TSE working paper series

11-243

Research Group: *Econometrics and Statistics*

May 13, 2011

Semiparametric transformation model with endogeneity: a control function approach

ANNE VANHEMS AND INGRID VAN KEILEGOM

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

Toulouse School of Economics

Semiparametric transformation model with endogeneity: a control function approach

Anne VANHEMS *   Ingrid VAN KEILEGOM §

May 13, 2011

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 small simulation study shows that the estimator behaves well in practice.

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

*Universite 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.
Research supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy), and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650.

1 Introduction

Suppose one is interested in modeling the relationship between the demand function of a certain good, the price of this good, and other variables like salary, education, ... Regressing the demand function on the price and the other variables via a nonparametric separable model will lead to skewed and possibly heteroscedastic errors, since the demand function is by definition positive. A possible way out is therefore to suitably transform the demand variable via some monotone *transformation* before applying nonparametric regression techniques. Another common issue in this context, is that some explanatory variables, e.g. the price of the good of interest, might be *endogenous* (due to measurement errors or important variables that are not included in the model). The objective of this paper is to develop a sound statistical theory for regression problems, in which one is confronted at the same time to an unknown transformation of the response and a problem of endogeneity in the explanatory variables.

Transformation models lie at the heart of many problems in structural econometrics. They take the general form:

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

where Y is a scalar dependent variable, (X, Z) is a vector of observed explanatory variables and ϵ is an unobserved random variable. The sub-sector X can be endogenous, that is correlated with U, whereas Z contains only exogenous variables. In a general setting, the functions Λ and ϕ can be either parametric or nonparametric.

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 (see the recent illustration in the two-sided market context by 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). Moreover, as stressed in Linton, Sperlich and Van Keilegom (2008), transformations have also been used to aid interpretability, to stabilize the variance of the error, to obtain errors that look more or less like normal errors, as well as to improve statistical performance. Often one prefers working with a parametric transformation, since they are easier to interpret than nonparametric ones. Well known examples of families of parametric transformations include the family of power transformations proposed by Box and Cox (1964), and the Bickel and Doksum (1981) class of transformations.

From a theoretical point of view, various papers have studied model (1.1) under different sets of assumptions. Horowitz (1996) considers a parametric function ϕ with fully exogenous explanatory variables whereas in the same context Linton, Sperlich and Van Keilegom (2008) impose a parametric form for Λ. In a still fully exogenous setting but with nonparametric forms for Λ and ϕ, we refer to Horowitz (2001), Jacho-Chavez, Lewbel and Linton (2008) or Ekeland, Heckman and Nesheim (2004) with an application to hedonic models.

More recently, papers have also studied the case where some explanatory variables X are endogenous. The issue of endogeneity is very crucial in econometrics and statistics and can arise as a result of e.g. omitted variables, autoregression with autocorrelated errors, or sample selection errors. In the specific setting of transformation models defined by equation (1.1), recent papers handling endogeneity are Florens and Sokullu (2011), Feve and Florens (2010) and Chiappori, Komunjer and Kristensen (2010). The first two papers consider a semiparametric form for the function ϕ and identify and estimate the model using an instrument W and 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). 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.

Our work stands in the line of Linton, Sperlich and Van Keilegom (2008) with a parametric transformation operator and a nonparametric function ϕ, and in addition some endogenous variables X. We prove identification of the structure $(\Lambda, \phi, F_\epsilon)$ using a control function approach, as in Imbens and Newey (2009). Indeed, as stressed in Matzkin (2003), transformation models can be viewed as particular cases of nonseparable models with $Y = \Lambda^{-1}(\phi(X, Z) + \epsilon)$ and results for nonseparable models with endogeneity can apply here. As in Linton, Sperlich and Van Keilegom (2008), we use a profile likelihood technique to estimate the parametric transformation, and with an additive structure for the function ϕ, prove the asymptotic normality with $\sqrt{n}$ rate of convergence. Some simulations in the end confirm the validity of our method.

The paper is organized as follows. In the next section we discuss the conditions under which our model is identified. Section 3 is devoted to the estimation part and the asymptotic results are stated in Section 4. The finite sample study is presented in Section 5, whereas some general conclusions are given in Section 6. Finally, the proofs are collected in the Appendix.

2 The model

2.1 Definitions and notations

Consider the following semiparametric transformation model:

$$\Lambda_\theta(Y) = \phi(X, Z) + \epsilon, \tag{2.2}$$

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 $(\Lambda_\theta, \phi, F_\epsilon)$, estimate θ and ϕ given a sample of observations and do inference on these estimators.

The approach we adopt to identify model (2.2) is based on a control function. The control function methodology has been studied in particular in Newey, Powell and Vella (1999), Blundell and Powell (2003), or Imbens and Newey (2009). As Imbens and Newey (2009) recall for the general setting of nonseparable models, a control variable is any observable or estimable variable V satisfying the following condition:

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

Since our model (2.2) is a particular case of a nonseparable model (as pointed out by Matzkin 2003) we also impose Assumption $(A.1)$ for our control function V.

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

$$X = \psi(Z, W) + V, \tag{2.3}$$

where W is a vector of instrumental variables taking values in $\mathbb{R}^{d_w}$ such that $\mathrm{E}(V|Z, W) = 0$.

A second option would be to consider a second nonseparable equation and a single endogenous variable X defined by:

$$X = \psi(Z, W, \eta), \tag{2.4}$$

where ψ is strictly monotone in η. Then $V = F_{X|Z,W}$ is a uniformly distributed variable satisfying Assumption $(A.1)$ under the following conditions: *(i)* (ϵ, η) *and* (Z, W) *are independent and (ii)* η *is a continuously distributed scalar with CDF that is strictly increasing 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).

