chen, Linton and Van Keilegom (2003). Hence, the asymptotic properties for $\widehat{\theta}$ and $\widehat{\phi}$ will be derived by verifying their general set of assumptions and under high level assumptions on the convergence of an estimator $\widehat{V} = \widehat{r}(X, Z, W)$ of V. Basically, we need to verify the conditions in both Theorem 1 (regarding consistency) and Theorem 2 (regarding asymptotic normality) in Chen, Linton and Van Keilegom (2003). In particular, to be able to apply their theorem on asymptotic normality, we need an estimator of the nonparametric function ϕ_0 that obeys a certain asymptotic expansion, which is in particular fulfilled when ϕ_0 is additively separable. Hence, from now on, we consider an additive structure on the true function $\phi_0(x, z)$:

$$\phi_0(x, z) = c + \sum_{\alpha=1}^{d_x} \phi_{x0}^{\alpha}(x_\alpha) + \sum_{\alpha=1}^{d_z} \phi_{z0}^{\alpha}(z_\alpha), \tag{4.1}$$

with $\mathrm{E}[\phi_{x0}^{\alpha}(X_\alpha)] = 0$ for $\alpha = 1, \ldots, d_x$ and $\mathrm{E}[\phi_{z0}^{\alpha}(Z_\alpha)] = 0$ for $\alpha = 1, \ldots, d_z$.

It is important to stress that consistency of $\widehat{\theta}$ could in principle be proved under the general non-additive form of ϕ_0 defined in the previous section in equation (3.4) (see Chen, Linton and Van Keilegom 2003, Theorem 1 for more details). However, to simplify the presentation and save space, we will consider from now on the additive form defined in equation (4.1). Note also that an additive structure for ϕ allows to circumvent the issue of curse of dimensionality. In what follows, we use marginal integration techniques (see e.g. Newey 1994, and Linton and Nielsen 1995)[3] and consider, for $\theta \in \Theta$, the function:

$$\phi_\theta^{add}(x, z) := c_\theta + \sum_{\alpha=1}^{d_x} \phi_{x\theta}^{\alpha}(x_\alpha) + \sum_{\alpha=1}^{d_z} \phi_{z\theta}^{\alpha}(z_\alpha)$$

with $\phi_{x\theta}^{\alpha}(x_\alpha) = \mathrm{E}(m_\theta(x_\alpha, X_{(-\alpha)}, Z, V)) - c_\theta$, where $X = (X_\alpha, X_{(-\alpha)})$, $\phi_{z\theta}^{\alpha}(z_\alpha) = \mathrm{E}(m_\theta(X, z_\alpha, Z_{(-\alpha)}, V)) - c_\theta$, where $Z = (Z_\alpha, Z_{(-\alpha)})$, and $c_\theta = \mathrm{E}[\Lambda_\theta(Y)]$.

Remark 4.1. *Note that for $\phi_\theta(x, z) = \mathrm{E}[m_\theta(x, z, V)]$ we have in general that $\phi_\theta^{add}(x, z) \neq \phi_\theta(x, z)$ except if $\theta = \theta_0$, since the additive structure of $m_\theta(x, z, v)$ only holds for $\theta = \theta_0$.*

[3] Note that other methods could have been used like smooth backfitting techniques (see Mammen, Linton and Nielsen 1999). We briefly comment on this at the end of Subsection 4.2.

Consider

$$\begin{aligned}
\widehat{\phi}^{\alpha}_{x\theta}(x_\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, \widehat{V}_i) - \widehat{c}_\theta \qquad (\alpha = 1,\ldots,d_x)\\
\widehat{\phi}^{\alpha}_{z\theta}(z_\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, \widehat{V}_i) - \widehat{c}_\theta \qquad (\alpha = 1,\ldots,d_z),
\end{aligned}$$

where $\widehat{c}_\theta = n^{-1}\sum_{i=1}^{n}\Lambda_\theta(Y_i)$. The nonparametric estimator of $\phi^{add}_\theta(x,z)$ is now given by:

$$\widehat{\phi}^{add}_\theta(x,z) = \widehat{c}_\theta + \sum_{\alpha=1}^{d_x}\widehat{\phi}^{\alpha}_{x\theta}(x_\alpha) + \sum_{\alpha=1}^{d_z}\widehat{\phi}^{\alpha}_{z\theta}(z_\alpha), \tag{4.2}$$

and $\widehat{\phi}^{add}(x,z) = \widehat{\phi}^{add}_{\widehat{\theta}}(x,z)$.

In the same spirit as in Section 2.1, let $s^{add}_\theta = (\phi^{add}_\theta, \dot{\phi}^{add}_\theta, f_{\epsilon(\theta)}, f'_{\epsilon(\theta)}, \dot{f}_{\epsilon(\theta)}, r)^t$ where $\dot{\phi}^{add}_\theta$ (respectively $\dot{f}_{\epsilon(\theta)}$) denotes the vector of partial derivatives of ϕ^{add}_θ (respectively $f_{\epsilon(\theta)}$) with respect to the components of θ and $f'_{\epsilon(\theta)}(y)$ denotes the derivative of $f_{\epsilon(\theta)}(y)$ with respect to y, and let $\widehat{s}^{add}_\theta = (\widehat{\phi}^{add}_\theta, \dot{\widehat{\phi}}^{add}_\theta, \widehat{f}_{\epsilon(\theta)}, \widehat{f}'_{\epsilon(\theta)}, \dot{\widehat{f}}_{\epsilon(\theta)}, \widehat{r})^t$.[4] Define

$$M(\theta, s^{add}_\theta, X, Z, W, Y) = \frac{1}{f_{\epsilon(\theta)}(\epsilon(\theta))}\Big[f'_{\epsilon(\theta)}(\epsilon(\theta))\big\{\dot{\Lambda}_\theta(Y) - \dot{\phi}^{add}_\theta(X,Z)\big\} + \dot{f}_{\epsilon(\theta)}(\epsilon(\theta))\Big] + \frac{\dot{\Lambda}'_\theta(Y)}{\Lambda'_\theta(Y)},$$

where $\epsilon(\theta) = \Lambda_\theta(Y) - \phi^{add}_\theta(X,Z)$ (again we omit 'add' for simplicity), and let $M_n(\theta, s^{add}_\theta) = n^{-1}\sum_{i=1}^{n} M(\theta, s^{add}_\theta, X_i, Z_i, W_i, Y_i)$. Then, $M_n(\theta, \widehat{s}^{add}_\theta)$ is the derivative (up to the multiplicative factor n^{-1}) of the log-likelihood defined in equation (3.7) with respect to θ . Moreover, let $G(\theta, s^{add}_\theta) = \mathrm{E}\{M(\theta, s^{add}_\theta, X, Z, W, Y)\}$, and

$$\Gamma = \dot{G}(\theta, s^{add}_\theta)\Big|_{\theta=\theta_0}, \tag{4.3}$$

where $s^{add}_0 = s^{add}_{\theta_0}$ and Γ is the matrix of partial derivatives of $G(\theta, s^{add}_\theta)$ with respect to θ. Note that $\|G(\theta, s^{add}_\theta)\|$ and $\|M_n(\theta, \widehat{s}^{add}_\theta)\|$ take their minimum at θ_0 and $\widehat{\theta}$ respectively, where $\|\cdot\|$ denotes the Euclidean norm. We also need to introduce the matrix

$$\Sigma = \mathrm{Var}\Big\{A(X,Z,W,Y)\Big\}, \tag{4.4}$$

where

$$\begin{aligned}
A(X,Z,W,Y) &= M(\theta_0, s^{add}_0, X, Z, W, Y) + \sum_{\alpha=1}^{d_x}\Big[v^{\alpha}_{1x}(\xi)K^{\alpha}_x(X_\alpha) + w^{\alpha}_{1x}(\xi)L^{\alpha}_x(X_\alpha)\Big]\\
&\quad + \sum_{\alpha=1}^{d_z}\Big[v^{\alpha}_{1z}(\xi)K^{\alpha}_z(Z_\alpha) + w^{\alpha}_{1z}(\xi)L^{\alpha}_z(Z_\alpha)\Big],
\end{aligned} \tag{4.5}$$

[4]Note that $f_{\epsilon(\theta)}$ (and also $f'_{\epsilon(\theta)}$ and $\dot{f}_{\epsilon(\theta)}$) depends on $\Lambda_\theta(Y) - \phi^{add}_\theta(X,Z)$, so in principle we should write $f^{add}_{\epsilon(\theta)}$, but for reasons of simplification of notation we continue to write $f_{\epsilon(\theta)}$ as before.

$\xi = (X, Z, V, Y)^t$, the functions v_{1x}^α, v_{1z}^α, w_{1x}^α and w_{1z}^α are defined in (7.4) and (7.5), and the functions K_x^α, K_z^α, L_x^α and L_z^α are defined in (7.6) and (7.7), all given in the Appendix.

For any $\ell \geq 1$ we let $\frac{\partial}{\partial e_\ell}$ denote the derivative with respect to the ℓth argument of a vector e, ∇_e denotes the gradient with respect to the vector e, and ∇_e^t is its transpose. The following regularity conditions are required for the asymptotic results:

(C.1) For $j = 1, 2$, k_j is a symmetric kernel of order $q_j \geq 4$, i.e. $\int u^m k_j(u)\, du = 0$ for $m = 1, \ldots, q_j - 1$ and $\int u^{q_j} k_j(u)\, du \neq 0$. Moreover, k_j has compact support and is twice continuously differentiable, and q_1 satisfies $q_1 > 2d_z + d_w + d_x + d_v + 1$.

(C.2) $nh^{4d_z+2d_w+2d_x+2d_v+2} \to \infty$, $nh^{2q_1} \to 0$, $ng^6(\log g^{-1})^{-2} \to \infty$ and $ng^{2q_2} \to 0$, where q_1 and q_2 are defined in condition (C.1).

(C.3) The density $f_{X,Z,V}$ exists and is bounded away from zero and infinity. Moreover, $f_{X,Z,V}$ is Lipschitz continuous and has a compact support $R_{X,Z,V}$.

(C.4) $m_\theta(x, z, v)$, $\dot{m}_\theta(x, z, v)$ and $\nabla_v m_\theta(x, z, v)$ exist and are q_1 times continuously differentiable with respect to the components of x, z and v on $R_{X,Z,V} \times \Theta$. In addition, all derivatives up to order q_1 are bounded, uniformly in (x, z, v, θ) in $R_{X,Z,V} \times \Theta$.

(C.5) $f_{Z,W}(z, w)$ and $F_{X|Z,W}(x|z, w)$ exist and are q_1 times continuously differentiable with respect to the components of z and w on $R_{Z,W}$. In addition, all derivatives up to order q_1 are bounded, uniformly in $(x, z, w) \in R_{X,Z,W}$, and $f_{Z,W}(z, w)$ is bounded away from zero, uniformly in z and w.

(C.6) $\Lambda_\theta(y)$ is three times continuously differentiable with respect to y and θ, and there exists a $\delta > 0$ such that

$$\mathrm{E}\Big[\sup_{\|\theta'-\theta\|\leq\delta}\Big|\frac{\partial^{k+l}}{\partial y^k \partial\theta_1^{l_1}\ldots\partial\theta_p^{l_p}}\Lambda_{\theta'}(Y)\Big|\Big] < \infty$$

for all θ in Θ and for all k and l such that $0 \leq k + l \leq 3$, where $l = l_1 + \ldots + l_p$ and $\theta = (\theta_1, \ldots, \theta_p)^t$. Moreover,

$$\sup_{\theta\in\Theta}\mathrm{E}\Big\|\dot{\Lambda}_\theta(Y)\Big\|^2 < \infty.$$

(C.7) $F_{\epsilon(\theta)}(y)$ is three times continuously differentiable with respect to y and θ, and

$$\sup_{\theta,y}\Big|\frac{\partial^{k+l}}{\partial y^k \partial\theta_1^{l_1}\ldots\partial\theta_p^{l_p}}F_{\epsilon(\theta)}(y)\Big| < \infty$$

for all k and l such that $0 \leq k + l \leq 2$, where $l = l_1 + \ldots + l_p$ and $\theta = (\theta_1, \ldots, \theta_p)^t$.

(C.8) For all $\eta > 0$, there exists $\epsilon(\eta) > 0$ such that

$$\inf_{\|\theta-\theta_0\|>\eta} \|G(\theta, s_\theta^{add})\| \geq \epsilon(\eta) > 0.$$

Moreover, the matrix Γ is of full rank.

