

Breunig, Christoph; Mammen, Enno; Simoni, Anna

**Working Paper**

## Nonparametric estimation in case of endogenous selection

SFB 649 Discussion Paper, No. 2015-050

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Breunig, Christoph; Mammen, Enno; Simoni, Anna (2015) : Nonparametric estimation in case of endogenous selection, SFB 649 Discussion Paper, No. 2015-050, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/146165>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

SFB 649 Discussion Paper 2015-050

# Nonparametric Estimation in case of Endogenous Selection

Christoph Breunig\*

Enno Mammen\*<sup>2</sup>

Anna Simoni\*<sup>3</sup>



\* Humboldt-Universität zu Berlin, Germany

\*<sup>2</sup> Heidelberg University, Germany and  
Higher School of Economics Moscow, Russian Federation

\*<sup>3</sup> CNRS and CREST, France

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

# Nonparametric Estimation in case of Endogenous Selection \*

CHRISTOPH BREUNIG \*

*Humboldt-Universität zu Berlin*

ENNO MAMMEN <sup>1</sup>

*Heidelberg University and*

*Higher School of Economics, Moscow*

ANNA SIMONI <sup>2</sup>

*CNRS and CREST*

November 10, 2015

This paper addresses the problem of estimation of a nonparametric regression function from selectively observed data when selection is endogenous. Our approach relies on independence between covariates and selection conditionally on potential outcomes. Endogeneity of regressors is also allowed for. In both cases, consistent two-step estimation procedures are proposed and their rates of convergence are derived. Also pointwise asymptotic distribution of the estimators is established. In addition, we propose a nonparametric specification test to check the validity of our independence assumption. Finite sample properties are illustrated in a Monte Carlo simulation study and an empirical illustration.

*Keywords:* Endogenous selection, instrumental variable, sieve minimum distance, regression estimation, convergence rate, asymptotic normality, hypothesis testing, inverse problem.

*JEL classification:* C14, C26

---

\*This work was supported by the DFG-SNF research group FOR916. The authors gratefully thank the Co-Editor Jianqing Fan, an Associate Editor, and two anonymous referees for their many constructive comments on the previous version of the paper. The authors are also grateful to Timothy Armstrong, Xiaohong Chen, Victor Chernozhukov, Elise Coudin, Laurent Davezies, Kirill Evdokimov, Xavier D'Haultfoeuille, Aureo de Paula, Christoph Rothe, Bernard Salanié and seminar participants at Bristol, Columbia University, CREST, Mannheim University, MIT and Yale University for useful comments and to the GIP team at SFB 884-Mannheim for providing data of the German Internet Panel. Christoph Breunig was supported by the DFG postdoctoral fellowship BR 4874/1-1. The author is also grateful for support and hospitality of the Cowles Foundation. Anna Simoni gratefully acknowledges financial support from ANR-13-BSH1-0004 (IPANEMA) and hospitality from University of Mannheim and Boston College.

\*Department of Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany, e-mail: [christoph.breunig@hu-berlin.de](mailto:christoph.breunig@hu-berlin.de)

<sup>1</sup>Institute for Applied Mathematics, Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany and Higher School of Economics, Pokrovskii Boulevard 11, 103012 Moscow, Russia, e-mail: [emammen@rumms.uni-mannheim.de](mailto:emammen@rumms.uni-mannheim.de)

<sup>2</sup>CREST, 15, Boulevard Gabriel Péri, 92240 Malakoff, France, e-mail: [simoni.anna@gmail.com](mailto:simoni.anna@gmail.com)

# 1. Introduction

This paper addresses the problem of estimation of a regression function from selectively observed data. To explain the problem at stake, consider a partially observed dependent variable  $Y^*$ , a vector of covariates  $X$  and a binary indicator  $\Delta$ . The econometrician observes a realization of  $\Delta$  and  $X$  for each individual in the random sample but only observes a realization of  $Y^*$  when  $\Delta = 1$ . In many applications it is important to learn about  $\mathbf{E}[Y^*|X]$  which, by the law of total expectation, can be written as

$$\mathbf{E}[Y^*|X] = \mathbf{E}[Y^*|X, \Delta = 1]\mathbf{P}(\Delta = 1|X) + \mathbf{E}[Y^*|X, \Delta = 0]\mathbf{P}(\Delta = 0|X).$$

The difficulty arises because the data available cannot identify  $\mathbf{E}[Y^*|X, \Delta = 0]$  nor  $\mathbf{E}[Y^*|X]$ . In this paper, we address this lack of identification by assuming independence between the regressors  $X$  and the selection mechanism  $\Delta$  conditionally on the selectively observed outcome  $Y^*$ . Relying on this assumption we propose a new methodology to consistently estimate the regression function  $\varphi(\cdot) \equiv \mathbf{E}[Y^*|X = \cdot]$ .

We also consider endogeneity of covariates by extending the nonparametric instrumental variable models of Newey and Powell [2003] and Darolles et al. [2011] to allow for selection. That is, we propose a method to estimate a structural function  $\psi$  which satisfies

$$Y^* = \psi(Z) + U$$

for some unobservables  $U$ , where  $Z$  is endogenous in the sense that  $\mathbf{E}[U|Z] \neq 0$  and an additional instrumental variable  $X$  is available such that  $\mathbf{E}[U|X] = 0$ . If the instrument  $X$  is independent of the selection given potential outcome  $Y^*$ , we show that  $\psi$  is identified and can be consistently estimated under commonly imposed assumptions.

Previous literature has proposed different solutions to overcome the problem of lack of identification of  $\mathbf{E}[Y^*|X]$ . One solution consists in assuming *missing-at-random* (MAR), namely, independence between the selection variable and the outcome conditional on the observed covariates, see Rubin [1976]. MAR implies  $\mathbf{E}[Y^*|X] = \mathbf{E}[Y^*|X, \Delta = 1] = \mathbf{E}[Y^*|X, \Delta = 0]$ . Unfortunately, the plausibility of this assumption may be questioned in many economic examples where missing observations arise due to self-selection, nonresponse or because counterfactual variables are unobservable (see the examples given in Heckman [1979]).

In his seminal work, Heckman [1974, 1979] relies on instruments that determine selection but not the outcome and proposes a consistent parametric estimation method. Point-identification comes from parametric restrictions. Ahn and Powell [1993] and Das et al. [2003] extend Heckman's approach to a semiparametric and nonparametric framework, respectively.

An alternative strategy relies on "identification at infinity", namely, on the fact that the selection problem becomes negligible for large values of the covariates. This strategy requires the existence of a covariate with a large support, see Chamberlain [1986]. Based on this idea Lewbel [2007] and D'Haultfoeuille and Maurel [2013] propose alternative identification strategies.

A completely different approach was proposed by Manski [1989] who poses, as the only restriction, a bound on the support of  $Y^*$  conditional on  $X$ . This implies a bound on  $\mathbf{E}[Y^*|X]$ . While such a weak restriction has the advantage of ensuring robust inference, only partial identification of  $\mathbf{E}[Y^*|X]$  can be achieved. Following Manski [1989], an extensive literature on bounds and partial identification in econometrics has flourished (see *e.g.* Chernozhukov

et al. [2013] and Tamer [2010] for a review).

In this paper, we solve the problem of endogenous selection by using a different type of instrumental strategy. We assume independence between selection  $\Delta$  and instruments  $X$ , conditional on the outcome  $Y^*$  (and possibly additional covariates), namely

$$\Delta \perp\!\!\!\perp X | Y^*. \tag{1.1}$$

This assumption is suitable when selection is driven by the outcome  $Y^*$ . For example, if  $Y^*$  denotes income and  $X$  expenditure then, typically in survey data, whether people report their income or not is primarily determined by the level of their income. Assumption (1.1) has been used in the previous literature on missing data (see *e.g.* Chen [2001], Tang et al. [2003] in the statistics literature and D’Haultfoeuille [2010], Davezies and D’Haultfoeuille [2013], Ramalho and Smith [2013] in the econometrics literature). This type of assumption is common in the finite mixtures literature (*e.g.* Henry et al. [2014]) as well as in the nonclassical measurement error literature (see *e.g.* Hu and Schennach [2008b]). Moreover, it is a particular case of Assumption (41) in Manski [1994]. Assumption (1.1) alone is not sufficient for nonparametric identification of the regression function  $\varphi$  or the selection probability  $\mathbb{P}(\Delta = 1|Y^*)$ . In this paper, however, we show that the function  $\varphi$  is identified if  $Y^*$  is complete for  $X$ . Moreover, Assumption (1.1) can be tested, see Theorem 2.4 in D’Haultfoeuille [2010].

Our paper builds on these results and goes beyond what has been proposed in the literature so far. We focus on  $\mathbb{E}[Y^*|X]$  which is point-identified despite the presence of endogenous selection. Then, we contribute to the literature in many directions. First, we propose a two-step estimator for  $\mathbb{E}[Y^*|X]$  that does not require  $\mathbb{P}(\Delta = 1|Y^*)$  to belong to a parametric family. Under our assumptions,  $\mathbb{P}(\Delta = 1|Y^*)$  is identified through a conditional moment restriction. Because  $\mathbb{P}(\Delta = 1|Y^*)$  is nonparametric, its estimation requires solving a linear inverse problem which is ill-posed in general. In the first step, we use a constraint sieve-type estimator for  $\mathbb{P}(\Delta = 1|Y^*)$  which has to account for the ill-posedness of the inverse problem (see *e.g.* Newey and Powell [2003], Ai and Chen [2003] for sieve estimators for inverse problems in econometrics). In the second step, we use this estimator to construct a plug-in series estimator for  $\mathbb{E}[Y^*|X]$ .

Second, we recover the rate of convergence of the estimator for the nonparametric regression function  $\mathbb{E}[Y^*|X]$  and show that it does not suffer from the ill-posedness of the underlying inverse problem. In contrast to the classical nonparametric rate of convergence, we get an additional bias due to estimation of the selection probability in the first step. Also Das et al. [2003] obtain an additional bias term in their convergence rate but which is due to estimation of a propensity score. Under additional smoothness conditions, this bias is asymptotically negligible and the usual nonparametric regression rate is obtained. The fact that the rate does not suffer from the ill-posedness of the estimation problem in the first step is because the second step of our estimation procedure performs a smoothing through integration with respect to the conditional distribution of  $Y^*$  given  $X$ .

We also study the case where regressors are potentially endogenous. In contrast to Das et al. [2003], who used a control function method to correct for endogeneity and selection, we rely on a nonparametric instrumental regression model where the dependent variable is selectively observed. In this model, we propose a two step estimator of the structural function and obtain a convergence rate that deviates from the usual ill-posed

inverse problem rate by an extra bias term.

Third, we establish asymptotic normality for our estimator of the regression function  $E[Y^*|X = \cdot]$  evaluated at some point in the support of  $X$ . We point out that, due to the nature of our estimation problem, both the rate of convergence and asymptotic normality are new and cannot be derived from the results in Newey [1997] and the related literature. Indeed, we derive an asymptotic variance formula that involves an additional part due to correction for sample selection in the first step.

Forth, we propose a testing procedure to test the identifying assumption (1.1). At the best of our knowledge a testing procedure for this assumption in the continuous- $(Y^*, X)$  case has not been developed yet in the literature. Some ideas about how to construct a test for assumption (1.1) are outlined in D’Haultfoeuille [2010] but without formal implementation. Our testing procedure is based on a different intuition.

The remainder of this paper is organized as follows. In Section 2, we present the setup and discuss identification. In Section 3, we present our two-step estimator for  $E[Y^*|X]$ , we give rates of convergence of the integrated squared error of our estimator and establish pointwise asymptotic normality of our estimator. Section 4 deals with possible endogenous covariates. Our testing procedure to test the instrumental variable restriction (1.1) is presented in Section 5. Its finite sample properties are investigate through a Monte Carlo experiments whose results are reported in Section 6. Section 7 presents an empirical application of our method to estimate the propensity to work in the German speaking population by using “German Internet Panel” data. All proofs are postponed to the appendix.

## 2. Identification

In this section, we provide assumptions under which the selection probability function  $\mathbb{P}(\Delta = 1|Y^* = \cdot)$  and the regression function  $E[Y^*|X = \cdot]$  are identified. We further motivate our estimation procedure.

### 2.1. Setup and Main Assumptions

Let  $(\Delta, Y^*, X^t)$  be a jointly distributed random vector where  $(Y^*, X^t)$  is a random vector which takes values in  $\mathbb{R}^{1+d_x}$  and  $\Delta$  is a random variable which takes values in  $\{0, 1\}$ . A realization of  $(\Delta, X^t)$  is observed for each individual in the random sample while a realization of the dependent variable  $Y^*$  is observed when  $\Delta = 1$  and missing when  $\Delta = 0$ . We write  $Y = \Delta Y^*$ .<sup>1</sup> We assume that the marginal distribution of  $Y^*$  (resp.  $X$ ) admits a probability density function  $p_Y$  (resp.  $p_X$ ) with respect to the Lebesgue measure. The following three assumptions are sufficient to identify the joint distribution of  $(\Delta, Y^*, X^t)$ .

**ASSUMPTION 1.** *It holds that*

$$\Delta \perp\!\!\!\perp X | Y^*.$$

---

<sup>1</sup>In our setting,  $Y^*$  is assumed to be a scalar. Our results would still hold if we extended this framework to allow for a  $p$ -dimensional vector  $Y^*$  of selectively observed variables. In this case  $\Delta = (\Delta^{(j)})_{1 \leq j \leq p}$  and the  $j$ -th component of  $Y^*$  would be observed when  $\Delta^{(j)} = 1$  and missing when  $\Delta^{(j)} = 0$ . This extension would require little modifications of our method but would burden the notation and the presentation. For this reason we do not consider it.

Assumption 1 states an exclusion restriction of the random vector  $X$  with respect to the selection variable  $\Delta$  given potential outcomes  $Y^*$ . The vector  $X$  is referred to as the vector of instruments. This assumption can be justified in many settings. An example is provided by measurement error models where  $Y^*$  is observed with error for some individuals. Then, for some error  $\varepsilon$ ,  $X = Y^* + \varepsilon$  can be interpreted as a proxy for  $Y^*$  and satisfies Assumption 1 if  $\varepsilon \perp\!\!\!\perp \Delta$ , see *e.g.* Chen et al. [2011]. Other examples are given by data with nonresponse. For instance, consider the case where  $Y^*$  is income and  $X$  is expenditure. It could be that people with high income are less likely to report it. Examples of such type of incomplete data sets are the French “Enquête Budget de famille” of INSEE or the British “Family expenditure Survey”. For further illustrations of Assumption 1 we refer to Ramalho and Smith [2013].

**ASSUMPTION 2.** For every function  $\phi$  that is bounded from below almost surely and satisfies  $\mathbf{E}|\phi(Y^*)| < \infty$  it holds that  $\mathbf{E}[\phi(Y^*)|X] = 0$  implies  $\phi(Y^*) = 0$ .<sup>2</sup>

Assumption 2 is weaker than  $L^1$ -completeness but stronger than bounded-completeness. Completeness conditions have been largely used in econometrics as identification assumptions, see *e.g.* Darolles et al. [2011], Newey and Powell [2003], Blundell et al. [2007], Hu and Schennach [2008a], D’Haultfoeuille [2011] and Hoderlein et al. [2012].

**ASSUMPTION 3.** It holds  $\mathbb{P}(\Delta = 1|Y^*) > 0$ .

This assumption can rule out a selection when it is a deterministic function of  $Y^*$  such as  $\mathbb{P}(\Delta = 1|Y^* = \cdot) = \mathbb{1}\{\cdot \geq c\}$  for some constant  $c$  belonging to the interval  $(\min(\text{Supp}(Y^*)), \infty)$  where  $\text{Supp}(Y^*)$  denotes the support of  $Y^*$ . Here  $\mathbb{1}$  denotes the indicator function. To understand Assumption 3, consider the example where  $\Delta = \xi(Y^*, \eta)$  for some function  $\xi(\cdot)$  and a random variable  $\eta$ . Then, Assumption 3 is verified if the distribution of  $\eta$  is such that the set  $\{\eta; \xi(y^*, \eta) = 1\}$  has positive probability for every  $y^*$  in the support of  $Y^*$  except, possibly, for a set of  $y^*$  with probability 0.

## 2.2. Identification and idea of the estimator

Our object of interest is  $\mathbf{E}[Y^*|X]$  while the selection probability  $\mathbb{P}(\Delta = 1|Y^*)$  is a nuisance (functional) parameter. However, knowledge of the latter allows us to identify and estimate  $\mathbf{E}[Y^*|X]$  in a way that we now explain. Let us introduce the function  $g(\cdot) \equiv 1/\mathbb{P}(\Delta = 1|Y^* = \cdot)$  for the inverse selection probability. Under Assumptions 1–3, the function  $g$  is identified through the conditional moment restriction

$$\mathbf{E}\left[\Delta g(Y^*)|X\right] = 1, \tag{2.1}$$

see Theorem 2.3 in D’Haultfoeuille [2010]. In the first step of our two-step procedure, we make use of (2.1) to estimate the inverse selection probability function  $g$ .

Since the function  $g$  is identified by equation (2.1), identification of the conditional expectation  $\mathbf{E}[Y^*|X]$  follows from

$$\mathbf{E}[Y^*|X] = \mathbf{E}[Y^* \mathbb{P}(\Delta = 1|Y^*)g(Y^*)|X] = \mathbf{E}\left[\mathbf{E}[Y^* \Delta g(Y^*)|Y^*]|X\right] = \mathbf{E}[Y^* \Delta g(Y^*)|X] = \mathbf{E}[Yg(Y)|X] \tag{2.2}$$

where the first equality follows from Assumption 3 and second to last equality follows from Assumption 1 and the fact that  $g(Y^*) = g(Y)$  whenever  $Y \equiv Y^* \Delta$  differs from zero. This

<sup>2</sup>Since conditional expectations are defined only up to equality a.s., all (in)equalities with conditional expectations and/or random variables are understood as (in)equalities a.s., even if we do not say so explicitly.



result shows that  $\mathbf{E}[Y^*|X]$  can be written as a weighted average of the observed  $Y$  where the weight is equal to the inverse selection probability function. We use equation (2.2) to construct an estimator for  $\mathbf{E}[Y^*|X]$  in the second step of our estimation procedure.

**REMARK 2.1 (Including additional covariates).** In empirical applications, only a subset of the covariates might be independent of selection given potential outcome. We can cover this case by slightly extending Assumption 1. More precisely, suppose that  $X = (X_1, X_2)$  and that Assumption 1 is modified as  $\Delta \perp\!\!\!\perp X_1|(Y^*, X_2)$  and hence,  $X_2$  can be correlated to  $\Delta$ . Under this assumption,  $\mathbf{E}[Y^*|X] = \mathbf{E}[Yg(Y, X_2)|X]$  where  $g(y, x) = 1/\mathbb{P}(\Delta = 1|Y^* = y, X_2 = x)$ . Moreover, if Assumptions 2 and 3 are modified with  $\phi(Y^*)$  replaced by  $\phi(Y^*, X_2)$  and  $\mathbb{P}(\Delta = 1|Y^*)$  replaced by  $\mathbb{P}(\Delta = 1|Y^*, X_2) > 0$ , respectively, then  $g$  is identified by  $\mathbf{E}[\Delta g(Y^*, X_2)|X] = 1$ .  $\square$

### 2.3. Notation

For a random vector  $V$  we use the corresponding calligraphic capital letter  $\mathcal{V}$  to denote its support. Let  $L_V^2 = \{\phi : \|\phi\|_V^2 \equiv \mathbf{E}|\phi(V)|^2 < \infty\}$  denote the space of square integrable functions of  $V$  with respect to the distribution of  $V$ . We denote by  $\langle \cdot, \cdot \rangle_V$  the inner product in  $L_V^2$  that induces  $\|\cdot\|_V^2$ . Moreover,  $\|\phi\|_\infty := \sup_{v \in \mathcal{V}} |\phi(v)|$  denotes the sup norm and  $\|\cdot\|$  is the usual Euclidean norm.