At last a natural extension to equation (2.4) when X is multidimensional, consists in considering the set of one dimensional independent equations:

$$\begin{cases} X_1 &= \psi_1(Z, W, V_1) \\ & \vdots \\ X_{d_x} &= \psi_{d_x}(Z, W, V_{d_x}), \end{cases} \tag{2.5}$$

and $V = (V_1, ..., V_{d_x})$.

In the next subsection, identification of (2.2) will be studied under the general characterization $(A.1)$ without specifying any particular form for V. Next, for the estimation and the asymptotic results, we will restrict ourselves to a control variable V defined using model (2.4).

2.2 Identification

We now address the identification issue of the structure $(\Lambda_\theta, \phi, F_\epsilon)$ from equation (2.2). In this section, we just assume there exists a variable V satisfying Assumption $(A.1)$, and V can be defined by any of the three equations (2.3), (2.4) or (2.5). Moreover, our result will prove identification for the full nonparametric structure $(\Lambda, \phi, F_\epsilon)$ and therefore, in this section, we omit the index θ for the operator Λ.
Hereafter, we make the following additional assumptions:

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

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

(A.4) Let $R_{X,Z}$ be the compact support of (X, Z). Then, for a.e. $(x, z) \in R_{X,Z}$, the density $f_{\epsilon|X=x,Z=z}$ 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) $E(\Lambda(Y)) = 1$, $\Lambda(0) = 0$, $\mathrm{E}(\epsilon) = 0$.

Assumption $(A.2)$ is the support condition introduced in Imbens and Newey (2009) that, combined with Assumption $(A.1)$, allows to identify ϕ. Assumptions $(A.3)$, $(A.4)$ and (A.5) give standard regularity conditions on the operator Λ, the function ϕ and the conditional density $f_{Y|X=x,Z=z}$. At last, assumption $(A.6)$ gives some normalization conditions in order to identify Λ. Our result is based on the standard 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, under Assumption $(A.2)$, we can integrate over the marginal distribution of V and apply iterative expectation to obtain:

$$F_{Y|X,Z}(y,x,z) = \int F_{Y|X,Z,V}(y,x,z,v)F_V(dv) = F_\epsilon(\Lambda(y) - \phi(x,z)). \tag{2.6}$$

Theorem 2.1. *Under Assumptions $(A.1)-(A.6)$, the structure $(\Lambda,\phi,F_\epsilon)$ is identified.*

The proof is given in the Appendix.

Remark 2.1.

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 Theorem 2.1 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:*

$$\mathrm{E}\left(\Lambda_{\theta_0}(Y)|X=x,Z=z,V=v\right) = \phi(x,z) + \lambda(v),$$

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 Estimation

In order to facilitate the interpretation, we prefer to work with parametric transformations of the response variable, 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. Denote by θ_0 and ϕ_0 the true unknown finite and infinite dimensional parameters of model (2.2). We impose the following additive structure on $\phi_0(x, z)$:

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

with $\mathrm{E}[\phi_{x0}(X)] = 0$ and $\mathrm{E}[\phi_{z0}^{\alpha}(Z_\alpha)] = 0$ for $\alpha = 1, \ldots, d_z$. Combined with Assumption $(A.1)$, it implies that

$$\begin{aligned} m_0(x,z,v) &:= \mathrm{E}\left(\Lambda_{\theta_0}(Y)|X = x, Z = z, V = v\right) \\ &= \phi_0(x,z) + \lambda(v) \\ &= c + \phi_{x0}(x) + \sum_{\alpha=1}^{d_z} \phi_{z0}^{\alpha}(z_\alpha) + \lambda(v), \end{aligned} \tag{3.7}$$

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

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

and $c = \mathrm{E}[\Lambda_{\theta_0}(Y)] = 1$. Moreover, from equation (2.6) we obtain:

$$f_{Y|X,Z}(y|x,z) = \int f_{Y|X,Z,V}(y|x,z,v)dF_V(v) = f_\epsilon\left(\Lambda_0(y) - \phi_0(x,z)\right).\ \Lambda_0'(y), \tag{3.8}$$

where f_ϵ, $f_{Y|X,Z}$ and $f_{Y|X,Z,V}$ are the probability density functions of ϵ and of Y given (X, Z) and (X, Z, V), respectively.

Consider a randomly drawn i.i.d. sample (X_i, Y_i, Z_i, W_i), $i = 1, \ldots, n$ from the random vector (X, Y, Z, W). Then, the log-likelihood function is given by:

$$\sum_{i=1}^{n} \Big\{ \log[f_\epsilon\left(\Lambda_0(Y_i) - \phi_0(X_i, Z_i)\right)] + \log[\Lambda_0'(Y_i)] \Big\}. \tag{3.9}$$

The idea is now to estimate θ be replacing all unknown quantities in the above log-likelihood by nonparametric estimators (depending on the unknown θ), and to maximize the so-obtained expression with respect to θ, following the same profile likelihood ideas as in Linton, Sperlich and Van Keilegom (2008). First of all, we need to estimate nonparametrically the control variable V. In what follows, we focus on the nonseparable model defined in equation (2.4) to characterize the control variable $V = F_{X|Z,W}$ and its nonparametric estimator:

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

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 $d = d_z$ or d_w, k_1 is a univariate kernel function, and h is a bandwidth converging to zero when n tends to infinity. For simplifying the presentation, we work with the same bandwidth for all variables.[1]

The next step is to estimate the function $\phi_\theta(x, z)$ for a fixed, arbitrary value of θ using marginal integration techniques (see e.g. Newey 1994, Linton and Nielsen 1995). See also Mammen, Rothe and Schienle (2010), who do inference for a marginal integration estimator in an additive model without transformation of the response, but with estimated covariates. The difference with their paper is that we need to deal with the uniformity in θ and that we need to be more precise regarding the behavior of the remainder term (see the proofs in the Appendix for more details). Note at last that the function $\phi_\theta(x, z)$ could have been estimated using smooth backfitting techniques (see Mammen, Linton and Nielsen 1999). We briefly comment on this at the end of section 4.