(C.9) The control function V_i and its estimate $\widehat{V}_i$ satisfy

$$\widehat{V}_i - V_i = \Big(n^{-1} \sum_{k=1}^n S_{ik}\Big)(1 + R_i),$$

where $(S_{ik})_{k=1,\ldots,n}$ have the same dimension as V_i and R_i is of dimension 1,

$$S_{ik} = Q(X_i, X_k, Z_i, W_i)K_h(Z_i - Z_k)K_h(W_i - W_k)$$

for some bounded function Q,

$$\max_{1\leq i,k\leq n} \Big\| \mathrm{E}(S_{ik}|Z_k, W_k, X_i, Z_i, W_i) \Big\| = O_P(h^{q_1}),$$

and $\max_{1\leq i\leq n} |R_i| = o_P(1)$.

Remark 4.2. *Note that condition (C.8) is needed to identify the true parameter θ_0. It is taken from the paper of Chen, Linton and Van Keilegom (2003), on which our proof is based, and it is needed to prove the weak consistency of $\widehat{\theta}$. Also note that, contrary to other papers in the literature, we explicitly show the consistency of $\widehat{\theta}$.*

4.2 Asymptotic properties

The following lemma gives an i.i.d. representation of the estimators $\widehat{\phi}_\theta^{add}(x,z)$ and $\dot{\widehat{\phi}}_\theta^{add}(x,z)$, uniformly in θ, x and z, and will be a key ingredient for obtaining the asymptotic limit of our estimator $\widehat{\theta}$.

Lemma 4.1. *Assume (A.1)-(A.6) and (C.1)–(C.9). Then,*

$$
\begin{aligned}
&\widehat{\phi}_\theta^{add}(x,z) - \phi_\theta^{add}(x,z) \\
&= n^{-1}\sum_{i=1}^{n}\Bigg(\sum_{\alpha=1}^{d_x} k_{1h}(x_\alpha - X_{\alpha i})\Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big] f^{-1}_{X_\alpha|X_{(-\alpha)},Z,V}(x_\alpha|X_{(-\alpha)i},Z_i,V_i) \\
&\qquad + \sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})\Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big] f^{-1}_{Z_\alpha|X,Z_{(-\alpha)},V}(z_\alpha|X_i,Z_{(-\alpha)i},V_i) \\
&\qquad + \mathrm{E}\Big[\Big\{\sum_{\alpha=1}^{d_x}\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)},Z_i,V) + \sum_{\alpha=1}^{d_z}\nabla_v^t m_\theta(X,z_\alpha,Z_{(-\alpha)i},V)\Big\} \\
&\qquad\qquad \times Q(X,X_i,Z_i,W_i)\Big|X_i,Z_i,W_i,Z=Z_i,W=W_i\Big] f_{ZW}(Z_i,W_i) \\
&\qquad + \Big[\sum_{\alpha=1}^{d_x} m_\theta(x_\alpha,X_{(-\alpha)i},Z_i,V_i) + \sum_{\alpha=1}^{d_z} m_\theta(X_i,z_\alpha,Z_{(-\alpha)i},V_i) \\
&\qquad - (d_z + d_x - 1)\Lambda_\theta(Y_i) - \phi_\theta^{add}(x,z)\Big]\Bigg) \\
&\quad + o_P(n^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\widehat{\dot\phi}_\theta^{add}(x,z) - \dot\phi_\theta^{add}(x,z) \\
&= n^{-1}\sum_{i=1}^{n}\Bigg(\sum_{\alpha=1}^{d_x} k_{1h}(x_\alpha - X_{\alpha i})\Big[\dot\Lambda_\theta(Y_i) - \dot m_\theta(X_i,Z_i,V_i)\Big] f^{-1}_{X_\alpha|X_{(-\alpha)},Z,V}(x_\alpha|X_{(-\alpha)i},Z_i,V_i) \\
&\qquad + \sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})\Big[\dot\Lambda_\theta(Y_i) - \dot m_\theta(X_i,Z_i,V_i)\Big] f^{-1}_{Z_\alpha|X,Z_{(-\alpha)},V}(z_\alpha|X_i,Z_{(-\alpha)i},V_i) \\
&\qquad + \mathrm{E}\Big[\Big\{\sum_{\alpha=1}^{d_x}\nabla_v^t \dot m_\theta(x_\alpha,X_{(-\alpha)},Z_i,V) + \sum_{\alpha=1}^{d_z}\nabla_v^t \dot m_\theta(X,z_\alpha,Z_{(-\alpha)i},V)\Big\} \\
&\qquad\qquad \times Q(X,X_i,Z_i,W_i)\Big|X_i,Z_i,W_i,Z=Z_i,W=W_i\Big] f_{ZW}(Z_i,W_i) \\
&\qquad + \Big[\sum_{\alpha=1}^{d_x} \dot m_\theta(x_\alpha,X_{(-\alpha)i},Z_i,V_i) + \sum_{\alpha=1}^{d_z} \dot m_\theta(X_i,z_\alpha,Z_{(-\alpha)i},V_i) \\
&\qquad - (d_z + d_x - 1)\dot\Lambda_\theta(Y_i) - \dot\phi_\theta^{add}(x,z)\Big]\Bigg) \\
&\quad + o_P(n^{-1/2}),
\end{aligned}
$$

uniformly in $(x,z) \in R_{X,Z}$ *and* $\theta \in \Theta$.

We are now ready to state the main result of this paper.

Theorem 4.1. *Assume (A.1)-(A.6) and (C.1)–(C.9). Then,*

$$n^{1/2}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega),$$

where

$$\Omega = \Gamma^{-1}\Sigma(\Gamma^t)^{-1},$$

and where Γ *and* Σ *are defined in (4.3) and (4.4).*

The following corollary is a by-product of the main result:

Corollary 4.1. *Assume (A.1)-(A.6) and (C.1)–(C.9). Then, for any* $(x, z) \in R_{X,Z}$*,*

$$(nh)^{1/2}\Big(\widehat{\phi}^{add}(x, z) - \phi_0(x, z)\Big) \xrightarrow{d} N(0, \sigma^2(x, z)),$$

where

$$\sigma^2(x, z) = \int k_1^2(u)du\Big\{\sum_{\alpha=1}^{d_x} f_{X_\alpha}(x_\alpha)s_x^\alpha(x_\alpha) + \sum_{\alpha=1}^{d_z} f_{Z_\alpha}(z_\alpha)s_z^\alpha(z_\alpha)\Big\},$$

$$s_x^\alpha(x_\alpha) = Var\Big\{\Big[\Lambda_0(Y) - m_{\theta_0}(X, Z, V)\Big] f^{-1}_{X_\alpha|X_{(-\alpha)},Z,V}(x_\alpha|X_{(-\alpha)}, Z, V)\Big|X_\alpha = x_\alpha\Big\}$$

for $\alpha = 1, \dots, d_x$*, and*

$$s_z^\alpha(z_\alpha) = Var\Big\{\Big[\Lambda_0(Y) - m_{\theta_0}(X, Z, V)\Big] f^{-1}_{Z_\alpha|X,Z_{(-\alpha)},V}(z_\alpha|X, Z_{(-\alpha)}, V)\Big|Z_\alpha = z_\alpha\Big\}$$

for $\alpha = 1, \dots, d_z$*.*

Let us comment on these asymptotic results:

1. It follows from the proof of Theorem 4.1 that the estimation of the control variable V by an estimator $\widehat{V}$ has no effect on the formula of the asymptotic variance Ω, i.e. the estimator $\widehat{\theta}$ behaves asymptotically as if the variables $V_1, \dots .V_n$ would be observed. As we already pointed out in Subsection 2.2, this is a very interesting property which has been observed in some cases, as in Mammen, Rothe and Schienle (2012b) or Hahn and Ridder (2013).

2. It can be seen from the proof of Theorem 4.1 that the extra terms in the formula of Σ come from the estimation of the nuisance functions ϕ_0, ϕ_0^{add}, $f_{\epsilon(\theta_0)}$, $f'_{\epsilon(\theta_0)}$ and $\dot{f}_{\epsilon(\theta_0)}$. Note that these terms would be equal to zero if (X, Z) and ϵ would be independent, which is the case in the exogenous model considered by Linton, Sperlich and Van Keilegom (2008). Another difference between the variance in the endogenous and the

exogenous case lies in the formula of $\phi_\theta^{add}(x,z)$ (denoted by $m_\theta(x)$ in their paper). Even for $\theta = \theta_0$, the function $\phi_0(x,z)$ is different in the two cases, namely in the exogenous case it equals $\mathrm{E}[\Lambda_0(Y)|X = x, Z = z]$, whereas in the endogenous case it is given by $\int \mathrm{E}[\Lambda_0(Y)|X = x, Z = z, V = v]dF_V(v)$.

3. Note that the asymptotic distribution of $\widehat{\phi}^{add}(x,z)$ in Corollary 4.1 is the same as that of $\widehat{\phi}_0^{add}(x,z)$, i.e. the asymptotic distribution is as if the parameter θ_0 were known. In addition, the estimation of the control variable V also vanishes asymptotically, similarly as in the case of the asymptotic distribution of $\widehat{\theta}$ (see point 1 above).

4. Instead of using the marginal integration method to estimate $\phi_0(x,z)$, we could as well use other estimation procedures, like e.g. the smooth backfitting method (see e.g. Mammen, Linton and Nielsen, 1999, and Mammen and Park, 2005). However, the proofs are considerably more complicated in that case. For the smooth backfitting, we expect that the asymptotic distribution of $\widehat{\theta}$ will be the same as for the marginal integration method, except that $\phi_\theta^{add}(x,z)$ is now given by the components depending on x and z of the function $m_\theta^{add}(x,z,v)$ defined as:

$$m_\theta^{add}(x,z,v) = \mathrm{argmin}_{m\in\mathcal{M}_{add}} \int \Big[m_\theta(x,z,v) - m(x,z,v)\Big]^2 dF_{X,Z,V}(x,z,v),$$

where

$$\mathcal{M}_{add} = \Big\{ m : m(x,z,v) = \sum_{\alpha=1}^{d_x} m_{x_\alpha}(x_\alpha) + \sum_{\alpha=1}^{d_z} m_{z_\alpha}(z_\alpha) + m_v(v) \\ \text{for some } m_{x_1},\ldots,m_{x_{d_x}}, m_{z_1},\ldots,m_{z_{d_z}}, m_v \Big\}.$$

Proving the asymptotic properties of this type of estimator is however not at all an easy task. We therefore restrict attention in this paper to the marginal integration estimator. The refinement of our method to smooth backfitting methods (or other methods to estimate an additive regression function) is left as a topic of future research.

5. At last, we would like to note that there exists an interesting literature on semiparametric efficiency bounds for two step estimation methods (see e.g. Ai and Chen 2012 and Chen, Hahn and Liao 2012), which is related to our paper. In particular, we could define our estimator $\widehat{\theta}$ by using a weighted estimation criterion and we could then choose the weights in order to minimize the asymptotic variance of this weighted estimator. This is an interesting path of future research, but it is beyond the scope of this paper.

4.3 Asymptotic variance and bootstrap

Note that although the asymptotic limit of $n^{1/2}(\widehat{\theta}-\theta_0)$ is explicitly defined and has a simple normal distribution, it cannot be directly applied in practice, since the covariance matrix contains a number of unknown quantities, namely the parameter vector θ_0, the error density $f_{\epsilon(\theta_0)}$, its derivative $f'_{\epsilon(\theta_0)}$, the function ϕ_0 and the derivatives of these functions with respect to θ. Each of these functions can be estimated by a kernel estimator, by taking the appropriate derivative of the kernel estimator of ϕ_0 and of $f_{\epsilon(\theta_0)}$ given in (3.5) and (3.6). This approach leads (under suitable conditions on the bandwidths) to a consistent estimator of the asymptotic variance, by using similar results as in Lemma 4.1 (for ϕ_0 and its derivatives) and Lemma 7.1 (for $f_{\epsilon(\theta_0)}$ and its derivatives). However, we do not recommend to follow this approach in practice since some of these unknown quantities are hard to estimate and require the introduction of new smoothing parameters.