The regression function of interest is denoted by:

$$\varphi(\cdot) = \mathbf{E}[Y^*|X = \cdot]$$

but we may use both notations depending on the context. Let  $\{f_j\}_{j \geq 1}$  (resp.  $\{e_j\}_{j \geq 1}$ ) be a sequence of approximating functions in  $L_X^2$  (resp.  $L_Y^2$ ). Then, we denote by  $\underline{f}_{m_n}(X) = (f_1(X), \dots, f_{m_n}(X))^t$  (resp.  $\underline{e}_{k_n}(Y) = (e_1(Y), \dots, e_{k_n}(Y))^t$ ) a vector of functions which are used to approximate the conditional expectation  $\mathbf{E}[Y^*|X]$  (resp. the inverse selection probability  $g(Y)$ ) and by  $\mathbf{X}_{m_n} = (\underline{f}_{m_n}(X_1), \dots, \underline{f}_{m_n}(X_n))^t$  (resp.  $\mathbf{Y}_{k_n} = (\Delta_1 \underline{e}_{k_n}(Y_1), \dots, \Delta_n \underline{e}_{k_n}(Y_n))^t$ ) the  $n \times m_n$  (resp.  $n \times k_n$ ) matrix obtained by putting together the  $n$  vectors  $\underline{f}_{m_n}(X_i)$ ,  $i = 1, \dots, n$  (resp.  $\Delta_i \underline{e}_{k_n}(Y_i)$ ,  $i = 1, \dots, n$ , where  $\Delta_i \underline{e}_{k_n}(Y_i)$  denotes the product of  $\Delta_i$  and the vector  $\underline{e}_{k_n}(Y_i)$ ). We denote by  $\mathcal{F}_{m_n} = \{\phi(\cdot) = \sum_{j=1}^{m_n} \beta_j f_j(\cdot) : \beta \in \mathbb{R}^{m_n}\}$  the linear sieve space of dimension  $m_n < \infty$  that becomes dense in  $L_X^2$  as  $n$  tends to infinity. For a matrix  $A$  we denote by  $A^-$  its generalized inverse. For a function  $\phi$  defined on  $\mathcal{Y}$ , we denote by  $M_\phi : L_Y^2 \rightarrow L_Y^2$  the multiplication operator  $M_\phi \psi = \phi \psi$  which is bounded if  $\phi$  is bounded on  $\mathcal{Y}$ . Then  $(M_{id} \psi)(y) = y \psi(y)$  for all  $y \in \mathcal{Y}$  where  $id$  denotes the identity function.

## 3. Nonparametric Regression with Sample Selection

In this section, we consider estimation of the regression function  $\varphi$ . The first step estimation procedure for the inverse selection probability  $g$  is based on constrained sieve minimum distance. In the second step, we use a plug-in series estimator of the conditional expectation  $\varphi(\cdot) = \mathbf{E}[Yg(Y)|X = \cdot]$ . We derive the rate of convergence in mean square error and the asymptotic distribution of our estimator of  $\varphi$ .



### 3.1. The Estimator and its Rate of Convergence

Let  $\mathcal{G}$  denote a separable Banach space which is endowed with a norm  $\|\cdot\|_{\mathcal{G}}$  and let  $g \in \mathcal{G}$ . Equation (2.1) can be written in a more compact form by using the following notation. Let  $T : \mathcal{G} \rightarrow L_X^2$  be the linear operator  $T\phi \equiv \mathbf{E}[\Delta\phi(Y)|X]$ . Thereby, equation (2.1) can be equivalently written as the operator equation

$$Tg = 1 \quad (3.1)$$

where the function  $g$  is identified under Assumptions 1–3. Note that since  $\mathbf{E}[g(Y)] = \mathbf{E}[g(Y^*\Delta)] = 1 + g(0)\mathbb{P}(\Delta = 0)$  the function  $g$  belongs to  $L_Y^1$  if  $g(0) < \infty$ . In the following we assume without loss of generality that  $g(0) < \infty$  and hence  $\mathcal{G} = L_Y^1$ .

For every  $\phi \in L_Y^1$ , denote  $\chi(\cdot, \phi) = \mathbf{E}[\Delta\phi(Y) - 1|X = \cdot]$ . The least squares estimator of  $\chi(\cdot, \phi)$  is given by

$$\widehat{\chi}_n(\cdot, \phi) = \underline{f}_{m_n}(\cdot)^t (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \sum_{i=1}^n (\Delta_i \phi(Y_i) - 1) \underline{f}_{m_n}(X_i) \quad (3.2)$$

for some integer  $m_n$  which increases with the sample size  $n$ . Under conditions given below,  $\mathbf{X}_{m_n}^t \mathbf{X}_{m_n}$  will be nonsingular with probability approaching one and hence its generalized inverse will be the standard inverse. We now introduce some assumptions.

**ASSUMPTION 4.** (i) We observe a sample  $((\Delta_1, X_1, Y_1), \dots, (\Delta_n, X_n, Y_n))$  of independent and identical distributed (iid.) copies of  $(\Delta, X, Y)$  where  $Y = \Delta Y^*$ . (ii) There exists a constant  $C > 0$  and a sequence of positive integers  $(m_n)_{n \geq 1}$  satisfying  $\sup_{x \in \mathcal{X}} \|f_{m_n}(x)\|^2 \leq Cm_n$  such that  $m_n^2/n = o(1)$ . (iii) The smallest eigenvalue of  $\mathbf{E}[f_m(X)f_m(X)^t]$  is bounded away from zero uniformly in  $m$ . (iv) Let  $\phi \in L_X^2$  and there is  $F_{m_n}\phi \in \mathcal{F}_{m_n}$  such that  $\|F_{m_n}\phi - \phi\|_X = O(m_n^{-\alpha/d_x})$  for some constant  $\alpha > 0$ .

Assumption 4 (ii) – (iii) restricts the magnitude of the approximating functions  $\{f_j\}_{j \geq 1}$  and impose nonsingularity of their second moment matrix. It is a standard assumption for series estimators (cf. e.g. Assumption 2 in Newey [1997]). Assumption 4 (ii) holds for instance for polynomial splines, Fourier series and wavelet bases but rules out orthogonal polynomials and power series sieves. Assumption 4 (iv) determines the sieve approximation error which in turn characterizes the bias of the estimated regression function  $\phi$ . In the following, we consider the linear sieve space  $\mathcal{G}_n = \{\phi(\cdot) = \sum_{j=1}^{k_n} \beta_j e_j(\cdot) : \beta \in \mathbb{R}^{k_n}\}$  of dimension  $k_n < \infty$  that becomes dense in the function space  $L_Y^1$  as  $n$  tends to infinity. We propose the following sieve minimum distance estimator

$$\widehat{g}_n \equiv \arg \min_{\{\phi \in \mathcal{G}_n : \phi(\cdot) \geq 1\}} \sum_{i=1}^n \widehat{\chi}_n^2(X_i, \phi). \quad (3.3)$$

The constraint  $\phi(\cdot) \geq 1$  imposed on the sieve space  $\mathcal{G}_n$  ensures that the estimated conditional probability of observing  $Y^*$  belongs to the unit interval. This estimator of  $g$  corresponds to the penalized sieve minimum distance estimator suggested by Chen and Pouzo [2012].

If no constraint is imposed then the sieve estimator  $\widehat{g}_n$  has an explicit solution given by

$$\widehat{g}_n(\cdot) \equiv e_{k_n}(\cdot)^t \widehat{\beta}_{k_n} \quad \text{and} \quad \widehat{\beta}_k = (\mathbf{Y}_k^t \mathbf{X}_m (\mathbf{X}_m^t \mathbf{X}_m)^{-1} \mathbf{X}_m^t \mathbf{Y}_k)^{-1} \mathbf{Y}_k^t \mathbf{X}_m (\mathbf{X}_m^t \mathbf{X}_m)^{-1} \mathbf{X}_m^t \mathbf{1}_n \quad (3.4)$$

where  $\mathbf{1}_n$  is a  $n$ -dimensional vector of ones and for some integer  $k_n \leq m_n$  which increases with the sample size  $n$ .

The second step of our estimation procedure consists in using the estimator  $\widehat{g}_n$  in (3.3) to construct an estimator for  $\varphi$ . Let  $\mathbf{G}_n = (Y_1 \widehat{g}_n(Y_1), \dots, Y_n \widehat{g}_n(Y_n))$ . Then, our estimator of the nonparametric regression function  $\varphi(\cdot)$  is given by

$$\widehat{\varphi}_n(\cdot) \equiv \underline{f}_{m_n}(\cdot)^t \widehat{\beta}_{m_n} \quad \text{where} \quad \widehat{\beta}_{m_n} = (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \mathbf{X}_{m_n}^t \mathbf{G}_n. \quad (3.5)$$

In the following, the sequence  $(\mathcal{R}_n)_{n \geq 1}$  denotes the rate of convergence of the estimator  $\widehat{g}_n$  w.r.t. to the norm  $\|\cdot\|_{\mathcal{G}}$ . The next assumption is used to recover the rate of convergence of  $\widehat{\varphi}_n$ .

**ASSUMPTION 5.** (i) For all  $\phi \in \mathcal{G}_n$  there exists  $F_{m_n} T\phi \in \mathcal{F}_{m_n}$  such that  $\|F_{m_n} T\phi - T\phi\|_{\infty} = O(m_n^{-\alpha/d_x})$ . (ii) There exists a sequence of positive integers  $(\xi_n)_{n \geq 1}$  satisfying  $\sup_{y \in \mathcal{Y}} \|e_{k_n}(y)\|^2 \leq \xi_n^2$  such that  $k_n \xi_n^2/n = o(1)$ . (iii) It holds  $k_n^2 \mathcal{R}_n = O(1)$ . (iv)  $\|TM_{id}\phi\|_X / \|T\phi\|_X$  is bounded uniformly over all  $\phi \in \mathcal{G}$  with  $\|T\phi\|_X \neq 0$ .

Assumption 5 (i) holds true, for example, for splines or power series if the family of functions  $\{T\phi : \phi \in \mathcal{G}_n\}$  contains only functions which are at least  $\alpha$ -times continuously differentiable (see also Assumption 3 of Blundell et al. [2007]). Assumption 5 (ii) is satisfied with  $\xi_n = \sqrt{k_n}$  when the approximating functions are for instance B-splines, Fourier series and wavelet bases. For Legendre polynomials this assumption is satisfied with  $\xi_n = k_n$ . Assumption 5 (iii) is a mild restriction on the rate of convergence of  $\widehat{g}_n$  which we illustrate below. Instead of a bound on  $\mathcal{Y}$ , Assumption 5 (iv) restricts the size of the multiplication operator  $M_{id}$  in the norm induced by  $T$ . Otherwise stated, it requires that the norms of the operators  $TM_{id}$  and  $T$  are equivalent. Assumption 5 (iv) is satisfied for instance under an additional link condition, like Assumption 6 (iii) below, if the basis functions coincide with Legendre polynomials or cardinal B-splines (see Example 3.1 below).

Estimation of  $g$  requires to "solve" a conditional moment restriction that is different from (2.2), namely  $\mathbf{E}[\Delta g(Y^*)|X] = 1$ . From Blundell et al. [2007] and Chen and Pouzo [2012] we obtain the rate of convergence of  $\|T(\widehat{g}_n - g)\|_X$ . Their result, however, is not enough to obtain the rate of convergence of  $\widehat{\varphi}_n$  as our case requires to determine the rate of  $\|\widehat{T}M_{id}\widehat{g}_n - TM_{id}g\|_X$  where  $\widehat{T}$  is a series least square estimator of  $T$ . Thereby, additional arguments are needed to obtain the rate of  $\|\widehat{\varphi}_n - \varphi\|_X$  given in the following theorem. In the following, for any  $\phi$  in  $\mathcal{G}$  let  $E_{k_n}\phi \in \mathcal{G}_n$  be such that  $\|E_{k_n}\phi - \phi\|_{\mathcal{G}} = o(1)$ .

**THEOREM 3.1.** *Let Assumptions 1 – 5 hold true. Then we have*

$$\|\widehat{\varphi}_n - \varphi\|_X^2 = O_p\left(\max\left(m_n^{-2\alpha/d_x}, \frac{m_n}{n}, \|T(E_{k_n}g - g)\|_X^2\right)\right).$$

As we see from Theorem 3.1, the rate of convergence of our estimator  $\widehat{\varphi}_n$  depends on both parameters  $k_n$  and  $m_n$  which correspond to the first and second estimation step, respectively. In addition to the usual nonparametric rate we obtain an additional bias term  $\|T(E_{k_n}g - g)\|_X^2$  which is due to the sieve approximation of the inverse selection probability function  $g$ . An additional bias occurs also in the convergence rate for estimating regression functions in Theorem 4.1 of Das et al. [2003]. In their case, however, the additional bias arises from nonparametric estimation of a propensity score. From Theorem 3.1 we see that  $\widehat{\varphi}_n$  attains the optimal nonparametric rate of convergence if  $\|T(E_{k_n}g - g)\|_X = O(m_n^{-\alpha/d_x})$ . If the inverse selection probability  $g$  is sufficiently smooth in the sense that  $\|E_{k_n}g - g\|_{\mathcal{G}} = O(m_n^{-\alpha/d_x})$  then, by Jensen's inequality, the optimal nonparametric rate is obtained. In the following,

we provide a more general treatment by incorporating mapping properties of the operator  $T$  in the Hilbert space case where  $\mathcal{G} = L_Y^2$ .

**ASSUMPTION 6.** (i)  $\mathbf{E}[Y^2] < \infty$  and  $\mathbf{E}[g^2(Y)] < \infty$ . (ii) Assume  $\|E_{k_n}g - g\|_Y = O(k_n^{-\beta})$  for some constant  $\beta > 0$ . (iii) There exists a sequence of non-increasing positive real numbers  $(\tau_j)_{j \geq 1}$  such that  $\|T\phi\|_X^2 \geq c \sum_{j=1}^{\infty} \tau_j \langle \phi, e_j \rangle_Y^2$  and  $\|T\phi\|_X^2 \leq C \sum_{j=1}^{\infty} \tau_j \langle \phi, e_j \rangle_Y^2$  for some constant  $C > 0$  and all  $\phi \in L_Y^2$ . (iv) The largest eigenvalue of  $(\tau_j^{1/2} \tau_l^{-1/2} \langle M_{id} e_j, e_l \rangle_Y)_{j,l \geq 1}$  is bounded away from infinity.

Assumption 6 (i) ensures that  $g$  belongs to the Hilbert space  $L_Y^2$  and  $Y$  has finite second moment while (ii) determines the sieve approximation error for estimating the function  $g$ . Assumption 6 (iii) is also known as a link condition and commonly used in the analysis of inverse problems (see, e.g. Chen and Reiß [2011]).

**COROLLARY 3.2.** Let Assumptions 1–4, 5 (i)–(iii), and 6 hold true. Then we have

$$\|\widehat{\varphi}_n - \varphi\|_X^2 = O_p\left(\max\left(m_n^{-2\alpha/d_x}, \frac{m_n}{n}, \tau_{k_n} k_n^{-2\beta}\right)\right). \quad (3.6)$$

**REMARK 3.1.** In the mildly ill-posed case where  $\tau_j \sim j^{-2t}$ ,  $t \geq 0$ , let  $k_n \sim n^{1/(2t+2\beta+1)}$  and  $m_n \sim n^{d_x/(2\alpha+d_x)}$ . Hence,

$$\|\widehat{\varphi}_n - \varphi\|_X^2 = O_p\left(\max\left(n^{-(2t+2\beta)/(2t+2\beta+1)}, n^{-2\alpha/(2\alpha+d_x)}\right)\right)$$

which is  $O_p(n^{-2\alpha/(2\alpha+d_x)})$  if  $\alpha \leq d_x(t+\beta)$ . In case of trigonometric basis functions, the operator  $T$  acts like integrating  $(t)$ -times which then automatically implies  $\alpha = d_x(t+\beta)$  (cf. page 12 in Breunig and Johannes [2011]). Also it holds  $\|\widehat{g}_n - g\|_Y^2 = O_p(n^{-2\beta/(2t+2\beta+1)})$  and in particular we have  $k_n^2 \mathcal{R}_n = O(n^{(2-2\beta)/(2t+2\beta+1)}) = o(1)$  if  $\beta > 1$ . In the severely ill-posed case where  $\tau_j \sim \exp(-j^{2t})$ ,  $t > 0$ , we let  $k_n \sim (\log n)^{1/2t}$  and obtain

$$\|\widehat{\varphi}_n - \varphi\|_X^2 = O_p\left(n^{-2\alpha/(2\alpha+d_x)}\right).$$

In this case,  $\|\widehat{g}_n - g\|_Y^2 = O_p(\log(n)^{-2\beta/t})$  and in particular,  $k_n^2 \mathcal{R}_n = O(\log(n)^{(2-2\beta)/t}) = o(1)$  again if  $\beta > 1$ . We conclude that under mild conditions on the smoothness of  $\varphi$  the optimal nonparametric rate of regression in mean squared error is obtained.  $\square$

The following example illustrates that Assumption 6 (iv) is automatically satisfied when  $\{e_j\}_{j \geq 1}$  coincides with Legendre polynomials. Similarly to this example, Assumption 6 (iv) can be verified for cardinal B-splines (cf. De Boor [1978]).

**EXAMPLE 3.1.** Assume that  $\mathcal{Y}$  is contained in  $[-1, 1]$  and consider the Hilbert space  $L_Y^2 = L_{[-1,1]}^2$  endowed with the usual norm. Let  $\{e_j\}_{j \geq 1}$  be the Legendre polynomials. That is, for  $y \in [-1, 1]$  we define  $e_1 \equiv 1$ ,  $e_2(y) = y$ , and

$$j e_{j+1}(y) = (2j-1)y e_j(y) - (j-1)e_{j-1}(y)$$

for  $j \geq 3$ . This recursion formula is equivalent to

$$y e_j(y) = \frac{j e_{j+1}(y) + (j-1)e_{j-1}(y)}{2j-1}$$

which implies

$$\begin{pmatrix} \langle M_{id}e_1, e_1 \rangle_Y & \tau_1^{1/2}\tau_2^{-1/2}\langle M_{id}e_1, e_2 \rangle_Y & \dots \\ \tau_2^{1/2}\tau_1^{-1/2}\langle M_{id}e_2, e_1 \rangle_Y & \langle M_{id}e_2, e_2 \rangle_Y & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 0 & \frac{2\tau_1^{1/2}}{3\tau_2^{1/2}} & 0 & 0 & \dots \\ \frac{2\tau_2^{1/2}}{3\tau_1^{1/2}} & 0 & \frac{4\tau_3^{1/2}}{15\tau_4^{1/2}} & 0 & \dots \\ 0 & \frac{4\tau_3^{1/2}}{15\tau_2^{1/2}} & 0 & \ddots & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The norm of the right hand side matrix is bounded by its Frobenius norm which is  $\sum_{j \geq 1} \frac{j^2}{(2j+1)^2} \frac{4}{(2(j-1)+1)^2}$  within a constant. This series is bounded and thereby Assumption 6 (iv) holds true for either the *mildly ill-posed case* or the *severely ill-posed case*.  $\square$

### 3.2. Pointwise Asymptotic Distribution of the Estimator

This subsection is about inference on the regression function  $\varphi$  evaluated at some point of the support of  $X$ . To establish the asymptotic distribution of our estimator we require the following additional assumptions. Let us introduce the matrices  $Q_n = \mathbf{E}[f_{m_n}(X)f_{m_n}(X)^t]$ ,  $\mathbf{T}_n = \mathbf{E}[\Delta f_{m_n}(X)e_{k_n}(Y)^t]$ , and  $\mathbf{T}_n^Y = \mathbf{E}[Y f_{m_n}(X)e_{k_n}(Y)^t]$ .