To do so, we first estimate the function $m_\theta(x, z, v) = \mathrm{E}[\Lambda_\theta(Y)|X = x, Z = z, V = v]$ by using a nonparametric kernel estimator based on $(X_i, Z_i, \widehat{V}_i, Y_i)$ $(i = 1, \ldots, n)$:

$$\begin{aligned}\widehat{m}_\theta(x, z, v) &= \widehat{\mathrm{E}}\left[\Lambda_\theta(Y)|X = x, Z = z, \widehat{V} = v\right] \\ &= \frac{\sum_{i=1}^n \Lambda_\theta(Y_i) k_1(\frac{x - X_i}{h}) K(\frac{z - Z_i}{h}) k_1(\frac{v - \widehat{V}_i}{h})}{\sum_{i=1}^n k_1(\frac{x - X_i}{h}) K(\frac{z - Z_i}{h}) k_1(\frac{v - \widehat{V}_i}{h})}.\end{aligned}$$

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

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

where $\widehat{c}_\theta = n^{-1}\sum_{i=1}^n \Lambda_\theta(Y_i)$. The nonparametric estimator of $\phi_\theta^{add}(x, z) := c_\theta + \phi_{x\theta}(x) +$

[1] When instead of working under model (2.4) we impose the separable model (2.3), then the error V_i can be estimated by

$$\widehat{V}_i = X_i - \widehat{\psi}(Z_i, W_i).$$

where $\widehat{\psi}$ is e.g. the Nadaraya-Watson estimator

$$\widehat{\psi}(z, w) = \frac{\sum_{j=1}^n X_j K(\frac{z - Z_j}{h}) K(\frac{w - W_j}{h})}{\sum_{j=1}^n K(\frac{z - Z_j}{h}) K(\frac{w - W_j}{h})}.$$

$\sum_{\alpha=1}^{d_z} \phi_{z\theta}^{\alpha}(z_\alpha)$ is now given by:

$$\widehat{\phi}_\theta^{add}(x,z) = \widehat{c}_\theta + \widehat{\phi}_{x\theta}(x) + \sum_{\alpha=1}^{d_z} \widehat{\phi}_{z\theta}^{\alpha}(z_\alpha). \tag{3.10}$$

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 $\theta = \theta_0$.

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

$$\widehat{f}_{\epsilon(\theta)}(e) = \frac{1}{ng} \sum_{i=1}^{n} k_2 \left(\frac{e - \widehat{\epsilon}_i(\theta)}{g} \right) \tag{3.11}$$

where $\widehat{\epsilon}_i(\theta) = \Lambda_\theta(Y_i) - \widehat{\phi}_\theta^{add}(X_i, Z_i)$, 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 (3.9):

$$\widehat{\theta} = \arg\max_\theta \sum_{i=1}^{n} \Big\{ \log[\widehat{f}_{\epsilon(\theta)}(\Lambda_\theta(Y_i) - \widehat{\phi}_\theta^{add}(X_i, Z_i))] + \log[(\Lambda_\theta'(Y_i))] \Big\}. \tag{3.12}$$

Once θ is estimated we can re-estimate the regression function $\phi_0(x,y)$, this time using $\widehat{\theta}$ instead of an arbitrary value of θ. This gives

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

for any x and z.

4 Asymptotic results

Some additional notations need to be introduced. The joint density of (X,Z,V) is denoted by $f_{X,Z,V}$ and its support by $R_{X,Z,V}$. Similar notations are used for the joint density and the support of any other random vector. Let $\dot{\phi}_\theta(x,z) = \frac{\partial}{\partial\theta}\phi_\theta(x,z)$ and similarly for the partial derivative with respect to θ of any other function. For any θ let $F_{\epsilon(\theta)}(y) = \Pr(\epsilon(\theta) \leq y)$, where $\epsilon(\theta) = \Lambda_\theta(Y) - \phi_\theta^{add}(X,Z)$. We use the notation $\|\cdot\|$ to denote the Euclidean norm, and for any $\ell \geq 1$ we let $\frac{\partial}{\partial e_\ell}$ denote the derivative with respect to the ℓth argument.

Let $s_\theta = (\phi_\theta^{add}, \dot{\phi}_\theta^{add}, f_{\epsilon(\theta)}, f_{\epsilon(\theta)}', \dot{f}_{\epsilon(\theta)})$ and $s_0 = s_{\theta_0}$, and define

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

Then $n^{-1}\sum_{i=1}^{n} M(\theta, s_\theta, X_i, Z_i, Y_i)$ is the derivative of the log-likelihood with respect to θ. Moreover, let $G(\theta, s_\theta) = \mathrm{E}\{M(\theta, s_\theta, X, Z, Y)\}$,

$$S = \mathrm{Var}\Big\{M(\theta_0, s_0, X, Z, Y)\Big\} \qquad \text{and} \qquad \Gamma = \frac{\partial}{\partial\theta} G(\theta, s_\theta)\Big|_{\theta=\theta_0}.$$

The following regularity conditions are required:

(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 $q_1 > d_z/2 + 2$ and $q_1 > (d_z + d_w)/2$.

