

Breunig, Christoph

**Working Paper**

## Testing missing at random using instrumental variables

SFB 649 Discussion Paper, No. 2015-016

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Breunig, Christoph (2015) : Testing missing at random using instrumental variables, SFB 649 Discussion Paper, No. 2015-016, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/119435>

**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.*

# Testing Missing at Random using Instrumental Variables

Christoph Breunig\*



\* Humboldt-Universität zu Berlin, Germany

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



# Testing Missing at Random using Instrumental Variables \*

CHRISTOPH BREUNIG \*

*Humboldt-Universität zu Berlin*

March 25, 2015

This paper proposes a test for *missing at random* (MAR). The MAR assumption is shown to be testable given instrumental variables which are independent of response given potential outcomes. A nonparametric testing procedure based on integrated squared distance is proposed. The statistic's asymptotic distribution under the MAR hypothesis is derived. We demonstrate that our results can be easily extended to a test of *missing completely at random* (MCAR) and *missing completely at random conditional on covariates X* (MCAR(X)). A Monte Carlo study examines finite sample performance of our test statistic. An empirical illustration concerns pocket prescription drug spending with missing values; we reject MCAR but fail to reject MAR.

*Keywords:* Incomplete data, missing-data mechanism, selection model, nonparametric hypothesis testing, consistent testing, instrumental variable, series estimation.

*JEL classification:* C12, C14

## 1. Introduction

When confronted with data sets with missing values it is often assumed in applied research that observations are *missing at random* (MAR) in the sense of Rubin [1976]. This condition requires that the probability of observing potential outcomes only depends on observed data. To help to decide whether MAR based techniques could be applied we develop

---

\*The author is grateful for very helpful discussions with Timothy Armstrong, Xiaohong Chen, Arthur Lewbel, Stefan Hoderlein, Enno Mammen, and Peter Robinson during the preparation of the manuscript. This research was supported by DFG postdoctoral fellowship BR 4874/1-1. The author is also grateful for support and hospitality of the Cowles Foundation.

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

in this paper a test for the MAR assumption. In general, MAR is not refutable without further assumptions and here we rely on instruments that are independent of the response mechanism given potential outcomes. We show that this condition is sufficient to ensure testability of MAR and derive the asymptotic distribution under MAR of a proposed test statistic. We provide two extensions of our testing procedure which are testing *missing completely at random* (MCAR) and *missing completely at random conditional on covariates  $X$*  (MCAR( $X$ )).

If the missing data mechanism does not follow MAR, a correction of the potential selection bias is necessary to ensure consistency of the estimation procedure. There exists two different instrumental variable approaches to overcome the problem of missing variables. The first approach relies on instruments that determine response but not the outcomes and was pioneered by Heckman [1974]. Such instruments, however, are difficult to find, in particular, when response is directly driven by the outcome. The second approach, also considered in this paper, relies on instruments that are independent of response given potential outcomes. This framework was used in parametric regression analysis by Chen [2001], Liang and Qin [2000], Tang et al. [2003], and Ramalho and Smith [2013]. A non-parametric extension was proposed by D'Haultfoeuille [2010] and Breunig et al. [2014]. While such instrumental variable methods reduce bias in general, if the data are missing at random, they unnecessarily increase variance. Indeed, D'Haultfoeuille [2010] showed that estimation of the distribution of the potential outcome leads to a statistical inverse problem that is *ill-posed* in general. This implies that the variance of the estimator becomes arbitrarily large relative to the degree of ill-posedness.

We also provide tests for the MCAR and MCAR( $X$ ) assumptions. Both impose stronger conditions on the response mechanism as MAR. Indeed, MCAR and MCAR( $X$ ) rule out any correlation between response and outcome. MCAR( $X$ ) is also known as the unconfoundedness assumption in the treatment effect literature (see, for instance, Imbens [2004]). When data are MAR but not MCAR various types of correction methods have been suggested so far and include weighted generalized estimating equations (Robins et al. [1994]), nonparametric estimation of the conditional estimating scores (Reilly and Pepe [1995]), and multiple imputation (Rubin [2004], Little and Rubin [2002]). For an overview and further references we refer to Ibrahim et al. [2005]. Either these methods make parametric model assumptions or have difficulties in dealing with continuous data. These methods reduce bias if MAR holds, under MCAR, however, they unnecessarily increase variance. Thus, it is of interest to examine the observed data for evidence whether the response mechanism satisfies not only MAR but also MCAR or MCAR( $X$ ).

We show that the MAR hypothesis is equivalent to an identified conditional moment equation. Based on this moment equation we construct our test statistic using integrated squared distance. Under the null hypothesis the test statistic converges to a series of independent,  $\chi^2$ -squared distributed random variables. The test statistic and its critical values can be easily implemented. Also only a slight modification is necessary to obtain a test for MCAR and MAR( $X$ ). Under a bounded completeness assumption, our testing procedure is shown to be consistent against any fixed alternative.

Besides a Monte Carlo simulation we demonstrate the finite sample properties in an empirical illustration using data from the Health and Retirement Study. In this survey, a fraction of participants does not report their exact expenditure of pocket prescription drugs. The assumption of MAR/MCAR seems problematic here as whether participants recall their exact expenditure might be related to the amount of expenditure itself. Using income as instrument we show that our test rejects the MCAR assumption; but fails to reject MAR.

In our instrumental variable framework, a test of MCAR has been proposed by Ramalho and Smith [2013]. Their Hausman type test statistic relies on a parametric model specification with discrete outcomes and differs from our method where no restriction on the marginal distribution of the outcome is imposed. Likelihood ratio tests to verify the hypothesis MCAR have been suggested by Fuchs [1982] and Little [1988], while Chen and Little [1999] considered a Wald-type test and Qu and Song [2002] proposed a generalized score type test based on quadratic inference functions. Kline and Santos [2013] develop a method for assessing the sensitivity of empirical conclusions to departures from MAR based on sharp bounds of conditional quantiles. As far as we know, a consistent test for MAR has not been proposed. We further emphasize that our testing procedure does not require knowledge of the conditional probability of observing potential outcomes up to a finite dimensional parameter.

The remainder of the paper is organized as follows. Section 2 provides sufficient conditions for testability of MAR, MCAR, and MCAR(X). The asymptotic distributions of the tests are derived and their consistency against local alternatives is established. Section 3 examines the finite sample performance of our test in a Monte Carlo study while Section 4 illustrates the usefulness of our procedure in an empirical application.

## 2. The Test Statistic and its asymptotic properties

This section is about testability of missing at random assumptions and the asymptotic behavior of proposed test statistics. First, we provide sufficient conditions on instruments to ensure testability of MAR, MCAR(X), and MCAR. Second, we build on identified conditional moment restrictions to construct test statistics. Third, the test statistics' asymptotic distributions under the null hypotheses are derived and we establish consistency of the tests against fixed alternatives.

### 2.1. Testability

Let  $Y^*$  denote a scalar depend variable and  $X$  a  $d_x$ -dimensional vector of covariates. Further,  $\Delta$  is a missing-data indicator for  $Y^*$ , such that  $\Delta = 1$  if a realization of  $Y^*$  is observed and  $\Delta = 0$  if  $Y^*$  is missing.<sup>1</sup> Throughout this paper, we write  $Y = \Delta Y^*$ . In the following, we discuss testability of the different hypothesis MAR, MCAR(X), and MCAR. First, we consider hypothesis MAR, whether missingness only depends on observed variables. More precisely, the response mechanism depends only on the observed realizations of  $Y^*$  and covariates  $X$ . That is, we consider the null hypothesis

$$MAR : \mathbb{P}(\Delta = 1|Y^*, X) = \Delta \mathbb{P}(\Delta = 1|Y, X) + (1 - \Delta) \mathbb{P}(\Delta = 1|X)$$

and the alternative  $\mathbb{P}(\mathbb{P}(\Delta = 1|Y^*, X) = \Delta \mathbb{P}(\Delta = 1|Y, X) + (1 - \Delta) \mathbb{P}(\Delta = 1|X)) < 1$ .<sup>2</sup> Second, we want to test the hypothesis whether the response mechanism only depends on covariates

<sup>1</sup>In our setting,  $Y^*$  is assumed to be a scalar. Our results could be easily extended to allow for a  $d_y$ -dimensional vector  $Y^*$  of potential outcome variables. In this case,  $\Delta = (\Delta^{(j)})_{1 \leq j \leq d_y}$  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 this multivariate case.

<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.