An alternative approach consists in approximating the variance, or even the whole distribution, of $\widehat{\theta}$ by means of a bootstrap procedure. The use of bootstrap techniques in the context of semiparametric inference has received a lot of attention in recent years. Chen, Linton and Van Keilegom (2003) propose a naive bootstrap procedure and give primitive conditions under which the bootstrap estimator converges to the same limit as the original estimator. Our estimator, which is a two-step semiparametric Z-estimator with s depending on θ, is a special case of the general estimator considered in their setting. In a closely related context of one-step semiparametric M-estimation with s independent of θ, Cheng and Huang (2010), respectively Cheng (2011), proposed an exchangeable bootstrap scheme for approximating the distribution, respectively the moments, of $\widehat{\theta}$, whereas Cheng and Pillai (2012) proposed a model based bootstrap procedure. Finally, instead of using a bootstrap procedure, one could also make use of Bayesian inference techniques to approximate the distribution of a semiparametric estimator. We refer to Cheng and Kosorok (2008) for more details.

Let us now focus on how the naive bootstrap proposed in Chen, Linton and Van Keilegom (2003) can be applied in our setting. Let $(X_i^*, Z_i^*, W_i^*, Y_i^*)$, $i=1,\ldots,n$, be drawn randomly with replacement from the original data (X_i, Z_i, W_i, Y_i), $i=1,\ldots,n$, and for any θ let $\widehat{s}^{add,*} = (\widehat{\phi}_\theta^{add,*}, \dot{\widehat{\phi}}_\theta^{add,*}, \widehat{f}^*_{\epsilon(\theta)}, \widehat{f}'^*_{\epsilon(\theta)}, \dot{\widehat{f}}^*_{\epsilon(\theta)}, \widehat{r}^*)^t$ be the same estimator as $\widehat{s}_\theta^{add}$ but based on the bootstrap data. For each (θ, s), define

$$M_n^*(\theta, s) = n^{-1}\sum_{i=1}^n M(\theta, s, X_i^*, Z_i^*, W_i^*, Y_i^*)$$

and define

$$\widehat{\theta}^* = \text{argmin}_{\theta\in\Theta}\Big\|M_n^*(\theta, \widehat{s}_\theta^{add,*})\Big\|.$$

Theorem B in Chen, Linton and Van Keilegom (2003) shows that under certain regularity conditions $n^{1/2}(\widehat{\theta}^* - \widehat{\theta})$ and $n^{1/2}(\widehat{\theta} - \theta_0)$ converge in distribution to the same normal limit. More precisely, using similar techniques as in the proof of Theorem 4.1, we conjecture that

$$\begin{aligned} & n^{1/2}(\widehat{\theta}^* - \widehat{\theta}) \\ &= -\Gamma^{-1} n^{-1/2} \sum_{i=1}^{n} \big[A(X_i^*, Z_i^*, W_i^*, Y_i^*) - A(X_i, Z_i, W_i, Y_i) \big] + o_{P^*}(1), \end{aligned} \tag{4.6}$$

where the function $A(X, Z, W, Y)$ is defined in (4.5) and where the $o_{P^*}(1)$-term goes to zero in probability, conditionally on the original data (X_i, Z_i, W_i, Y_i), $i = 1, \ldots, n$. From this claim together with the central limit theorem and Theorem 4.1 the result would follow. However, a detailed proof of (4.6) is beyond the scope of this paper, since it requires elaborate, lengthy and sophisticated calculations which are too space consuming. Instead we will check the validity of the proposed bootstrap procedure by means of a simulation study (see Section 5).

4.4 Examples of control variables

Different candidates can be proposed to characterize the control variable V. In the line of Newey, Powell and Vella (1999), or Blundell and Powell (2003), V can be defined as the error of a separable equation in a triangular nonparametric model:

$$X = \psi(Z, W) + V, \tag{4.7}$$

where W is a vector of instrumental variables taking values in $\mathbb{R}^{d_w}$ such that (ϵ, V) and (Z, W) are independent, in order to satisfy Assumption (A.1) (like in the nonseparable model defined below). A second option would be to consider a second nonseparable equation and a single endogenous variable X defined by:

$$X = \psi(Z, W, \eta), \tag{4.8}$$

where ψ is strictly monotone in V. Then, $V = F_{X|Z,W}(X|Z, W) = F_\eta(\eta)$ is a uniformly distributed control variable under the following conditions: (i) (ϵ, η) and (Z, W) are independent, and (ii) η is a continuously distributed random variable with strictly increasing distribution function on the support of η and $\psi(Z, W, t)$ is strictly monotone in t with probability 1 (see Imbens and Newey 2009 for more details).

A natural extension of equation (4.8) when X is multidimensional, consists in considering the set of one dimensional independent equations:

$$\begin{cases} X_1 &= \psi_1(Z, W, \eta_1) \\ & \vdots \\ X_{d_x} &= \psi_{d_x}(Z, W, \eta_{d_x}), \end{cases} \tag{4.9}$$

and $\eta = (\eta_1, ..., \eta_{d_x})$.

Let us consider in more detail the second case with nonseparable equation (4.8). A nonparametric estimator of V is then derived as follows:

$$\begin{aligned} \widehat{V}_i &= \widehat{F}_{X|Z,W}(X_i|Z_i, W_i) \\ &= \frac{\sum_{j=1}^n 1(X_j \leq X_i) K_h(Z_i - Z_j) K_h(W_i - W_j)}{\sum_{j=1}^n K_h(Z_i - Z_j) K_h(W_i - W_j)}. \end{aligned}$$

Let us check briefly that Condition (C.9) is satisfied. We have:

$$\begin{aligned} &\widehat{V}_i - V_i \\ &= \frac{\sum_{k=1}^n \Big[I(X_k \leq X_i) - F_{X|ZW}(X_i|Z_i, W_i)\Big] K_h(Z_i - Z_k) K_h(W_i - W_k)}{\sum_{k=1}^n K_h(Z_i - Z_k) K_h(W_i - W_k)} \\ &= \frac{n^{-1}\sum_{k=1}^n \Big[I(X_k \leq X_i) - F_{X|ZW}(X_i|Z_i, W_i)\Big] K_h(Z_i - Z_k) K_h(W_i - W_k)}{f_{ZW}(Z_i, W_i)} \\ &\quad + O_P((nh^{d_z+d_w})^{-1}) + O(h^{2q_1}) \\ &:= \Big(n^{-1}\sum_{k=1}^n S_{jk}\Big)(1 + o_P(1)). \end{aligned}$$

It can be easily seen that $\mathrm{E}(S_{ik}|Z_k, W_k, X_i, Z_i, W_i) = O_P(h^{q_1})$ uniformly in i and k which proves the result. A similar sketch of proof could be derived for the separable case.

5 Finite sample study

5.1 Simulations

We consider the following data generating process:

$$\Lambda_\theta(Y) = b_0 + b_1 X + \epsilon,$$

where Λ_θ is the Box-Cox transformation, that is $\Lambda_\theta(y) = \frac{y^\theta - 1}{\theta}$ $(\theta \neq 0)$, $\Lambda_\theta(y) = \log(y)$ $(\theta = 0)$, and ϵ is drawn from $N(0, \sigma_e^2)$. In this setting, we omit the exogenous variable Z. The variable X is generated from the following generating process:

$$X = a_0 + a_1 W + a_2 \epsilon + U,$$

where W, ϵ and U are mutually independent, W is drawn from $N(0, \sigma_w^2)$ and U from $N(0, \sigma_u^2)$. The regressor X is then correlated with the error term ϵ and the instrumental variable W is correlated with X but not with ϵ in order to correct for this endogeneity issue. The control function V is identified as the residual of the regression of X on W. We present here the results for the case where $b_0 = 1, b_1 = 0.25, a_0 = 1, a_1 = -0.5, a_2 = 2$, $\sigma_w = 1, \sigma_e = 0.25$ and $\sigma_u = 0.2$. The parameter θ_0 is set equal to 1, 2 and 3 and is estimated using the package "optimize" in R. We use the gaussian kernel and fix the bandwidth parameters as follows: $h_X = h_W = 0.1$, $h_V = 0.04$ and $h_\epsilon = 0.05$. Note that optimizing the bandwidth parameters in order to minimize the mean squared error should give better results but we believe this is beyond the scope of this paper. The Monte Carlo study has been performed with $mc = 500$ replications and a sample size $n = 100$. We provide each time the mean, the standard deviation and the mean squared error (mse hereafter) of $\widehat{\theta}$. We also provide the bias, the standard deviation and mse for the nonparametric estimator $\widehat{\phi}(x)$ evaluated at the median value of X. Moreover we also present the same results when the true value of V is used.

The results are summarized in Table 1 and show first that the method works well for reasonable sample size, that is the bias and variance value are relatively small. Moreover, we can see from the numbers between parentheses that the impact of estimating the control variable V is negligible.

θ_0	mean($\widehat{\theta}$)	sd($\widehat{\theta}$)	mse($\widehat{\theta}$)	bias($\widehat{\phi}$)	sd($\widehat{\phi}$)	mse($\widehat{\phi}$)
1	0.94 (0.96)	0.69 (0.64)	0.48 (0.41)	0.09 (0.09)	0.44 (0.42)	0.20 (0.18)
2	1.91 (1.95)	0.76 (0.74)	0.58 (0.54)	0.06 (0.06)	0.36 (0.38)	0.14 (0.14)
3	2.89 (2.93)	0.81 (0.79)	0.66 (0.63)	0.05 (0.05)	0.33 (0.34)	0.11 (0.11)

Table 1: Simulation results for θ_0 and $\phi_0(x)$ evaluated at the median of X. The numbers between parentheses correspond to the values computed using the true control function V.

At last, in order to evaluate the accuracy of the normal approximation and the usefulness of Theorem 4.1 for conducting inference, we also provide some QQ-plot for the estimated values of θ_0 and $\phi_0(x)$ (estimated at the median point) for $n = 100$. All QQ-plots given in Figures 1 and 2 confirm that the normal approximation is reasonable.

5.2 Bootstrap

In order to check the validity of the bootstrap procedure proposed in Subsection 4.3, we continue to use the same model as in Subsection 5.1. For each sample of observations $(X_i, Y_i, W_i)_{i=1,\dots,n}$ of size $n = 100$, we generate $B = 100$ bootstrapped samples $(X_i^*, Y_i^*, W_i^*)_{i=1,\dots,n}$ of the same size, drawn randomly with replacement from the original data. Then, from these bootstrapped samples, B estimators $(\widehat{\theta}^{b,*})_{b=1,\dots,B}$ are computed as well as the mean

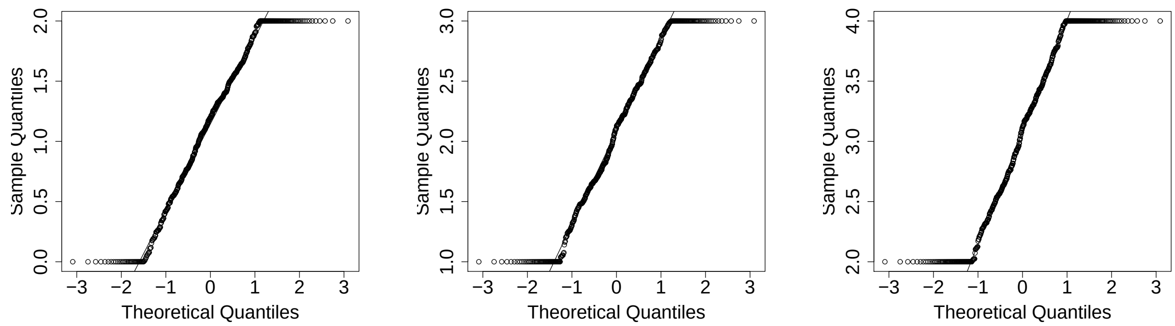

Figure 1: QQ-plots of mc=500 simulated values of $\widehat{\theta}$ for $\theta_0 = 1,\ 2,\ 3$ and $n = 100$.