(C.2) $nh^{d_z+2} \to \infty$, $nh^{d_z+d_w} \to \infty$, $nh^{2q_1} \to 0$, $ng^6(\log g^{-1})^{-2} \to \infty$ 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 $\frac{\partial m_\theta}{\partial v}(x, z, v)$ 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) $f_{ZW}(z, w)$ and $F_{X|ZW}(x|z, w)$ 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 $f_{ZW}(z, w)$ 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

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

for all θ in Θ and all $0 \leq k + l \leq 3$.

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

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

for all $0 \leq k + l \leq 2$.

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

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

Moreover, the matrix Γ is of full rank.

The following lemma gives an i.i.d. representation of the estimator $\widehat{\phi}_\theta^{add}(x,z)$, 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.5). Then,*

$$
\begin{aligned}
&\widehat{\phi}_\theta^{add}(x,z) - \phi_\theta^{add}(x,z) \\
&= n^{-1}\sum_{i=1}^n \Big(k_{1h}(x-X_i)\Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big] f_{X|Z}^{-1}(x|Z_i) \\
&\qquad + \sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})\Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big] f_{Z_\alpha|X,Z_{(-\alpha)}}^{-1}(z_\alpha|X_i,Z_{(-\alpha)i}) \\
&\qquad + \mathrm{E}\Big[\Big\{\frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_i,V) + \sum_{\alpha=1}^{d_z}\frac{\partial m_\theta}{\partial e_{d_z+2}}(F_{X|ZW}^{-1}(V|Z_i,W_i),z_\alpha,Z_{(-\alpha)i},V)\Big\} \\
&\qquad\qquad \times\{I(V_i\le V) - V\}\Big| Z_i,V_i,W_i\Big] \\
&\qquad + \Big[m_\theta(x,Z_i,V_i) + \sum_{\alpha=1}^{d_z} m_\theta(X_i,z_\alpha,Z_{(-\alpha)i},V_i) - d_z\Lambda_\theta(Y_i) - \phi_\theta^{add}(x,z)\Big]\Big) \\
&\quad + o_P(n^{-1/2}),
\end{aligned}
$$

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.8). Then,*

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

where

$$\Omega = \Gamma^{-1}S(\Gamma^T)^{-1}.$$

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

Corollary 4.1. *Assume (A.1)-(A.6) and (C.1)–(C.8). Then, for any $(x,z)\in R_{X,Z}$,*

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

where

$$
\begin{aligned}
\Sigma &= \int k_1^2(u)du\ \ f_X(x)Var\Big\{\Big[\Lambda_{\theta_0}(Y) - m_{\theta_0}(X,Z,V)\Big] f_{X|Z}^{-1}(x|Z)\Big| X=x\Big\} \\
&\quad + \int k_1^2(u)du\ \ \sum_{\alpha=1}^{d_z} f_{Z_\alpha}(z_\alpha)Var\Big\{\Big[\Lambda_{\theta_0}(Y) - m_{\theta_0}(X,Z,V)\Big] f_{Z_\alpha|X,Z_{(-\alpha)}}^{-1}(z_\alpha|X,Z_{(-\alpha)})\Big| Z_\alpha = z_\alpha\Big\}.
\end{aligned}
$$

Remark 4.1

1. Note that the asymptotic variance of $\widehat{\theta}$ in Theorem 4.1 equals the variance of the estimator of θ_0 that is based on the true unknown values of the nuisance functions $\phi_0^{add}, \dot{\phi}_0^{add}, f_{\epsilon(\theta_0)}, f'_{\epsilon(\theta_0)}$ and $\dot{f}_{\epsilon(\theta_0)}$. The estimation of these unknown functions does not show up in the asymptotic variance, which is a very nice feature of our method ! Hence, the problem of estimating θ_0 is asymptotically a parametric problem.

2. 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). For the smooth backfitting, 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:
$$m_\theta^{add}(x, z, v) = \text{argmin}_{m \in \mathcal{M}_{add}} \int \Big[m_\theta(x, z, v) - m(x, z, v)\Big]^2 dF_{X,Z,V}(x, z, v),$$
where
$$\mathcal{M}_{add} = \Big\{m : m(x, z, v) = m_x(x) + \sum_{\alpha=1}^{d_z} m_{z_\alpha}(z_\alpha) + m_v(v) \text{ for some } m_x, m_{z_1}, \ldots, m_{z_{d_z}}, m_v\Big\}.$$
We expect that the estimator $\widehat{\theta}$ is semiparametrically efficient in this case.

3. Theorem 4.1 is an extension of the work of Linton, Sperlich and Van Keilegom (2008), who showed the asymptotic normality of $\widehat{\theta}$ when all explanatory variables are exogenous. The variance of their estimator looks very similar to ours (see their Theorem 4.1), except that the definition of $\phi_\theta^{add}(x, z)$ (denoted by $m_\theta(x)$ in their paper) is intrinsically different, since they are in the exogenous case. Note that 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 not). However, apart from this difference in interpretation of these two functions, the variances are exactly the same. In particular, note that the estimation of the control variable V vanishes asymptotically, i.e. the estimator $\widehat{\theta}$ behaves asymptotically as if the variables $V_1, \ldots . V_n$ would be observed !

4. Although the asymptotic variance of $\widehat{\theta}$ has a simple structure and does not depend on the estimators of the nuisance functions, nor on the estimation of the control variable V, its estimation in practice might be cumbersome, since it involves the estimation of

the density $f_{\epsilon(\theta_0)}$ and of its derivatives $f'_{\epsilon(\theta_0)}$ and $\dot{f}_{\epsilon(\theta_0)}$. A bootstrap procedure might therefore be a useful alternative. We refer to Chen, Linton and Van Keilegom (2003), who propose a bootstrap procedure for general semiparametric estimation problems. They justify the use of the ordinary bootstrap under certain high-level conditions, which need to be verified for our particular model. We refer to their paper for more details.