X. This condition is stronger than MAR as it rules out any correlation between response and outcome. In this case, the null hypothesis under consideration is given by

$$MCAR(X) : \mathbb{P}(\Delta = 1|Y^*, X) = \mathbb{P}(\Delta = 1|X)$$

and the alternative by  $\mathbb{P}(\mathbb{P}(\Delta = 1|Y^*, X) = \mathbb{P}(\Delta = 1|X)) < 1$ . Third, we consider the MCAR hypothesis whether response is completely at random. As this hypothesis rules out any correlation between response and observed data, MCAR is stronger than MCAR(X) and, in particular, MAR. The hypothesis under consideration is

$$MCAR : \mathbb{P}(\Delta = 1|Y^*, X) = \mathbb{P}(\Delta = 1)$$

and the alternative is  $\mathbb{P}(\mathbb{P}(\Delta = 1|Y^*, X) = \mathbb{P}(\Delta = 1)) < 1$ .

We now provide sufficient conditions for testability of the above hypotheses. A key requirement is that an additional vector  $W$ , an instrument, is available which satisfies the following conditions.

**ASSUMPTION 1.** *For each unit we observe  $\Delta$ ,  $Y$ ,  $X$ , and  $W$ .*

Assumption 1 is satisfied when only observations of  $Y^*$  are missing. In the following, we assume that the random vector  $W$  is independent of the response variable conditional on potential outcomes and covariates.

**ASSUMPTION 2.** *It holds*

$$\Delta \perp\!\!\!\perp W | (Y^*, X).$$

Assumption 2 requires missingness to be primarily determined by the potential outcome  $Y^*$  and covariates  $X$ . In particular, this exclusion restriction requires any influence of  $W$  on  $\Delta$  to be carried solely through  $(Y^*, X)$ . Conditional independence assumptions of this type are quite familiar in the econometrics and statistics literature. Examples are treatment effects (cf. Imbens [2004]) or non-classical measurement error (cf. Hu and Schennach [2008]). In case of nonresponse, Assumption 2 (without covariates) was exploited by Ramalho and Smith [2013]. This assumption was also made by D'Haultfoeulle [2010] where further illustrative examples in case of the counterfactual issue are given. We further emphasize that Assumption 2 is a testable condition (see Theorem 2.4 of D'Haultfoeulle [2010]).

**ASSUMPTION 3.** *For all bounded measurable functions  $\phi$ ,  $\mathbb{E}[\phi(Y^*, X)|X, W] = 0$  implies that  $\phi(Y^*, X) = 0$ .*

Assumption 3 is known as bounded completeness. In contrast, to ensure identification in nonparametric instrumental variable models, stronger versions of Assumption 3, such as  $L^2$ -completeness, are required. This type of completeness condition requires Assumption 3 to hold for any measurable function  $\phi$  with  $\mathbb{E}|\phi(Y^*, X)|^2 < \infty$ .  $L^2$ -completeness is also a common assumption in nonparametric hypothesis testing in instrumental variable models, see, for instance, Blundell and Horowitz [2007] or Fève et al. [2012]. There are only a few examples in the nonparametric instrumental regression literature where it is sufficient to assume completeness only for bounded functions. One example is estimation of Engel curves as in Blundell et al. [2007] which, by definition, are bounded between zero and one. We emphasize that bounded completeness is much less restrictive than  $L^2$  completeness. Sufficient conditions for bounded completeness have been provided by Mattner [1993] or D'Haultfoeulle [2011] among others. We see below that inference under the considered hypotheses does not require bounded completeness. On the other hand, we need to impose Assumption 3 to ensure consistency against fixed alternatives.

If a valid instrumental variable  $W$  is available then consistent density estimation and regression is possible even if MAR does not hold true. On the other hand, using instrumental variable estimation methods when MAR holds can be inappropriate as the following two examples illustrate.

**EXAMPLE 2.1 (Density Estimation).** The joint probability density function of  $(Y^*, X)$  satisfies

$$p_{Y^*X}(\cdot, \cdot) = \frac{p_{\Delta Y^* X}(1, \cdot, \cdot)}{\mathbb{P}(\Delta = 1|Y^* = \cdot, X = \cdot)}$$

assuming that the conditional probability in the denominator is bounded away from zero. The conditional probability  $\mathbb{P}(\Delta = 1|Y^*, X)$  is not identified in general. On the other hand, if instrumental variables  $W$  are available that are independent of  $\Delta$  conditional on  $(Y^*, X)$  then this probability is identified (cf. D'Haultfoeuille [2010]) through the conditional moment restriction

$$\mathbb{E}\left(\frac{\Delta}{\mathbb{P}(\Delta = 1|Y^*, X)} \middle| X, W\right) = 1. \quad (2.1)$$

Estimating  $\mathbb{P}(\Delta = 1|Y^*, X)$  via this equation leads to a large variance relative to the ill-posedness of the underlying inverse problem and the accuracy of this estimator can be very low. If the data, however, reveals that MAR holds true then  $\mathbb{P}(\Delta = 1|Y^*, X) = \Delta\mathbb{P}(\Delta = 1|Y, X) + (1 - \Delta)\mathbb{P}(\Delta = 1|X)$  which can be directly estimated from the data.  $\square$

**EXAMPLE 2.2 (Regression).** Consider estimation of  $\mathbb{E}(\phi(Y^*)|X)$  for some known function  $\phi$ . Either  $\phi$  is the identity function in case of mean regression or  $\phi(Y^*) = \mathbb{1}\{Y^* \leq q\}$  in quantile regression for some quantile  $q \in (0, 1)$ . Let the conditional probability  $\mathbb{P}(\Delta = 1|Y^*, X)$  be bounded away from zero. As in Breunig et al. [2014] (p. 5) it holds

$$\mathbb{E}(\phi(Y^*)|X) = \mathbb{E}\left(\frac{\Delta\phi(Y^*)}{\mathbb{P}(\Delta = 1|Y^*, X)} \middle| X\right)$$

where  $\mathbb{P}(\Delta = 1|Y^*, X)$  can be estimated via the conditional mean restriction 2.1. As shown in Breunig et al. [2014], the first step estimation of  $\mathbb{P}(\Delta = 1|Y^*, X)$  leads to an additional bias term which can reduce accuracy of estimation. In contrast, under MAR it holds

$$\mathbb{E}(\phi(Y^*)|X) = \mathbb{E}\left(\frac{\Delta\phi(Y^*)}{\mathbb{P}(\Delta = 1|Y, X)} \middle| X\right)$$

where the right hand side is identified from the data and  $\mathbb{P}(\Delta = 1|Y, X)$  can be directly estimated. Similarly, when interest lies in quantile/mean regressing of  $Y^*$  on  $W$  where  $W \perp\!\!\!\perp \Delta|Y^*$  (cf. Breunig et al. [2014]) then under MAR (without covariates  $X$ ) it holds

$$\mathbb{E}(\phi(Y^*)|W) = \mathbb{E}\left(\frac{\Delta\phi(Y^*)}{\mathbb{P}(\Delta = 1|Y)} \middle| W\right).$$

Also in this case, imposing MAR is desirable to simplify the estimation procedure and increase estimation precision.  $\square$

**EXAMPLE 2.3 (Relation to Triangular Models).** Assumptions 2 and 3 hold true in the triangular model

$$\begin{aligned} \Delta &= \varphi(Y^*, X, \eta) \quad \text{with} \quad \eta \perp (W, \varepsilon) \\ Y^* &= \phi(\psi(X, W) + \varepsilon) \quad \text{with} \quad W \perp \varepsilon \end{aligned}$$

under a large support condition of  $\psi(X, W)$ , regularity assumptions for  $\varepsilon$ , and if the conditional characteristic function of  $\varepsilon$  given  $X$  is infinitely often differentiable and does not vanish on the real line. See D'Haultfoeuille [2011] page 462–463 for further details. Requiring this characteristic function to be nonvanishing is a standard assumption in the deconvolution literature. The normal, Student,  $\chi^2$ , gamma, and double exponential distributions all satisfy this assumption while the uniform and the triangular distributions are the only common distributions to violate this restriction.

In this triangular model, MCAR( $X$ ) requires the structural function  $\varphi$  to be dependent on  $X$  and  $\eta$  only; that is,  $\Delta = \varphi(X, \eta)$ . Under MCAR,  $\varphi$  depends neither on  $Y^*$  nor on  $X$  and hence, the structural equation simplifies to  $\Delta = \varphi(\eta)$ . The triangular model illustrates the difference to Heckman's approach (cf. its nonparametric version in Das et al. [2003]) where an instrument enters only the selection equation.  $\square$

The following result states that the null hypothesis MAR is testable under the previous conditions. Further, exploiting the properties of the instrument  $W$  shows that MAR is equivalent to an identified conditional moment restriction.