**ASSUMPTION 7.** (i) Let  $\mathbf{E}[|Yg(Y) - \varphi(X)|^4|X] \leq C$  and  $\text{Var}(Yg(Y)|X) \geq c$  for some constants  $c, C > 0$ . (ii) The matrix  $\mathbf{T}_n^t \mathbf{T}_n$  has full rank. (iii) The function  $g$  is uniformly bounded away from 1. (iv) There exist some constants  $c, C > 0$  such that  $c \leq l_n/\tau_{k_n} \leq C$ , where  $l_n$  is the minimum eigenvalue of  $\mathbf{T}_n^t Q_n^{-1} \mathbf{T}_n$ . (v) It holds  $\|E_{k_n}g - g\|_\infty = O(k_n^{-\beta+1/2})$  for  $\beta > 1/2$  and  $\|F_{m_n}\varphi - \varphi\|_\infty = O(m_n^{-\alpha/d_x+1/2})$  for  $\alpha > d_x/2$ .

A bounded fourth moment of the error was also assumed by Newey [1997] to establish asymptotic normality of series estimators in the regression context. The consistency result established in Theorem 3.1 together with Assumption 7 (iv) imply that the constraint is not binding asymptotically and hence the estimator  $\widehat{g}_n$  given in (3.4) coincides with the one in (3.3). To prove consistency of the asymptotic variance formula we require uniform convergence of the approximation biases at a certain rate. Examples of approximating functions that satisfy this condition are Fourier series, splines series and wavelet series (see also Belloni et al. [2015]).

For the asymptotic distribution result we introduce the asymptotic variance

$$\begin{aligned} \mathcal{V}_n(x) = & \text{Var}\left(f_{m_n}(x)^t Q_n^{-1} \left(f_{m_n}(X)(Yg(Y) - \varphi(X))\right)\right) \\ & + \text{Var}\left(f_{m_n}(x)^t Q_n^{-1} \mathbf{T}_n^Y \left(\mathbf{T}_n^t Q_n^{-1} \mathbf{T}_n\right)^{-1} \mathbf{T}_n^t Q_n^{-1} \left(f_{m_n}(X) - \mathbf{E}[f_{m_n}(X)]\right)\right). \end{aligned}$$

In contrast to Newey [1997], the asymptotic variance is driven by an additional summand that arises from the first step estimation of the inverse selection probability  $g$ . An additional part in the asymptotic variance formula, due to sample selection correction, is obtained also in Das et al. [2003] but which is due to the estimation of a propensity score. In the next

result we replace the variance  $\mathcal{V}_n(x)$  by the estimator

$$\begin{aligned} \widehat{\mathcal{V}}_n(x) = & n^{-1} \sum_{i=1}^n \left| f_{\underline{m}_n}(x)^t (\mathbf{X}_{\underline{m}_n}^t \mathbf{X}_{\underline{m}_n} / n)^{-} f_{\underline{m}_n}(X_i) (Y_i \widehat{g}_n(Y_i) - \widehat{\varphi}_n(X_i)) \right|^2 + \\ & n^{-1} \sum_{i=1}^n \left| f_{\underline{m}_n}(x)^t (\mathbf{X}_{\underline{m}_n}^t \mathbf{X}_{\underline{m}_n})^{-} \mathbf{X}_{\underline{m}_n}^t \text{diag}(\mathbf{Y}) \mathbf{Y}_{k_n} (\mathbf{Y}_{k_n}^t \mathbf{X}_{\underline{m}_n} (\mathbf{X}_{\underline{m}_n}^t \mathbf{X}_{\underline{m}_n})^{-} \mathbf{X}_{\underline{m}_n}^t \mathbf{Y}_{k_n})^{-} \times \right. \\ & \left. \mathbf{Y}_{k_n}^t \mathbf{X}_{\underline{m}_n} (\mathbf{X}_{\underline{m}_n}^t \mathbf{X}_{\underline{m}_n} / n)^{-} \left( f_{\underline{m}_n}(X_i) - \sum_{i'=1}^n f_{\underline{m}_n}(X_{i'}) \right) \right|^2. \end{aligned}$$

where  $\text{diag}(\mathbf{Y})$  is the diagonal matrix with entries on the diagonal given by  $(Y_1, \dots, Y_n)$ . The next result establishes the asymptotic distribution of the estimator  $\widehat{\varphi}_n$  evaluated at some point  $x$  in the support of  $X$ .

**THEOREM 3.3.** *Let Assumptions 1 – 5 and 7 be satisfied. If*

$$n \|F_{m_n} \varphi - \varphi\|_X^2 = o(1) \quad \text{and} \quad n \|T(E_{k_n} g - g)\|_X^2 = o(1) \quad (3.7)$$

then we have

$$\sqrt{n/\widehat{\mathcal{V}}_n(x)} (\widehat{\varphi}_n(x) - \varphi(x)) \xrightarrow{d} \mathcal{N}(0, 1).$$

Moreover, if  $\mathcal{Y}$  is bounded, and  $m_n k_n = o(\tau_{k_n}^4 n)$  then

$$\sqrt{n/\widehat{\mathcal{V}}_n(x)} (\widehat{\varphi}_n(x) - \varphi(x)) \xrightarrow{d} \mathcal{N}(0, 1).$$

In the setting of Corollary 3.2, condition (3.7) is satisfied if  $\max(n m_n^{-2\alpha/d_x}, n \tau_{k_n} k_n^{-2\beta}) = o(1)$ . Thereby, in both estimation steps, undersmoothed estimators have to be considered. This ensures that also the sieve approximation bias for estimating the function  $g$  becomes asymptotically negligible. Moreover, the second result of the theorem requires the additional rate  $m_n k_n = o(\tau_{k_n}^4 n)$ . In the setting of Corollary 3.2, this is equivalent to  $m_n = o(k_n^{-(8t+1)} n)$  in the mildly ill-posed case and to  $m_n = o(n \exp(-4k_n^{2t}))$  in the severely ill-posed case.

## 4. Sample Selection with Endogenous Covariates

In many economic applications, it is necessary to correct for both sample selection of the dependent variable and endogeneity of (some) covariates. In this section, we show that under the assumptions of Section 2 identification of the corresponding reduced form equation can be achieved. Under further conditions, which are common in the nonparametric instrumental variable literature, identification of the structural function is also obtained. An estimator of the nonparametric structural function is proposed and we establish its rate of convergence as well as its asymptotic distribution.

### 4.1. Model and Identification

In this section, we consider the instrumental variable model under selectively observed outcomes given by

$$Y^* = \psi(Z) + U \quad \text{where} \quad \mathbb{E}[U|X] = 0 \quad \text{and} \quad Y = \Delta Y^* \quad (4.1)$$

where  $Z$  is a  $d_z$ -vector of possibly endogenous regressors in the sense that  $\mathbf{E}[U|Z] \neq 0$  and hence  $\psi(Z)$  need not to coincide with  $\mathbf{E}[Y^*|Z]$ . Here,  $X$  is a vector of instruments used to identify the structural function  $\psi$ . The instrument  $X$  is also assumed to satisfy Assumption 1; that is,  $\Delta \perp\!\!\!\perp X|Y^*$ . An example is the estimation of Engel curves, where  $Y^*$  denotes budget share allocated to alcohol which is often not reported (see for instance the British FES) and  $Z$  is total expenditure. Expenditure is commonly thought of as endogenous and typically instrumented for with labor income  $X$ . In this case, the instrument certainly influences  $Y^*$  through  $Z$  but is unlikely to directly influence survey nonresponse. The reduced form equation of the structural model (4.1) is given by

$$\mathbf{E}[Y^*|X] = \mathbf{E}[\psi(Z)|X]$$

where the left hand side is not identified. By making use of equation (2.2), we obtain the reduced form

$$\mathbf{E}[Yg(Y)|X] = \mathbf{E}[\psi(Z)|X] \quad (4.2)$$

where the left hand side is identified under Assumptions 1–3. Thereby, completes of  $Z$  with respect to  $X$  ensures identification of the structural function  $\psi$ . In the following example, we see that Assumptions 1 and 2 are satisfied in a triangular model.

**EXAMPLE 4.1.** Let us rewrite model (4.1) in reduced form and additionally specify a selection equation. Then the assumption  $\Delta \perp\!\!\!\perp X|Y^*$  is satisfied in the triangular model

$$\begin{aligned} Y^* &= \mathbf{E}[\psi(Z)|X] + \varepsilon \quad \text{where } \mathbf{E}[\varepsilon|X] = 0 \\ \Delta &= \xi(Y^*, \eta) \end{aligned}$$

and  $\eta \perp\!\!\!\perp (X, \varepsilon)$ . As in D'Haultfoeuille [2011] it can be argued that, under some assumptions on the distribution of  $(X, \varepsilon)$ ,  $Y^*$  is complete for  $X$ .  $\square$

## 4.2. The Estimator and its Rate of Convergence

In this section, we propose an estimator for the structural function  $\psi$  and derive its rate of convergence. For any  $\phi \in L_Z^2$  we introduce the function  $\varrho(\cdot, g, \phi) = \mathbf{E}[Yg(Y) - \phi(Z)|X = \cdot]$ . The least squares estimator of  $\varrho(\cdot, g, \phi)$  is given by

$$\widehat{\varrho}_n(\cdot, g, \phi) = \underline{f}_{m_n}(\cdot)^t (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \sum_{i=1}^n (Y_i g(Y_i) - \phi(Z_i)) \underline{f}_{m_n}(X_i).$$

Let us now propose a plug-in minimum distance estimator of  $\psi$  which involves the estimator  $\widehat{g}_n$  given in (3.3) of the inverse selection probability  $g$ . That is, we estimate  $\psi$  by

$$\widehat{\psi}_n \equiv \arg \min_{\phi \in \Psi_n} \sum_{i=1}^n \widehat{\varrho}_n^2(X_i, \widehat{g}_n, \phi). \quad (4.3)$$

Here, we consider the linear sieve space  $\Psi_n = \{\phi(\cdot) = \sum_{j=1}^{k_n} \beta_j p_j(\cdot) : \beta \in \mathbb{R}^{k_n}\}$  of dimension  $k_n < \infty$  for some basis functions  $\{p_j\}_{j \geq 1}$  in  $L_Z^2$ . In particular, we have the least squares solution

$$\widehat{\psi}_n(\cdot) \equiv \underline{p}_{k_n}(\cdot)^t \widehat{\vartheta}_{k_n} \quad \text{and} \quad \widehat{\vartheta}_{k_n} = (\mathbf{Z}_k^t \mathbf{X}_m (\mathbf{X}_m^t \mathbf{X}_m)^{-1} \mathbf{X}_m^t \mathbf{Z}_k)^{-1} \mathbf{Z}_k^t \mathbf{X}_m (\mathbf{X}_m^t \mathbf{X}_m)^{-1} \mathbf{X}_m^t \mathbf{G}_n \quad (4.4)$$

where  $\mathbf{G}_n = (Y_1 \widehat{g}_n(Y_1), \dots, Y_n \widehat{g}_n(Y_n))^t$  and  $\mathbf{Z}_k = (p_{k_n}(Z_1), \dots, p_{k_n}(Z_n))^t$ .



**ASSUMPTION 8.** (i) We observe a sample  $((\Delta_1, Y_1, Z_1, X_1), \dots, (\Delta_n, Y_n, Z_n, X_n))$  of iid. copies of  $(\Delta, Y, Z, X)$  where  $Y = \Delta Y^*$ . (ii) There exists a constant  $C \geq 1$  and a sequence of positive integers  $(k_n)_{n \geq 1}$  satisfying  $\sup_{z \in \mathcal{Z}} \|p_{k_n}(z)\|^2 \leq Ck_n$  such that  $k_n^2/n = o(1)$ . (iii) The smallest eigenvalue of  $\mathbf{E}[p_{k_n}(Z)p_{k_n}(Z)^t]$  is bounded away from zero uniformly in  $k$ . (iv) For every  $\psi \in L_Z^2$  there exists  $\Pi_{k_n}\psi \in \Psi_n$  such that  $\|\Pi_{k_n}\psi - \psi\|_Z = O(k_n^{-\gamma/d_Z})$  for some constant  $\gamma > 0$ .

Let us introduce the linear conditional expectation operator  $K : L_Z^2 \rightarrow L_X^2$  with  $K\phi = \mathbf{E}[\phi(Z)|X]$  for all  $\phi \in L_Z^2$ . We introduce the following assumption which ensures identification of the  $\psi$  in the model (4.1).

**ASSUMPTION 9.** (i) For some  $\alpha > 0$  and for every  $\phi \in \Phi_{k_n}$  there exists  $F_{m_n}K\phi \in \mathcal{F}_{m_n}$  such that  $\|F_{m_n}K\phi - K\phi\|_\infty = O(m_n^{-\alpha/d_X})$ . (ii) For every function  $\phi \in L_Z^2$ ,  $\mathbf{E}[\phi(Z)|X] = 0$  implies  $\phi(Z) = 0$ . (iii) There exists a sequence  $(\kappa_j)_{j \geq 1}$  such that  $\|K\phi\|_X^2 \leq C \sum_{j=1}^\infty \kappa_j \langle \phi, p_j \rangle_Z^2$  and  $\|K\phi\|_X^2 \geq c \sum_{j=1}^\infty \kappa_j \langle \phi, p_j \rangle_Z^2$  for all  $\phi \in L_Z^2$  and some constants  $c, C > 0$ .

The next result establishes the rate of convergence of the estimator  $\widehat{\psi}_n$ .

**THEOREM 4.1.** Let Assumptions 1–5, 8 and 9 hold true. Then we have

$$\|\widehat{\psi}_n - \psi\|_Z^2 = O_p\left(\max\left(k_n^{-2\gamma/d_Z}, \frac{k_n}{n\kappa_{k_n}}, \kappa_{k_n}^{-1}\|T(E_{k_n}g - g)\|_X^2\right)\right). \quad (4.5)$$

In contrast to Theorem 3.1, the additional bias due to sample selection is also effected by the potential ill-posed coming from endogeneity of covarites  $Z$ . Under the conditions of Corollary 3.2, the bias  $\kappa_{k_n}^{-1}\|T(E_{k_n}g - g)\|_X^2$  can be bounded by  $\kappa_{k_n}^{-1}\tau_{k_n}k_n^{-2\beta}$ . Thereby, the usual rate in nonparametric instrumental regression (see Chen and Reiß [2011]) can be only obtained if  $\tau_{k_n}k_n^{2\gamma/d_Z} \leq \text{const.} \kappa_{k_n}k_n^{2\beta}$  for all  $n$  sufficiently large.

**REMARK 4.1.** To conclude this section it is worth to mention that with our estimation method we can deal with another type of endogeneity, different from the one just considered. Suppose that the random vector  $X$  that satisfies Assumption 1 is endogenous, in the sense that the relationship of interest is the structural function  $\psi$  satisfying

$$Y^* = \psi(X) + U \quad \text{where} \quad \mathbf{E}[U|X] \neq 0 \quad \text{and} \quad Y = \Delta Y^*. \quad (4.6)$$

This situation can be easily dealt with by assuming that there exists another vector  $W$  of instruments such that  $\mathbf{E}[U|W] = 0$  and  $\Delta \perp\!\!\!\perp X | (Y^*, W)$ . The latter assumption replaces Assumption 1 and corresponds to the one in Remark 2.1. For simplicity, we assume that  $W$  is observed for all the individuals so that the conditional distribution of  $X|W$  is identified from the data. Moreover, we have to assume that  $\mathbb{P}(\Delta = 1|Y^*, W) > 0$  a.s. and that Assumption 2 holds with  $\phi(Y^*)$  replaced by  $\phi(Y^*, W)$ . Then  $\psi$  is identified through (4.6) and

$$\mathbf{E}[Y^*|W] = \mathbf{E}[Y^*\mathbb{P}(\Delta = 1|Y^*, W)g(Y^*, W)|W] = \mathbf{E}[Y^*\Delta g(Y^*, W)|W] = \mathbf{E}[Yg(Y, W)|W].$$

Consequently, we obtain the identified reduced form equation

$$\mathbf{E}[Yg(Y, W)|W] = \mathbf{E}[\psi(X)|W]$$

and identification follows as above.



### 4.3. Pointwise Asymptotic Distribution of the Estimator

This subsection is about inference on the structural function  $\psi$  evaluated at some point of the support of  $Z$ . To establish the asymptotic distribution of our estimator we require the following additional assumptions. Let us introduce the matrix  $\mathbf{K}_n = \mathbf{E}[f_{m_n}(X)p_{k_n}(Z)^t]$  and recall  $\mathbf{Q}_n = \mathbf{E}[f_{m_n}(X)f_{m_n}(X)^t]$ ,  $\mathbf{T}_n = \mathbf{E}[\Delta f_{m_n}(X)e_{k_n}(Y)^t]$ , and  $\mathbf{T}_n^Y = \mathbf{E}[Yf_{m_n}(X)e_{k_n}(Y)^t]$ .

**ASSUMPTION 10.** (i) Let  $\mathbf{E}[|Yg(Y) - \psi(Z)|^4|X] \leq C$  and  $\text{Var}(Yg(Y) - \psi(Z)|X) \geq c$  for some constants  $c, C > 0$ . (ii) The matrix  $\mathbf{T}_n^t \mathbf{T}_n$  has full rank. (iii) There exist some constants  $c, C > 0$  such that  $c \leq \lambda_n / \kappa_{k_n} \leq C$  where  $\lambda_n$  is the minimum eigenvalue of  $\mathbf{K}_n^t \mathbf{Q}_n^{-1} \mathbf{K}_n$ . (iv) It holds  $\|E_{k_n} g - g\|_\infty = O(k_n^{-\beta+1/2})$  for  $\beta > 1/2$  and  $\|\Pi_{k_n} \psi - \psi\|_\infty = O(k_n^{-\gamma/d_z+1/2})$  for  $\gamma > d_z/2$ .

Since  $\mathbf{E}[|Y^*|^2|X] = \mathbf{E}[|Y|^2 g(Y)|X] \leq \mathbf{E}[|Yg(Y)|^2|X]$  it holds

$$\begin{aligned} \text{Var}(Yg(Y) - \psi(Z)|X) &\geq \mathbf{E}[|Yg(Y)|^2|X] - 2\sqrt{\mathbf{E}[|Yg(Y)|^2|X]}\sqrt{\mathbf{E}[\psi^2(Z)|X]} + \mathbf{E}[\psi^2(Z)|X] \\ &= \left(\sqrt{\mathbf{E}[|Yg(Y)|^2|X]} - \sqrt{\mathbf{E}[\psi^2(Z)|X]}\right)^2 \\ &\geq \mathbf{E}[|Y^*|^2|X]/2 - \mathbf{E}[\psi^2(Z)|X] \\ &\geq \text{Var}(U|X)/4 - 3\mathbf{E}[\psi^2(Z)|X]/2 \end{aligned}$$

where we used twice the basic inequality  $(a - b)^2 \geq a^2/2 - b^2$ . Thus,  $\text{Var}(Yg(Y) - \psi(Z)|X)$  is bounded from below if  $\text{Var}(U|X)$  is sufficiently large, more precisely, if  $\text{Var}(U|X) \geq 2(c + 3\mathbf{E}[\psi^2(Z)|X])$ . On the other hand, assuming a bounded conditional fourth moment of the structural disturbance  $U$  ( $\mathbf{E}[U^4|X] \leq \text{const.}$ ) implies  $\mathbf{E}[|Yg(Y) - \psi(Z)|^4|X] \leq \text{const.}$  due to the finite support of  $Y$ . This condition can be similarly motivated as in Example 3.1. Assumption 10 (ii) further ensures that the asymptotic distribution is not degenerate. Let us introduce

$$\begin{aligned} \mathcal{W}_n(z) &= \text{Var}\left(p_{k_n}(z)^t (\mathbf{K}_n^t \mathbf{Q}_n^{-1} \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{Q}_n^{-1} (f_{m_n}(X)(Yg(Y) - \psi(Z)))\right) \\ &\quad + \text{Var}\left(p_{k_n}(z)^t (\mathbf{K}_n^t \mathbf{Q}_n^{-1} \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{Q}_n^{-1} \mathbf{T}_n)^{-1} \mathbf{T}_n^t \mathbf{Q}_n^{-1} (f_{m_n}(X) - \mathbf{E}[f_{m_n}(X)])\right). \end{aligned}$$

If there is no endogenous selection (that is,  $g \equiv 1$ ) then the first part of the asymptotic variance formula coincides to the one obtained by Chen and Pouzo [2013] in nonparametric instrumental regression. Under endogenous selection, however, an additional term in the asymptotic variance arises due to the first step estimation of the inverse selection probability  $g$ .