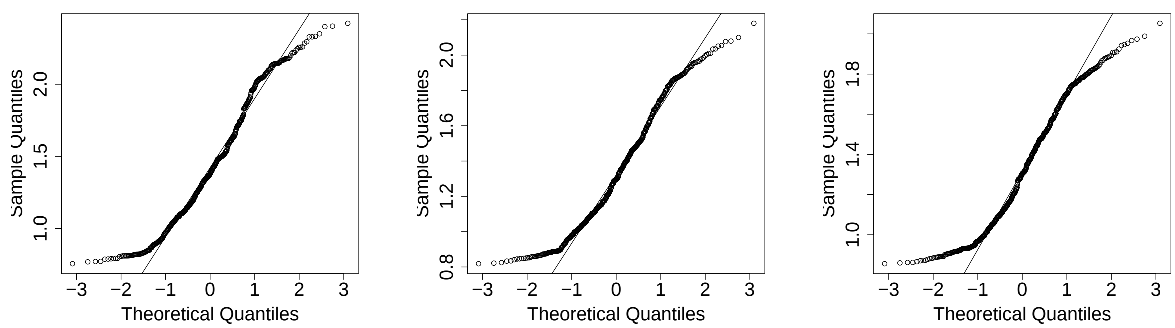

Figure 2: QQ-plots of mc=500 simulated values of $\widehat{\phi}(x)$ evaluated at the median of X for $\theta_0 = 1,\ 2,\ 3$ and $n = 100$.

and the variance of these B bootstrapped estimators. We simulate $mc = 100$ initial samples $(X_i, Y_i, W_i)_{i=1,\dots,n}$ in order to obtain a total of mc bootstrapped means and bootstrapped variances. At last, we provide the histograms of these bootstrapped means and bootstrapped variances for different values of θ_0 (see Figures 3, 4 and 5). In order to provide an empirical proof of the validity of our bootstrap procedure, we check that each histogram is centered around the mean and the variance of the 100 estimated values of θ_0. This is indeed the case for each of the 6 figures which therefore suggests that the bootstrap procedure proposed in Subsection 4.3 works well in practice.

5.3 Real data analysis

We conclude this finite sample study by considering the estimation of Engel curves based on the UK Family Expenditure Survey as in Blundell, Chen and Kristensen (2007). The Engel curve relationship describes the expansion path for commodity demands as the household's budget increases. The motivation for a control function approach derives from the endogene-

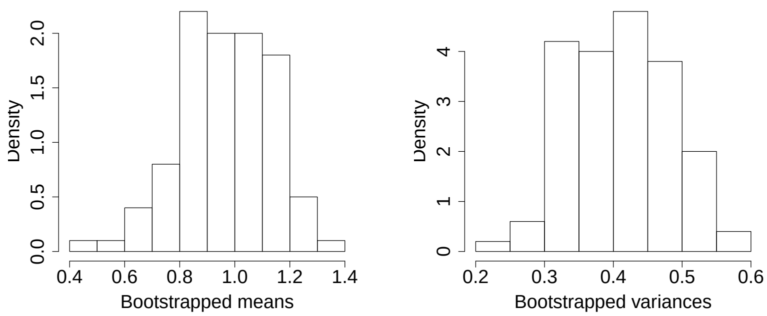

Figure 3: Histograms of bootstrapped means and variances for $\theta_0 = 1$. The corresponding values for the original samples are $Mean(\widehat{\theta}) = 1.04$ and $Var(\widehat{\theta}) = 0.41$.

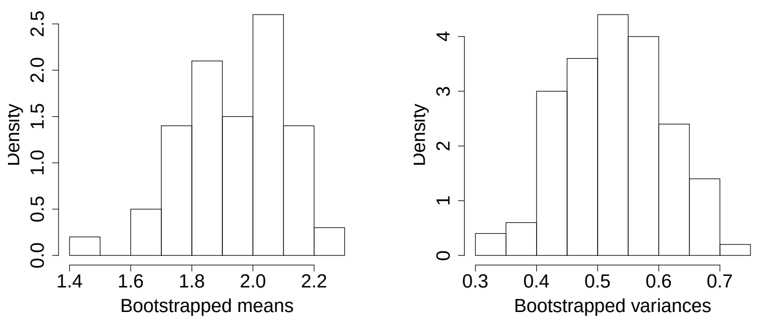

Figure 4: Histograms of bootstrapped means and variances for $\theta_0 = 2$. The corresponding values for the original samples are $Mean(\widehat{\theta}) = 2.01$ and $Var(\widehat{\theta}) = 0.57$.

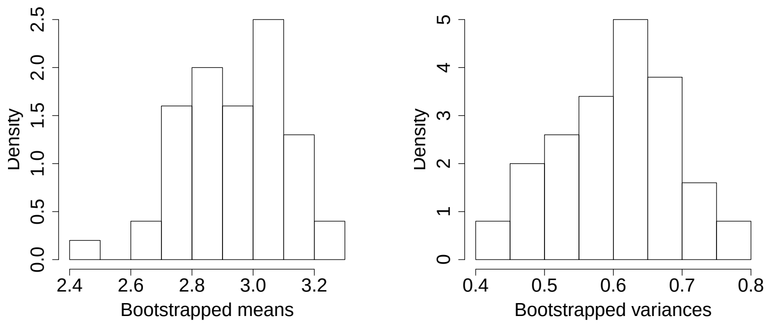

Figure 5: Histograms of bootstrapped means and variances for $\theta_0 = 3$. The corresponding values for the original samples are $Mean(\widehat{\theta}) = 2.99$ and $Var(\widehat{\theta}) = 0.67$.

ity of the total budget variable. As total expenditure is endogenous for individual commodity demands, we use gross earnings of the household head as an instrument (see Blundell, Chen and Kristensen 2007 for a detailed discussion). In this application, we consider a single year of study, 1995, and 3 broad categories of nondurables and services: (1) leisure goods and services, (2) travel and (3) household goods and services. To preserve some demographic homogeneity, we consider couples where the head of household is aged between 20 and 55 and at work and among them select a subset of couples with 3 children. We first present some descriptive statistics for this subsample in Table 2.

	Mean	Sd.
Leisure goods	0.129	0.105
Travel	0.190	0.098
Household goods	0.114	0.085
log nondurable expenditure	5.810	0.637
log gross earnings	5.769	0.644
Sample size	294	

Table 2: Data descriptives

The objective is to estimate the model defined in (1.1) where Y represents a budget share (leisure, travel or household) and X the log of nondurable expenditure. There is no exogenous variable Z in the application. The instrumental variable W used to identify and estimate the model is the log of gross earnings. The operator Λ_θ is chosen as the Box-Cox transformation. The control variable V is identified as the conditional distribution of X given W and the bandwidth parameters are fixed as follows: $h_X = h_W = 0.5$, $h_V = 0.02$ and $h_\epsilon = 0.3$. The same remark as in Subsection 5.1 applies, that is optimizing the bandwidth parameters in order to minimize the mse should give better results but this is beyond the scope of this paper. Figure 6 presents the estimated curves of ϕ_0 for the three goods and the corresponding 95% pointwise confidence bands obtained using the naive bootstrap described in Subsection 4.3 and based on 100 resamples. The results for the estimation of θ_0 are presented in Table 3 with the values of the mean and the standard deviation obtained by the same bootstrap procedure. The results show small standard deviations and relatively small confidence intervals.

6 Conclusion

In this work we have studied a semiparametric transformation model with a parametric transformation operator Λ_θ, a nonparametric regression function ϕ and some endogenous

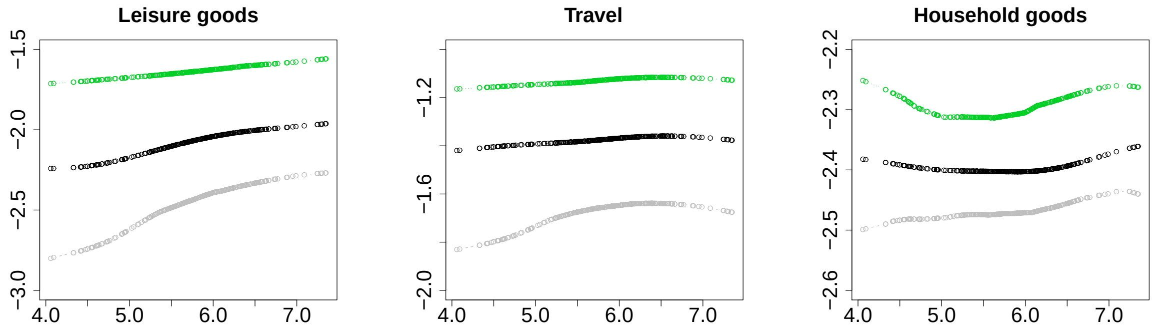

Figure 6: Estimation of the function ϕ_0 for the 3 budget shares together with the 95% pointwise confidence intervals, based on 100 resamples.

	$\hat{\theta}$	Mean($\hat{\theta}^*$)	Sd($\hat{\theta}^*$)
Leisure goods	0.120	0.107	0.087
Travel	0.303	0.314	0.126
Household goods	0.001	0.003	0.012

Table 3: Estimation of θ_0 for the 3 budget shares together with bootstrapped means and standard deviations based on 100 resamples.

explanatory variables. We use a control function approach to identify the nonparametric structure $(\Lambda, \phi, F_\epsilon)$. A profile likelihood method is proposed to estimate the parametric transformation, and by imposing an additive structure on the function ϕ, we showed the asymptotic normality of the proposed estimator with $\sqrt{n}$ rate of convergence. The asymptotic results show that the estimation of the control variable disappears asymptotically, both for the estimation of the transformation parameter as for the estimation of the function ϕ. Some finite sample simulations confirm the validity of our method. Finally, we illustrated our method using data from the UK Family Expenditure Survey.

7 Appendix: Proofs

Proof of Proposition 3.1. To prove identification of the structure $(\Lambda, \phi, F_\epsilon)$, we proceed in two steps: we first establish identification of Λ and then prove that ϕ and F_ϵ are identified.

1. *Identification of* Λ. This first step is inspired by the proof of Chiappori, Komunjer and Kristensen (2010). Under the regularity assumptions $(A.3)$ and $(A.4)$, we can differentiate equation (3.1) with respect to y and x_1 (the first coordinate of x) to

obtain:

$$\begin{aligned}\frac{\partial}{\partial y}\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv) &= f_\epsilon(\Lambda(y)-\phi(x,z)).\Lambda'(y)\\ \frac{\partial}{\partial x_1}\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv) &= -f_\epsilon\left(\Lambda(y)-\phi\left(x,z\right)\right).\frac{\partial}{\partial x_1}\phi(x,z).\end{aligned}$$

Let $A=\{(x,z)\in R_{X,Z}:\frac{\partial}{\partial x_1}\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv)\neq 0 \text{ for every } y\in R_Y\}$. Under Assumptions $(A.4)$ and $(A.5)$, the set A has a nonempty interior. Then, for any point $(x,z)\in A$ and for every $y\in R_Y$, we have:

$$-\frac{\Lambda'(y)}{\frac{\partial}{\partial x_1}\phi(x,z)}=s(y,x,z),$$

where $s(y,x,z)=\frac{\frac{\partial}{\partial y}\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv)}{\frac{\partial}{\partial x_1}\int F_{Y|X,Z,V}(y|x,z,v)F_V(dv)}$. Note that $s(y,x,z)$ is non zero and keeps a constant sign for all $y\in R_Y$. Integrating from 0 to y and under Assumption $(A.6)$ we get:

$$\Lambda(y)=-\frac{\partial}{\partial x_1}\phi(x,z).S(y,x,z),$$

where $S(y,x,z)=\int_0^y s(t,x,z)dt$. Again, $S(y,x,z)$ is nonzero and keeps a constant sign for all $y\in R_Y$. Hence, $\mathrm{E}[S(Y,x,z)]\neq 0$. Using again Assumption $(A.6)$ we get:

$$\frac{\partial}{\partial x_1}\phi(x,z)=-\frac{1}{\mathrm{E}[S(Y,x,z)]},$$

and finally we obtain that:

$$\Lambda(y)=\frac{S(y,x,z)}{\mathrm{E}[S(Y,x,z)]}.$$

Hence, Λ is identified.