**THEOREM 2.1.** *Under Assumptions 1–3 the null hypothesis MAR is testable.*

*Proof.* We rewrite the null hypothesis MAR as

$$\mathbb{E}[\Delta - \mathbb{E}(\Delta|X)|Y^*, X] - \Delta(g(Y^*, X) - \mathbb{E}(\Delta|X)) = 0 \quad (2.2)$$

where  $g(y, x) := \mathbb{E}[\Delta|Y = y, X = x]$  and making use of  $\Delta g(Y, X) = \Delta g(Y^*, X)$ . The left hand side of equation (2.2) is a bounded and measurable function with respect to the  $\sigma$ -algebra generated by  $(Y^*, X)$ . Now by Assumption 3 the hypothesis MAR is equivalent to

$$\mathbb{E}[\mathbb{E}(\Delta|Y^*, X) - \Delta g(Y^*, X) - (1 - \Delta) \mathbb{E}(\Delta|X)|X, W] = 0. \quad (2.3)$$

Further, Assumption 2 implies  $\mathbb{E}[\mathbb{E}(\Delta|Y^*, X)|X, W] = \mathbb{E}[\Delta|X, W]$ . Thereby, equation (2.3) is equivalent to

$$\mathbb{E}[\Delta(1 - \mathbb{E}(\Delta|Y, X)) - (1 - \Delta) \mathbb{E}(\Delta|X)|X, W] = 0 \quad (2.4)$$

where the left hand side is identified.  $\square$

Let us now turn to testability of the hypothesis MCAR( $X$ ); that is, whether response only depends on covariates  $X$ . As we see in the following, testability of MCAR( $X$ ) follows as in the proof of Theorem 2.1.

**COROLLARY 2.2.** *Under Assumptions 1–3 the null hypothesis MCAR( $X$ ) is testable.*

*Proof.* Due to Assumption 3 the null hypothesis MCAR( $X$ ) is equivalent to

$$\mathbb{E}[\mathbb{E}(\Delta|Y^*, X) - \mathbb{E}(\Delta|X)|X, W] = 0.$$

Assumption 2 yields  $\mathbb{E}[\mathbb{E}(\Delta|Y^*, X)|X, W] = \mathbb{E}[\Delta|X, W]$  and hence hypothesis MCAR( $X$ ) is equivalent to

$$\mathbb{E}[\Delta - \mathbb{E}(\Delta|X)|X, W] = 0 \quad (2.5)$$

where the left hand side is identified.  $\square$



The following corollary provides a testability result for the hypothesis MCAR. The result follows as in the proof of Corollary 2.2 by replacing  $\mathbf{E}(\Delta|X)$  with  $\mathbf{E}(\Delta)$ .

**COROLLARY 2.3.** *Let Assumptions 1–3 hold true. Then the null hypothesis MCAR is equivalent to  $\mathbf{E}[\Delta - \mathbf{E}(\Delta)|X, W] = 0$  and hence, is testable.*

## 2.2. The Test Statistic

In the previous section, we observed that each null hypothesis is equivalent to a conditional moment restriction

$$\mathbf{E}[r(\Delta, Y, X)|X, W] = 0$$

for some bounded function  $r$ , which is equivalent to

$$\int \mathbf{E}[r(\Delta, Y, X)|X = x, W = w]^2 \pi(x, w) d(x, w) = 0$$

for some weight function  $\pi$  which is strictly positive almost surely (a.s.) on  $\mathcal{X} \times \mathcal{W}$  ( $\mathcal{X}$  and  $\mathcal{W}$  denote the supports of  $X$  and  $W$ , respectively). Let  $p_{XW}$  denote the joint probability density function of  $(X, W)$ . Further, let  $\nu$  be an a.s. strictly positive density function on  $\mathcal{X} \times \mathcal{W}$ . Let us introduce approximating functions  $\{f_j\}_{j \geq 1}$  which are assumed to form an orthonormal basis in the Hilbert space  $L_\nu^2 := \{\phi : \int |\phi(x, w)|^2 \nu(x, w) d(x, w) < \infty\}$ . Now choosing  $\pi(x, w) = p_{XW}^2(x, w)/\nu(x, w)$  together with Parseval's identity yields

$$\begin{aligned} 0 &= \int \left| \mathbf{E}[r(\Delta, Y, X)|X = x, W = w] p_{XW}(x, w)/\nu(x, w) \right|^2 \nu(x, w) d(x, w) \\ &= \sum_{j=1}^{\infty} \left( \mathbf{E}[r(\Delta, Y, X) f_j(X, W)] \right)^2. \end{aligned}$$

Given a strictly positive sequence of weights  $(\tau_j)_{j \geq 1}$  the last equation is equivalent to

$$\sum_{j=1}^{\infty} \tau_j \left( \mathbf{E}[r(\Delta, Y, X) f_j(X, W)] \right)^2 = 0. \quad (2.6)$$

Our test statistic is based on an empirical analog of the left hand side of (2.6) given  $(\Delta_1, Y_1, X_1, W_1), \dots, (\Delta_n, Y_n, X_n, W_n)$  of independent and identical distributed (iid.) copies of  $(\Delta, Y, X, W)$  where  $Y = \Delta Y^*$ . For a random vector  $V$  and some integer  $k \geq 1$ , we denote by  $e_k(V) := (e_1(V), \dots, e_k(V))^t$  a vector of basis functions which are used to approximate the conditional expectations  $\mathbf{E}[\Delta|V]$ . In the multivariate case, we consider a tensor-product linear sieve basis, which is the product of univariate linear sieves. Further, let us denote  $\mathbf{Y}_{k_n} = (e_{k_n}(Y_1, X_1), \dots, e_{k_n}(Y_n, X_n))^t$  and  $\mathbf{X}_{l_n} = (e_{l_n}(X_1), \dots, e_{l_n}(X_n))^t$ . We introduce the functions  $g(y, x) := \mathbf{E}[\Delta|Y = y, X = x]$  and  $h(x) := \mathbf{E}[\Delta|X = x]$ . We estimate the functions  $g$  and  $h$ , respectively, by the series least square estimators

$$\widehat{g}_n(y, x) := e_{k_n}(y, x)^t (\mathbf{Y}_{k_n}^t \mathbf{Y}_{k_n})^{-1} \mathbf{Y}_{k_n}^t \Delta_n$$

and

$$\widehat{h}_n(x) := e_{l_n}(x)^t (\mathbf{X}_{l_n}^t \mathbf{X}_{l_n})^{-1} \mathbf{X}_{l_n}^t \Delta_n$$

where  $\Delta_n = (\Delta_1, \dots, \Delta_n)$ .

Consider the null hypothesis MAR. From the proof of Theorem 2.1, we deduce  $r(\Delta, Y, X) = \Delta(1 - g(Y, X)) - (1 - \Delta)h(X)$ . Replacing  $g$  and  $h$  by the proposed estimators we obtain our test statistic

$$S_n^{MAR} = \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_{i=1}^n \left( \Delta_i - \Delta_i \widehat{g}_n(Y_i, X_i) - (1 - \Delta_i) \widehat{h}_n(X_i) \right) f_j(X_i, W_i) \right|^2 \quad (2.7)$$

where  $m_n$  increases with sample size  $n$  and  $(\tau_j)_{j \geq 1}$  is a strictly positive sequence of weights which is nonincreasing. Additional weighting of the testing procedure was also used by Horowitz [2006], Blundell and Horowitz [2007], and Breunig [2015].

Let us now turn to a test of the null hypothesis MCAR( $X$ ). From Corollary 2.2 we have  $r(\Delta, X) = \Delta - h(X)$  where  $h(\cdot) = \mathbb{E}(\Delta|X = \cdot)$ . Hence, replacing  $h$  by  $\widehat{h}_n$  we obtain the test statistic

$$S_n^{MCAR(X)} = \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_{i=1}^n \left( \Delta_i - \widehat{h}_n(X_i) \right) f_j(X_i, W_i) \right|^2. \quad (2.8)$$

For the null hypothesis MCAR, Corollary 2.3 gives  $r(\Delta) = \Delta - \mathbb{E} \Delta$ . Again, following the derivation of the statistic  $S_n^{MAR}$  we obtain a statistic for MCAR given by

$$S_n^{MCAR} = \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_{i=1}^n \left( \Delta_i - \widehat{\Delta}_n \right) f_j(X_i, W_i) \right|^2 \quad (2.9)$$

where  $\widehat{\Delta}_n = n^{-1} \sum_{i=1}^n \Delta_i$ .