5. Note that the asymptotic distribution of $\widehat{\phi}(x,z)$ in Corollary 4.1 is the same as that of $\widehat{\phi}_0(x,z)$, i.e. the asymptotic distribution is as if the parameters θ_0 were known.

5 Finite sample study

We consider the following data generating process:

$$\Lambda_\theta(Y) = b_0X^2 + b_1 + \epsilon,$$

where Λ_θ is the Box-Cox transformation, that is $\Lambda_\theta(y) = \frac{y^\theta - 1}{\theta}$ $(\theta \neq 0)$ and $\Lambda_\theta(y) = \log(y)$ $(\theta = 0)$. ϵ is drawn from $N(0, \sigma_e^2)$ but restricted to $[-b_1; +\infty[$. In this setting, we omit the exogenous variable Z. The variable X is generated from the following generating process:

$$X = a_0W + a_1W^2 + a_2\epsilon + a_3 + U,$$

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. We present here the results for the following model where $b_0 = 0.3, b_1 = 20, a_0 = -2, a_1 = -3, a_2 = 2,$ $\sigma_w^2 = \sigma_e^2 = 2$ and $\sigma_u^2 = 0.2$. The parameter θ_0 is set equal to 2 and 3. Note that $\Lambda_\theta(Y)$ is by construction always positive in our simulation.
We estimated θ by a grid search on $[0.1, 5]$ with a step length of 0.1. We use the gaussian kernel and apply the cross-validation method to select the bandwidth parameters. The Monte-Carlo study has been performed with $B = 500$ replications for two different sample sizes $n = 200$ and $n = 300$. We provide each time the mean and standard deviation of $\hat{\theta}$ and the mean squared error (mse hereafter).

The results are summarized in Table 1 and show that the method works well for reasonable sample sizes. In particular, we note that as the sample size increases, both the bias and the variance decrease.

n	200			300		
θ_0	mean($\hat{\theta}$)	sd($\hat{\theta}$)	mse($\hat{\theta}$)	mean($\hat{\theta}$)	sd($\hat{\theta}$)	mse($\hat{\theta}$)
2	2.16	0.41	0.19	1.90	0.25	0.07
3	3.21	0.52	0.31	2.85	0.31	0.11

Table 1: Simulation results for different sample sizes with bandwidth chosen by cross-validation.

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 explanatory variables. Using a control function approach, we prove identification of the structure $(\Lambda, \phi, F_\epsilon)$. As in Linton, Sperlich and Van Keilegom (2008), we use profile likelihood techniques to estimate the parametric transformation, and by imposing an additive structure on the function ϕ, we prove the asymptotic normality of the proposed estimator with $\sqrt{n}$ rate of convergence. Some simulations confirm the validity of our method.

7 Appendix: Proofs

Proof of Theorem 2.1. 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 follows from the proof of Chiappori, Komunjer and Kristensen (2010). Under the regularity assumptions $(A.3)$ and $(A.4)$, we can differentiate equation (2.6) with respect to y and x (or x_1, the first coordinate of x) to obtain:

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

Let $A = \{(x,z) \in R_{X,Z} : \frac{\partial}{\partial x_1} F_{Y|X,Z}(y,x,z) \neq 0 \text{ for every } y \in R_Y\}$. Under Assumptions $(A.4)$ and $(A.5)$, the set A has a nonempty interior. Then, for any point $(x,z) \in R_{X,Z}$ 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 $s(y,x,z) = \frac{\frac{\partial}{\partial y} F_{Y|X,Z}(y,x,z)}{\frac{\partial}{\partial x_1} F_{Y|X,Z}(y,x,z)}$. 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 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 (2.6). This finishes the proof. □

Proof of Lemma 4.1. Consider the following operator:

$$A(u,v)(x,z) = \frac{1}{n}\sum_{i=1}^n u(x,Z_i,v(X_i,Z_i,W_i)) + \sum_{\alpha=1}^{d_z}\left[\frac{1}{n}\sum_{i=1}^n u(X_i,z_\alpha,Z_{(-\alpha)i},v(X_i,Z_i,W_i))\right]$$
$$-d_z\frac{1}{n}\sum_{i=1}^n \Lambda_\theta(Y_i) - \phi_\theta^{add}(x,z).$$

defined on the Hilbert space of square integrable and twice continuously differentiable functions associated with the L_2-norm $\|\cdot\|_{L_2}$. A is continuously differentiable with respect to the L_2-norm and we can compute its Fréchet derivative $dA(u,v)[h_1,h_2]$ of A in the direction $[h_1,h_2]$:

$$\begin{aligned}
&dA(u,v)[h_1,h_2](x,z) \\
&= \frac{1}{n}\sum_{i=1}^n h_1(x,Z_i,v(X_i,Z_i,W_i)) + \sum_{\alpha=1}^{d_z}\left[\frac{1}{n}\sum_{i=1}^n h_1(X_i,z_\alpha,Z_{(-\alpha)i},v(X_i,Z_i,W_i))\right] \\
&\quad +\frac{1}{n}\sum_{i=1}^n \frac{\partial\, u}{\partial e_{d_z+2}}(x,Z_i,v(X_i,Z_i,W_i)).h_2(X_i,Z_i,W_i) \\
&\quad +\sum_{\alpha=1}^{d_z}\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial\, u}{\partial e_{d_z+2}}(X_i,z_\alpha,Z_{(-\alpha)i},v(X_i,Z_i,W_i)).h_2(X_i,Z_i,W_i)\right].
\end{aligned}$$

Moreover, since A is twice continuously differentiable,

$$A(u+h_1, v+h_2) = A(u,v) + dA(u,v)[h_1,h_2] + O_P\left(\max(\|h_1\|_{L_2}^2, \|h_2\|_{L_2}^2)\right).$$