Let  $\widehat{\mathcal{A}}_n = (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \mathbf{X}_{m_n}^t \mathbf{Z}_{k_n} / n)$ . In the next result we replace the variance  $\mathcal{W}_n(z)$  by the estimator

$$\begin{aligned} \widehat{\mathcal{W}}_n(z) &= n^{-1} \sum_{i=1}^n \left| p_{k_n}(z)^t \widehat{\mathcal{A}}_n (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n}) (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} f_{m_n}(X_i) (Y_i \widehat{g}_n(Y_i) - \widehat{\psi}_n(Z_i)) \right|^2 + \\ &\quad n^{-1} \sum_{i=1}^n \left| p_{k_n}(z)^t \widehat{\mathcal{A}}_n (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n} / n) \mathbf{X}_{m_n}^t \text{diag}(\mathbf{Y}) \mathbf{Y}_{k_n} (\mathbf{Y}_{k_n}^t \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \mathbf{X}_{m_n}^t \mathbf{Y}_{k_n})^{-1} \times \right. \\ &\quad \left. \mathbf{Y}_{k_n}^t \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} (f_{m_n}(X_i) - \sum_{i'=1}^n f_{m_n}(X_{i'})) \right|^2 \end{aligned}$$

where  $\text{diag}(\mathbf{Y})$  is the diagonal matrix with entries on the diagonal given by  $(Y_1, \dots, Y_n)$ . The next result establishes the asymptotic distribution of the estimator  $\widehat{\psi}_n$  evaluated at some point  $z$  in the support  $\mathcal{Z}$  of  $Z$ .

**THEOREM 4.2.** *Let Assumptions 1 – 5, and 8 – 10 be satisfied. If*

$$n\|\Pi_{k_n}\psi - \psi\|_Z^2 = o(1) \quad \text{and} \quad n\|T(E_{k_n}g - g)\|_X^2 = o(1) \quad (4.7)$$

then we have

$$\sqrt{n/\widehat{\mathcal{W}}_n(z)}(\widehat{\psi}_n(z) - \psi(z)) \xrightarrow{d} \mathcal{N}(0, 1).$$

Moreover, if  $\mathcal{Y}$  is bounded and  $m_n = o(\sqrt{n} \min(\sqrt{\tau_{k_n}} \kappa_{k_n}^2, \sqrt{\kappa_{k_n}} \tau_{k_n}^2))$  then

$$\sqrt{n/\widehat{\mathcal{W}}_n(z)}(\widehat{\psi}_n(z) - \psi(z)) \xrightarrow{d} \mathcal{N}(0, 1).$$

## 5. A Model Specification Test

Our estimation procedure crucially relies on the conditional independence between selection and covariates given potential outcomes (see Assumption 1). Hence, it would be desirable to test the validity of this assumption before conducting the estimation procedure. An attractive feature of Assumption 1 is that it is indeed testable (cf. Theorem 2.4 in D’Haultfoeuille [2010]). In this section we construct a test for this assumption. As seen in Section 2, given Assumptions 2 and 3, Assumption 1 is equivalent to the operator equation  $Tg = 1$ . Let us consider a reasonable class of functions for  $g$  namely  $\mathcal{F} = \{\phi \in L_Y^2 : \phi(\cdot) \geq 1 \text{ and } \|\phi - E_{k_n}\phi\|_Y \leq Ck_n^{-\beta} \text{ for any } n \geq 1\}$  for some  $\beta > 0$ . The null hypothesis under consideration is

$$H_0 : \text{there exists a function } g \in \mathcal{F} \text{ such that } Tg = 1.$$

**The test statistic.** Our testing procedure is based on the criterion in (3.3). We verify whether  $\sum_{i=1}^n \widehat{\chi}_n^2(X_i, \widehat{g}_n)$  does not become too large, which is the case if the true inverse conditional probability function  $g$  does not satisfy the minimal smoothness conditions imposed by  $H_0$ . By reformulating the quantity  $\sum_{i=1}^n \widehat{\chi}_n^2(X_i, \widehat{g}_n)$  we obtain our test statistic

$$S_n = \left( \sum_{i=1}^n (\widehat{g}_n(Y_i)\Delta_i - 1) f_{\underline{m}_n}(X_i) \right)^t (\mathbf{X}_{\underline{m}_n}^t \mathbf{X}_{\underline{m}_n})^{-1} \left( \sum_{i=1}^n (\widehat{g}_n(Y_i)\Delta_i - 1) f_{\underline{m}_n}(X_i) \right) \quad (5.1)$$

where the dimension  $m_n$  coincides with the first step dimension used for the estimator  $\widehat{\varphi}_n$ . Our testing procedure builds on Breunig [2015]. But as we consider a constraint estimation procedure we cannot apply the method of Breunig [2015] directly. A constraint sieve testing procedure was proposed by Breunig [2013] but for the specific situation of quantile versions of instrumental variable models. In addition, note that in these two papers the basis functions used to construct the test statistics are assumed to be orthonormal, which is not required in the following.

**Asymptotic distribution of the statistic.** Our test statistic  $S_n$  is asymptotically normally distributed if it is standardized by appropriate mean and variance, which are introduced in the next definition.

**DEFINITION 5.1.** Let us introduce the matrix

$$\Sigma_{m_n} = \mathbf{E} \left[ \left( g(Y)\Delta - 1 \right)^2 Q_n^{-1/2} \underline{f}_{m_n}(X) \underline{f}_{m_n}(X)^t Q_n^{-1/2} \right]$$

Then the trace and the Frobenius norm of  $\Sigma_{m_n}$  are respectively denoted by  $\mu_{m_n}$  and  $\varsigma_{m_n}$ .

**ASSUMPTION 11.** There exist constants  $c, C > 0$  such that  $\mathbb{V}\text{ar}(g(Y)\Delta|X) \geq c$  and  $\mathbf{E}[(g(Y)\Delta - 1)^4|X] \leq C$ .

Due to Assumption 11 (iii) it holds  $\varsigma_{m_n} \geq C\sqrt{m_n}$  for some constant  $C > 0$ . The next result establishes asymptotic normality of  $S_n$  after standardization.

**THEOREM 5.1.** Let Assumptions 1–6, and 11 be satisfied. If

$$m_n^3 = o(n), \quad k_n m_n^2 = O(n\tau_{k_n}), \quad \text{and} \quad \max(k_n, n\tau_{k_n} k_n^{-2\beta}) = o(\sqrt{m_n}) \quad (5.2)$$

then it holds under  $H_0$

$$(\sqrt{2\varsigma_{m_n}})^{-1} (nS_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

**Estimation of Critical Values.** For the estimation of critical values of Theorem 5.1, let us define  $\mathbf{U}_n = (\Delta_1 \widehat{g}_n(Y_1) - 1, \dots, \Delta_n \widehat{g}_n(Y_n) - 1)^t$ . We estimate the matrix  $\Sigma_{m_n}$  by  $\widehat{\Sigma}_{m_n} \equiv (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1/2} \mathbf{X}_{m_n}^t \text{diag}(\mathbf{U}_n)^2 \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1/2}$ . The asymptotic result of Theorem 5.1 continues to hold if we replace  $\varsigma_{m_n}$  by the Frobenius norm of  $\widehat{\Sigma}_{m_n}$ , denoted by  $\widehat{\varsigma}_{m_n}$ , and  $\mu_{m_n}$  by the trace of  $\widehat{\Sigma}_{m_n}$ , denoted by  $\widehat{\mu}_{m_n}$ .

**THEOREM 5.2.** Let the assumptions of Theorem 5.1 be satisfied. Then it holds under  $H_0$

$$(\sqrt{2\widehat{\varsigma}_{m_n}})^{-1} (nS_n - \widehat{\mu}_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

**Limiting behavior under local alternatives.** In the following, we study the power of the test, that is, the probability to reject a false hypothesis against a sequence of linear local alternatives that tends to zero as the sample size tends to infinity. We consider alternative models defined through a sequence of functions  $g_n$  that satisfies

$$\|Tg_n - 1 - m_n^{1/4} n^{-1/2} \delta\|_X = o(m_n^{1/4} n^{-1/2}) \quad (5.3)$$

for some function  $\delta \in L_X^4$ . Due to (5.3), for any  $n \geq 1$  the function  $g_n$  does not satisfy  $Tg_n = 1$ . The next result is a direct consequence of Proposition 2.4 of Breunig [2013] and thus, its proof is omitted.

**PROPOSITION 5.3.** Given the conditions of Theorem 5.1 it holds under (5.3)

$$(\sqrt{2\varsigma_{m_n}})^{-1} (nS_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2} \sum_{j=1}^{\infty} \left( \mathbf{E}[\delta(X) f_j(X)] \right)^2, 1\right).$$

## 6. Monte Carlo simulation

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo simulation. There are 1000 Monte Carlo replications in each experiment. Let  $X = \Phi(\chi)$  where  $\chi \sim \mathcal{N}(0, 1)$ . Further, generate  $Y^*$  from the model  $Y^* = \varphi(X) + c_V V$  where  $\varphi(x) = \Phi(8(x - 0.5))$  with standard normal distribution function  $\Phi$ ,  $c_V = 0.4$ , and  $V \sim \mathcal{N}(0, 1)$ .

**Estimation of conditional expectation.** We consider estimation of the conditional expectation of  $Y^*$  given  $X$ . We generate realizations of the selection variable  $\Delta$  from  $\Delta \sim$

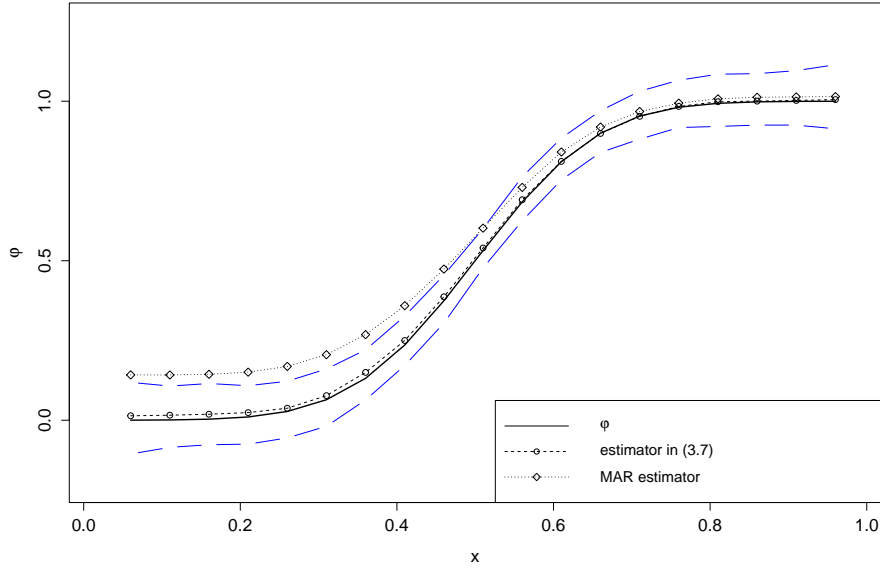


Figure 1: The regression function  $\varphi$ , the median of  $\widehat{\varphi}_n$  with 90% confidence intervals and an estimator under MAR assumption

Binomial( $1, h(Y^*)$ ) where  $h(y) = 0.4 * \mathbb{1}\{y \leq 0.4\} + \mathbb{1}\{y > 0.4\}$ . We estimate the function  $\varphi$  by using the constraint estimator  $\widehat{\varphi}_n$  given in (3.5). As basis functions we use B-splines of order 3 with 2 knots (hence  $k_n = 6$ ) and for the criterion function we use orthogonal B-splines of order 3 with 6 knots (hence  $m_n = 10$ ). Figure 1 depicts the median of the estimator  $\widehat{\varphi}_n$  together with its 90% pointwise confidence bands and an estimator under the missing at random (MAR) assumption. The MAR estimator becomes more biased for small values of  $x$  and lies even outside of the pointwise confidence bands of  $\widehat{\varphi}_n$ .

**Nonparametric Specification Test.** Results are presented for the nominal level 0.05.

Let us now study the finite sample behavior of our nonparametric specification test. We construct the observations of  $\Delta$  via the function  $h(y) = 1/(10y + 2) + 0.5$  for  $y \in [0, 1]$ . If  $H_0$  holds true we generate  $\Delta \sim \text{Binomial}(1, h(Y^*))$ .

In the experiments where  $H_0$  fails,  $X$  is not a valid instrument in the sense that it influences the endogenous selection. In this case, we generate realizations of  $\Delta$  from

$$\Delta \sim \text{Binomial}(1, (1 - \nu)h(Y^*) + \nu\rho(X))$$

for some constant  $\nu > 0$  and where  $\rho(x) = 1 - (2x - 1)^2$ . Clearly, if  $\nu = 0$  then the null hypothesis  $H_0$  is true. We estimate the regression function  $\varphi$  as in the previous paragraph.

Model		Empirical Rejection probability
$\rho$	$\nu$	$(\sqrt{2}\widehat{\zeta}_{m_n})^{-1}(nS_n - \widehat{\mu}_{m_n})$
$H_0$ true		0.046
$\rho_1$	0.2	0.221
	0.3	0.389
	0.5	0.562
$\rho_2$	0.2	0.374
	0.3	0.677
	0.5	0.814
$\rho_3$	0.2	0.659
	0.3	0.909
	0.5	0.942

Table 1: Empirical Rejection probabilities for Nonparametric Specification Test

In Table 1, we depict the empirical rejection probabilities of our test statistic with critical value 0.05. The critical values of these statistics are estimated as described in Theorem 5.2.

## 7. Empirical Illustration

In this section, we apply our estimation procedure to study the way in which the level of expenditure of an individual affects his/her expected “propensity to work”.

We use data from the German Internet Panel (GIP)<sup>3</sup>. This data set contains data about individual attitudes and preferences which are relevant for political and economic decision-making processes. The survey represents the German speaking population aged 16 to 75 in Germany.

In our application we measure the “propensity to work” by the “number of desired hours” which is present in our data set. The latter variable is the number of weekly hours a person would like to work by taking into account that the income would change according to the hours of work. Let  $Y^*$  denote this variable and  $X$  denote the variable “expenditure”. The latter measures the total average expenditure in one month of a person.

The object of interest in our study is the regression function of  $Y^*$  given  $X$ , that is, the expected number of desired hours given a level of monthly total expenditure.

In our data set we have 1118 observations, a small number of missing in the variable “expenditure” (56 observations) and a large number of missing in the variable “number of desired hours” (392 observations). As the number of missing values in “expenditure” is small we eliminate these observations from our data set (since the bias is going to be negligible) so that the sample size we work with becomes  $n = 1056$  and the missing values in the variable “number of desired hours” are now 365. This shows that only less than half

<sup>3</sup>The GIP is a survey part of the Collaborative Research Centre “Political Economy of Reforms” (SFB 884) based at the University of Mannheim which studies the determinants and perceptions of political reforms and their consequences for the economy.

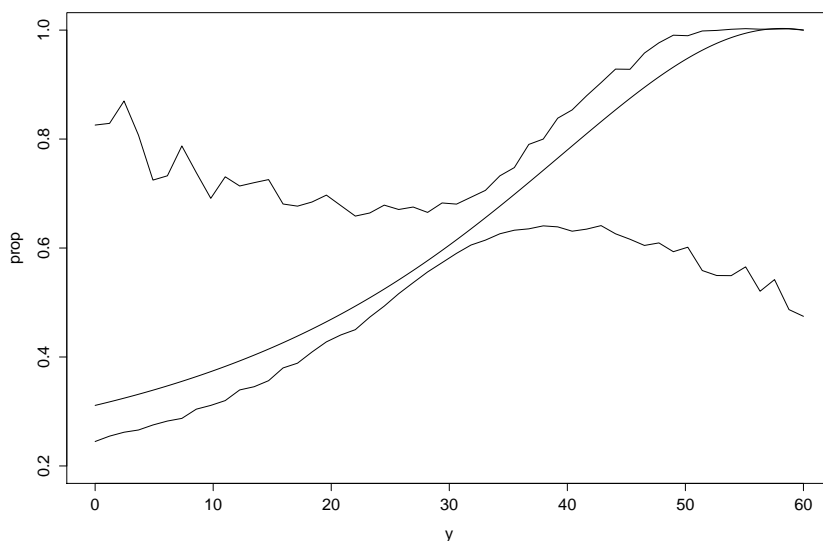


Figure 2: Estimator of  $\mathbb{P}(\Delta = 1|Y^* = y)$  with 90% percent confidence intervals.

of the missing in “expenditure” corresponds to missing in “number of desired hours”. The fact that  $Y^*$  is not observed is likely to be endogenous since one could think that in “extreme” situations, where  $Y^*$  is either excessively low or excessively high, a person would be more likely to not provide this information. While the variable “expenditure” is statistically related to  $Y^*$ , it is reasonable to assume that it is independent of the selection mechanism once  $Y^*$  is accounted for. In particular, we have implemented our specification test proposed in Section 5. We have computed the test statistics for a grid of values for  $m_n$  and the maximum absolute value of the (standardized) test statistics is obtained for  $m_n = 42$  and is 0.1757. Therefore, our test fails to reject  $H_0$  at the level 5%.

Figure 2 depicts our estimator for the conditional probability  $\mathbb{P}(\Delta = 1|Y^*)$  and we observe that this estimated probability of reporting increases with potential desired hours of working. The pointwise confidence intervals, however, are too wide to make significant statements. A reason for this is that the conditional probability function is a solution to an inverse problem and hence estimators can be imprecise in finite samples.

	<i>Mean</i>	<i>Median</i>	<i>Min</i>	<i>Max</i>	<i>Std.</i>	<i>n</i>	<i>missing</i>
$Y = \Delta Y^*$	22.05	30	0	60	18.2392	1056	373
$X$	1439	1225	0	6000	899.2948	1056	0

Table 2: Descriptive Statistics for  $Y = \Delta Y^*$  and  $X$ .