### 2.3. Assumptions for inference

In the following,  $\mathcal{Y}$ ,  $\mathcal{X}$ , and  $\mathcal{W}$  denote the supports of  $Y$ ,  $X$ , and  $W$ , respectively. The usual Euclidean norm is denoted by  $\|\cdot\|$  and  $\|\cdot\|_\infty$  is the supremum norm.

**ASSUMPTION 4.** (i) The functions  $\{f_j\}_{j \geq 1}$  form an orthonormal basis in  $L_v^2$ . (ii) There exists some constant  $C > 0$  such that  $\sup_{(x,w) \in \mathcal{X} \times \mathcal{W}} \{p_{XW}(x, w)/v(x, w)\} \leq C$ .

In our simulations, we used trigonometric basis functions or orthonormalized Hermite polynomials where Assumption 4 (i) is automatically satisfied if, respectively,  $v$  is Lebesgue measure on  $[0, 1]$  or  $v$  is the standard normal density. Assumption 4 (ii) is a mild restriction on the density of  $(X, W)$  relative to  $v$ . Assumption 4 implies  $\mathbb{E}|f_j(X, W)|^2 \leq C$ . The next assumption involves the linear sieve space  $\mathcal{H}_n := \{\phi : \phi(x) = \beta'_n e_{l_n}(x) \text{ where } \beta_n \in \mathbb{R}^{l_n}\}$ .

**ASSUMPTION 5.** (i) There exists  $E_{l_n} h \in \mathcal{H}_n$  such that  $\|E_{l_n} h - h\|_\infty^2 = O(1/\gamma_{l_n}^h)$  for some nondecreasing sequences  $(\gamma_l^h)_{l \geq 1}$ . (ii) It holds  $\sup_{x \in \mathcal{X}} \|e_{l_n}(x)\|^2 = O(l_n)$  such that  $l_n^2 \log(n) = o(n)$ . (iii) The smallest eigenvalue of  $\mathbb{E}[e_{l_n}(X) e_{l_n}(X)^t]$  is bounded away from zero uniformly in  $n$ .

Assumption 5 (i) determines the sieve approximation error for estimating the function  $h$  in the supremum norm and is used to control the bias of the estimator of  $h$ . This assumption was also imposed by Newey [1997] (for the relation to  $L^2$  approximation conditions see Belloni et al. [2012] p. 10–16). An excellent review of approximating properties of different sieve bases is given in Chen [2007]. Assumption 5 (ii) and (iii) restrict the magnitude of the approximating functions  $\{e_j\}_{j \geq 1}$  and impose nonsingularity of

their second moment matrix (cf. Newey [1997]). The next assumption involves the linear sieve space  $\mathcal{G}_n := \{\phi : \phi(x, y) = \beta_n' e_{k_n}(x, y) \text{ where } \beta_n \in \mathbb{R}^{k_n}\}$ .

**ASSUMPTION 6.** (i) There exists  $E_{k_n}g \in \mathcal{G}_n$  such that  $\|E_{k_n}g - g\|_\infty^2 = O(1/\gamma_{k_n}^s)$  for some nondecreasing sequences  $(\gamma_k^s)_{k \geq 1}$ . (ii) It holds  $\sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \|e_{k_n}(x, y)\|^2 = O(k_n)$  such that  $k_n^2 \log(n) = o(n)$ . (iv) The smallest eigenvalue of  $\mathbf{E}[e_{k_n}(X, Y)e_{k_n}(X, Y)']$  is bounded away from zero uniformly in  $n$ .

Assumption 6 determines the sieve approximation error for estimating the function  $g$  and restrictions on the basis functions  $\{e_j\}_{j \geq 1}$  when their multivariate extension is considered.

## 2.4. Asymptotic distribution of the test statistic under MAR

Before establishing the asymptotic distribution of the test statistic  $S_n^{MAR}$  under MAR, we require the following definitions. Recall that in case of MAR we have  $r(\Delta, Y, X) = \Delta(1 - g(Y, X)) - (1 - \Delta)h(X)$ . Let  $\varepsilon(\Delta, Y, X, W)$  be an infinite dimensional vector with  $j$ -th entry

$$\varepsilon_j(\Delta, Y, X, W) := \sqrt{\tau_j} r(\Delta, Y, X) f_j(X, W) - \varepsilon_j^s(\Delta, Y, X) - \varepsilon_j^h(\Delta, X)$$

where

$$\varepsilon_j^s(\Delta, Y, X) := \sqrt{\tau_j} (\Delta - g(X, Y)) \sum_{l=1}^{\infty} \mathbf{E} [\Delta f_j(X, W) e_l(Y, X)] e_l(Y, X)$$

and

$$\varepsilon_j^h(\Delta, X) := \sqrt{\tau_j} (\Delta - h(X)) \sum_{l=1}^{\infty} \mathbf{E} [(1 - \Delta) f_j(X, W) e_l(X)] e_l(X).$$

We have  $\mathbf{E}[\varepsilon_j(\Delta, Y, X, W)] = 0$  under MAR. We assume  $\mathbf{E}|\varepsilon_j^s(\Delta, Y, X)|^2 < \infty$  and  $\mathbf{E}|\varepsilon_j^h(\Delta, X)|^2 < \infty$  which is satisfied, for instance, if  $\{e_l\}_{l \geq 1}$  forms an orthonormal basis. Thereby, under MAR the covariance matrix  $\Sigma = \mathbf{E}[\varepsilon(\Delta, Y, X, W)\varepsilon(\Delta, Y, X, W)']$  of  $\varepsilon(\Delta, Y, X, W)$  is well defined. The ordered eigenvalues of  $\Sigma$  are denoted by  $(\lambda_j)_{j \geq 1}$ . Furthermore, we introduce a sequence  $\{\chi_{1j}^2\}_{j \geq 1}$  of independent random variables that are distributed as chi-square with one degree of freedom. The proof of the next theorem can be found in the appendix.

**THEOREM 2.4.** Let Assumptions 1, 2, 4, 5, and 6 hold true. If

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad n = o(\gamma_{k_n}^s), \quad n = o(\gamma_{l_n}^h), \quad \text{and} \quad m_n^{-1} = o(1) \quad (2.10)$$

then under MAR

$$n S_n^{MAR} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2.$$

The rates  $n = o(\gamma_{k_n}^s)$  and  $n = o(\gamma_{l_n}^h)$  ensure that biases for estimating the functions  $g$  and  $h$  vanish sufficiently fast. Below, we show that under classical smoothness assumptions these rates require undersmoothed estimators for  $g$  and  $h$ . We also like to emphasize that for the asymptotic result in Theorem 2.4, the bounded completeness condition stated in Assumption 3 is not required. Below we write  $a_n \sim b_n$  when there exist constants  $c, c' > 0$  such that  $cb_n \leq a_n \leq c'b_n$  for all sufficiently large  $n$ .

EXAMPLE 2.4. Let  $Y$  and  $X$  be continuously distributed. Let  $e_1, \dots, e_{k_n}$  be spline basis functions and  $p$  be the number of continuous derivatives of  $g$ . Then Assumption 6 (i) holds true with  $\gamma_j^g \sim j^{2p/(1+d_x)}$  (cf. Newey [1997]). Condition  $n = o(\gamma_{k_n}^g)$  and Assumption 6 (ii) is satisfied if  $k_n \sim n^\kappa$  with  $(1 + d_x)/(2p) < \kappa < 1/(2 + \varepsilon)$  for any small  $\varepsilon > 0$ . Here, the required smoothness of  $g$  is  $p > (2 + \varepsilon)(1 + d_x)/2$ . Hence, the estimator of  $g$  needs to be undersmoothed. Similarly, also the estimator for  $h$  needs to be undersmoothed.  $\square$