Then, computed for $(u,v) = (m_\theta, F_{X|Z,W})$ and $(h_1,h_2) = (\widehat{m}_\theta - m_\theta, \widehat{F}_{X|Z,W} - F_{X|Z,W})$ we obtain:

$$\begin{aligned}
&\widehat{\phi}_\theta(x,z) - \phi_\theta(x,z) \\
&= A(u+h_1, v+h_2)(x,z) \\
&= A(u,v)(x,z) + dA(u,v)[h_1,h_2](x,z) + O_P\left(\max(\|h_1\|_{L_2}^2, \|h_2\|_{L_2}^2)\right) \\
&= \frac{1}{n}\sum_{i=1}^n \left[m_\theta(x, Z_i, V_i) + \sum_{\alpha=1}^{d_z} m_\theta(X_i, z_\alpha, Z_{(-\alpha)i}, V_i) - d_z \Lambda_\theta(Y_i) - \phi_\theta^{add}(x,z) \right] \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left(\widehat{m}_\theta - m_\theta\right)(x, Z_i, F_{X|Z,W}(X_i|Z_i,W_i)) \\
&\quad + \sum_{\alpha=1}^{d_z}\left[\frac{1}{n}\sum_{i=1}^n \left(\widehat{m}_\theta - m_\theta\right)(X_i, z_\alpha, Z_{(-\alpha)i}, F_{X|Z,W}(X_i|Z_i,W_i))\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^n \frac{\partial\, m_\theta}{\partial e_{d_z+2}}(x, Z_i, F_{X|Z,W}(X_i|Z_i,W_i)).\left(\widehat{F}_{X|Z,W} - F_{X|Z,W}\right)(X_i|Z_i,W_i) \\
&\quad + \sum_{\alpha=1}^{d_z}\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial\, m_\theta}{\partial e_{d_z+2}}(X_i, z_\alpha, Z_{(-\alpha)i}, F_{X|Z,W}(X_i|Z_i,W_i)).\left(\widehat{F}_{X|Z,W} - F_{X|Z,W}\right)(X_i|Z_i,W_i)\right] \\
&\quad + O_P\left(\max(\|\widehat{m}_\theta - m_\theta\|_{L_2}^2, \|\widehat{F}_{X|Z,W} - F_{X|Z,W}\|_{L_2}^2)\right) \\
&= R_1 + R_2 + R_3 + R_4 + R_5 + R_6.
\end{aligned}$$

We will prove below that $R_6 = o_P(n^{-1/2})$, R_2 and R_3 have $(nh)^{-1/2}$-rate of convergence, whereas R_1, R_4 and R_5 converge at $n^{-1/2}$-rate. We start with R_2. Write

$$(\widehat{m}_\theta - m_\theta)(x, Z_i, V_i) = (\widehat{m}_\theta - \widetilde{m}_\theta)(x, Z_i, V_i) + (\widetilde{m}_\theta - m_\theta)(x, Z_i, V_i),$$

where

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

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 $n^{-1}\sum_{i=1}^n (\widetilde{m}_\theta - m_\theta)(x, Z_i, V_i)$ 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

$$n^{-1}\sum_{i=1}^n \left[\Lambda_\theta(Y_i) - m_\theta(X_i, Z_i, V_i)\right] k_{1h}(x - X_i) f_{X|Z}^{-1}(x|Z_i) + o_P(n^{-1/2}).$$

Now consider

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

where $\widehat{N}_{ij} = \Lambda_\theta(Y_j)k_{1h}(x-X_j)K_h(Z_i-Z_j)k_{1h}(V_i-\widehat{V}_j)$, $\widehat{D}_{ij} = k_{1h}(x-X_j)K_h(Z_i-Z_j)k_{1h}(V_i-\widehat{V}_j)$, and similarly for $\widetilde{N}_{ij}$ and $\widetilde{D}_{ij}$. In analogy with these notations, we define $N_i = \mathrm{E}(\Lambda_\theta(Y)|x, Z_i, V_i)f_{XZV}(x, Z_i, V_i)$ and $D_i = f_{XZV}(x, Z_i, V_i)$. Next, write

$$\begin{aligned}
& n^{-1}\sum_{i=1}^n(\widehat{m}_\theta - \widetilde{m}_\theta)(x, Z_i, V_i) \\
&= n^{-1}\sum_{i=1}^n\sum_{j=1}^n(\widehat{N}_{ij} - \widetilde{N}_{ij})\frac{1}{\sum_{j=1}^n \widehat{D}_{ij}} + n^{-1}\sum_{i=1}^n\sum_{j=1}^n \widetilde{N}_{ij}\Big(\frac{1}{\sum_{j=1}^n \widehat{D}_{ij}} - \frac{1}{\sum_{j=1}^n \widetilde{D}_{ij}}\Big) \\
&= \Big[n^{-2}\sum_{i=1}^n\sum_{j=1}^n(\widehat{N}_{ij} - \widetilde{N}_{ij})\frac{1}{D_i} - n^{-2}\sum_{i=1}^n\sum_{j=1}^n(\widehat{D}_{ij} - \widetilde{D}_{ij})\frac{N_i}{D_i^2}\Big](1 + o_P(1)) \\
&= \Big[n^{-2}\sum_{i=1}^n D_i^{-1}\sum_{j=1}^n\Big\{N'_{ij} - D'_{ij}\frac{N_i}{D_i}\Big\}(V_j - \widehat{V}_j)\Big](1 + o_P(1)) \\
&= -\Big[n^{-3}\sum_{i=1}^n D_i^{-1}\sum_{j=1}^n\Big\{N'_{ij} - D'_{ij}\frac{N_i}{D_i}\Big\}\sum_{k=1}^n S_{jk}\Big](1 + o_P(1)), \qquad (7.13)
\end{aligned}$$

where $N'_{ij} = \Lambda_\theta(Y_j)k_{1h}(x - X_j)K_h(Z_i - Z_j)h^{-1}k'_{1h}(V_i - V_j)$, $D'_{ij} = k_{1h}(x - X_j)K_h(Z_i - Z_j)h^{-1}k'_{1h}(V_i - V_j)$, and