Figure 3 shows the estimated regression function of “number of desired hours” on “expenditure” together with the 90% percent confidence intervals. The estimator is based on the nonparametric methodology described in Section 3 where we use B-Splines where the first step was penalized by a smoothness matrix as described in Blundell et al. [2007] page 1638. The additional penalization avoids to stabilize the estimates even if the number

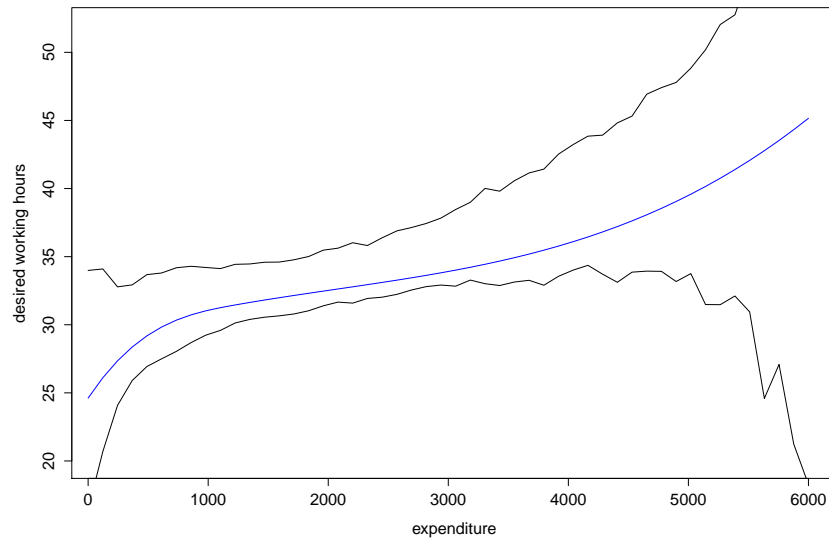


Figure 3: Regression curve of “number of desired hours” on “expenditure” with 90% percent confidence intervals.

of B-spline basis functions might be chosen too large. The B-splines used here are order 4 with 3 knots, hence  $k_n = 8$ . We also report the estimated regression function of “number of desired hours” on “expenditure” estimated under the MAR-assumption in Figure 4. As we see from the figures the MAR estimator is only significantly different from our estimator when expenditure is small.



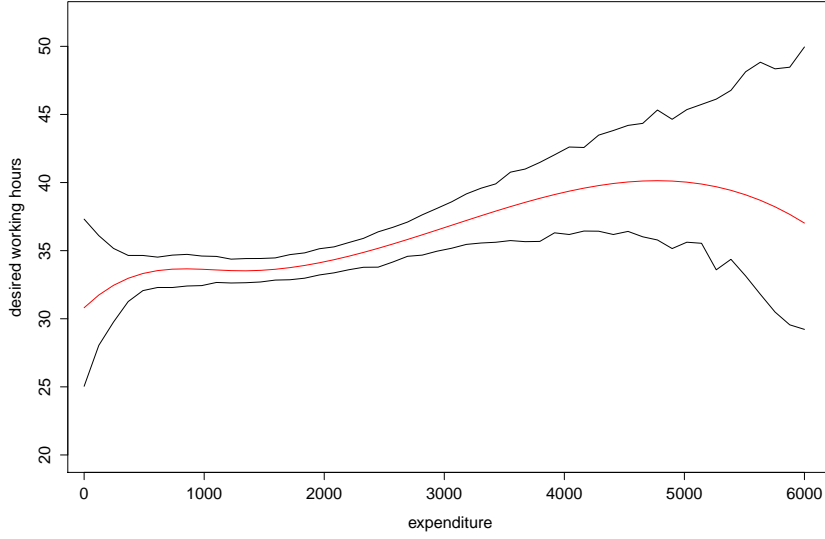


Figure 4: Regression under MAR of “number of desired hours” on “expenditure” with 90% percent confidence intervals.

## A. Appendix

Throughout the proofs, we will use  $C > 0$  to denote a generic finite constant that may be different in different uses.  $I_{m_n}$  denotes the  $m_n \times m_n$  identity matrix. Further, for ease of notation we write  $\sum_i$  for  $\sum_{i=1}^n$  and  $\sum_{i' < i}$  for  $\sum_i \sum_{i'=1}^{i-1}$ . For a matrix  $A$ , we denote by  $\|A\|$  its operator norm.

In the following, we denote  $\widehat{Q}_n = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t$ . By Assumption 4, the eigenvalues of  $\mathbf{E}[f_{m_n}(X) f_{m_n}(X)^t]$  are bounded away from zero and hence, it may be assumed that  $\mathbf{E}[f_{m_n}(X) f_{m_n}(X)^t] = I_{m_n}$  (cf. Newey [1997], p. 161).

### A.1. Proofs of Section 3.

**PROOF OF THEOREM 3.1.** The proof is based on the following decomposition

$$\begin{aligned} \|\widehat{\varphi}_n - \varphi\|_X^2 &\leq 2 \|\widehat{Q}_n^{-1} (I_{m_n} - \widehat{Q}_n) (\mathbf{X}_{m_n}^t \mathbf{G}_n / n - \widehat{Q}_n \mathbf{E}[\varphi(X) f_{m_n}(X)])\|^2 \\ &\quad + 4 \|\mathbf{X}_{m_n}^t \mathbf{G}_n / n - \widehat{Q}_n \mathbf{E}[\varphi(X) f_{m_n}(X)]\|^2 + 4 \|F_{m_n} \varphi - \varphi\|_X^2 \\ &= 2I_n + 4II_n + 4III_n \quad (\text{say}). \quad (\text{A.1}) \end{aligned}$$

First observe that

$$I_n \leq \|\widehat{Q}_n^{-1}\|^2 \|\widehat{Q}_n - I_{m_n}\|^2 \|\mathbf{X}_{m_n}^t \mathbf{G}_n / n - \widehat{Q}_n \mathbf{E}[\varphi(X) f_{m_n}(X)]\|^2.$$

We have  $\|\widehat{Q}_n - I_{m_n}\|^2 = m_n^2/n$  and  $\|\widehat{Q}_n^{-1}\|^2 = 1 + o_p(1)$  as Newey [1997] page 162. Further, we observe

$$\begin{aligned}
II_n &\leq 4 \sum_{j=1}^{m_n} \left| n^{-1} \sum_i Y_i (\widehat{g}_n(Y_i) - (E_{k_n}g)(Y_i)) f_j(X_i) - \langle TM_{id}(\widehat{g}_n - E_{k_n}g), f_j \rangle_X \right|^2 \\
&\quad + 8 \sum_{j=1}^{m_n} \left| n^{-1} \sum_i Y_i (E_{k_n}g)(Y_i) f_j(X_i) - \mathbf{E}[Y(E_{k_n}g)(Y) f_j(X)] \right|^2 \\
&\quad + 4 \sum_{j=1}^{m_n} \left| n^{-1} \sum_i \sum_{l=1}^{m_n} (f_j(X_i) f_l(X_i) - \mathbf{E}[f_j(X) f_l(X)]) \mathbf{E}[Y f_l(X)] \right|^2 \\
&\quad + 2 \sum_{j=1}^{m_n} \langle TM_{id}(\widehat{g}_n - E_{k_n}g), f_j \rangle_X^2 + 8 \sum_{j=1}^{m_n} \langle TM_{id}(E_{k_n}g - g), f_j \rangle_X^2 \\
&= 4B_{n1} + 8B_{n2} + 4B_{n3} + 2B_{n4} + 8B_{n5} \quad (\text{say}).
\end{aligned}$$

Consider  $B_{n1}$ . The Cauchy Schwarz inequality implies

$$\begin{aligned}
B_{n1} &\leq \|\widehat{\beta}_n - \beta_{k_n}\|^2 \sum_{j=1}^{m_n} \left\| n^{-1} \sum_i Y_i e_{k_n}(Y_i) f_j(X_i) - \mathbf{E}[Y e_{k_n}(Y) f_j(X)] \right\|^2 \\
&= \|\widehat{\beta}_n - \beta_{k_n}\|^2 O_p(k_n m_n/n) \\
&= O_p(m_n/n)
\end{aligned}$$

where we used that  $k_n \|\widehat{\beta}_n - \beta_{k_n}\|^2 = O_p(1)$ . Consider  $B_{n2}$ . Since  $\text{Var}(Y E_{k_n}g(Y))$  is bounded we have

$$\begin{aligned}
\mathbf{E} B_{n2} &\leq n^{-1} \sum_{j=1}^{m_n} \mathbf{E} \left| Y E_{k_n}g(Y) f_j(X) \right|^2 \\
&\leq n^{-1} \sup_{x \in \mathcal{X}} \|f_{\underline{m}_n}(x)\|^2 \mathbf{E} \left| Y E_{k_n}g(Y) \right|^2 \\
&\leq 2n^{-1} \sup_{x \in \mathcal{X}} \|f_{\underline{m}_n}(x)\|^2 \left( \text{Var}(Y E_{k_n}g(Y)) + \|TM_{id}E_{k_n}g\|_{\mathcal{G}}^2 \right) \\
&= O(m_n/n).
\end{aligned}$$

Consider  $B_{n3}$ . It holds

$$\mathbf{E} B_{n3} = n^{-1} \sum_{j=1}^{m_n} \mathbf{E} |f_j(X)(F_{m_n}\varphi)(X)|^2 \leq 2n^{-1} \sup_{x \in \mathcal{X}} \|f_{\underline{m}_n}(x)\|^2 \left( \|F_{m_n}\varphi - \varphi\|_X^2 + \|\varphi\|_X^2 \right) = O(m_n/n).$$

Consider  $B_{n4}$ . We observe

$$\begin{aligned}
B_{n4} &= \|F_{m_n} TM_{id}(\widehat{g}_n - E_{k_n}g)\|_X^2 \leq \|TM_{id}(\widehat{g}_n - E_{k_n}g)\|_X^2 \\
&\leq \sup_{\{\phi \in \mathcal{G}_{k_n} : \phi(\cdot) \geq 1\}} \left\{ \frac{\|TM_{id}(\phi - E_{k_n}g)\|_X^2}{\|T(\phi - E_{k_n}g)\|_X^2} \right\} \|T(\widehat{g}_n - E_{k_n}g)\|_X^2
\end{aligned}$$

Due to Assumption 5 (iv) we have

$$\sup_{\{\phi \in \mathcal{G}_{k_n} : \phi(\cdot) \geq 1\}} \left\{ \frac{\|TM_{id}(\phi - E_{k_n}g)\|_X^2}{\|T(\phi - E_{k_n}g)\|_X^2} \right\} < \infty.$$

From Blundell et al. [2007] page 1659 we deduce

$$\|T(\widehat{g}_n - E_{k_n}g)\|_X^2 = O_p\left(m_n^{-2\alpha/d_x} + m_n/n + \|T(E_{k_n}g - g)\|_X^2\right).$$

Consequently, we have

$$B_{n4} = O_p\left(m_n^{-2\alpha/d_x} + m_n/n + \|T(E_{k_n}g - g)\|_X^2\right).$$

Further,

$$B_{n5} \leq \|F_{m_n}T(E_{k_n}g - g)\|_X^2 = O\left(\|T(E_{k_n}g - g)\|_X^2\right),$$

which proves the result.  $\square$

**PROOF OF COROLLARY 3.2.** It is sufficient to check that

$$\sup_{\phi \in \mathcal{G}_{k_n}} \frac{\|TM_{id}\phi\|_X}{\|T\phi\|_X} < \infty. \quad (\text{A.2})$$

Let  $T^*$  denote the adjoint operator of  $T$  which is given by  $T^*\phi = \mathbf{E}[\Delta\phi(W)|Y]$ . Since the multiplication operator  $M_{id}$  is a selfadjoint operator we obtain

$$\|TM_{id}\phi\|_X^2 = \langle TM_{id}\phi, TM_{id}\phi \rangle_X = \langle T\phi, (T^*)^{-1}M_{id}T^*TM_{id}\phi \rangle_X \leq \|T\phi\|_X \|(T^*)^{-1}M_{id}T^*TM_{id}\phi\|_X$$

From the link condition  $\|T\phi\|_X^2 \geq c \sum_{j=1}^{\infty} \tau_j \langle \phi, e_j \rangle_Y^2$  we infer by a duality argument  $\|(T^*)^{-1}\phi\|_X^2 \leq c^{-1} \sum_{j=1}^{\infty} \tau_j^{-1} \langle \phi, e_j \rangle_Y^2$ . Let  $L$  be a selfadjoint operator acting on  $L_Y^2$  with eigenvalue decomposition  $\{\tau_j^{1/2}, e_j\}_{j \geq 1}$ . Then we conclude

$$\|TM_{id}\phi\|_X^2 \leq c^{-1} \|T\phi\|_X \|L^{-1}M_{id}T^*TM_{id}\phi\|_X \leq c^{-1} \|T\phi\|_X \|TM_{id}L^{-1}\|_X \|TM_{id}\phi\|_X$$

which gives

$$\frac{\|TM_{id}\phi\|_X}{\|T\phi\|_X} \leq c^{-1} \|TM_{id}L^{-1}\|_X.$$

Consequently, (A.2) follows since  $\|TM_{id}L^{-1}\|_X$  is bounded.  $\square$

**PROOF OF THEOREM 3.3.** To bound the asymptotic variance  $\mathcal{V}_n(x)$  from below we observe that assumption  $\text{Var}(Yg(Y)|X) \geq C$  yields

$$\mathcal{V}_n(x) \geq \underline{f}_{m_n}(x)^t \mathbf{E} \left[ \underline{f}_{m_n}(X) \text{Var}(Yg(Y)|X) \underline{f}_{m_n}(X)^t \right] \underline{f}_{m_n}(x) / 2 \geq C \|\underline{f}_{m_n}(x)\|^2.$$

In the following, we also make use of

$$\begin{aligned} \text{Var}(Yg(Y) - \varphi(X)|X) &= \mathbf{E}(|Yg(Y) - \varphi(X)|^2|X) \\ &= \mathbf{E}(|Yg(Y)|^2|X) - \varphi^2(X) \\ &= \text{Var}(Yg(Y)|X). \end{aligned}$$

The proof is based on the decomposition

$$\begin{aligned}
\widehat{\varphi}_n(x) - \varphi(x) &= \underline{f}_{m_n}(x)^t (n\widehat{Q}_n)^{-1} \sum_i \underline{f}_{m_n}(X_i) (Y_i g(Y_i) - \varphi(X_i)) \\
&\quad + \underline{f}_{m_n}(x)^t (n\widehat{Q}_n)^{-1} \sum_i \underline{f}_{m_n}(X_i) Y_i (\widehat{g}_n(Y_i) - E_{k_n} g(Y_i)) \\
&\quad + \underline{f}_{m_n}(x)^t (n\widehat{Q}_n)^{-1} \sum_i \underline{f}_{m_n}(X_i) Y_i (E_{k_n} g(Y_i) - g(Y_i)) \\
&\quad + \underline{f}_{m_n}(x)^t (n\widehat{Q}_n)^{-1} \sum_i \underline{f}_{m_n}(X_i) \varphi(X_i) - \varphi(x) \\
&= I_n + II_n + III_n + IV_n \quad (\text{say}).
\end{aligned}$$

Let us denote  $\beta_{k_n} = (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t \mathbf{E}[f_{m_n}(X)]$ . As in the proof of Theorem 3.1 and since  $\mathbf{E}[f_{m_n}(X)] = \mathbf{E}[\Delta g(Y) f_{m_n}(X)]$  it is easily seen

$$II_n = \sqrt{n} \underline{f}_{m_n}(x)^t \mathbf{T}_n^Y (\widehat{\beta}_{k_n} - \beta_{k_n}) + \sqrt{n} \underline{f}_{m_n}(x)^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t \mathbf{E} [f_{m_n}(X) \Delta(g(Y) - E_{k_n} g(Y))] + o_p(1).$$

From Assumption 5 (iv) we infer

$$\|\mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t\| \leq \sup_{\phi \in \mathcal{G}} \left\{ \frac{\|TM_{id}\phi\|_X}{\|F_{m_n} T\phi\|_X} \right\} \|\mathbf{T}_n (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t\| \leq C \|\mathbf{T}_n (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t\| = C$$

where we used the fact that projection matrices are bounded in the operator norm. The Cauchy Schwarz inequality together with  $n\|T(g - E_{k_n} g)\|_X^2 = o(1)$  implies

$$II_n = \sqrt{n} \underline{f}_{m_n}(x)^t \mathbf{T}_n^Y (\widehat{\beta}_{k_n} - \beta_{k_n}) + o_p(1).$$

Consequently, we obtain

$$\begin{aligned}
\sqrt{n/\mathcal{V}_n(x)} (I_n + II_n) &= \sum_i (n\mathcal{V}_n(x))^{-1/2} \underline{f}_{m_n}(x)^t (f_{m_n}(X_i) (Y_i g(Y_i) - \varphi(X_i)) \\
&\quad - \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t (f_{m_n}(X_i) - \mathbf{E}[f_{m_n}(X)])) + o_p(1) \\
&= \sum_i s_{in} + o_p(1)
\end{aligned}$$

Moreover,  $s_{in}$ ,  $1 \leq i \leq n$  satisfy the Lindeberg condition which can be seen as follows. It holds  $\mathbf{E}[s_{in}] = 0$  and  $n\mathbf{E}[s_{in}^2] = 1$ . Following Newey [1997] (p. 164), for all  $\varepsilon > 0$  due to  $\mathbf{E}|Yg(Y) - \varphi(X)|^4 \leq C$  we observe

$$\sum_i \mathbf{E}[s_{in}^2 \mathbb{1}_{\{|s_{in}| > \varepsilon\}}] \leq n\varepsilon^2 \mathbf{E}|s_{in}/\varepsilon|^4 \leq Cn^{-1}\varepsilon^{-2}m_n^2 = o(1).$$

Again from Assumption 5 (iv) together with  $n\|T(g - E_{k_n} g)\|_X^2 = o(1)$  yields  $nIII_n = o_p(1)$ . Finally,  $nIV_n = o_p(1)$  follows from  $n\|F_{m_n}\varphi - \varphi\|_X^2 = o(1)$ , which completes the proof of the first statement in the theorem.