REMARK 2.1 (Estimation of Critical Values). The asymptotic distribution of our test statistic derived in Theorem 2.4 depends on unknown population quantities. As we see in the following, the critical values can be easily estimated. Let us define

$$\widehat{\varepsilon}_j(\Delta, Y, X, W) := \sqrt{\tau_j} \left( \Delta - \widehat{g}_n(Y, X) - (1 - \Delta) \widehat{h}_n(X) \right) f_j(X, W) - \widehat{\varepsilon}_j^g(\Delta, Y, X) - \widehat{\varepsilon}_j^h(\Delta, X)$$

where

$$\widehat{\varepsilon}_j^g(\Delta, Y, X) := \sqrt{\tau_j} \left( \Delta - \widehat{g}_n(X, Y) \right) \sum_{j'=1}^{k_n} \left( n^{-1} \sum_{i=1}^n \Delta_i f_j(X_i, W_i) e_{j'}(Y_i, X_i) \right) e_{j'}(Y, X)$$

and

$$\widehat{\varepsilon}_j^h(\Delta, X) := \sqrt{\tau_j} \left( \Delta - \widehat{h}_n(X) \right) \sum_{j'=1}^{k_n} \left( n^{-1} \sum_{i=1}^n (1 - \Delta_i) f_j(X_i, W_i) e_{j'}(X_i) \right) e_{j'}(X).$$

and  $\widehat{\varepsilon}_{m_n} = (\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_{m_n})^t$ . We replace  $\Sigma$  by the  $m_n \times m_n$  dimensional matrix

$$\widehat{\Sigma}_{m_n} := n^{-1} \sum_{i=1}^n \left( \widehat{\varepsilon}_1(\Delta_i, Y_i, X_i, W_i), \dots, \widehat{\varepsilon}_{m_n}(\Delta_i, Y_i, X_i, W_i) \right)^t \left( \widehat{\varepsilon}_1(\Delta_i, Y_i, X_i, W_i), \dots, \widehat{\varepsilon}_{m_n}(\Delta_i, Y_i, X_i, W_i) \right).$$

Let  $(\widehat{\lambda}_j)_{1 \leq j \leq m_n}$  denote the ordered eigenvalues of  $\widehat{\Sigma}_{m_n}$ . We approximate  $\sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{m_n} \widehat{\lambda}_j \chi_{1j}^2$ . Indeed, we have  $\max_{1 \leq j \leq m_n} |\widehat{\lambda}_j - \lambda_j| = O_p(\mathbf{E} \|\widehat{\Sigma}_{m_n} - \Sigma_{m_n}\|) = o_p(1)$ , where  $\Sigma_{m_n}$  denote the upper  $m_n \times m_n$  matrix of  $\Sigma$ .  $\square$

## 2.5. Asymptotic distribution of the test statistic under MCAR(X)

In the following, we derive the asymptotic distribution of the test statistic  $S_n^{MCAR(X)}$  under the null hypothesis  $H_{MCAR(X)}$ . Let  $\mu(\Delta, X, W)$  be an infinite dimensional vector with  $j$ -th entry

$$\mu_j(\Delta, X, W) := \sqrt{\tau_j} \left( \Delta - h(X) \right) \left( f_j(X, W) - \sum_{l=1}^{\infty} \mathbf{E} [f_j(X, W) e_l(X)] e_l(X) \right).$$

We have  $\mathbf{E}[\mu_j(\Delta, X, W)] = 0$  under MCAR and we assume as in the previous subsection that  $\mathbf{E}[\sum_{l \geq 1} \mathbf{E}[f_j(X, W) e_l(X)]^2] < \infty$ . Let  $\Sigma_\mu$  be the covariance matrix of  $\mu(\Delta, X, W)$ ; that is,  $\Sigma_\mu = \mathbf{E}[\mu(\Delta, X, W) \mu(\Delta, X, W)^t]$ .

For the next result,  $(\lambda_l)_{l \geq 1}$  denote the ordered eigenvalues of  $\Sigma_\mu$ . Recall that  $\{\chi_{1l}^2\}_{l \geq 1}$  is a sequence of independent random variables that are distributed as chi-square with one degree of freedom. The next result is a direct consequence of Theorem 2.4 and hence, we omit its proof.

COROLLARY 2.5. Let Assumptions 1, 2, 4, and 5 hold true. If

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad n = o(\gamma_{l_n}^h), \quad \text{and} \quad m_n^{-1} = o(1) \quad (2.11)$$

then under MCAR(X)

$$n S_n^{\text{MCAR}(X)} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2.$$

REMARK 2.2 (Estimation of Critical Values). Estimation of critical values in case of Corollary 2.5 follows easily from Remark 2.1. Let us define

$$\widehat{\mu}_j(\Delta, X, W) = \sqrt{\tau_j}(\Delta - \widehat{h}_n(X)) \left\{ f_j(X, W) - \sum_{l=1}^{k_n} \left( n^{-1} \sum_{i=1}^n f_j(X_i, W_i) e_l(X_i) \right) e_l(X) \right\}.$$

We replace  $\Sigma_\mu$  by the  $m_n \times m_n$  dimensional matrix

$$\widehat{\Sigma}_{m_n} := n^{-1} \sum_{i=1}^n \left( \widehat{\mu}_1(\Delta_i, X_i, W_i), \dots, \widehat{\mu}_{m_n}(\Delta_i, X_i, W_i) \right)^t \left( \widehat{\mu}_1(\Delta_i, X_i, W_i), \dots, \widehat{\mu}_{m_n}(\Delta_i, X_i, W_i) \right).$$

Let  $(\widehat{\lambda}_j)_{1 \leq j \leq m_n}$  denote the ordered eigenvalues of  $\widehat{\Sigma}_{m_n}$ . We approximate  $\sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{m_n} \widehat{\lambda}_j \chi_{1j}^2$ . Consistency follows as in Remark 2.1.  $\square$

## 2.6. Asymptotic distribution of the test statistic under MCAR

We now derive the asymptotic distribution of the statistic for testing  $S_n^{\text{MCAR}}$  under the null hypothesis MCAR. Let us introduce an infinite dimensional vector  $v(\Delta, X, W)$  with  $j$ -th entry

$$v_j(\Delta, X, W) := \sqrt{\tau_j}(\Delta - \mathbf{E} \Delta) \left( f_j(X, W) - \mathbf{E}[f_j(X, W)] \right).$$

We have  $\mathbf{E}[v_j(\Delta, X, W)] = 0$  under MCAR. Let  $\Sigma_v$  be the covariance matrix of  $v(\Delta, X, W)$ ; that is,  $\Sigma_v = \mathbf{E}[v(\Delta, X, W)v(\Delta, X, W)^t]$ .

In this subsection, the ordered eigenvalues of  $\Sigma_v$  are denoted by  $(\lambda_l)_{l \geq 1}$ . The next result is a direct consequence of Theorem 2.4 and hence, we omit its proof.

COROLLARY 2.6. Let Assumptions 1, 2, and 4 hold true. If

$$\sum_{j=1}^{m_n} \tau_j = O(1), \quad \text{and} \quad m_n^{-1} = o(1)$$

then under MCAR

$$n S_n^{\text{MCAR}} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2.$$

REMARK 2.3 (Estimation of Critical Values). Estimation of critical values in case of Corollary 2.6 follows easily from Remark 2.1. Let us define

$$\widehat{v}_j(\Delta, X, W) := \sqrt{\tau_j}(\Delta - \widehat{\Delta}_n) \left\{ f_j(X, W) - n^{-1} \sum_{i=1}^n \Delta_i f_j(X_i, W_i) \right\}$$

with  $\widehat{\Delta}_n = n^{-1} \sum_{i=1}^n \Delta_i$ . We replace  $\Sigma_v$  by the  $m_n \times m_n$  dimensional matrix

$$\widehat{\Sigma}_{m_n} := n^{-1} \sum_{i=1}^n \left( \widehat{v}_1(\Delta_i, X_i, W_i), \dots, \widehat{v}_{m_n}(\Delta_i, X_i, W_i) \right)^t \left( \widehat{v}_1(\Delta_i, X_i, W_i), \dots, \widehat{v}_{m_n}(\Delta_i, X_i, W_i) \right).$$

Let  $(\widehat{\lambda}_j)_{1 \leq j \leq m_n}$  denote the ordered eigenvalues of  $\widehat{\Sigma}_{m_n}$ . We approximate  $\sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2$  by the finite sum  $\sum_{j=1}^{m_n} \widehat{\lambda}_j \chi_{1j}^2$ . Consistency follows as in Remark 2.1.  $\square$

## 2.7. Consistency against fixed alternatives

Under each null hypothesis, the asymptotic distribution results remain valid if  $(Y^*, X)$  is not bounded complete for  $(X, W)$ ; that is, Assumption 3 does not hold true. On the other hand, we see in the following that bounded completeness is a necessary condition to obtain consistency of our tests against fixed alternatives. To establish this property we require the following additional assumption.

ASSUMPTION 7. *The function  $p_{XW}/v$  is uniformly bounded away from zero.*

If MAR fails, Assumption 7 together with Assumption 3 ensures that the generalized Fourier coefficients  $E[r(\Delta, Y, X)f_j(X, W)]$  are non-zero for some integer  $j \geq 1$ . The following proposition shows that our test has the ability to reject a false null hypothesis with probability 1 as the sample size grows to infinity. For the next results, let us introduce a sequence  $(a_n)_{n \geq 1}$  satisfying  $a_n = o(n)$ . The proof of the next proposition can be found in the appendix.