2. *Identification of ϕ and F_ϵ.* The identification of ϕ is a direct consequence of Assumptions $(A.1)$ and $(A.2)$ following Imbens and Newey (2009). Identification of F_ϵ eventually follows from equation (3.1). This finishes the proof. □

Proof of Lemma 4.1. We restrict attention to proving the first result of Lemma 4.1, since the second one can be shown in a very similar way. We first decompose $\widehat{\phi}_\theta^{add}(x,z)-\phi_\theta^{add}(x,z)$

as follows:

$$
\begin{aligned}
&\widehat{\phi}_\theta^{add}(x,z) - \phi_\theta^{add}(x,z) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{\alpha=1}^{d_x}\widehat{m}_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, \widehat{V}_i) + \sum_{\alpha=1}^{d_z}\widehat{m}_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, \widehat{V}_i) - (d_z + d_x - 1)\Lambda_\theta(Y_i)\right] - \phi_\theta^{add}(x,z) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{\alpha=1}^{d_x} m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, V_i) + \sum_{\alpha=1}^{d_z} m_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, V_i) - (d_z + d_x - 1)\Lambda_\theta(Y_i) - \phi_\theta^{add}(x,z)\right] \\
&\quad + \sum_{\alpha=1}^{d_x}\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{m}_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, \widehat{V}_i) - m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, V_i)\right] \\
&\quad + \sum_{\alpha=1}^{d_z}\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{m}_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, \widehat{V}_i) - m_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, V_i)\right] \\
&= R_1 + \sum_{\alpha=1}^{d_x} R_2^\alpha + \sum_{\alpha=1}^{d_z} R_3^\alpha.
\end{aligned}
$$

Then, using a Taylor expansion on R_2^α, we have:

$$
\begin{aligned}
R_2^\alpha &= \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{m}_\theta - m_\theta\right)(x_\alpha, X_{(-\alpha)i}, Z_i, \widehat{V}_i) + \frac{1}{n}\sum_{i=1}^{n}\Big(m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, \widehat{V}_i) - m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, V_i)\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{m}_\theta - m_\theta\right)(x_\alpha, X_{(-\alpha)i}, Z_i, V_i) + \frac{1}{n}\sum_{i=1}^{n}\nabla_v^t m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, V_i)(\widehat{V}_i - V_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\nabla_v^t\left(\widehat{m}_\theta - m_\theta\right)(x_\alpha, X_{(-\alpha)i}, Z_i, \xi_i)(\widehat{V}_i - V_i) \\
&\quad + \frac{1}{2n}\sum_{i=1}^{n}(\widehat{V}_i - V_i)^t\nabla_{vv} m_\theta(x_\alpha, X_{(-\alpha)i}, Z_i, \xi_i')(\widehat{V}_i - V_i) \\
&= R_{21}^\alpha + R_{22}^\alpha + R_{23}^\alpha + R_{24}^\alpha,
\end{aligned}
$$

where $\xi_i = \lambda_i V_i + (1-\lambda_i)\widehat{V}_i$ for some $\lambda_i \in [0,1]$, $\xi_i' = \lambda_i' V_i + (1-\lambda_i')\widehat{V}_i$ for $\lambda_i' \in [0,1]$, $(\widehat{V}_i - V_i)^t$ is the transpose of the vector $\widehat{V}_i - V_i$, $\nabla_v m_\theta$ represents the gradient of m_θ, i.e. the vector of partial derivatives of m_θ with respect to the components of v, $\nabla_v^t m_\theta$ its transpose, and $\nabla_{vv} m_\theta$ is the Hessian matrix.

A similar expansion can be derived for R_3^α:

$$\begin{aligned}R_3^\alpha &= \frac{1}{n}\sum_{i=1}^n (\widehat{m}_\theta - m_\theta)(X_i, z_\alpha, Z_{(-\alpha)i}, V_i) + \frac{1}{n}\sum_{i=1}^n \nabla_v^t m_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, V_i)(\widehat{V}_i - V_i) \\ &\quad + \frac{1}{n}\sum_{i=1}^n \nabla_v^t (\widehat{m}_\theta - m_\theta)(X_i, z_\alpha, Z_{(-\alpha)i}, \eta_i)(\widehat{V}_i - V_i) \\ &\quad + \frac{1}{2n}\sum_{i=1}^n (\widehat{V}_i - V_i)^t \nabla_{vv} m_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, \eta_i^{'})(\widehat{V}_i - V_i) \\ &= R_{31}^\alpha + R_{32}^\alpha + R_{33}^\alpha + R_{34}^\alpha,\end{aligned}$$

with similar definitions for η_i and $\eta_i^{'}$. Comparing the decomposition of R_2^α and R_3^α, it is clear that for $k = 1, ..., 4$, R_{2k}^α has the same behavior as R_{3k}^α. In what follows we concentrate on R_{21}^α and R_{22}^α, since it is easily seen that R_{23}^α and R_{24}^α are of lower order.

We start with R_{21}^α. Write

$$\begin{aligned}R_{21}^\alpha &= \frac{1}{n}\sum_{i=1}^n (\widehat{m}_\theta - \widetilde{m}_\theta)(x_\alpha, X_{(-\alpha)i}, Z_i, V_i) + \frac{1}{n}\sum_{i=1}^n (\widetilde{m}_\theta - m_\theta)(x_\alpha, X_{(-\alpha)i}, Z_i, V_i) \\ &= R_{211}^\alpha + R_{212}^\alpha,\end{aligned}$$

where

$$\widetilde{m}_\theta(x, z, v) = \frac{\sum_{i=1}^n \Lambda_\theta(Y_i) K_h(x - X_i) K_h(z - Z_i) K_h(v - V_i)}{\sum_{i=1}^n K_h(x - X_i) K_h(z - Z_i) K_h(v - V_i)},$$

i.e. with respect to $\widehat{m}_\theta(x, z, v)$ we have replaced the $\widehat{V}_i$'s by the true (but unknown) V_i's. The term R_{212}^α can be worked out similarly as in e.g. Linton and Nielsen (1995), since this is the ordinary marginal integration estimator. Hence, this term equals

$$n^{-1}\sum_{i=1}^n \Big[\Lambda_\theta(Y_i) - m_\theta(X_i, Z_i, V_i)\Big] k_{1h}(x_\alpha - X_{\alpha i}) f^{-1}_{X_\alpha | X_{(-\alpha)}, Z, V}(x_\alpha | X_{(-\alpha)i}, Z_i, V_i) + o_P(n^{-1/2}).$$

Now consider

$$(\widehat{m}_\theta - \widetilde{m}_\theta)(x_\alpha, X_{(-\alpha)i}, Z_i, V_i) = \frac{\sum_{j=1}^n \widehat{N}_{ij}}{\sum_{j=1}^n \widehat{D}_{ij}} - \frac{\sum_{j=1}^n \widetilde{N}_{ij}}{\sum_{j=1}^n \widetilde{D}_{ij}},$$

where $\widehat{N}_{ij} = \Lambda_\theta(Y_j) k_{1h}(x_\alpha - X_{\alpha j}) K_h(X_{(-\alpha)i} - X_{(-\alpha)j}) K_h(Z_i - Z_j) K_h(V_i - \widehat{V}_j)$, $\widehat{D}_{ij} = k_{1h}(x_\alpha - X_{\alpha j}) K_h(X_{(-\alpha)i} - X_{(-\alpha)j}) K_h(Z_i - Z_j) K_h(V_i - \widehat{V}_j)$, and similarly for $\widetilde{N}_{ij}$ and $\widetilde{D}_{ij}$. In analogy with these notations, we define $N_i = \mathrm{E}(\Lambda_\theta(Y) | x_\alpha, X_{(-\alpha)i}, Z_i, V_i) f_{X,Z,V}(x_\alpha, X_{(-\alpha)i}, Z_i, V_i)$ and $D_i = f_{X,Z,V}(x_\alpha, X_{(-\alpha)i}, Z_i, V_i)$. In order to simplify the notation, we have omitted the dependence on θ and x_α, but of course it will be a crucial point in the proof. Next, write

$$\begin{aligned}R_{211}^\alpha &= n^{-1}\sum_{i=1}^n \sum_{j=1}^n (\widehat{N}_{ij} - \widetilde{N}_{ij}) \frac{1}{\sum_{j=1}^n \widehat{D}_{ij}} + n^{-1}\sum_{i=1}^n \sum_{j=1}^n \widetilde{N}_{ij} \Big(\frac{1}{\sum_{j=1}^n \widehat{D}_{ij}} - \frac{1}{\sum_{j=1}^n \widetilde{D}_{ij}}\Big) \\ &= \Big[n^{-2}\sum_{i=1}^n \sum_{j=1}^n (\widehat{N}_{ij} - \widetilde{N}_{ij}) \frac{1}{D_i} - n^{-2}\sum_{i=1}^n \sum_{j=1}^n (\widehat{D}_{ij} - \widetilde{D}_{ij}) \frac{N_i}{D_i^2}\Big](1 + o_P(1)),\end{aligned}$$

where the $o_P(1)$ term is uniform in x_α (from assumption (C.3)) and in θ. The latter equals

$$\Big[n^{-2}\sum_{i=1}^n D_i^{-1}\sum_{j=1}^n \Big\{\nabla_v^t \widetilde{N}_{ij} - \nabla_v^t \widetilde{D}_{ij}\frac{N_i}{D_i}\Big\}(V_j - \widehat{V}_j)\Big](1+o_P(1))$$
$$= -\Big[n^{-3}\sum_{i=1}^n D_i^{-1}\sum_{j=1}^n\sum_{k=1}^n \Big\{\nabla_v^t \widetilde{N}_{ij} - \nabla_v^t \widetilde{D}_{ij}\frac{N_i}{D_i}\Big\}S_{jk}\Big](1+o_P(1)), \qquad (7.1)$$

where the $o_P(1)$ term is again uniform in x_α and θ, and where $\nabla_v \widetilde{N}_{ij} = \Lambda_\theta(Y_j)k_{1h}(x_\alpha - X_{\alpha j})K_h(X_{(-\alpha)i} - X_{(-\alpha)j})K_h(Z_i - Z_j)h^{-1}\nabla_v K_h(V_i - V_j)$, $\nabla_v \widetilde{D}_{ij} = k_{1h}(x_\alpha - X_{\alpha j})K_h(X_{(-\alpha)i} - X_{(-\alpha)j})K_h(Z_i - Z_j)h^{-1}\nabla_v K_h(V_i - V_j)$, and

$$\widehat{V}_j - V_j = \Big(n^{-1}\sum_{k=1}^n S_{jk}\Big)(1+o_P(1)),$$

uniformly in $1 \le j \le n$, by condition (C.9). Ignoring the factor $(1+o_P(1))$, (7.1) is a V-process of order three depending on x_α, θ and h, which can be rewritten as:

$$n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n q(T_i, T_j, T_k, x_\alpha, \theta, h)$$

where $T_i = (X_i, Z_i, W_i, Y_i)^t$ and

$$q(T_i, T_j, T_k, x_\alpha, \theta, h) = -D_i^{-1}\Big\{\nabla_v^t \widetilde{N}_{ij} - \nabla_v^t \widetilde{D}_{ij}\frac{N_i}{D_i}\Big\}S_{jk}.$$

We denote $p(T_i, T_j, T_k, x_\alpha, \theta, h) = h^{2d_z+d_w+d_x+d_v+1}q(T_i, T_j, T_k, x_\alpha, \theta, h)$ and consider the following V-process:

$$\mathcal{V}_n(x_\alpha, \theta, h) = n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n p(T_i, T_j, T_k, x_\alpha, \theta, h).$$

Since a V-process can be written as a U-process plus negligible terms, following Sherman (1994), we introduce the associated U-process $\mathcal{U}_n(x_\alpha, \theta, h)$ which can be decomposed as (see equation (6) on page 449 in Sherman):

$$\begin{aligned}
&\mathcal{U}_n(x_\alpha, \theta, h)\\
&= \frac{1}{n(n-1)(n-2)}\sum_{i,j,k\neq} p(T_i, T_j, T_k, x_\alpha, \theta, h)\\
&= n^{-1}\sum_{i=1}^n \mathrm{E}[p(T_i, T, T', x_\alpha, \theta, h)|T_i] + n^{-1}\sum_{j=1}^n \mathrm{E}[p(T, T_j, T', x_\alpha, \theta, h)|T_j]\\
&\quad + n^{-1}\sum_{k=1}^n \mathrm{E}[p(T, T', T_k, x_\alpha, \theta, h)|T_k] - 2\mathrm{E}[p(T, T', T'', x_\alpha, \theta, h)] + \mathcal{R}_n(x_\alpha, \theta, h), \qquad (7.2)
\end{aligned}$$

where T, T', T'' are i.i.d and have the same distribution as $T_1, ..., T_n$. The last term $\mathcal{R}_n(x_\alpha, \theta, h)$ is by construction the sum of two degenerate U-processes, one of order 2, denoted by $\mathcal{R}_{n2}(x_\alpha, \theta, h)$, and one of order 3, denoted by $\mathcal{R}_{n3}(x_\alpha, \theta, h)$. In what follows, we concentrate on $\mathcal{R}_{n2}(x_\alpha, \theta, h)$, which will be dominant. In order to control uniformly in x_α, θ and h the term $\mathcal{R}_{n2}(x_\alpha, \theta, h)$, we will apply Corollary 4 in Sherman (1994). Let us first introduce some notations. We define the following functional class associated to the U-process $\mathcal{U}_n(x_\alpha, \theta, h)$:

$$\mathcal{F} = \{(t, t', t'') \to p(t, t', t'', x_\alpha, \theta, h) : x_\alpha \in R_{X_\alpha}, \theta \in \Theta, h > 0\}.$$