By Lemma A.1 we have

$$\sqrt{n/\widehat{\mathcal{V}}_n(x)} (\widehat{\varphi}_n(x) - \varphi(x)) = \sqrt{n/\mathcal{V}_n(x)} (\widehat{\varphi}_n(x) - \varphi(x)) (\mathcal{V}_n(x)/\widehat{\mathcal{V}}_n(x))^{1/2} \xrightarrow{d} \mathcal{N}(0, 1)$$

which proves the second result of the theorem.  $\square$

**LEMMA A.1.** Let Assumptions 1 – 5 and 7 be satisfied. Moreover, assume that  $\mathcal{Y}$  is bounded. Then,

$$\mathcal{V}_n(x)^{-1}\widehat{\mathcal{V}}_n(x) = 1 + o_p(1)$$

where  $\mathcal{V}_n(x)$  and  $\widehat{\mathcal{V}}_n(x)$  are as defined in Theorem 3.3

*Proof.* Let us introduce some notation. We denote  $\Sigma_n = \mathbf{E}[f_{m_n}(X) \text{Var}(Yg(Y)|X)f_{m_n}(X)^t]$ ,  $\Sigma_{n,f} = I_{m_n} - \mathbf{E}[f_{m_n}(X)]\mathbf{E}[f_{m_n}(X)^t]$ ,  $\widehat{\Sigma}_n = \sum_{i=1}^n f_{m_n}(X_i)f_{m_n}(X_i)^t(Y_i\widehat{g}_n(Y_i) - \widehat{\varphi}_n(X_i))^2/n$ ,  $\widehat{\Sigma}_{n,f} = \widehat{Q}_n - (\mathbf{X}_m^t \mathbf{1}_n/n)(\mathbf{1}_n^t \mathbf{X}_m/n)$  and  $\widetilde{\Sigma}_n = \sum_i f_{m_n}(X_i)f_{m_n}(X_i)^t(Y_i g(Y_i) - \varphi(X_i))^2/n$ . Moreover, let  $\widehat{h}_1^t = (\mathcal{V}_n(x))^{-1/2} f_{m_n}(x)^t \widehat{Q}_n^{-1}$ ,  $h_1^t = (\mathcal{V}_n(x))^{-1/2} f_{m_n}(x)^t$ ,  $\widehat{\mathbf{T}}_n = \mathbf{X}_{m_n}^t \mathbf{Y}_{k_n}/n$

$$\widehat{h}_2^t = (\mathcal{V}_n(x))^{-1/2} f_{m_n}(x)^t \widehat{Q}_n^{-1} (\mathbf{X}_{m_n}^t \text{diag}(\mathbf{Y}) \mathbf{Y}_{k_n}/n) (\widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1} \widehat{\mathbf{T}}_n)^{-1} \widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1}$$

and  $h_2^t = (\mathcal{V}_n(x))^{-1/2} f_{m_n}(x)^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t$ . Hence,

$$\mathcal{V}_n(x) (\widehat{h}_1^t \widehat{\Sigma}_n \widehat{h}_1 + \widehat{h}_2^t \widehat{\Sigma}_{n,f} \widehat{h}_2) = \widehat{\mathcal{V}}_n(x)$$

and by noticing that  $h_1^t \Sigma_n h_1 + h_2^t \Sigma_{n,f} h_2 = 1$ , the triangle inequality gives

$$\begin{aligned} \left| (\mathcal{V}_n(x))^{-1/2} \widehat{\mathcal{V}}_n(x) (\mathcal{V}_n(x))^{-1/2} - 1 \right| &\leq \left| \widehat{h}_1^t (\widehat{\Sigma}_n - \widetilde{\Sigma}_n) \widehat{h}_1 \right| + \left| \widehat{h}_1^t (\widetilde{\Sigma}_n - \Sigma_n) \widehat{h}_1 \right| + \left| \widehat{h}_1^t \Sigma_n \widehat{h}_1 - h_1^t \Sigma_n h_1 \right| \\ &\quad + \left| \widehat{h}_2^t (\widehat{\Sigma}_{n,f} - \Sigma_{n,f}) \widehat{h}_2 \right| + \left| \widehat{h}_2^t \Sigma_{n,f} \widehat{h}_2 - h_2^t \Sigma_{n,f} h_2 \right|. \end{aligned} \quad (\text{A.3})$$

Remark that  $\|\widehat{h}_1 - h_1\| = o_p(1)$  and  $\|\widehat{h}_1\| = O_p(1)$ . Consider  $\widehat{h}_2$ . We have

$$\begin{aligned} \|(\widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1} \widehat{\mathbf{T}}_n)^{-1} - (\mathbf{T}_n^t \mathbf{T}_n)^{-1}\| &\leq \|(\widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1} \widehat{\mathbf{T}}_n)^{-1}\| \|(\mathbf{T}_n^t \mathbf{T}_n) - (\widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1} \widehat{\mathbf{T}}_n)\| \|(\mathbf{T}_n^t \mathbf{T}_n)^{-1}\| \\ &= O_p(\tau_{k_n}^{-2} n^{-1/2} \sqrt{m_n k_n}), \end{aligned} \quad (\text{A.4})$$

as we see in the following. It holds

$$\begin{aligned} \|\widehat{\mathbf{T}}_n^t \widehat{Q}_n^{-1} \widehat{\mathbf{T}}_n - \mathbf{T}_n^t \mathbf{T}_n\| &\leq \|\widehat{\mathbf{T}}_n^t (\widehat{Q}_n - I_{m_n}) \widehat{\mathbf{T}}_n\| + \|\widehat{\mathbf{T}}_n^t (\widehat{\mathbf{T}}_n - \mathbf{T}_n) + \mathbf{T}_n^t (\widehat{\mathbf{T}}_n - \mathbf{T}_n)\| \\ &= \|\widehat{\mathbf{T}}_n^t (\widehat{Q}_n - I_{m_n}) \widehat{\mathbf{T}}_n\| + \|(\widehat{\mathbf{T}}_n - \mathbf{T}_n)^t (\widehat{\mathbf{T}}_n - \mathbf{T}_n) + 2\mathbf{T}_n^t (\widehat{\mathbf{T}}_n - \mathbf{T}_n)\| \\ &\leq \|\widehat{\mathbf{T}}_n^t (\widehat{Q}_n - I_{m_n}) \mathbf{T}_n\|^2 + \|\widehat{\mathbf{T}}_n - \mathbf{T}_n\|^2 (1 + \|\widehat{Q}_n - I_{m_n}\|) + 2\|\widehat{\mathbf{T}}_n - \mathbf{T}_n\| (\|\mathbf{T}_n\| + \|\widehat{Q}_n - I_{m_n}\|) \\ &= \|\mathbf{T}_n^t (\widehat{Q}_n - I_{m_n}) \mathbf{T}_n\|^2 + o_p(1). \end{aligned}$$

Moreover,

$$\begin{aligned} \|\mathbf{T}_n^t (\widehat{Q}_n - I_{m_n}) \mathbf{T}_n\|^2 &\leq n^{-1} \sum_{j,l=1}^{k_n} \mathbf{E} \left| \mathbf{E}[\Delta e_l(Y) f_{m_n}(X)]^t f_{m_n}(X) f_{m_n}(X)^t \mathbf{E}[\Delta e_j(Y) f_{m_n}(X)] \right|^2 \\ &= n^{-1} \mathbf{E} \left( \sum_{j,l=1}^{k_n} |(F_{m_n} T e_l)(X) (F_{m_n} T e_j)(X)|^2 \right) \\ &\leq n^{-1} \left( \sup_{x \in \mathcal{X}} \|(T e_{k_n})(x)\|^2 \right) \\ &\leq n^{-1} \left( \sup_{y \in \mathcal{Y}} \|e_{k_n}(y)\|^2 \right) \\ &= O(n^{-1} k_n^2) \end{aligned}$$

since  $\|(\widehat{\mathbf{T}}_n \widehat{\mathbf{Q}}_n^{-1} \widehat{\mathbf{T}}_n)^{-1}\| \leq \|(\mathbf{T}_n^t \mathbf{T}_n)^{-1}\| + o_p(1)$  and  $\|(\mathbf{T}_n^t \mathbf{T}_n)^{-1}\| \leq C\tau_{k_n}^{-1}$  by assumption 7 (iv),  $\|\widehat{\mathbf{Q}}_n^{-1} - \mathbf{I}\|^2 = O_p(m_n^2/n)$ ,  $\|X_{m_n}^t \text{diag}(\mathbf{Y})\mathbf{Y}_{k_n}/n - T_n^Y\|^2 = O_p(m_n k_n/n)$ ,  $\|\widehat{\mathbf{T}}_n - \mathbf{T}_n\|^2 = O_p(m_n k_n/n)$ . Then,  $\|\widehat{h}_2 - h_2\|^2 = o_p(1)$  by Assumptions 4 (ii) and 5 (ii), (iv), and so  $\|\widehat{h}_2\| = O_p(1)$ .

Under Assumptions 4 (ii)-(iii) and 7 (i), we can show similarly as in Newey [1997] page 165 – 166 that:

$$\left| \widehat{h}_1^t \widehat{\Sigma}_n \widehat{h}_1 - h_1^t \Sigma_n h_1 \right| = o_p(1) \quad \text{and} \quad \left| \widehat{h}_1^t (\widehat{\Sigma}_n - \Sigma_n) \widehat{h}_1 \right| = o_p(1). \quad (\text{A.5})$$

Moreover, denote  $\widehat{\mathfrak{E}}_n = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t |Y_i g(Y_i) - \varphi(X_i)|$ ,  $\mathfrak{E}_n = \mathbf{E}[f_{m_n}(X) f_{m_n}(X)^t |Y g(Y) - \varphi(X)|]$ ,  $D_{\widehat{g}_n}(\cdot) = \widehat{g}_n(\cdot) - g(\cdot)$  and  $D_{\widehat{\varphi}_n}(\cdot) = \widehat{\varphi}_n(\cdot) - \varphi(\cdot)$  and remark that  $\mathbf{E}\|\widehat{\mathfrak{E}}_n - \mathfrak{E}_n\|^2 \leq Cm_n^2/n$  under Assumption 7 (i). Hence,

$$\begin{aligned} \left| \widehat{h}_1^t (\widehat{\Sigma}_n - \widehat{\Sigma}_n) \widehat{h}_1 \right| &= \left| n^{-1} \sum_i \left( \widehat{h}_1^t f_{m_n}(X_i) \right)^2 \left\{ \left( Y_i \widehat{g}_n(Y_i) - \widehat{\varphi}_n(X_i) \right)^2 - \left( Y_i g(Y_i) - \varphi(X_i) \right)^2 \right\} \right| \\ &\leq \left| n^{-1} \sum_i \left( \widehat{h}_1^t f_{m_n}(X_i) \right)^2 \left( Y_i D_{\widehat{g}_n}(Y_i) - D_{\widehat{\varphi}_n}(X_i) \right)^2 \right| \\ &\quad + 2 \left| n^{-1} \sum_i \left( \widehat{h}_1^t f_{m_n}(X_i) \right)^2 \left( Y_i g(Y_i) - \varphi(X_i) \right) \left( Y_i D_{\widehat{g}_n}(Y_i) - D_{\widehat{\varphi}_n}(X_i) \right) \right| \\ &\leq o_p(1) \widehat{h}_1^t [\widehat{\mathbf{Q}}_n - \mathbf{I}_{m_n}] \widehat{h}_1 + 2o_p(1) \widehat{h}_1^t [\widehat{\mathfrak{E}}_n - \mathfrak{E}_n] \widehat{h}_1 + o_p(1) \|\widehat{h}_1\|^2 + o_p(1) \widehat{h}_1^t \mathfrak{E}_n \widehat{h}_1 \end{aligned} \quad (\text{A.6})$$

since  $\sup_{y \in \mathcal{Y}} |y|^2 (\sup_{y \in \mathcal{Y}} |D_{\widehat{g}_n}(y)|)^2 = O_p\left(\max\left(k_n^2/(n\tau_{k_n}), k_n^{-2\beta+1}\right)\right)$  and

$$\left( \sup_{x \in \mathcal{X}} |D_{\widehat{\varphi}_n}(x)| \right)^2 = O_p\left(\max\left(m_n^{-2\alpha/d_x+1}, \frac{m_n^2}{n}, m_n \|T(E_{k_n} g - g)\|_X^2\right)\right)$$

by the Cauchy Schwartz inequality and both converge to zero under the assumptions of the theorem.

Next, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| \widehat{h}_2^t \widehat{\Sigma}_{n,f} \widehat{h}_2 - h_2^t \Sigma_{n,f} h_2 \right| &\leq |(\widehat{h}_2 - h_2)^t \Sigma_{n,f} (\widehat{h}_2 - h_2)| + 2\widehat{h}_2^t \Sigma_{n,f} \widehat{h}_2 - 2h_2^t \Sigma_{n,f} h_2 \\ &\leq |(\widehat{h}_2 - h_2)^t \Sigma_{n,f} (\widehat{h}_2 - h_2)| + 2|(\widehat{h}_2 - h_2)^t \Sigma_{n,f} h_2| \\ &= o_p(1). \end{aligned} \quad (\text{A.7})$$

Furthermore, by  $\mathbf{E}\|\widehat{\Sigma}_{n,f} - \Sigma_{n,f}\|^2 \leq Cm_n^2/n$ ,

$$\left| \widehat{h}_2^t (\widehat{\Sigma}_{n,f} - \Sigma_{n,f}) \widehat{h}_2 \right| \leq \left\| \widehat{\Sigma}_{n,f} - \Sigma_{n,f} \right\| \|\widehat{h}_2\|^2 = o_p(1). \quad (\text{A.8})$$

Then, by (A.3), the triangle inequality and (A.5)–(A.8), the result of the lemma follows.  $\square$

## A.2. Proofs of Section 4.

**PROOF OF THEOREM 4.1.** The proof is based on the inequality

$$\|\widehat{\psi}_n - \psi\|_Z \leq \|\Pi_{k_n} \psi - \psi\|_Z + \|\widehat{\psi}_n - \Pi_{k_n} \psi\|_Z.$$

By assumption, we have  $\|\Pi_{k_n}\psi - \psi\|_Z = O(k_n^{-\gamma/d_z})$  and thus, it is sufficient to bound  $\|\widehat{\psi}_n - \Pi_{k_n}\psi\|_Z$ . By Lemma B.2 of Chen and Pouzo [2012] it holds  $\|\widehat{\psi}_n - \Pi_{k_n}\psi\|_Z^2 \leq C\kappa_{k_n}^{-1}\|K(\widehat{\psi}_n - \Pi_{k_n}\psi)\|_X^2$ . From the proof of Theorem 3.1 we have

$$\sum_{j=1}^{m_n} \left| n^{-1} \sum_i Y_i \widehat{g}_n(Y_i) f_j(X_i) - \mathbf{E}[Yg(Y)f_j(X)] \right|^2 = O_p(r_n) \quad (\text{A.9})$$

where we denote  $r_n = \max(m_n^{-2\alpha/d_x}, n^{-1}m_n, \|T(g - E_{k_n}g)\|_X^2)$ . Consequently, we observe

$$\begin{aligned} \left\| n^{-1} \sum_i (Y_i \widehat{g}_n(Y_i) - (\Pi_{k_n}\psi)(Z_i)) p_{\underline{m}_n}(X_i) \right\|^2 \\ \leq 2 \left\| \mathbf{E} \left[ (Yg(Y) - (\Pi_{k_n}\psi)(Z)) p_{\underline{m}_n}(X) \right] \right\|^2 + O_p(r_n) \\ \leq 2 \|K(\Pi_{k_n}\psi - \psi)\|_X^2 + O_p(r_n + \|F_{\underline{m}_n}^\perp K(\Pi_{k_n}\psi - \psi)\|_X^2). \end{aligned}$$

Further, using the elementary inequality  $(a - b)^2 \geq a^2/2 - b^2$  and again applying relation (A.9) gives uniformly in  $\phi$

$$\begin{aligned} \left\| n^{-1} \sum_i (Y_i \widehat{g}_n(Y_i) - \phi(Z_i)) f_{\underline{m}_n}(X_i) \right\|^2 &\geq \left\| \mathbf{E} \left[ (Y^* - \phi(Z)) f_{\underline{m}_n}(X) \right] \right\|^2 / 2 \\ &\quad - \sum_{j=1}^{k_n} \max_{\phi \in \mathcal{B}_{k_n}} \left| n^{-1} \sum_i (Y_i \widehat{g}_n(Y_i) - \phi(Z_i)) f_j(X_i) - \mathbf{E} \left[ (Y^* - \phi(Z)) f_j(X) \right] \right|^2 \\ &\geq C \|K(\Pi_{k_n}\psi - \psi)\|_X^2 - O_p(r_n + \|F_{\underline{m}_n}^\perp K(\Pi_{k_n}\psi - \psi)\|_X^2). \end{aligned}$$

For some  $\varepsilon > 0$  let us denote  $\widetilde{\Psi}_n = \{\phi \in \Psi_n : \|K(\phi - \psi)\|_X^2 \geq \varepsilon \widetilde{r}_n\}$  where  $\widetilde{r}_n = r_n + \|F_{\underline{m}_n}^\perp K(\Pi_{k_n}\psi - \psi)\|_X^2$ . Therefore, following the proof of Lemma B.1 of Chen and Pouzo [2012] we obtain

$$\begin{aligned} \mathbb{P}(\|K(\widehat{\psi}_n - \psi)\|_X^2 \geq \varepsilon \tau_n) \\ \leq \mathbb{P}\left( \min_{\phi \in \widetilde{\Psi}_n} \left\| \sum_i (Y_i \widehat{g}_n(Y_i) - \phi(Z_i)) f_{\underline{m}_n}(X_i) \right\|^2 \leq \left\| \sum_i (Y_i \widehat{g}_n(Y_i) - (\Pi_{k_n}\psi)(Z_i)) f_{\underline{m}_n}(X_i) \right\|^2 \right) \\ \leq \mathbb{P}\left( \min_{\phi \in \widetilde{\Psi}_n} \|K(\phi - \psi)\|_X^2 \leq \|K(\Pi_{k_n}\psi - \psi)\|_X^2 + O_p(\widetilde{r}_n) \right) \end{aligned}$$

which goes to zero for all  $n \geq 1$  as  $\varepsilon \rightarrow \infty$ . This shows  $\|K(\widehat{\psi}_n - \Pi_{k_n}\psi)\|_X = O_p(\widetilde{r}_n)$  and thus proves the result.  $\square$

**PROOF OF THEOREM 4.2.** Observe that the asymptotic variance  $\mathcal{W}_n(z)$  is bounded from below. Indeed, from condition  $\text{Var}(Yg(Y) - \psi(Z)|X) \geq C$  we infer

$$\begin{aligned} \mathcal{W}_n(z) &\geq p_{\underline{k}_n}(z)^t (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{E} \left[ f_{\underline{m}_n}(X) \text{Var}(Yg(Y) - \psi(Z)|X) f_{\underline{m}_n}(X)^t \right] \mathbf{K}_n (\mathbf{K}_n^t \mathbf{K}_n)^{-1} p_{\underline{k}_n}(z) \\ &\geq C p_{\underline{k}_n}(z)^t (\mathbf{K}_n^t \mathbf{K}_n)^{-1} p_{\underline{k}_n}(z) \end{aligned}$$



which we use in the following. Let  $\widehat{A}_n^{-1} = (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1} \mathbf{X}_{m_n}^t \mathbf{Z}_{k_n} / n)^{-1} \mathbf{Z}_{k_n}^t \mathbf{X}_{m_n} (\mathbf{X}_{m_n}^t \mathbf{X}_{m_n})^{-1}$  and  $A_n^{-1} = (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t$ . Then, we make use of the decomposition

$$\begin{aligned} \widehat{\psi}_n(z) - \psi(z) &= p_{\underline{k}_n}(z)^t n^{-1} \widehat{A}_n^{-1} \sum_i f_{\underline{m}_n}(X_i) (Y_i g(Y_i) - \psi(Z_i)) \\ &\quad + p_{\underline{k}_n}(z)^t n^{-1} \widehat{A}_n^{-1} \sum_i f_{\underline{m}_n}(X_i) Y_i (\widehat{g}_n(Y_i) - E_{k_n} g(Y_i)) \\ &\quad + p_{\underline{k}_n}(z)^t n^{-1} \widehat{A}_n^{-1} \sum_i f_{\underline{m}_n}(X_i) Y_i (E_{k_n} g(Y_i) - g(Y_i)) \\ &\quad + p_{\underline{k}_n}(z)^t n^{-1} \widehat{A}_n^{-1} \sum_i f_{\underline{m}_n}(X_i) \psi(Z_i) - \psi(z) \\ &= I_n + II_n + III_n + IV_n \quad (\text{say}). \end{aligned}$$

Let us denote  $\vartheta_{k_n} = (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{E}[Y g(Y) f_{\underline{m}_n}(X)]$ . As in the proof of Theorem 3.1 and since  $\mathbf{E}[f_{\underline{m}_n}(X)] = \mathbf{E}[\Delta g(Y) f_{\underline{m}_n}(X)]$  it is easily seen

$$\begin{aligned} II_n &= \sqrt{n} p_{\underline{k}_n}(z)^t (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\widehat{\vartheta}_{k_n} - \vartheta_{k_n}) \\ &\quad + \sqrt{n} p_{\underline{m}_n}(z)^t (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t \mathbf{E} [f_{\underline{m}_n}(X) \Delta (g(Y) - E_{k_n} g(Y))] + o_p(1). \end{aligned}$$

Since the largest eigenvalue of  $\mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t$  is bounded and Assumption 5 (iv) together with  $n \|T(g - E_{k_n} g)\|_X^2 = o(1)$  the Cauchy Schwarz inequality implies

$$II_n = \sqrt{n} p_{\underline{k}_n}(z)^t (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\widehat{\vartheta}_{k_n} - \vartheta_{k_n}) + o_p(1).$$

Consequently, we obtain

$$\begin{aligned} \sqrt{n/\mathcal{W}_n(z)} (I_n + II_n) &= \sum_i (n \mathcal{W}_n(z))^{-1/2} p_{\underline{k}_n}(z)^t A_n^{-1} (f_{\underline{m}_n}(X_i) (Y_i g(Y_i) - \psi(Z_i)) \\ &\quad - (\mathbf{K}_n^t \mathbf{K}_n)^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t (f_{\underline{m}_n}(X_i) - \mathbf{E}[f_{\underline{m}_n}(X)])) + o_p(1) \\ &= \sum_i s_{in} + o_p(1) \end{aligned}$$

Moreover,  $s_{in}$ ,  $1 \leq i \leq n$  satisfy the Lindeberg condition which can be seen as follows. It holds  $\mathbf{E}[s_{in}] = 0$  and  $n \mathbf{E}[s_{in}^2] = 1$ . From Assumption 5 (iv) we infer  $\|\mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t\| \leq C \|\mathbf{T}_n (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t\| = C$  and due to  $\mathbf{E}|Y g(Y) - \psi(X)|^4 \leq C$  we observe

$$\sum_i \mathbf{E}[s_{in}^2 \mathbb{1}_{\{|s_{in}| > \varepsilon\}}] \leq n \varepsilon^2 \mathbf{E}|s_{in}/\varepsilon|^4 \leq C n^{-1} \varepsilon^{-2} m_n^2 = o(1).$$

Again from Assumption 5 (iv) together with  $n \|T(g - E_{k_n} g)\|_X^2 = o(1)$  yields  $n III_n = o_p(1)$ . Finally,  $n IV_n = o_p(1)$  follows from  $n \|II_n \psi - \psi\|_Z^2 = o(1)$ , which completes the proof of the first statement in the theorem.