PROPOSITION 2.7. *Assume that MAR does not hold. Let Assumptions 1–7 be satisfied. Then*

$$\mathbb{P}(n S_n^{MAR} > a_n) = 1 + o(1).$$

The rate  $(a_n)_{n \geq 1}$  is arbitrarily close to the parametric rate  $n^{-1}$  which is due the weighting sequence  $(\tau_j)_{j \geq 1}$  with  $\sum_{j=1}^{m_n} \tau_j = O(1)$ . The next two results are direct consequences of Proposition 2.7 and hence, their proofs are omitted.

COROLLARY 2.8. *Assume that MCAR(X) does not hold. Let Assumptions 1–5 and 7 be satisfied. Then*

$$\mathbb{P}(n S_n^{MCAR(X)} > a_n) = 1 + o(1).$$

COROLLARY 2.9. *Assume that MCAR does not hold. Let Assumptions 1–4 and 7 be satisfied. Then*

$$\mathbb{P}(n S_n^{MCAR} > a_n) = 1 + o(1).$$

### 3. Monte Carlo simulation

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo simulation. The experiments use a sample size of 500 and there are 1000 Monte Carlo replications in each experiment. Results are presented for the nominal level 0.05.

As basis functions  $\{f_j\}_{j \geq 1}$  used to construct our test statistic, we use throughout the experiments orthonormalized Hermite polynomials. Hermite polynomials form an orthonormal basis of  $L^2_\omega$  with a weighting function being the density of the standard normal distribution; that is,  $\omega(x) = \exp(-x^2)/\sqrt{2\pi}$ . They can be obtained by applying the Gram-Schmidt procedure to the polynomial series  $1, x, x^2, \dots$  under the inner product  $\langle \phi, \psi \rangle_\omega = (2\pi)^{-1/2} \int \phi(x)\psi(x) \exp(-x^2)dx$ . That is,  $H_1(x) = 1$  and for all  $j = 2, 3, \dots$

$$H_j(x) = \frac{x^{j-1} - \sum_{k=1}^{j-1} \langle id^{j-1}, p_j \rangle_\omega p_k(x)}{\int (x^{j-1} - \sum_{k=1}^{j-1} \langle id^{j-1}, p_j \rangle_\omega p_k(x)) \omega(x) dx}. \quad (3.1)$$

Our testing procedure is now build up on the basis functions

$$f_j(\cdot) = \frac{H_{j+1}(\cdot)}{\sqrt{\langle H_j, H_j \rangle_\omega}}$$

for all  $j = 1, 2, \dots$ . If the support of the instrument  $W$  or its transformation lies in the interval  $[0, 1]$  then one could also use, for instance, cosine basis functions

$$f_j(x) = \sqrt{2} \cos(\pi j x)$$

for  $j = 1, 2, \dots$ . We also implemented our test statistic with these cosine functions in the settings studied below. But as the results are very similar to the ones with Hermite polynomials presented below we do not report them here. Throughout our simulation study, the number of orthonormalized Hermite polynomials is 10. Due to the weighting sequence  $(\tau_j)_{j \geq 1}$ , results not too sensitive to the number of Hermite polynomials. When implementing the test with cosine basis functions we used 100 basis functions. In contrast, results might be more sensitive to the choice of basis functions used to estimate  $g$  and  $h$ . Below we use cross validation to choose the appropriate number of basis functions for these functions.

**Testing MCAR** Realizations of  $(Y^*, W)$  were generated by  $W \sim \mathcal{N}(0, 1)$  and  $Y^* \sim \rho W + \sqrt{1 - \rho^2} \varepsilon$  where  $\varepsilon \sim \mathcal{N}(0, 1)$ . The constants  $\rho$  characterizes the "strength" of the instrument  $W$  and is varied in the experiments. For a random variable  $V$ , introduce the function  $\phi_{.2}(V) = \mathbb{1}\{V \geq q\} + 0.1 * \mathbb{1}\{V \leq q\}$  where  $q$  is the 0.2 quantile of the empirical distribution of  $V$ . In each experiment, realizations of the response variable  $\Delta$  were generated by

$$\Delta \sim \text{Bin}\left(1, \phi_{.2}\left(\nu Y^* + \sqrt{1 - \nu^2} \xi\right)\right)$$

for some constant  $0 \leq \nu \leq 1$  and where  $\xi \sim \mathcal{N}(0, 1)$ . If  $\nu = 0$  then response  $\Delta$  does not depend on  $Y^*$  and hence the null hypothesis MCAR holds true.

The critical values are estimated as in Remark 2.3. For  $m = 100$  we observed that the estimated eigenvalues  $\hat{\lambda}_j$  are sufficiently close to zero for all  $j \geq m$ . To provide a basis for



<i>Model</i>		<i>Empirical Rejection probability of <math>nS_n</math> with</i>			<i>Little's test</i>
$\rho$	$\nu$	$\tau_j = j^{-2}$	$\tau_j = j^{-3}$	$\tau_j = j^{-4}$	
0.2	0.0	0.055	0.057	0.056	0.062
	0.3	0.148	0.159	0.162	0.155
	0.5	0.290	0.297	0.304	0.338
	0.7	0.505	0.529	0.530	0.568
0.3	0.0	0.055	0.057	0.056	0.062
	0.3	0.253	0.265	0.268	0.273
	0.5	0.559	0.588	0.592	0.611
	0.7	0.839	0.855	0.857	0.853
0.4	0.0	0.055	0.057	0.056	0.062
	0.3	0.387	0.393	0.398	0.430
	0.5	0.813	0.831	0.840	0.843
	0.7	0.985	0.988	0.986	0.988

Table 1: Empirical Rejection probabilities for Testing MCAR

judging whether the power of our test is high or low, we also provide the empirical rejection probabilities when using a test of MCAR for normal data proposed by Little [1988]. The empirical rejection probabilities of test statistic  $S_n^{MCAR}$  using different weightings and Little's test are depicted in Table 1. First, we observe, not surprisingly, that the power of all tests increase as the correlation between  $Y^*$  and  $W$  (measured by  $\rho$ ) becomes larger. Second, power also increases with constant  $\nu$ . From Table 1 we also see that our tests with different weighting sequences have similar power properties and our tests behave similar as Little's test.

**Testing MAR** Realizations of  $(Y^*, W)$  were generated by  $W \sim \mathcal{N}(0, 1)$  and  $Y^* \sim \rho W + \sqrt{1 - \rho^2} \varepsilon$  where  $\varepsilon \sim \mathcal{N}(0, 1)$  and the constant  $\rho \in (0, 1)$  is varied in the experiments. In each experiment, realizations of  $\Delta$  were generated by

$$\Delta \sim \begin{cases} 1, & \text{if } Y^* \in (0, 0.5), \\ \text{Bin}(1, \phi_{.2}(\nu Y^* + \sqrt{1 - \nu^2} \xi)), & \text{otherwise.} \end{cases}$$

for some constant  $0 \leq \nu \leq 1$  and where  $\xi \sim \mathcal{N}(0, 1)$ . If  $\nu = 0$  then response  $\Delta$  only depends on observed realizations  $(0, 0.5)$  and thus, the null hypothesis MAR holds true.

To construct the test statistic, we estimate the function  $g(\cdot) = \mathbf{E}(\Delta|Y = \cdot)$  using B-splines. The number of knots and orders is chosen via cross validation. Computational procedures were implemented using the statistical software R using the crs Package Hayfield and Racine [2007]. In our experiments, cross validation tended to undersmooth the estimator of  $g$  which implied a sufficiently small bias of this estimator. On the other hand, to obtain appropriate undersmoothing one could also use a data driven choice of basis functions suggested by Picard and Tribouley [2000]. The critical values are estimated as described in Remark 2.1. In Table 2 we depict the empirical rejection probabilities when using different weightings.