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

Note that (7.13) is a V-statistic of order three (ignoring the factor $1 + o_P(1)$), with kernel depending on n. Write the V-statistic as

$$\mathcal{V}_n = n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n p_n(T_i, T_j, T_k),$$

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

$$p_n(T_i, T_j, T_k) = -D_i^{-1}\Big\{N'_{ij} - D'_{ij}\frac{N_i}{D_i}\Big\}S_{jk}.$$

Since a V-statistic can be written as a U-statistic plus negligible terms, we can apply the generalization of the Hoefding decomposition of a U-statistic to the case where the kernel depends on n (see e.g. Lemma 3.1 in Powell, Stock and Stoker (1989)), which leads to

$$\begin{aligned}\mathcal{V}_n = n^{-1}\sum_{i=1}^n \mathrm{E}[p_n(T_i,T,T')|T_i] + n^{-1}\sum_{j=1}^n \mathrm{E}[p_n(T,T_j,T')|T_j] \\ +n^{-1}\sum_{k=1}^n \mathrm{E}[p_n(T,T',T_k)|T_k] - 2\mathrm{E}[p_n(T',T',T'')] + o_P(n^{-1/2}),\end{aligned}$$

where T,T',T'',T_j $(j=1,\ldots,n)$ are i.i.d. Now, it can be easily seen that $\mathrm{E}(S_{jk}|T_j) = O_P(h^{q_1})$ uniformly in j and that $\mathrm{E}(N'_{ij} - D'_{ij}\frac{N_i}{D_i}|T_i) = O_P(h^{q_1})$ uniformly in i. Hence, it follows that

$$\mathrm{E}[p_n(T_i,T,T')|T_i] = O_P(h^{2q_1}) = o_P(n^{-1/2}),$$

since $nh^{4q_1} \to 0$,

$$\mathrm{E}[p_n(T,T_j,T')|T_j] = O_P(h^{q_1}(nh^{d_z+4})^{-1/2} + h^{2q_1}) = o_P(n^{-1/2})$$

provided $q_1 > d_z/2+2$, and

$$\mathrm{E}[p_n(T,T',T_k)|T_k] = O_P(h^{q_1}(nh^{d_z+d_w})^{-1/2} + h^{2q_1}) = o_P(n^{-1/2})$$

since $q_1 > (d_z+d_w)/2$. Hence, we also have that $\mathrm{E}[p_n(T,T',T'')] = o(n^{-1/2})$. This shows that $\mathcal{V}_n = o_P(n^{-1/2})$, and so R_2 equals

$$R_2 = n^{-1}\sum_{i=1}^n \Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big]k_{1h}(x-X_i)f^{-1}_{X|Z}(x|Z_i) + o_P(n^{-1/2}).$$

In a similar way we can show that

$$\begin{aligned}R_3 = n^{-1}\sum_{i=1}^n\sum_{\alpha=1}^{d_z} \Big[\Lambda_\theta(Y_i) - m_\theta(X_i,Z_i,V_i)\Big]k_{1h}(z_\alpha - Z_{\alpha i})f^{-1}_{Z_\alpha|X,Z_{(-\alpha)}}(z_\alpha|X_i,Z_{(-\alpha)i}) \\ +o_P(n^{-1/2}).\end{aligned}$$

Next, consider R_4. Using again the Hoefding decomposition for U-statistics with kernel depending on n (but this time for U-statistics of order 2), we obtain:

$$
\begin{aligned}
R_4 &= n^{-1}\sum_{i=1}^n \frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_i,V_i)(\widehat{V}_i - V_i) \\
&= n^{-2}\sum_{i=1}^n\sum_{k=1}^n \frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_i,V_i)S_{ik} + o_P(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_i,V_i)f_{ZW}^{-1}(Z_i,W_i)\Big\{I(X\le X_i) - F_{X|ZW}(X_i|Z_i,W_i)\Big\} \\
&\qquad \times K_h(Z_i - Z)K_h(W_i - W)\Big|T_i\Big] \\
&\quad + n^{-1}\sum_{k=1}^n \mathrm{E}\Big[\frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z,V)f_{ZW}^{-1}(Z,W)\Big\{I(X_k\le X) - F_{X|ZW}(X|Z,W)\Big\} \\
&\qquad \times K_h(Z - Z_k)K_h(W - W_k)\Big|T_k\Big] - \mathrm{E}\Big[\frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_1,V_1)S_{12}\Big] + o_P(n^{-1/2}) \\
&= n^{-1}\sum_{k=1}^n \mathrm{E}\Big[\frac{\partial m_\theta}{\partial e_{d_z+2}}\Big(x,Z_k,F_{X|ZW}(X|Z_k,W_k)\Big)\Big\{I(X_k\le X) - F_{X|ZW}(X|Z_k,W_k)\Big\}\Big|T_k\Big] \\
&\qquad + o_P(n^{-1/2}) \\
&= n^{-1}\sum_{k=1}^n \mathrm{E}\Big[\frac{\partial m_\theta}{\partial e_{d_z+2}}(x,Z_k,V)\{I(V_k\le V) - V\}\Big|T_k\Big] + o_P(n^{-1/2}).
\end{aligned}
$$

provided $nh^{2q_1}\to 0$. Similarly, R_5 can be decomposed in a sum of i.i.d. terms plus a term of smaller order. Finally, $R_6 = O_P((nh^{d_z/2+1})^{-1}) + O_P((nh^{(d_z+d_w)/2})^{-1}) + O(h^{2q_1}) = o_P(n^{-1/2})$. This finishes the proof. □