To prove the second statement of the theorem, remark that by Lemma A.2,  $\mathcal{W}_n(z)^{-1}\widehat{\mathcal{W}}_n(z) = 1 + o_p(1)$ . Therefore, we obtain

$$\sqrt{n/\widehat{\mathcal{W}}_n(z)}(\widehat{\psi}_n(z) - \psi(z)) = \sqrt{n/\mathcal{W}_n(z)}(\widehat{\psi}_n(z) - \psi(z))(\mathcal{W}_n(z)/\widehat{\mathcal{W}}_n(z))^{1/2} \xrightarrow{d} \mathcal{N}(0, 1)$$

which proves the second result of the theorem.  $\square$

**LEMMA A.2.** *Let Assumptions 1 – 5, and 8 – 10 be satisfied. Moreover, assume that  $\mathcal{Y}$  is bounded and  $m_n = o(\sqrt{n} \min(\sqrt{\tau_{k_n}} \kappa_{k_n}^2, \sqrt{\kappa_{k_n}} \tau_{k_n}^2))$ . Then,*

$$\mathcal{W}_n(z)^{-1}\widehat{\mathcal{W}}_n(z) = 1 + o_p(1)$$

where  $\mathcal{W}_n(z)$  and  $\widehat{\mathcal{W}}_n(z)$  are as defined in Theorem 3.3.

*Proof.* Let  $\Sigma_{n,f}$  and  $\widehat{\Sigma}_{n,f}$  as defined in the proof of Lemma A.1. We denote  $\Sigma_n^\psi = \mathbf{E} \left[ f_{m_n}(X) \text{Var}(Yg(Y) - \psi(Z)|X) f_{m_n}(X)^t \right]$ ,  $\widehat{\Sigma}_n^\psi = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t (Y_i \widehat{g}_n(Y_i) - \widehat{\psi}_n(Z_i))^2$  and  $\widetilde{\Sigma}_n^\psi = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t (Y_i g(Y_i) - \psi(X_i))^2$ . Moreover, let  $\widehat{\mathcal{A}}_n = (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n}/n \widehat{\mathbf{Q}}_n^{-1} \mathbf{X}_{m_n}^t \mathbf{Z}_{k_n}/n)$ ,  $\mathcal{A}_n = \mathbf{K}_n^t \mathbf{K}_n$ ,

$$\widehat{h}_3^t = (\mathcal{W}_n(z))^{-1/2} p_{k_n}(z)^t \widehat{\mathcal{A}}_n^{-1} (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n}/n) \widehat{\mathbf{Q}}_n^{-1}, \quad h_3^t = (\mathcal{W}_n(z))^{-1/2} p_{k_n}(z)^t \mathcal{A}_n^{-1} \mathbf{K}_n^t,$$

$$\widehat{h}_4^t = (\mathcal{W}_n(z))^{-1/2} p_{k_n}(z)^t \widehat{\mathcal{A}}_n^{-1} (\mathbf{Z}_{k_n}^t \mathbf{X}_{m_n}/n) \mathbf{X}_{m_n}^t \text{diag}(\mathbf{Y}) \mathbf{Y}_{k_n} (\widehat{\mathbf{T}}_n \widehat{\mathbf{Q}}_n^{-1} \widehat{\mathbf{T}}_n)^{-1} \widehat{\mathbf{T}}_n \widehat{\mathbf{Q}}_n^{-1}$$

and  $h_4^t = (\mathcal{W}_n(z))^{-1/2} p_{k_n}(z)^t \mathcal{A}_n^{-1} \mathbf{K}_n^t \mathbf{T}_n^Y (\mathbf{T}_n^t \mathbf{T}_n)^{-1} \mathbf{T}_n^t$ . Hence,  $\mathcal{W}_n(z) (\widehat{h}_3^t \widehat{\Sigma}_n^\psi \widehat{h}_3 + \widehat{h}_4^t \widehat{\Sigma}_{n,f} \widehat{h}_4) = \widehat{\mathcal{W}}_n(z)$  and by noticing that  $h_3^t \Sigma_n^\psi h_3 + h_4^t \Sigma_{n,f} h_4 = 1$  and by the triangle inequality

$$\begin{aligned} \left| (\mathcal{W}_n(z))^{-1/2} \widehat{\mathcal{W}}_n(z) (\mathcal{W}_n(z))^{-1/2} - 1 \right| &\leq \left| \widehat{h}_3^t (\widehat{\Sigma}_n^\psi - \widetilde{\Sigma}_n^\psi) \widehat{h}_3 \right| + \left| \widehat{h}_3^t (\widetilde{\Sigma}_n^\psi - \Sigma_n^\psi) \widehat{h}_3 \right| \\ &\quad + \left| \widehat{h}_3^t \Sigma_n^\psi \widehat{h}_3 - h_3^t \Sigma_n^\psi h_3 \right| + \left| \widehat{h}_4^t (\widehat{\Sigma}_{n,f} - \Sigma_{n,f}) \widehat{h}_4 \right| + \left| \widehat{h}_4^t \Sigma_{n,f} \widehat{h}_4 - h_4^t \Sigma_{n,f} h_4 \right|. \end{aligned} \quad (\text{A.10})$$

Remark that

$$\|\widehat{\mathcal{A}}_n^{-1} - \mathcal{A}_n^{-1}\| \leq \|\widehat{\mathcal{A}}_n^{-1}\| \|\widehat{\mathcal{A}}_n - \mathcal{A}_n\| \|\mathcal{A}_n^{-1}\| = O_p(\max(\frac{\sqrt{k_n m_n}}{\sqrt{n}}, \frac{m_n}{\sqrt{n}}) \kappa_{k_n}^{-2})$$

since  $\|\widehat{\mathcal{A}}_n^{-1}\|^2 \leq 2\|\mathcal{A}_n^{-1}\|^2 + o_p(1)$ ,  $\|\mathcal{A}_n^{-1}\|^2 = \lambda_n^{-1} \leq C \kappa_{k_n}^{-2}$ ,  $\|\mathbf{X}_{m_n}^t \mathbf{Z}_{k_n}/n - \mathbf{K}_n\|^2 = O_p(k_n m_n/n)$  and  $\|\widehat{\mathbf{Q}}_n^{-1} - \mathbf{I}\|^2 = O_p(m_n^2/n)$ . Then,  $\|\widehat{h}_3 - h_3\| = o_p(1)$  and  $\|\widehat{h}_3\| = O_p(1)$ . By these results, (A.4),  $\|\mathbf{X}_{m_n}^t \text{diag}(\mathbf{Y}) \mathbf{Y}_{k_n}/n - \mathbf{T}_n^Y\|^2 = O_p(m_n k_n/n)$ ,  $\|\widehat{\mathbf{T}}_n - \mathbf{T}_n\|^2 = O_p(m_n k_n/n)$ , Assumptions 4 (ii) and 5 (ii), (iv) and the assumptions of the lemma we conclude that  $\|\widehat{h}_4 - h_4\|^2 = o_p(1)$  and so  $\|\widehat{h}_4\| = O_p(1)$ .

Under Assumptions 4 (ii) – (iii) and 10 (i), we can show similarly as in Newey [1997] page 165 – 166 that:

$$\left| \widehat{h}_3^t \Sigma_n^\psi \widehat{h}_3 - h_3^t \Sigma_n^\psi h_3 \right| = o_p(1) \quad \text{and} \quad \left| \widehat{h}_3^t (\widetilde{\Sigma}_n^\psi - \Sigma_n^\psi) \widehat{h}_3 \right| = o_p(1). \quad (\text{A.11})$$

Moreover, denote  $\widehat{\xi}_i^\psi = (Y_i \widehat{g}_n(Y_i) - \widehat{\psi}_n(X_i))$ ,  $\xi_i^\psi = (Y_i g(Y_i) - \psi(X_i))$ ,  $\widehat{S}_n = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t |\xi_i^\psi|$ ,  $S_n^\psi = \mathbf{E}[f_{m_n}(X) f_{m_n}(X)^t |\xi^\psi|]$ ,  $D_{\widehat{g}_n}(Y_i) = \widehat{g}_n(Y_i) - g(Y_i)$  and  $D_{\widehat{\psi}_n}(X_i) = \widehat{\psi}_n(X_i) - \psi(X_i)$  and remark that  $\mathbf{E} \|\widehat{S}_n^\psi - S_n^\psi\|^2 \leq Cm_n^2/n$  under Assumption 10 (i). Hence,

$$\begin{aligned} \left| \widehat{h}_3^t (\widehat{\Sigma}_n^\psi - \widetilde{\Sigma}_n^\psi) \widehat{h}_3 \right| &= \left| n^{-1} \sum_{i=1}^n (\widehat{h}_3^t f_{m_n}(X_i))^2 ((\widehat{\xi}_i^\psi)^2 - (\xi_i^\psi)^2) \right| \\ &\leq \left| n^{-1} \sum_{i=1}^n (\widehat{h}_3^t f_{m_n}(X_i))^2 (Y_i D_{\widehat{g}_n}(Y_i) - D_{\widehat{\psi}_n}(X_i))^2 \right| + 2 \left| n^{-1} \sum_{i=1}^n (\widehat{h}_3^t f_{m_n}(X_i))^2 \xi_i (Y_i D_{\widehat{g}_n}(Y_i) - D_{\widehat{\psi}_n}(X_i)) \right| \\ &\leq o_p(1) \widehat{h}_3^t [\widehat{Q}_n - I_{m_n}] \widehat{h}_3 + 2o_p(1) \widehat{h}_3^t [\widehat{S}_n^\psi - S_n^\psi] \widehat{h}_3 + o_p(1) \|\widehat{h}_3\|^2 + o_p(1) \widehat{h}_3^t S_n^\psi \widehat{h}_3 \end{aligned} \quad (\text{A.12})$$

since  $\sup_{y \in \mathcal{Y}} |y|^2 (\sup_{y \in \mathcal{Y}} |D_{\widehat{g}_n}(y)|)^2 = O_p\left(\max\left(k_n^2/(n\tau_{k_n}), k_n^{-2\beta+1}\right)\right)$  and

$$\left( \sup_{z \in \mathcal{Z}} |D_{\widehat{\psi}_n}(x)| \right)^2 = O_p\left(\max\left(k_n^{-2\gamma/d_z+1}, \frac{k_n^2}{n\kappa_{k_n}}, k_n \kappa_{k_n}^{-1} \|T(E_{k_n} g - g)\|_X^2\right)\right)$$

and both converge to zero under the assumptions of the lemma. Next, by the Cauchy-Schwarz inequality and similarly as in the proof of Lemma A.1 we get

$$\begin{aligned} |\widehat{h}_4^t \Sigma_{n,f} \widehat{h}_4 - h_4^t \Sigma_{n,f} h_4| &\leq \|\widehat{h}_4 - h_4\|^2 + 2 \left[ (\widehat{h}_4 - h_4)^t \Sigma_{n,f} (\widehat{h}_4 - h_4) \right]^{1/2} \left[ h_4^t \Sigma_{n,f} h_4 \right]^{1/2} \\ &\leq C \|\widehat{h}_4 - h_4\|^2 \\ &= o_p(1). \end{aligned} \quad (\text{A.13})$$

Furthermore, by  $\mathbf{E} \|\widehat{\Sigma}_{n,f} - \Sigma_{n,f}\|^2 \leq Cm_n^2/n$ ,

$$|\widehat{h}_4^t (\widehat{\Sigma}_{n,f} - \Sigma_{n,f}) \widehat{h}_4| \leq \left\| \widehat{\Sigma}_{n,f} - \Sigma_{n,f} \right\| \|\widehat{h}_4\|^2 = o_p(1). \quad (\text{A.14})$$

Then, by (A.10) and (A.11)-(A.14), the result of the lemma follows.  $\square$

### A.3. Proofs of Section 5.

**PROOF OF THEOREM 5.1.** Since we have  $\|\widehat{Q}_n - I_{m_n}\|^2 = o_p(m_n^2/n)$  it is sufficient to prove that  $(\sqrt{2\zeta_{m_n}})^{-1} \left( \sum_{j=1}^{m_n} |n^{-1/2} \sum_i (\Delta_i \widehat{g}_n(Y_i) - 1) f_j(X_i)|^2 - \mu_{m_n} \right) \xrightarrow{d} \mathcal{N}(0, 1)$ . The proof of this statement is based on the decomposition

$$\begin{aligned} \sum_{j=1}^{m_n} \left| n^{-1} \sum_i (\Delta_i \widehat{g}_n(Y_i) - 1) f_j(X_i) \right|^2 &= \sum_{j=1}^{m_n} \left| n^{-1} \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right|^2 \\ &- \frac{2}{n^2} \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \left( \sum_i \Delta_i (\widehat{g}_n(Y_i) - g(Y_i)) f_j(X_i) \right) \\ &+ \sum_{j=1}^{m_n} \left| n^{-1} \sum_i \Delta_i (\widehat{g}_n(Y_i) - g(Y_i)) f_j(X_i) \right|^2 = I_n - 2II_n + III_n. \end{aligned} \quad (\text{A.15})$$

Consider  $I_n$ . We calculate

$$\begin{aligned} \varsigma_{m_n}^{-1}(nI_n - \mu_{m_n}) &= \frac{1}{\varsigma_{m_n} n} \sum_i \sum_{j=1}^{m_n} \left( |(\Delta_i g(Y_i) - 1) f_j(X_i)|^2 - \mathbf{E} \left[ (\Delta g(Y) - 1)^2 f_j^2(X) \right] \right) \\ &\quad + \frac{1}{\varsigma_{m_n} n} \sum_{i \neq i'} \sum_{j=1}^{m_n} (\Delta_i g(Y_i) - 1) (\Delta_{i'} g(Y_{i'}) - 1) f_j(X_i) f_j(X_{i'}) \end{aligned}$$

where the first summand tends in probability to zero as  $n \rightarrow \infty$ . Indeed, we have

$$\begin{aligned} \mathbf{E} \left| \frac{1}{\varsigma_{m_n} n} \sum_i \sum_{j=1}^{m_n} \left( |(\Delta_i g(Y_i) - 1) f_j(X_i)|^2 - \mathbf{E} \left[ (\Delta g(Y) - 1)^2 f_j^2(X) \right] \right) \right|^2 \\ \leq \frac{1}{\varsigma_{m_n}^2 n} \mathbf{E} \left| \sum_{j=1}^{m_n} \left( |(\Delta g(Y) - 1) f_j(X)|^2 - \mathbf{E} \left[ (\Delta g(Y) - 1)^2 f_j^2(X) \right] \right) \right|^2 \\ \leq \frac{1}{\varsigma_{m_n}^2 n} \sup_{x \in \mathcal{X}} \|f_{m_n}(x)\|^4 \mathbf{E} |\Delta g(Y) - 1|^4 \\ \leq \frac{C m_n^2}{\varsigma_{m_n}^2 n} = o(1). \end{aligned}$$

Therefore, to establish  $(\sqrt{2}\varsigma_{m_n})^{-1}(nI_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1)$  it is sufficient to show

$$\frac{\sqrt{2}}{\varsigma_{m_n} n} \sum_{i \neq i'} \sum_{j=1}^{m_n} (\Delta_i g(Y_i) - 1) (\Delta_{i'} g(Y_{i'}) - 1) f_j(X_i) f_j(X_{i'}) \xrightarrow{d} \mathcal{N}(0, 1).$$

This follows from Lemma A.2 of Breunig [2015].

Consider  $II_n$ . We observe

$$\begin{aligned} nII_n &= \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \left( n^{-1} \sum_i \Delta_i (\widehat{g}_n(Y_i) - g(Y_i)) f_j(X_i) \right) \\ &= \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \left( n^{-1} \sum_i \Delta_i (\widehat{g}_n(Y_i) - E_{k_n} g(Y_i)) f_j(X_i) \right) \\ &\quad + \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \left( n^{-1} \sum_i \Delta_i (E_{k_n} g_n(Y_i) - g(Y_i)) f_j(X_i) \right) \\ &= C_{n1} + C_{n2}. \end{aligned}$$

Consider  $C_{n1}$ . We have

$$\begin{aligned} C_{n1} &= \| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{1/2} (\widehat{\beta}_{k_n} - \beta_{k_n}) \| \\ &\quad \times \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{-1/2} \mathbf{E} [\Delta e_{k_n}(Y) f_j(X)] \| + o_p(1). \end{aligned}$$

Using

$$\begin{aligned}
& \mathbf{E} \left| \sum_{j=1}^{m_n} \left( \sum_i (\Delta_i g(Y_i) - 1) f_j(X_i) \right) \right\| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{-1/2} \mathbf{E}[\Delta e_{\underline{k}_n}(Y) f_j(X)] \Big\|^2 \\
& \leq n \sum_{j=1}^{m_n} \mathbf{E} \left[ \left( (\Delta g(Y) - 1) f_j(X) \right)^2 \right] \left\| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{-1/2} \mathbf{E}[\Delta e_{\underline{k}_n}(Y) f_j(X)] \right\|^2 \\
& \leq Cn \left\| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{-1/2} \mathbf{E}[\Delta e_{\underline{k}_n}(Y) f_{\underline{m}_n}(X)^t] \right\|^2.
\end{aligned}$$

Since  $\left\| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{-1/2} \mathbf{E}[\Delta e_{\underline{k}_n}(Y) f_{\underline{m}_n}(X)^t] \right\|^2 = O(k_n)$  it holds

$$C_{n1} = \left\| \text{Diag}(\tau_1, \dots, \tau_{k_n})^{1/2} (\widehat{\beta}_{k_n} - \beta_{k_n}) \right\| O_p(\sqrt{k_n n}) = O_p(k_n) = o_p(\sqrt{m_n}).$$

Further, we have

$$\begin{aligned}
\mathbf{E}|C_{n2}| & \leq \sum_{j=1}^{m_n} \sqrt{\mathbf{E}[(E_{k_n} g - g)(Y) f_j(X)]^2} \sqrt{\mathbf{E}[(\Delta_i g(Y) - 1) f_j(X)]^2} \\
& \quad + Cn^{1/2} \mathbf{E} \left| \sum_{j=1}^{m_n} \langle T(E_{k_n} g - g), f_j \rangle_X f_j(X) \right| \\
& \leq C \left( m_n \|E_{k_n} g - g\|_\infty + \sqrt{n} \|T(E_{k_n} g - g)\|_X \right) \\
& = O(m_n k_n^{1/2-\beta} + \sqrt{n \tau_{k_n}} k_n^{-\beta}) = o(\sqrt{m_n})
\end{aligned}$$

where we used that  $m_n k_n^{1-2\beta} = o(1)$ . Consider  $III_n$ . It holds true that

$$\begin{aligned}
III_n & \leq Cn \left\| (\widehat{\beta}_{k_n} - \beta_{k_n})^t \mathbf{E}[\Delta e_{\underline{k}_n}(Y) f_{\underline{m}_n}(X)^t] \right\|^2 \\
& \quad + Cn \|T(E_{k_n} g - g)\|_X^2 \\
& = O_p(k_n + n \tau_{k_n} k_n^{-2\beta}) \\
& = o_p(\sqrt{m_n})
\end{aligned}$$

which completes the proof of the first result in the theorem.  $\square$

**PROOF OF THEOREM 5.2.** Remark that  $(\sqrt{2\widehat{\varsigma}_{m_n}})^{-1}(nS_n - \widehat{\mu}_{m_n}) = (\sqrt{2\varsigma_{m_n}})^{-1}(nS_n - \mu_{m_n}) \frac{\widehat{\varsigma}_{m_n}}{\varsigma_{m_n}} + (\sqrt{2\varsigma_{m_n}})^{-1}(\mu_{m_n} - \widehat{\mu}_{m_n}) \frac{\widehat{\varsigma}_{m_n}}{\varsigma_{m_n}}$ . The statement of the theorem follows from the results of Theorem 5.1, Lemma A.3 and Lemma A.4.  $\square$

**LEMMA A.3.** Let Assumptions 1 – 6 be satisfied. Then,

$$\varsigma_{m_n}^{-1} \widehat{\varsigma}_{m_n} = 1 + o_p(1)$$

where  $\varsigma_{m_n}$  and  $\widehat{\varsigma}_{m_n}$  are as defined in Theorem 5.1.