In order to apply Corollary 4, we need to check that $\mathcal{F}$ is Euclidean (see Sherman 1994 or Pakes and Pollard 1989 for a precise definition). Using conditions (C.5), (C.6) and (C.9), and Lemma 2.14 and Example 2.10 in Pakes and Pollard (1989), it follows that $\mathcal{F}$ is Euclidean and so is the class of functions associated to $\mathcal{R}_{n2}(x_\alpha, \theta, h)$ (see Lemma 6 in Sherman 1994). Then, using Corollary 4 in Sherman (1994), it follows that

$$\sup_{x_\alpha, \theta, h} |\mathcal{R}_{n2}(x_\alpha, \theta, h)| = O_P(n^{-1})$$

and hence, using Assumption (C.2),

$$\begin{aligned} h_n^{-(2d_z+d_w+d_x+d_v+1)} \sup_{x_\alpha, \theta} |\mathcal{R}_{n2}(x_\alpha, \theta, h_n)| &= O_P(n^{-1} h_n^{-(2d_z+d_w+d_x+d_v+1)}) \\ &= o_P(n^{-1/2}), \end{aligned}$$

where h_n denotes (here) the smoothing parameter associated to the sample size n (in order to make the distinction with the parameter h of the U-process). Let us now go back to the first term on the right hand side of equation (7.2) evaluated at $h = h_n$:

$$n^{-1} \sum_{i=1}^{n} \mathrm{E}[p(T_i, T, T', x_\alpha, \theta, h_n)|T_i] := n^{-1} h_n^{2d_z+d_w+d_x+d_v+1} \sum_{i=1}^{n} \mathrm{E}[q(T_i, T, T', x_\alpha, \theta, h_n)|T_i].$$

By definition, we have:

$$n^{-1} \sum_{i=1}^{n} \mathrm{E}[q(T_i, T_j, T_k, x_\alpha, \theta, h_n)|T_i] \;=\; -n^{-1} \sum_{i=1}^{n} D_i^{-1} \mathrm{E}\Big[\Big\{\nabla_v^t \widetilde{N}_{ij} - \nabla_v^t \widetilde{D}_{ij} \frac{N_i}{D_i}\Big\} S_{jk} \Big| T_i\Big].$$

From condition (C.9) we know that $\|E(S_{jk}|T_j)\| = O_P(h_n^{q_1})$ uniformly in j. Then, it easily follows that

$$n^{-1} \sum_{i=1}^{n} \mathrm{E}[q(T_i, T, T', x_\alpha, \theta, h_n)|T_i] = O_P(h_n^{q_1}) = o_P(n^{-1/2}),$$

since $nh_n^{2q_1} \to 0$ and by using assumptions (C.3), (C.5) and (C.6). In reality the order is even smaller than $O_P(h_n^{q_1})$, but it is not necessary to do a more detailed order calculation, since we reach already the required $o_P(n^{-1/2})$-rate based on this simple argument. In a similar way we can show the order of the second and third term on the right hand side of (7.2) (with p replaced by q) :

$$n^{-1}\sum_{j=1}^{n} \mathrm{E}[q(T,T_j,T',x_\alpha,\theta,h_n)|T_j] = O_P(h_n^{q_1}) = o_P(n^{-1/2}),$$

and

$$n^{-1}\sum_{k=1}^{n} \mathrm{E}[q(T,T',T_k,x_\alpha,\theta,h_n)|T_k] = O_P(h_n^{q_1}) = o_P(n^{-1/2})$$

uniformly in x_α and θ. Moreover, we also have that $\mathrm{E}[q(T,T',T'',x_\alpha,\theta,h_n)] = o(n^{-1/2})$. This shows that $R_{211}^\alpha = o_P(n^{-1/2})$, and so R_{21}^α equals

$$R_{21}^\alpha = n^{-1}\sum_{i=1}^{n}\Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big]k_{1h}(x_\alpha - X_{\alpha i})f^{-1}_{X_\alpha|X_{(-\alpha)},Z,V}(x_\alpha|X_{(-\alpha)i},Z_i,V_i) + o_P(n^{-1/2}).$$

Next, consider R_{22}^α. Using again Sherman (1994)'s result on degenerate U-processes, we can prove in a very similar way as for R_{21}^α that

$$\begin{aligned}
R_{22}^\alpha &= n^{-1}\sum_{i=1}^{n}\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)i},Z_i,V_i)(\widehat{V}_i - V_i)\\
&= \Big\{n^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n}\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)i},Z_i,V_i)S_{ik}\Big\}(1+o_P(1))\\
&= \Big\{n^{-1}\sum_{i=1}^{n}\mathrm{E}\Big[\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)i},Z_i,V_i)Q(X_i,X,Z_i,W_i)K_h(Z_i-Z)K_h(W_i-W)\Big|T_i\Big]\\
&\quad + n^{-1}\sum_{k=1}^{n}\mathrm{E}\Big[\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)},Z,V)Q(X,X_k,Z,W)K_h(Z-Z_k)K_h(W-W_k)\Big|T_k\Big]\\
&\quad - \mathrm{E}\Big[\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)},Z_1,V_1)S_{12}\Big]\Big\}(1+o_P(1)) + o_P(n^{-1/2})\\
&= n^{-1}\sum_{k=1}^{n}\mathrm{E}\Big[\nabla_v^t m_\theta(x_\alpha,X_{(-\alpha)},Z_k,V)Q(X,X_k,Z_k,W_k)\Big|T_k,Z=Z_k,W=W_k\Big]f_{ZW}(Z_k,W_k)\\
&\quad + O(h_n^{q_1}) + o_P(n^{-1/2}),
\end{aligned}$$

provided $nh_n^{2q_1} \to 0$. This finishes the proof. □

Proof of Theorem 4.1. In order to prove the asymptotic properties of our estimator, we need to check the high level assumptions of Theorems 1 and 2 in Chen, Linton and Van Keilegom (2003). Note that our setting is very different from Linton, Sperlich and Van Keilegom (2008) due to the fact that (X,Z) and ϵ are not independent in our case and that we also have a generated covariate $\widehat{V}$ to take into account. However the structure of our proof is somewhat similar to the structure of the proof of their Theorem 4.1.

A crucial assumption of their Theorem 4.1 is assumption A.8 given in the Appendix of their paper, which gives the properties that the estimator $\widehat{\phi}^{add}_\theta(x,z)$ (denoted by $\widehat{m}_\theta(x)$ in their paper) needs to satisfy. In addition, to check condition (2.6) of Theorem 2 in Chen, Linton and Van Keilegom (2003), they use the results of 11 lemmas given in their Appendix A.2.

In our setting, using the conditions $(C.1)-(C.9)$ which already include their assumptions A.1.-A.7., everything boils down to checking an analogue of their assumption A.8. and an analogue of their lemmas. Let's start with the analogue of their assumption A.8, which in our case corresponds to the following:

(i) The estimator $\widehat{\phi}^{add}_0$ can be written as

$$\begin{aligned}&\widehat{\phi}^{add}_0(x,z)-\phi^{add}_0(x,z)\\&=n^{-1}\sum_{i=1}^n\Big[\sum_{\alpha=1}^{d_x}k_{1h}(x_\alpha-X_{\alpha i})v_{01\alpha}(x_\alpha,T_i)+\sum_{\alpha=1}^{d_z}k_{1h}(z_\alpha-Z_{\alpha i})v_{02\alpha}(z_\alpha,T_i)\Big]\\&\quad+n^{-1}\sum_{i=1}^n v_{03}(x,z,T_i)+\widehat{v}_0(x,z),\end{aligned}$$

where $T_i=(X_i,Z_i,W_i,Y_i)^t$, $\sup_{x,z}|\widehat{v}_0(x,z)|=o_P(n^{-1/2})$, $E(v_{01\alpha}(x_\alpha,T)|X_\alpha=x_\alpha)=0$, $E(v_{02\alpha}(z_\alpha,T)|Z_\alpha=z_\alpha)=0$ and $E(v_{03}(x,z,T))=0$. Moreover, a similar expansion holds for the estimator $\dot{\widehat{\phi}}^{add}_0$.

(ii) Define $\mathcal{M}=\sum_{\alpha=1}^{d_x}C^1_a(R_{X_\alpha})+\sum_{\alpha=1}^{d_z}C^1_a(R_{Z_\alpha})$, where $C^b_a(R)$ $(0<a<\infty,\ 0<b\le 1$, $R\subset I\!R^k$ for some k) is the set of all continuous functions $f:R\to I\!R$ for which

$$\sup_y|f(y)|+\sup_{y,y'}\frac{|f(y)-f(y')|}{\|y-y'\|^b}\le a.$$

We equip the space $\mathcal{M}$ with the L_2-norm $\|\cdot\|_{L_2}$. Then, $P(\widehat{\phi}^{add}_\theta,\dot{\widehat{\phi}}^{add}_\theta\in\mathcal{M}$ for all $\theta\in\Theta)\to 1$ as $n\to\infty$.

(iii) The space $\mathcal{M}$ satisfies $\int\sqrt{\log N(\lambda,\mathcal{M},\|\cdot\|_\infty)}\,d\lambda<\infty$, where $N(\lambda,\mathcal{M},\|\cdot\|_\infty)$ is the covering number with respect to the norm $\|\cdot\|_\infty$ of the class $\mathcal{M}$, i.e. the minimal number of balls of $\|\cdot\|_\infty$-radius λ needed to cover $\mathcal{M}$.

(iv) $\sup_{\theta\in\Theta}\|\widehat{\phi}^{add}_\theta-\phi^{add}_\theta\|=o_P(1)$, $\sup_{\theta\in\Theta}\|\dot{\widehat{\phi}}^{add}_\theta-\dot{\phi}^{add}_\theta\|=o_P(1)$.

(v) Uniformly over all θ with $\|\theta-\theta_0\|=o(1)$, $\|\widehat{\phi}^{add}_\theta-\phi^{add}_\theta\|=o_P(n^{-1/4})$ and $\|\dot{\widehat{\phi}}^{add}_\theta-\dot{\phi}^{add}_\theta\|=o_P(n^{-1/4})$.

(vi) For all θ with $\|\theta - \theta_0\| = o(1)$,

$$\sup_{x,z} \left| (\widehat{\dot{\phi}}_\theta^{add} - \dot{\phi}_\theta^{add})(x,z) - (\widehat{\dot{\phi}}_0^{add} - \dot{\phi}_0^{add})(x,z) \right| = o_P(1)\|\theta - \theta_0\| + O_P(n^{-1/2}).$$

First, for point (i), note that the i.i.d. representations for $\widehat{\phi}_0^{add}(x,z) - \phi_0^{add}(x,z)$ and $\widehat{\dot{\phi}}_0^{add}(x,z) - \dot{\phi}_0^{add}(x,z)$ are given in Lemma 4.1.

Next, let us check that $P(\widehat{\phi}_\theta^{add}, \widehat{\dot{\phi}}_\theta^{add} \in \mathcal{M}$ for all $\theta \in \Theta) \to 1$ as $n \to \infty$. We have to prove that $\widehat{\phi}_\theta^{add}$ and $\widehat{\dot{\phi}}_\theta^{add}$ are uniformly bounded in x, z and θ as well as their first derivatives with respect to the components of x and z. Using condition (C.2), the decomposition in Lemma 4.1 allows to uniformly bound $\widehat{\phi}_\theta^{add} - \phi_\theta^{add}$ and $\widehat{\dot{\phi}}_\theta^{add} - \dot{\phi}_\theta^{add}$. As for the first derivatives of these estimators, it suffices to show that they converge in probability to the true functions, uniformly in x, z and θ. The proof for these derivatives is somewhat similar in structure to the proof of Lemma 4.1, and we therefore restrict to explaining the main differences. In fact, the proof is even much simpler than that of Lemma 4.1, since the remainder terms are only required to be $o_P(1)$, instead of the much sharper bound $o_P(n^{-1/2})$ that is required in the aforementioned proof. In particular, contrary to the proof of Lemma 4.1, we do not need to develop expansions of U-processes and we do not need to perform detailed order calculations. Hence, the uniform boundedness of these derivatives follows, which shows point (ii) above.