Model		Empirical Rejection probability of $nS_n$ with		
$\rho$	$\nu$	$\tau_j = j^{-2}$	$\tau_j = j^{-3}$	$\tau_j = j^{-4}$
0.3	0.0	0.044	0.048	0.050
	0.3	0.199	0.205	0.206
	0.5	0.295	0.300	0.302
	0.7	0.399	0.408	0.414
0.5	0.0	0.055	0.053	0.050
	0.3	0.493	0.496	0.493
	0.5	0.701	0.710	0.714
	0.7	0.860	0.869	0.869
0.7	0.0	0.097	0.076	0.069
	0.3	0.817	0.803	0.789
	0.5	0.954	0.951	0.950
	0.7	0.994	0.995	0.995

Table 2: Empirical Rejection probabilities for Testing MAR

**Testing MCAR(X)** Realizations of  $(Y^*, X, W)$  were generated by  $W \sim \mathcal{N}(0, 1)$ ,  $X \sim 0.2W + \sqrt{1 - 0.2^2} \xi$  and  $Y^* \sim \rho W + \sqrt{1 - \rho^2} \xi + \varepsilon$  where  $\xi, \varepsilon \sim \mathcal{N}(0, 0.25)$ . The constant  $\rho$  is varied in the experiments. The critical values are estimated as described in Remark 2.2. In each experiment, realizations of response  $\Delta$  were generated by

$$\Delta \sim \text{Binomial}\left(1, \phi_{.2}\left(\nu Y^* + \sqrt{1 - \nu^2} X\right)\right)$$

for some constant  $0 \leq \nu \leq 1$ . Clearly, if  $\nu = 0$  then the null hypothesis MCAR(X) holds true. We estimate the functions  $h$  using B-splines. Again, the number of knots and orders is chosen via cross validation. Table 3 depicts the empirical rejection probabilities of the test  $S_n^{MCAR(X)}$  when using different weightings.

## 4. Empirical Illustration

We now apply our testing procedure to analyze response mechanisms in a data set from the Health and Retirement Study (HRS). In this survey, respondents were asked about their out of pocket prescription drug spending. Those who were not able to report point values for these were asked to provide brackets, giving point values for some observations and intervals of different sizes for others. This censoring problem might violate the MAR hypothesis: the variable is censored only for those who do not recall how much they spent, and remembering how much one spent might be correlated with the level of spending itself. We refer to Armstrong [2015] who constructed confidence intervals for partially identified regression of prescription drug spending on income.

In this empirical illustration, we consider the 1996 wave of the survey and restrict attention to women with less than 25,000\$ of yearly income who report using prescription medications. This results in a data set with 943 observations. Of these observations, participants do not report the exact prescription expenditures but rather an interval of nonzero width

<i>Model</i>		<i>Empirical Rejection probability of <math>nS_n</math> with</i>		
$\rho$	$\nu$	$\tau_j = j^{-2}$	$\tau_j = j^{-3}$	$\tau_j = j^{-4}$
0.3	0.0	0.041	0.046	0.048
	0.3	0.061	0.060	0.059
	0.5	0.118	0.123	0.122
	0.7	0.199	0.211	0.205
0.5	0.0	0.041	0.046	0.048
	0.3	0.180	0.186	0.185
	0.5	0.545	0.561	0.575
	0.7	0.823	0.851	0.864
0.7	0.0	0.041	0.046	0.048
	0.3	0.406	0.417	0.426
	0.5	0.876	0.895	0.904
	0.7	0.980	0.990	0.995

Table 3: Empirical Rejection probabilities for Testing MCAR(X)

with finite endpoints or provide no information on their prescription expenditures at all. The length of these intervals is 10, 80, and 400. If the participants do not report their exact amount but an interval of length 10 we consider the center of this interval as observation (these are only six observations). Intervals with width larger or equal to 80 are treated as missing values. Thereby, 55 observations are missing (roughly 6%). The results presented below are essentially the same when intervals with length 10 are treated as missing observations.

Whether we observe prescription drug expenditure  $Y^*$  is assumed to be independent of the yearly income  $W$  conditional on  $Y^*$ . We thus use yearly income as instrumental variable. The test statistics are constructed as described in the previous section. More precisely, we truncate  $W$  to  $[0, 1]$  and use cosine basis functions with  $m_n = 100$ . As we see from Table 4 our test statistics reject the hypothesis MCAR but fail to reject the hypothesis MAR.

		$\tau_j = j^{-1}$	$\tau_j = j^{-2}$	$\tau_j = j^{-3}$
MCAR	Value of $nS_n^{MCAR}$	0.273	0.233	0.230
	Critical Values	0.214	0.170	0.163
MAR	Value of $nS_n^{MAR}$	0.194	0.143	0.138
	Critical Values	0.231	0.188	0.183

Table 4: Values of  $nS_n^{MCAR}$  and  $nS_n^{MAR}$  together with their critical values

In Figure 1, we estimate the conditional probability  $\mathbb{P}(\Delta = 1|Y^*)$ . This conditional probability is identified through the conditional mean equation  $\mathbb{E}[\Delta/\mathbb{P}(\Delta = 1|Y^*)|X] = 1$  (cf. D'Haultfoeuille [2010]). We use a sieve minimum distance estimator based on B-splines as in Breunig et al. [2014] to estimate  $\mathbb{P}(\Delta = 1|Y^*)$ . From Figure 1 we see that the estimator

for  $\mathbb{P}(\Delta = 1|Y^*)$  is strictly decreasing with  $Y^*$ . On the other hand, if data is MCAR then  $\mathbb{P}(\Delta = 1|Y^*)$  would be constant. Significant statements about the shape of the curve, however, are hard to make as the solution to the conditional moment restriction is an ill-posed problem and hence confidence intervals can be very wide.

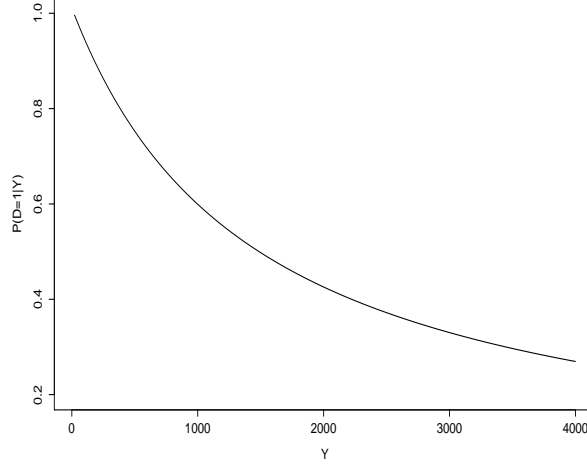


Figure 1: Graph of  $\mathbb{P}(\Delta = 1|Y^* = \cdot)$

## A. Appendix

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

Throughout the Appendix, let  $C > 0$  denote a generic constant that may be different in different uses. For ease of notation let  $\sum_i = \sum_{i=1}^n$  and  $\sum_{i' < i} = \sum_{i=1}^n \sum_{i'=1}^{i-1}$ . Further, to keep notation simple we define  $V := (Y, X)$  and  $Z := (X, W)$ . In the following,  $\varepsilon_{\underline{m}_n}(\Delta, V, W)$ ,  $\varepsilon_{\underline{m}_n}^g(\Delta, V)$ , and  $\varepsilon_{\underline{m}_n}^h(\Delta, X)$  denote  $m_n$ -dimensional vectors with  $j$ -th entries given by  $\varepsilon_j(\Delta, V, W)$ ,  $\varepsilon_j^g(\Delta, V)$ , and  $\varepsilon_j^h(\Delta, X)$ , respectively. In the appendix,  $f_{\underline{m}_n}^\tau$  denotes a  $m_n$  dimensional vector with entries  $\sqrt{\tau_j} f_j$  for  $1 \leq j \leq m_n$ .

**PROOF OF THEOREM 2.4.** The proof is based on the decomposition

$$\begin{aligned}
& n^{-1/2} \sum_i \left( \Delta_i - \Delta_i \widehat{g}_n(V_i) - (1 - \Delta_i) \widehat{h}_n(X_i) \right) f_{\underline{m}_n}(Z_i) \\
&= n^{-1/2} \sum_i \varepsilon_{\underline{m}_n}(\Delta_i, V_i, W_i) \\
&+ n^{-1/2} \sum_i \left( \Delta_i (g(V_i) - \widehat{g}_n(V_i)) f_{\underline{m}_n}(Z_i) + \varepsilon_{\underline{m}_n}^g(\Delta_i, V_i) \right) \\
&+ n^{-1/2} \sum_i \left( (1 - \Delta_i) (h(X_i) - \widehat{h}_n(X_i)) f_{\underline{m}_n}(Z_i) + \varepsilon_{\underline{m}_n}^h(\Delta_i, X_i) \right) \\
&= I_n + II_n + III_n \quad (\text{say}).
\end{aligned}$$

Consider  $I_n$ . Consider some fixed integer  $m \geq 1$ . Using Cramer Wold device it is easily seen that

$$n^{-1/2} \sum_i \varepsilon_{\underline{m}}(\Delta_i, V_i, W_i) \xrightarrow{d} \mathcal{N}(0, \Sigma_m)$$

where  $\Sigma_m$  is the upper  $m \times m$  submatrix of  $\Sigma$ . Hence, we have

$$\sum_{j=1}^m \left| n^{-1/2} \sum_i \varepsilon_j(\Delta_i, V_i, W_i) \right|^2 \xrightarrow{d} \sum_{j=1}^m \lambda_j \chi_{1j}^2$$

with  $\lambda_j$ ,  $1 \leq j \leq m$ , being eigenvalues of  $\Sigma_m$ . On the other hand, observe

$$\sum_{j>m} \mathbf{E} \left| n^{-1/2} \sum_i \varepsilon_j(\Delta_i, V_i, W_i) \right|^2 = \sum_{j>m} \mathbb{V}\text{ar} \left( n^{-1/2} \sum_i \varepsilon_j(\Delta_i, V_i, W_i) \right) = \sum_{j>m} \mathbf{E} \varepsilon_j^2(\Delta, V, W)$$

which becomes sufficiently small for large  $m$  as  $\mathbf{E} \varepsilon_j^2(\Delta, V, W)/\tau_j \leq C$  for all  $j \geq 1$ . Hence, from page 199 in Serfling [1981] we infer that  $I_n \xrightarrow{d} \sum_{j \geq 1} \lambda_j \chi_{1j}^2$ .