Proof of Theorem 4.1. The proof is based on Theorem 4.1 in Linton, Sperlich and Van Keilegom (2008). In the latter paper the authors prove the asymptotic normality of $\widehat{\theta}$ when no endogeneity is present. 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 $\widehat{\phi}_\theta^{add}(x,z)$ (denoted by $\widehat{m}_\theta(x)$ in their paper) needs to satisfy. In can be easily seen that the proof of their Theorem 4.1 remains valid in our case, provided we can show that our estimator $\widehat{\phi}_\theta^{add}(x,z)$ satisfies their assumption A.8. The remaining assumptions A.1–A.7 are in fact given by our conditions (C.1)–(C.4) and (C.6)–(C.8). In what follows, we check this assumption in detail.

First, note that the i.i.d. representation for $\widehat{\phi}_0^{add}(x,z) - \phi_0^{add}(x,z)$ is given in Lemma 4.1. In a similar way, $\dot{\widehat{\phi}}_0^{add}(x,z) - \dot{\phi}_0^{add}(x,z)$ can also be decomposed in a sum of i.i.d. terms plus negligible terms of order $o_P(n^{-1/2})$. Next, define $\mathcal{M} = C_1^1(R_X) + \sum_{\alpha=1}^{d_z} C_1^1(R_{Z_\alpha})$, where $C_a^b(R)$ ($0<a<\infty$, $0<b\le 1$, $R\subset I\!R^k$ for some k) is the set of all continuous functions

$f : R \to I\!R$ for which

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

We equip the space $\mathcal{M}$ with the L_2-norm $\|\cdot\|_{L_2}$. It can be easily seen that $P(\widehat{\phi}_\theta^{add}, \dot{\widehat{\phi}}_\theta^{add} \in \mathcal{M}) \to 1$. Moreover, the covering number $N(\lambda, \mathcal{M}, \|\cdot\|_{L_2})$ satisfies $\log N(\lambda, \mathcal{M}, \|\cdot\|_{L_2}) \leq K\lambda^{-1}$, and hence

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

(see Corollary 2.7.2 in Van der Vaart and Wellner, 1996). Next, using Lemma 4.1 it is easy to show that $\sup_{\theta\in\Theta} \|\widehat{\phi}_\theta^{add} - \phi_\theta^{add}\|_{L_2} = O_P((nh^{(d_z+1)/2})^{-1/2}) = o_P(n^{-1/4})$, since $nh^{d_z+1} \to \infty$ (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 $\sup_{\theta\in\Theta} \|\dot{\widehat{\phi}}_\theta^{add} - \dot{\phi}_\theta^{add}\|_{L_2} = o_P(n^{-1/4})$. Finally, we need to prove that

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

for all θ such that $\|\theta - \theta_0\| = o(1)$. For this, note that (again using the extension of Lemma 4.1 to $\dot{\widehat{\phi}}_\theta^{add}(x,z) - \dot{\phi}_\theta^{add}(x,z)$) it suffices to control (for all i)

$$\left\| \dot{\Lambda}_\theta(Y_i) - \dot{m}_\theta(X_i, Z_i, V_i) - \dot{\Lambda}_{\theta_0}(Y_i) + \dot{m}_{\theta_0}(X_i, Z_i, V_i) \right\|,$$

and this is bounded by

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

which is of the required order. This finishes the proof of assumption A.8 in Linton, Sperlich and Van Keilegom (2008), and hence the result follows. □

Proof of Corollary 4.1. Write

$$\widehat{\phi}(x,z) - \phi_0(x,z) = \left[\widehat{\phi}_{\widehat{\theta}}^{add}(x,z) - \widehat{\phi}_{\theta_0}^{add}(x,z)\right] + \left[\widehat{\phi}_{\theta_0}^{add}(x,z) - \phi_{\theta_0}^{add}(x,z)\right]. \quad (7.14)$$

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

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

and hence the first term of (7.14) is $O_P(n^{-1/2}) = o_P((nh)^{-1/2})$ by Theorem 4.1. For the second term of (7.14) we apply Lemma 4.1, which yields that

$$
\begin{aligned}
&\widehat{\phi}^{add}_{\theta_0}(x,z) - \phi^{add}_{\theta_0}(x,z) \\
&= n^{-1}\sum_{i=1}^{n} k_{1h}(x-X_i)\Big[\Lambda_{\theta_0}(Y_i) - m_{\theta_0}(X_i,Z_i,V_i)\Big] f^{-1}_{X|Z}(x|Z_i) \\
&\quad + n^{-1}\sum_{i=1}^{n}\sum_{\alpha=1}^{d_z} k_{1h}(z_\alpha - Z_{\alpha i})\Big[\Lambda_{\theta_0}(Y_i) - m_{\theta_0}(X_i,Z_i,V_i)\Big] f^{-1}_{Z_\alpha|X,Z_{(-\alpha)}}(z_\alpha|X_i,Z_{(-\alpha)i}) \\
&\quad + o_P((nh)^{-1/2}).
\end{aligned}
$$

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

References

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

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. To appear in *Econometrics Journal*.

Florens, J.P. and Sokullu, S. (2011). Nonparametric estimation of semiparametric transformation models. Technical report.

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. (2008). Identification and nonparametric estimation of a transformation additively separable model. Technical report.

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. (2010). Nonparametric regression with nonparametrically 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.

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

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.