For point (iii), note that the covering number $N(\lambda, \mathcal{M}, \|\cdot\|_{L_2})$ satisfies $\log N(\lambda, \mathcal{M}, \|\cdot\|_{L_2}) \leq K\lambda^{-1}$ (see Corollary 2.7.2 in Van der Vaart and Wellner, 1996), and hence

$$\int_0^\infty \sqrt{\log N(\lambda, \mathcal{M}, \|\cdot\|_{L_2})}\, d\lambda < \infty.$$

Next, using Lemma 4.1 it is easy to show that $\sup_{\theta\in\Theta} \|\widehat{\phi}_\theta^{add} - \phi_\theta^{add}\|_{L_2} = O_P((nh^{1/2})^{-1/2} + h^{q_1}) = o_P(n^{-1/4})$ (the uniformity in θ can be shown using standard arguments based on partitioning the compact set Θ in small subsets, and the rate of the L_2-distance can be proved following e.g. the method of proof in Härdle and Mammen, 1993). In a similar way we can show that $\sup_{\theta\in\Theta} \|\widehat{\dot{\phi}}_\theta^{add} - \dot{\phi}_\theta^{add}\|_{L_2} = o_P(n^{-1/4})$. This shows (iv) and (v).

Finally, for point (vi), note that (again using the second part of Lemma 4.1) it suffices to control (for all i)

$$\left\| \dot{\Lambda}_\theta(Y_i) - \dot{m}_\theta(X_i, Z_i, V_i) - \dot{\Lambda}_0(Y_i) + \dot{m}_0(X_i, Z_i, V_i) \right\|,$$

and this is bounded by

$$\left\| \ddot{\Lambda}_0(Y_i) - \ddot{m}_0(X_i, Z_i, V_i) \right\| \|\theta - \theta_0\|(1 + o_P(1)) = o_P(1)\|\theta - \theta_0\|,$$

which is of the required order, and where $\ddot{\Lambda}_0$ represents the Hessian matrix with respect to θ_0. This finishes the proof of results (i)-(vi).

The next step is to present the analogues of the 11 lemmas given in Linton, Sperlich and Van Keilegom (2008). Their lemmas A.1-A-3, A.5 and A.9 concern results about the density estimation of the error ϵ and its derivatives and correspond to our Lemma 7.1. Their lemmas A.4, A.6-A.8, A.10-A.11 concern results about the functions M, M_n and their derivatives and correspond to our Lemma 7.2. The statement and the proof of these two lemmas are given at the end of this appendix.

Conditions (C.1)-(C.9), the results (i)-(vi) stated above and these last two lemmas allow us to conclude. In particular, Lemma 7.2 is crucial for calculating the asymptotic variance of $\hat{\theta}$, which is equal to the asymptotic variance of $\Gamma^{-1}\{M_n(\theta_0, s_0^{add}) + \Delta(\theta_0, s_0^{add})[\hat{s_0}^{add} - s_0^{add}]\}$, with $\Delta(\theta_0, s_0^{add})[\hat{s_0}^{add} - s_0^{add}]$ defined in the paragraph above Lemma 7.2 (see condition (2.6) in Theorem 2 in Chen, Linton and Van Keilegom 2003). The asymptotic normality of $\widehat{\theta}$ then follows. Note that the fact that our asymptotic variance does not depend on the estimation of the control variable V is proved in Lemma 7.2. □

Proof of Corollary 4.1. Write

$$\widehat{\phi}^{add}(x,z) - \phi_0(x,z) = \Big[\widehat{\phi}^{add}_{\widehat{\theta}}(x,z) - \widehat{\phi}^{add}_0(x,z)\Big] + \Big[\widehat{\phi}^{add}_0(x,z) - \phi^{add}_0(x,z)\Big]. \tag{7.3}$$

The first term on the right hand side equals $(\dot{\widehat{\phi}}^{add}_{\theta}(x,z)|_{\theta=\xi})^t(\widehat{\theta} - \theta_0)$ for some ξ on the line segment between $\widehat{\theta}$ and θ_0. From the proof of Theorem 4.1 it follows that

$$\sup_{\theta\in\Theta} \|\dot{\widehat{\phi}}^{add}_{\theta}(x,z)\| \leq \sup_{\theta\in\Theta} \|\dot{\widehat{\phi}}^{add}_{\theta}(x,z) - \dot{\phi}^{add}_{\theta}(x,z)\| + \sup_{\theta\in\Theta} \|\dot{\phi}^{add}_{\theta}(x,z)\| = O_P(1),$$

and hence the first term of (7.3) is $O_P(n^{-1/2}) = o_P((nh)^{-1/2})$ by Theorem 4.1. For the second term of (7.3) we apply Lemma 4.1, which yields that

$$\begin{aligned}
&\widehat{\phi}^{add}_0(x,z) - \phi^{add}_0(x,z) \\
&= n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} k_{1h}(x_\alpha - X_{\alpha i})\Big[\Lambda_0(Y_i) - m_0(X_i,Z_i,V_i)\Big] f^{-1}_{X_\alpha|X_{(-\alpha)},Z,V}(x_\alpha|X_{(-\alpha)i},Z_i,V_i) \\
&\quad + n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})\Big[\Lambda_0(Y_i) - m_0(X_i,Z_i,V_i)\Big] f^{-1}_{Z_\alpha|X,Z_{(-\alpha)},V}(z_\alpha|X_i,Z_{(-\alpha)i},V_i) \\
&\quad + o_P((nh)^{-1/2}).
\end{aligned}$$

The result now follows from e.g. Lindeberg's central limit theorem, together with standard variance calculations. □

We end this Appendix with two lemmas, that were needed in the above proofs. For this, we first write the result of Lemma 4.1 for $\theta = \theta_0$ using the following abbreviated notations :

$$\widehat{\phi}_0^{add}(x,z) - \phi_0^{add}(x,z) \tag{7.4}$$
$$= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x} k_{1h}(x_\alpha - X_{\alpha i})v_{1x}^{\alpha}(\xi_i) + \sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})v_{1z}^{\alpha}(\xi_i) + v_2(\xi_i)\Big\} + o_P(n^{-1/2}),$$

and

$$\widehat{\dot{\phi}}_0^{add}(x,z) - \dot{\phi}_0^{add}(x,z) \tag{7.5}$$
$$= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x} k_{1h}(x_\alpha - X_{\alpha i})w_{1x}^{\alpha}(\xi_i) + \sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})w_{1z}^{\alpha}(\xi_i) + w_2(\xi_i)\Big\} + o_P(n^{-1/2}),$$

uniformly in $(x,z) \in R_{X,Z}$ and $\theta \in \Theta$, where $\xi_i = (X_i, Z_i, V_i, Y_i)^t$ for $i = 1, \ldots, n$.

Lemma 7.1. *Assume (A.1)-(A.6) and (C.1)–(C.9). Then,*

$$\begin{aligned}\widehat{f}_{\epsilon(\theta_0)}(y) - f_{\epsilon(\theta_0)}(y) &= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)\frac{\partial}{\partial y}f_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i}) \\ &\quad + \sum_{\alpha=1}^{d_z} v_{1z}^{\alpha}(\xi_i)\frac{\partial}{\partial y}f_{\epsilon(\theta_0),Z_\alpha}(y, Z_{\alpha i}) + v_2(\xi_i)f'_{\epsilon(\theta_0)}(y)\Big\} \\ &\quad + n^{-1}\sum_{i=1}^{n} k_{2g}(y - \epsilon_i(\theta_0)) - f_{\epsilon(\theta_0)}(y) + R_{n1}(y),\end{aligned}$$

where $\sup_y |R_{n1}(y)| = o_P(n^{-1/2})$,

$$\begin{aligned}\widehat{f}'_{\epsilon(\theta_0)}(y) - f'_{\epsilon(\theta_0)}(y) &= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)\frac{\partial^2}{\partial y^2}f_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i}) \\ &\quad + \sum_{\alpha=1}^{d_z} v_{1z}^{\alpha}(\xi_i)\frac{\partial^2}{\partial y^2}f_{\epsilon(\theta_0),Z_\alpha}(y, Z_{\alpha i}) + v_2(\xi_i)f''_{\epsilon(\theta_0)}(y)\Big\} \\ &\quad + (ng)^{-1}\sum_{i=1}^{n} k'_{2g}\big(y - \epsilon_i(\theta_0)\big) - f'_{\epsilon(\theta_0)}(y) + R_{n2}(y),\end{aligned}$$

where $\sup_y |R_{n2}(y)| = o_P(n^{-1/2})$, *and*

$$
\begin{aligned}
&\dot{\widehat{f}}_{\epsilon(\theta_0)}(y) - \dot{f}_{\epsilon(\theta_0)}(y) \\
&= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)\frac{\partial}{\partial y}\dot{f}_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i}) \\
&\quad + \sum_{\alpha=1}^{d_z} v_{1z}^{\alpha}(\xi_i)\frac{\partial}{\partial y}\dot{f}_{\epsilon(\theta_0),Z_\alpha}(y, Z_{\alpha i}) + v_2(\xi_i)\dot{f}'_{\epsilon(\theta_0)}(y) \\
&\quad + \sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)\frac{\partial}{\partial y}f_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i}) \\
&\quad + \sum_{\alpha=1}^{d_z} w_{1z}^{\alpha}(\xi_i)\frac{\partial}{\partial y}f_{\epsilon(\theta_0),Z_\alpha}(y, Z_{\alpha i}) + w_2(\xi_i)f'_{\epsilon(\theta_0)}(y)\Big\} \\
&\quad +(ng)^{-1}\sum_{i=1}^{n} k'_{2g}\big(y - \epsilon_i(\theta_0)\big)\big(\dot{\Lambda}_0(Y_i) - \dot{\phi}_0^{add}(X_i, Z_i)\big) - \dot{f}_{\epsilon(\theta_0)}(y) + R_{n3}(y),
\end{aligned}
$$

where $\sup_y |R_{n3}(y)| = o_P(n^{-1/2})$.

The proof of Lemma 7.1 is similar to that of Lemmas A.1–A.3 in Linton, Sperlich and Van Keilegom (2008), and is therefore omitted. The only difference is that here ϵ and (X, Z) are not independent, which has an effect on the main term in the above representations.

For the second lemma, we say for any $\theta \in \Theta$ that $G(\theta, s)$ is pathwise differentiable at s in the direction $[\bar{s} - s]$ if the limit $\lim_{\tau\to 0}\big[G\big\{\theta, s + \tau(\bar{s} - s)\big\} - G(\theta, s)\big]/\tau$ exists. The limit is in that case denoted by $\Delta(\theta, s)[\bar{s} - s]$. This limit places an important role in the calculation of the asymptotic variance of $\widehat{\theta}$.

Lemma 7.2. *Assume (A.1)-(A.6) and (C.1)–(C.9). Then,*

$$
\begin{aligned}
&\Delta(\theta_0, s_0^{add})[\widehat{s}_0^{add} - s_0^{add}] \\
&= n^{-1}\sum_{i=1}^{n}\Big\{\sum_{\alpha=1}^{d_x}\Big[v_{1x}^{\alpha}(\xi_i)K_x^{\alpha}(X_{\alpha i}) + w_{1x}^{\alpha}(\xi_i)L_x^{\alpha}(X_{\alpha i})\Big] \\
&\quad + \sum_{\alpha=1}^{d_z}\Big[v_{1z}^{\alpha}(\xi_i)K_z^{\alpha}(Z_{\alpha i}) + w_{1z}^{\alpha}(\xi_i)L_z^{\alpha}(Z_{\alpha i})\Big]\Big\} + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$
\begin{aligned}
K_x^{\alpha}(X_{\alpha i}) \;=\; & E\Big[-\frac{\partial}{\partial y}\Big(\frac{1}{f_{\epsilon(\theta_0)}(y)}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(y)\Big)\big|_{y=\epsilon(\theta_0)}f_{X_\alpha|\epsilon(\theta_0)}(X_{\alpha i}|\epsilon(\theta_0)) \\
& -\frac{1}{f^2_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{\partial}{\partial y}f_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i})\big|_{y=\epsilon(\theta_0)}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0)) \\
& +\frac{1}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}\frac{\partial}{\partial y}f_{\epsilon(\theta_0),X_\alpha}(y, X_{\alpha i})\big|_{y=\epsilon(\theta_0)}\Big|X_{\alpha i}\Big],
\end{aligned}
\tag{7.6}
$$

and

$$L_x^\alpha(X_{\alpha i}) = E\Big[\frac{\partial}{\partial y} f_{X_\alpha|\epsilon(\theta_0)}(X_{\alpha i}|y)|_{y=\epsilon(\theta_0)}\Big|X_{\alpha i}\Big], \tag{7.7}$$

and similarly for $K_z^\alpha(Z_{\alpha i})$ *and* $L_z^\alpha(Z_{\alpha i})$ $(i = 1, \ldots, n)$.
In particular, the main term in the expansion of $\Delta(\theta_0, s_0^{add})[\widehat{s}_0^{add} - s_0^{add}]$ *comes from the estimation of the nuisance functions* $\phi_0^{add}, \dot\phi_0^{add}, f_{\epsilon(\theta_0)}, f'_{\epsilon(\theta_0)}$ *and* $\dot f_{\epsilon(\theta_0)}$*, and not from the estimation of the control variable* V*. The main term would in fact be exactly the same in the case* V *would be observed.*