*Proof.* Let  $\|\cdot\|_F$  denote the Frobenius norm of a matrix. Then  $\varsigma_{m_n} = \|\Sigma_{m_n}\|_F$  and  $\widehat{\varsigma}_{m_n} = \|\widehat{\Sigma}_{m_n}\|_F$ . Let us denote  $\widetilde{\Sigma}_{m_n} = n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t (E_{k_n} g(Y_i) \Delta_i - 1)^2$ . Observe that

$$\begin{aligned}
& \|\widehat{\Sigma}_{m_n} - \widetilde{\Sigma}_{m_n}\|_F^2 \\
&= \left\| n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t \left[ |(\widehat{g}_n - E_{k_n} g)(Y_i)|^2 \Delta_i + 2(E_{k_n} g(Y_i) \Delta_i - 1) \Delta_i (\widehat{g}_n - E_{k_n} g)(Y_i) \right] \right\|_F^2 \\
&\leq 2 \left\| n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t |(\widehat{g}_n - E_{k_n} g)(Y_i)|^2 \Delta_i \right\|_F^2 \\
&+ 2 \left\| n^{-1} \sum_i f_{m_n}(X_i) f_{m_n}(X_i)^t (E_{k_n} g(Y_i) \Delta_i - 1) \Delta_i (\widehat{g}_n - E_{k_n} g)(Y_i) \right\|_F^2 \\
&= 2I_n + 2II_n.
\end{aligned}$$

We further calculate

$$\begin{aligned}
I_n &\leq \left\| \frac{1}{n} \sum_i \Delta_i (\widehat{\beta}_{k_n} - \beta_{k_n})^t e_{k_n}(Y_i) f_{m_n}(X_i) f_{m_n}(X_i)^t e_{k_n}(Y_i)^t (\widehat{\beta}_{k_n} - \beta_{k_n}) \right\|_F^2 \\
&\leq \left\| (\widehat{\beta}_{k_n} - \beta_{k_n})^t \mathbf{E}[\Delta e_{k_n}(Y) f_{m_n}(X) f_{m_n}(X)^t e_{k_n}(Y)^t] (\widehat{\beta}_{k_n} - \beta_{k_n}) \right\|_F^2 + o_p(1) \\
&\leq \|\widehat{\beta}_{k_n} - \beta_{k_n}\|^4 \sum_{j,l=1}^{m_n} \mathbf{E}[\|\Delta e_{k_n}(Y)\|^2 |f_j(X) f_l(X)|^2] + o_p(1) \\
&\leq C m_n^2 \|\widehat{\beta}_{k_n} - \beta_{k_n}\|^4 \mathbf{E}[\|\Delta e_{k_n}(Y)\|^2]^2 + o_p(1) \\
&= O_p(m_n^2 k_n^4 / (\tau_{k_n} n)^2) = o_p(1)
\end{aligned}$$

by using  $k_n = o(\sqrt{m_n})$ . Similarly, we conclude

$$\begin{aligned}
II_n &\leq \left\| \frac{1}{n} \sum_i (E_{k_n} g(Y_i) \Delta_i - 1) f_{m_n}(X_i) f_{m_n}(X_i)^t e_{k_n}(Y_i)^t (\widehat{\beta}_{k_n} - \beta_{k_n}) \right\|_F^2 \\
&\leq \left\| \mathbf{E}[(E_{k_n} g(Y) \Delta - 1) \Delta f_{m_n}(X) f_{m_n}(X)^t e_{k_n}(Y)^t] (\widehat{\beta}_{k_n} - \beta_{k_n}) \right\|_F^2 + o_p(1) \\
&\leq \|\widehat{\beta}_{k_n} - \beta_{k_n}\|^2 \sum_{j,l=1}^{m_n} \sum_{l'=1}^{k_n} \left( \mathbf{E}[(E_{k_n} g(Y) \Delta - 1) \Delta e_{l'}(Y) f_j(X) f_l(X)] \right)^2 + o_p(1) \\
&= \|\widehat{\beta}_{k_n} - \beta_{k_n}\|^2 \sum_{j=1}^{m_n} \sum_{l'=1}^{k_n} \|F_{m_n} T((E_{k_n} g - 1) \cdot e_{l'}) \cdot f_j\|_X^2 + o_p(1) \\
&= O_p(m_n k_n^2 / (\tau_{k_n} n)) = o_p(1)
\end{aligned}$$

by using  $k_n = o(\sqrt{m_n})$ . Next,

$$\begin{aligned} \mathbf{E} \|\widehat{\Sigma}_{m_n} - \Sigma_{m_n}\|_F^2 &= \sum_{j,l=1}^{m_n} \mathbf{E} \left( \frac{1}{n} \sum_i (f_j(X_i) f_l(X_i) (E_{k_n} g(Y_i) \Delta_i - 1)^2 - \mathbf{E} [f_j(X) f_l(X) (g(Y) \Delta - 1)^2]) \right)^2 \\ &\leq \frac{1}{n} \sum_{j,l=1}^{m_n} \mathbf{E} [f_j^2(X) f_l^2(X) (E_{k_n} g(Y) \Delta - 1)^4] + \sum_{j,l=1}^{m_n} \left( \mathbf{E} [f_j(X) f_l(X) \Delta (E_{k_n} g(Y) - g(Y))^2] \right)^2 \\ &\leq C \frac{m_n^2}{n} + \sum_{j=1}^{m_n} \|F_{m_n} T(E_{k_n} g - g) \cdot f_j\|_X^2 = O(m_n^2 n^{-1} + m_n k_n^{1-2\beta}) = o(1) \end{aligned}$$

Finally, by these results and the reverse triangle inequality we conclude that

$$\left| \zeta_{m_n}^{-1} \widehat{\zeta}_{m_n} - 1 \right| = \zeta_{m_n}^{-1} \left| \|\widehat{\Sigma}_{m_n}\|_F - \|\Sigma_{m_n}\|_F \right| \leq \zeta_{m_n}^{-1} \|\widehat{\Sigma}_{m_n} - \widetilde{\Sigma}_{m_n}\|_F + \zeta_{m_n}^{-1} \|\widetilde{\Sigma}_{m_n} - \Sigma_{m_n}\|_F = o_p(1)$$

which proves the result.  $\square$

**LEMMA A.4.** *Let Assumptions 1–6 be satisfied. Then,*

$$\widehat{\mu}_{m_n} = \mu_{m_n} + o_p(\zeta_{m_n})$$

where  $\mu_{m_n}$  and  $\widehat{\mu}_{m_n}$  are as defined in Theorem 5.1.

*Proof.* The proof of Lemma A.3 establishes  $\|\widehat{\Sigma}_{m_n} - \Sigma_{m_n}\|_F = o_p(1)$ . In particular, convergence of the trace of  $\widehat{\Sigma}_{m_n}$  to the trace of  $\Sigma_{m_n}$  follows by using  $|\widehat{\mu}_{m_n} - \mu_{m_n}| \leq \sqrt{m_n} \|\widehat{\Sigma}_{m_n} - \Sigma_{m_n}\|_F = o_p(\zeta_{m_n})$ .  $\square$

## References

- H. Ahn and J. L. Powell. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1):3–29, 1993.
- C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71:1795–1843, 2003.
- A. Belloni, V. Chernozhukov, D. Chetverikov, and K. Kato. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2): 345 – 366, 2015.
- R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. 75(6):1613–1669, 2007.
- C. Breunig. Specification testing in nonparametric instrumental quantile regression (submitted). Technical report, 2013.
- C. Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.
- C. Breunig and J. Johannes. Adaptive estimation of functionals in nonparametric instrumental regression. *Econometric Theory*, pages 1–43, 2011.



- G. Chamberlain. Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics*, 32:189–218, 1986.
- K. Chen. Parametric models for response-biased sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):775–789, 2001.
- X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth moments. *Econometrica*, 80(1):277–322, 2012.
- X. Chen and D. Pouzo. Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. Technical report, Cowles Foundation for Research in Economics, Yale University, 2013.
- X. Chen and M. Reiß. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27:497–521, 2011.
- X. Chen, H. Hong, and D. Nekipelov. Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4):901–937, 2011.
- V. Chernozhukov, S. Lee, and A. M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81:667–737, 2013.
- S. Darolles, Y. Fan, J.-P. Florens, and E. Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.
- M. Das, W. K. Newey, and F. Vella. Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1):33–58, 2003.
- L. Davezies and X. D’Haultfoeuille. Partial identification with missing data. *mimeo*, 2013.
- C. De Boor. A practical guide to splines. *Mathematics of Computation*, 1978.
- X. D’Haultfoeuille. A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, 154(1):1–15, 2010.
- X. D’Haultfoeuille. On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3):460, 2011.
- X. D’Haultfoeuille and A. Maurel. Another look at identification at infinity of sample selection models. *Econometric Theory*, 29, 2013.
- J. Heckman. Shadow prices, market wages, and labor supply. *Econometrica*, 70:679–694, 1974.
- J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47:153–161, 1979.
- M. Henry, Y. Kitamura, and B. Salanié. Partial identification of finite mixtures in econometric models. *Quantitative Economics*, 5:123–144, 2014.
- S. Hoderlein, L. Nesheim, and A. Simoni. Semiparametric estimation of random coefficients in structural economic models. *CEMMAP Working paper, CWP09/12*, 2012.
- Y. Hu and S. Schennach. instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76:195–216, 2008a.

- Y. Hu and S. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008b.
- A. Lewbel. Endogenous selection or treatment model estimation. *Journal of Econometrics*, 141:777–806, 2007.
- C. F. Manski. Anatomy of the selection problem. *The Journal of Human Resources*, 24:343–360, 1989.
- C. F. Manski. The selection problem. In C. E. Sims, editor, *Advances in Econometrics, Sixth World Congress*. Cambridge University Press., 1994.
- W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147 – 168, 1997.
- W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71:1565–1578, 2003.
- E. A. Ramalho and R. J. Smith. Discrete choice non-response. *The Review of Economic Studies*, 80(1):343–364, 2013.
- D. B. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.
- E. Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2:167–195, 2010.
- G. Tang, R. J. Little, and T. E. Raghunathan. Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika*, 90(4):747–764, 2003.

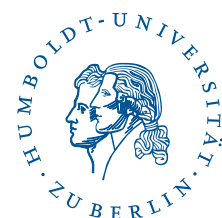
# SFB 649 Discussion Paper Series 2015

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Pricing Kernel Modeling" by Denis Belomestny, Shujie Ma and Wolfgang Karl Härdle, January 2015.
- 002 "Estimating the Value of Urban Green Space: A hedonic Pricing Analysis of the Housing Market in Cologne, Germany" by Jens Kolbe and Henry Wüstemann, January 2015.
- 003 "Identifying Berlin's land value map using Adaptive Weights Smoothing" by Jens Kolbe, Rainer Schulz, Martin Wersing and Axel Werwatz, January 2015.
- 004 "Efficiency of Wind Power Production and its Determinants" by Simone Pieralli, Matthias Ritter and Martin Odening, January 2015.
- 005 "Distillation of News Flow into Analysis of Stock Reactions" by Junni L. Zhang, Wolfgang K. Härdle, Cathy Y. Chen and Elisabeth Bommers, January 2015.
- 006 "Cognitive Bubbles" by Ciril Bosch-Rosay, Thomas Meissner and Antoni Bosch-Domènech, February 2015.
- 007 "Stochastic Population Analysis: A Functional Data Approach" by Lei Fang and Wolfgang K. Härdle, February 2015.
- 008 "Nonparametric change-point analysis of volatility" by Markus Bibinger, Moritz Jirak and Mathias Vetter, February 2015.
- 009 "From Galloping Inflation to Price Stability in Steps: Israel 1985–2013" by Rafi Melnick and Till Strohsal, February 2015.
- 010 "Estimation of NAIRU with Inflation Expectation Data" by Wei Cui, Wolfgang K. Härdle and Weining Wang, February 2015.
- 011 "Competitors In Merger Control: Shall They Be Merely Heard Or Also Listened To?" by Thomas Giebe and Miyu Lee, February 2015.
- 012 "The Impact of Credit Default Swap Trading on Loan Syndication" by Daniel Streitz, March 2015.
- 013 "Pitfalls and Perils of Financial Innovation: The Use of CDS by Corporate Bond Funds" by Tim Adam and Andre Guettler, March 2015.
- 014 "Generalized Exogenous Processes in DSGE: A Bayesian Approach" by Alexander Meyer-Gohde and Daniel Neuhoff, March 2015.
- 015 "Structural Vector Autoregressions with Heteroskedasticity" by Helmut Lütkepohl and Aleksei Netšunajev, March 2015.
- 016 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, March 2015.
- 017 "Loss Potential and Disclosures Related to Credit Derivatives – A Cross-Country Comparison of Corporate Bond Funds under U.S. and German Regulation" by Dominika Paula Gałkiewicz, March 2015.
- 018 "Manager Characteristics and Credit Derivative Use by U.S. Corporate Bond Funds" by Dominika Paula Gałkiewicz, March 2015.
- 019 "Measuring Connectedness of Euro Area Sovereign Risk" by Rebekka Gätjen Melanie Schienle, April 2015.
- 020 "Is There an Asymmetric Impact of Housing on Output?" by Tsung-Hsien Michael Lee and Wenjuan Chen, April 2015.
- 021 "Characterizing the Financial Cycle: Evidence from a Frequency Domain Analysis" by Till Strohsal, Christian R. Proaño and Jürgen Wolters, April 2015.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



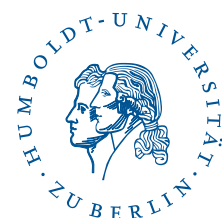
## SFB 649 Discussion Paper Series 2015

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 022 "Risk Related Brain Regions Detected with 3D Image FPCA" by Ying Chen, Wolfgang K. Härdle, He Qiang and Piotr Majer, April 2015.
- 023 "An Adaptive Approach to Forecasting Three Key Macroeconomic Variables for Transitional China" by Linlin Niu, Xiu Xu and Ying Chen, April 2015.
- 024 "How Do Financial Cycles Interact? Evidence from the US and the UK" by Till Strohsal, Christian R. Proaño, Jürgen Wolters, April 2015.
- 025 "Employment Polarization and Immigrant Employment Opportunities" by Hanna Wielandt, April 2015.
- 026 "Forecasting volatility of wind power production" by Zhiwei Shen and Matthias Ritter, May 2015.
- 027 "The Information Content of Monetary Statistics for the Great Recession: Evidence from Germany" by Wenjuan Chen and Dieter Nautz, May 2015.
- 028 "The Time-Varying Degree of Inflation Expectations Anchoring" by Till Strohsal, Rafi Melnick and Dieter Nautz, May 2015.
- 029 "Change point and trend analyses of annual expectile curves of tropical storms" by P.Burdejova, W.K.Härdle, P.Kokoszka and Q.Xiong, May 2015.
- 030 "Testing for Identification in SVAR-GARCH Models" by Helmut Lutkepohl and George Milunovich, June 2015.
- 031 "Simultaneous likelihood-based bootstrap confidence sets for a large number of models" by Mayya Zhilova, June 2015.
- 032 "Government Bond Liquidity and Sovereign-Bank Interlinkages" by Sören Radde, Cristina Checherita-Westphal and Wei Cui, July 2015.
- 033 "Not Working at Work: Loafing, Unemployment and Labor Productivity" by Michael C. Burda, Katie Genadek and Daniel S. Hamermesh, July 2015.
- 034 "Factorisable Sparse Tail Event Curves" by Shih-Kang Chao, Wolfgang K. Härdle and Ming Yuan, July 2015.
- 035 "Price discovery in the markets for credit risk: A Markov switching approach" by Thomas Dimpfl and Franziska J. Peter, July 2015.
- 036 "Crowdfunding, demand uncertainty, and moral hazard - a mechanism design approach" by Roland Strausz, July 2015.
- 037 ""Buy-It-Now" or "Sell-It-Now" auctions : Effects of changing bargaining power in sequential trading mechanism" by Tim Grebe, Radosveta Ivanova-Stenzel and Sabine Kröger, August 2015.
- 038 "Conditional Systemic Risk with Penalized Copula" by Ostap Okhrin, Alexander Ristig, Jeffrey Sheen and Stefan Trück, August 2015.
- 039 "Dynamics of Real Per Capita GDP" by Daniel Neuhoff, August 2015.
- 040 "The Role of Shadow Banking in the Monetary Transmission Mechanism and the Business Cycle" by Falk Mazelis, August 2015.
- 041 "Forecasting the oil price using house prices" by Rainer Schulz and Martin Wersing, August 2015.
- 042 "Copula-Based Factor Model for Credit Risk Analysis" by Meng-Jou Lu, Cathy Yi-Hsuan Chen and Karl Wolfgang Härdle, August 2015.
- 043 "On the Long-run Neutrality of Demand Shocks" by Wenjuan Chen and Aleksei Netsunajev, August 2015.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



## SFB 649 Discussion Paper Series 2015

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 044 "The (De-)Anchoring of Inflation Expectations: New Evidence from the Euro Area" by Laura Pagenhardt, Dieter Nautz and Till Strohsal, September 2015.
- 045 "Tail Event Driven ASset allocation: evidence from equity and mutual funds' markets" by Wolfgang Karl Härdle, David Lee Kuo Chuen, Sergey Nasekin, Xinwen Ni and Alla Petukhina, September 2015.
- 046 "Site assessment, turbine selection, and local feed-in tariffs through the wind energy index" by Matthias Ritter and Lars Deckert, September 2015.
- 047 "TERES - Tail Event Risk Expectile based Shortfall" by Philipp Gschöpf, Wolfgang Karl Härdle and Andrija Mihoci, September 2015.
- 048 "CRIX or evaluating Blockchain based currencies" by Wolfgang Karl Härdle and Simon Trimborn, October 2015.
- 049 "Inflation Co-movement across Countries in Multi-maturity Term Structure: An Arbitrage-Free Approach" by Shi Chen, Wolfgang Karl Härdle, Weining Wang, November 2015.
- 050 "Nonparametric Estimation in case of Endogenous Selection" by Christoph Breunig, Enno Mammen and Anna Simoni, November 2015.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