Proof. Consider an arbitrary θ. Straightforward calculations show that

$$\begin{aligned}
&\Delta(\theta, s_\theta^{add})[\widehat{s}_\theta^{add} - s_\theta^{add}]\\
&= E\Big[\Big\{\frac{f'_{\epsilon(\theta)}(\epsilon(\theta))}{f^2_{\epsilon(\theta)}(\epsilon(\theta))}(\widehat{\phi}_\theta^{add} - \phi_\theta^{add})(X,Z) - \frac{(\widehat{f}_{\epsilon(\theta)} - f_{\epsilon(\theta)})(\epsilon(\theta))}{f^2_{\epsilon(\theta)}(\epsilon(\theta))}\Big\}\\
&\quad\times\Big\{f'_{\epsilon(\theta)}(\epsilon(\theta))\big[\dot\Lambda_\theta(Y) - \dot\phi_\theta^{add}(X,Z)\big] + \dot f_{\epsilon(\theta)}(\epsilon(\theta))\Big\}\\
&\quad+\frac{1}{f_{\epsilon(\theta)}(\epsilon(\theta))}\Big\{-f''_{\epsilon(\theta)}(\epsilon(\theta))\big[\dot\Lambda_\theta(Y) - \dot\phi_\theta^{add}(X,Z)\big](\widehat{\phi}_\theta^{add} - \phi_\theta^{add})(X,Z)\\
&\quad+(\widehat{f}'_{\epsilon(\theta)} - f'_{\epsilon(\theta)})(\epsilon(\theta))\big[\dot\Lambda_\theta(Y) - \dot\phi_\theta^{add}(X,Z)\big]\\
&\quad-f'_{\epsilon(\theta)}(\epsilon(\theta))(\dot{\widehat{\phi}}_\theta^{add} - \dot\phi_\theta^{add})(X,Z)\\
&\quad+(\dot{\widehat{f}}_{\epsilon(\theta)} - \dot f_{\epsilon(\theta)})(\epsilon(\theta)) - \dot f'_{\epsilon(\theta)}(\epsilon(\theta))(\widehat{\phi}_\theta^{add} - \phi_\theta^{add})(X,Z)\Big\}\Big].
\end{aligned}$$

In order to calculate this expression for $\theta = \theta_0$, we make use of the expansions given in (7.4) and (7.5) and of Lemma 7.1. We will develop i.i.d. expansions for the terms involving v_{1x}^α, v_2, w_{1x}^α and w_2. The calculations for v_{1z}^α and w_{1z}^α are similar to those for v_{1x}^α and w_{1x}^α.

Note that it follows from the proof of Lemma 4.1 that the estimation of V has an impact on the expression of the terms v_2 and w_2 (namely the terms involving ∇_v), but not on the terms v_{1x}^α, v_{1z}^α, w_{1x}^α and w_{1z}^α. Hence, if we want to show that the estimation of V has no effect on the expansion of $\Delta(\theta_0, s_0^{add})[\widehat{s}_0^{add} - s_0^{add}]$, we should show that the terms involving ∇_v (or more generally the whole term $n^{-1}\sum_{i=1}^n v_2(\xi_i)$ and $n^{-1}\sum_{i=1}^n w_2(\xi_i)$) cancel out.

We start with w_{1x}^α. The terms that contribute to w_{1x}^α are those involving $\big(\dot{\widehat{\phi}}_0^{add} - \dot\phi_0^{add}\big)(X,Z)$ and $\big(\dot{\widehat{f}}_{\epsilon(\theta_0)} - \dot f_{\epsilon(\theta_0)}\big)(\epsilon(\theta_0))$. More precisely, from the i.i.d. representations of these expressions, we get

$$n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)E\Big[-\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}k_{1h}(X_\alpha - X_{\alpha i}) + \frac{\frac{\partial}{\partial\epsilon}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big|X_{\alpha i}\Big]$$

$$= n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)E\Big[-\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}f_{X_\alpha|\epsilon(\theta_0)}(X_{\alpha i}|\epsilon(\theta_0)) + \frac{\frac{\partial}{\partial\epsilon}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big|X_{\alpha i}\Big]$$

$$= n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)E\Big[\frac{\partial}{\partial\epsilon}\Big(\frac{f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big)\Big|X_{\alpha i}\Big]$$

$$= n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)E\Big[\frac{\partial}{\partial\epsilon}f_{X_\alpha|\epsilon(\theta_0)}(X_{\alpha i}|\epsilon(\theta_0))\Big|X_{\alpha i}\Big]$$

$$= n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} w_{1x}^{\alpha}(\xi_i)L_x^{\alpha}(X_{\alpha i}). \tag{7.8}$$

Note that the terms in this sum have mean zero, since $E[w_{1x}^{\alpha}(\xi)|X_\alpha] = 0$.

We now consider the terms involving v_{1x}^{α}. Note that

$$\frac{d}{d\theta}f_{\epsilon(\theta)}(\epsilon(\theta)) = f'_{\epsilon(\theta)}(\epsilon(\theta))\big[\dot{\Lambda}_\theta(Y) - \dot{\phi}_\theta^{add}(X,Z)\big] + \dot{f}_{\epsilon(\theta)}(\epsilon(\theta))$$

and that

$$\frac{d}{d\theta}f_{\epsilon(\theta),X_\alpha}(\epsilon(\theta), X_\alpha) = \frac{\partial}{\partial\epsilon}f_{\epsilon(\theta),X_\alpha}(\epsilon(\theta), X_\alpha)\big[\dot{\Lambda}_\theta(Y) - \dot{\phi}_\theta^{add}(X,Z)\big] + \dot{f}_{\epsilon(\theta),X_\alpha}(\epsilon(\theta), X_\alpha).$$

The terms that involve v_{1x}^{α} can hence be written as

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)E\Big[&\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}k_{1h}(X_\alpha - X_{\alpha i})\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0)) \\ &-\frac{\frac{\partial}{\partial\epsilon}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f^2_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0)) - \frac{\frac{d}{d\theta}f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}k_{1h}(X_\alpha - X_{\alpha i}) \\ &+\frac{\frac{\partial^2}{\partial\epsilon^2}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}\epsilon(\theta_0) + \frac{\frac{\partial}{\partial\epsilon}\dot{f}_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big] \\ = n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)E\Big[&-\frac{\partial}{\partial\epsilon}\Big(\frac{\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big)f_{X_\alpha|\epsilon(\theta_0)}(X_{\alpha i}|\epsilon(\theta_0)) \\ &-\frac{\frac{\partial}{\partial\epsilon}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f^2_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0)) + \frac{\frac{d}{d\theta}\frac{\partial}{\partial\epsilon}f_{\epsilon(\theta_0),X_\alpha}(\epsilon(\theta_0), X_{\alpha i})}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big] \\ = n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_x} v_{1x}^{\alpha}(\xi_i)&K_x^{\alpha}(X_{\alpha i}). \end{aligned} \tag{7.9}$$

Again, note that the above expression has mean zero since $E[v_{1x}^{\alpha}(\xi)|X_\alpha] = 0$.

We now turn to the calculation of the expressions involving w_2, which are given by

$$n^{-1}\sum_{i=1}^{n} w_2(\xi_i)E\Big[-\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}+\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big]=0. \tag{7.10}$$

Finally, we consider the terms involving v_2, and those are given by

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n} v_2(\xi_i)E\Big[\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f^2_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0))-\frac{f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f^2_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}f_{\epsilon(\theta_0)}(\epsilon(\theta_0))\\&\qquad-\frac{\frac{d}{d\theta}f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}+\frac{f''_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\frac{d}{d\theta}\epsilon(\theta_0)+\frac{\dot f'_{\epsilon(\theta_0)}(\epsilon(\theta_0))}{f_{\epsilon(\theta_0)}(\epsilon(\theta_0))}\Big]\\&=0.\end{aligned} \tag{7.11}$$

It now suffices to combine (7.8), (7.9), (7.10) and (7.11). □

References

Ai, C. and Chen, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, **170**, 442-457.

Bickel, P.J. and Doksum, K. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.*, **76**, 296-311.

Blundell, R. and Powell, J. (2003). Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econometrics*, ed. M. Dewatripont, L.P. Hansen, and S.J. Turnovsky. Cambridge: Cambridge Univ. Press.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J. Royal Statist. Soc. - Series B*, **26**, 211-252.

Carrasco, M., Florens, J.P and Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In *Handbook of Econometrics*, **6**, eds. Heckman, J.J, Leamer, E.E.. North-Holland, Amsterdam.

Chen, X., Hahn, J. and Liao, Z. (2012). Asymptotic efficiency of semiparametric two-step GMM. Cowles Foundation Discussion Paper n. 1880.

Chen, X., Linton, O.B. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, **71**, 1591-1608.

Cheng, G. (2011). A note on bootstrap moment consistency for semiparametric M-estimation. Technical report.

Cheng, G. and Huang, J.Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Annals of Statistics*, **38**, 2884–2915.

Cheng, G. and Kosorok, M.R. (2008). General frequentist properties of the posterior profile distribution. *Annals of Statistics*, **36**, 1819–1853.

Cheng, G. and Pillai, N. (2012). Semiparametric model based bootstrap. *Working paper*.

Chiappori, P.-A., Komunjer, I. and Kristensen, D. (2010). Nonparametric identification and estimation of transformation models. Technical report.

Ekeland, I., Heckman, J.J. and Nesheim, L. (2004). Identification and estimation of hedonic models. *The Journal of Political Economy*, **112**, S60-S109.

Feve, F. and Florens, J.P. (2010). The practice of nonparametric estimation by solving inverse problems: the example of transformation models. *Econometrics Journal*, **13**, S1-S27.

Florens, J.P. and Sokullu, S. (2012). Nonparametric estimation of semiparametric transformation models. Technical report.

Hahn, J. and Ridder, G. (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, **81**, 315-340.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926-1947.

Heckman, J.J., Matzkin, R.L. and Nesheim, L. (2005). Estimation and simulations of hedonic models. In *Frontiers in Applied General Equilibrium*, ed. T. Kehoe, T. Srinivasan and J. Whalley. Cambridge University Press, Cambridge.

Heckman, J.J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, **52**, 271-320.

Horowitz, J.L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103-137.

Horowitz, J.L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, **69**, 499-513.

Imbens, G., and Newey, W. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77**, 1481-1512.

Jacho-Chavez, D., Lewbel, A. and Linton, O. (2010). Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics*, **156**, 392-407.

Keifer, N.M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, **26**, 646-679.

Linton, O.B. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression using marginal integration. *Biometrika*, **82**, 93-100.

Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation on a semiparametric transformation model. *Ann. Statist.*, **36**, 686-718.

Mammen, E., Linton, O.B. and Nielsen, J.P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.*, **27**, 1443-1490.

Mammen, E. and Park, B.U. (2005). Bandwidth selection for smooth backfitting in additive models. *Ann. Statist.*, **33**, 1260-1294.

Mammen, E., Rothe, C. and Schienle, M. (2012a). Nonparametric regression with nonparametrically generated regressors. *Ann. Statist.*, **40**, 1132-1170.

Mammen, E., Rothe, C. and Schienle, M. (2012b). Semiparametric estimation with generated covariates. Technical report.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica,* **71** , 1339-1375.

Newey, W.K. (1994). Kernel estimation of partial means. *Econometric Theory,* **10**, 233-253.

Newey, W.K., Powell, J.L. and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equation models. *Econometrica*, **67**, 565-603.

Nielsen, G.G., Gill, D., Andersen, P.K. and Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, **19**, 25-43.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, **57**, 1027-1057.

Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **57**, 1403-1430.

Sherman, R. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics*, **22**, 439-459.

Sokullu, S. (2011). Nonparametric analysis of two-sided markets. Technical report.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Wooldridge, J. (2008). *Introductory Econometrics: A Modern Approach.* South-Western College Pub.