Consider  $II_n$ . We have

$$\begin{aligned} \|II_n\|^2 &\leq 2 \sum_{j=1}^{m_n} \tau_j \left| n^{-1/2} \sum_i \Delta_i (\widehat{g}_n(V_i) - (E_{k_n} g)(V_i)) f_j(Z_i) - \varepsilon_j^g(\Delta_i, V_i) \right|^2 \\ &\quad + 2 \sum_{j=1}^{m_n} \tau_j \left| n^{-1/2} \sum_i \Delta_i (E_{k_n} g - g)(V_i) f_j(Z_i) \right|^2 \\ &=: A_{n1} + A_{n2}. \end{aligned}$$

Consider  $A_{n1}$ . In the following, we denote  $Q_n := n^{-1} \sum_i e_{\underline{k}_n}(V_i) e_{\underline{k}_n}(V_i)^t$ . It holds  $\widehat{g}_n(\cdot) = e_{\underline{k}_n}(\cdot)^t (n Q_n)^{-1} \sum_i \Delta_i e_{\underline{k}_n}(V_i)$ . By Assumption 5, the eigenvalues of  $\mathbf{E}[e_{\underline{k}_n}(V) e_{\underline{k}_n}(V)^t]$  are bounded away from zero and hence, it may be assumed that  $\mathbf{E}[e_{\underline{k}_n}(V) e_{\underline{k}_n}(V)^t] = \mathbf{I}_{k_n}$  where  $\mathbf{I}_{k_n}$  is the  $k_n$  dimensional identity matrix (cf. Newey [1997], p. 161). We observe

$$\begin{aligned} A_{n1} &\leq 2 \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)] Q_n^{-1} n^{-1/2} \sum_i (\Delta_i - E_{k_n} g(V_i)) e_{\underline{k}_n}(V_i) - \varepsilon_j^g(\Delta_i, V_i) \Big|^2 \\ &\quad + 2 \|E_{k_n} g - \widehat{g}_n\|_V^2 \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \left| n^{-1/2} \sum_i \Delta_i e_l(V_i) f_j(Z_i) - \mathbf{E}[\Delta e_l(V) f_j(Z)] \right|^2 \\ &=: 2B_{n1} + 2B_{n2} \quad (\text{say}). \quad (\text{A.1}) \end{aligned}$$

For  $B_{n1}$  we evaluate due to the relation  $Q_n^{-1} = \mathbf{I}_{k_n} - Q_n^{-1}(Q_n - \mathbf{I}_{k_n})$  that

$$\begin{aligned} B_{n1} &\leq 2 \left\| \mathbf{E}[\Delta f_{\underline{m}_n}^\tau(Z) e_{\underline{k}_n}(V)^t] n^{-1/2} \sum_i (\Delta_i - E_{k_n} g(V_i)) e_{\underline{k}_n}(V_i) - \varepsilon_{\underline{m}_n}^g(\Delta_i, V_i) \right\|^2 \\ &\quad + 2 \left\| \mathbf{E}[\Delta f_{\underline{m}_n}^\tau(Z) e_{\underline{k}_n}(V)^t] \right\|^2 \|Q_n - \mathbf{I}_{k_n}\|^2 \|Q_n^{-1}\|^2 \left\| n^{-1/2} \sum_i (\Delta_i - E_{k_n} g(V_i)) e_{\underline{k}_n}(V_i) \right\|^2 \\ &=: 2C_{n1} + 2C_{n2} \quad (\text{say}). \end{aligned}$$

Further, from  $\mathbf{E}[(\Delta - E_{k_n}g(V))e_{\underline{k_n}}(V)] = 0$ ,  $\mathbf{E}[(g - E_{k_n}g)(V)e_{\underline{k_n}}(V)] = 0$ , and  $\mathbf{E}[\varepsilon_{m_n}(\Delta, V, W)] = 0$  we deduce

$$\begin{aligned}
C_{n1} &\leq 2 \sum_{j=1}^{m_n} \tau_j \mathbf{E} \left| \sum_{l > k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)] (g(V) - \Delta) e_l(V) \right|^2 \\
&\quad + 2 \sum_{j=1}^{m_n} \tau_j \mathbf{E} \left| \sum_{l=1}^{k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)] (E_{k_n}g - g)(V) e_l(V) \right|^2 \\
&\leq 2 \sum_{j=1}^{m_n} \tau_j \mathbf{E} \left| \sum_{l > k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)] e_l(V) \right|^2 \\
&\quad + C k_n \|E_{k_n}g - g\|_V^2 \sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)]^2 \\
&= o(1)
\end{aligned}$$

using that  $\mathbf{E}[(g(V) - \Delta)^2 | V]$  is bounded,  $\sum_{j=1}^{m_n} \tau_j \sum_{l=1}^{k_n} \mathbf{E}[\Delta f_j(Z) e_l(V)]^2 = O(1)$ , and by assumption

$$k_n^2 \|E_{k_n}g - g\|_V^2 = O(k_n^2 / \gamma_{k_n}^g) = O(k_n^2 / n) = o(1).$$

Consider  $C_{n2}$ . Further, by Rudelson's matrix inequality (see Rudelson [1999] and also Lemma 6.2 of Belloni et al. [2012]) it holds

$$\|Q_n - \mathbf{I}_{k_n}\|^2 = O_p(n^{-1} \log(n) k_n).$$

Moreover, since the difference of eigenvalues of  $Q_n$  and  $\mathbf{I}_{k_n}$  is bounded by  $\|Q_n - \mathbf{I}_{k_n}\|$ , the smallest eigenvalue of  $Q_n$  converges in probability to one and hence,  $\|Q_n^{-1}\|^2 = 1 + o_p(1)$ . Further,

$$\sum_{l=1}^{k_n} \mathbf{E} \left| n^{-1/2} \sum_i (\Delta_i - E_{k_n}g(V_i)) e_l(V_i) \right|^2 = \sum_{l=1}^{k_n} \mathbf{E} \left| (\Delta - E_{k_n}g(V)) e_l(V) \right|^2 = O(k_n)$$

and hence  $C_{n2} = O_p(n^{-1} \log(n) k_n^2) = o_p(1)$ . Consequently,  $B_{n1} = o_p(1)$ . Consider  $B_{n2}$ . It holds

$$\sum_{l=1}^{k_n} \left| n^{-1} \sum_i \Delta_i e_l(V_i) f_j(Z_i) - \mathbf{E}[\Delta e_l(V) f_j(Z)] \right|^2 = O_p(k_n/n).$$

Since  $\|E_{k_n}g - \widehat{g}_n\|_V^2 = O_p(k_n/n)$  (cf. Theorem 1 of Newey [1997]) and  $k_n^2/n = o(1)$  it follows that  $B_{n2} = o(1)$ . Thus, we conclude  $A_{n1} = o_p(1)$ . For  $A_{n2}$  we observe that

$$\mathbf{E} A_{n2} \leq n \sum_{j=1}^{m_n} \tau_j \mathbf{E} |(E_{k_n}g(V) - g(V)) f_j(Z)|^2 \leq n \|E_{k_n}g - g\|_\infty^2 \sum_{j=1}^{m_n} \tau_j \mathbf{E} |f_j(Z)|^2 = O(n / \gamma_{k_n}^g) = o(1).$$

Consequently, we have  $II_n = o_p(1)$ . The proof of  $III_n = o_p(1)$  is analogous.  $\square$

**PROOF OF PROPOSITION 2.7.** Let us introduce a smoothing operator  $K$  which has an eigenvalue decomposition  $\{\sqrt{\tau_j}, f_j\}_{j \geq 1}$  and a conditional expectation operator  $T$  defined by  $T\phi = \mathbb{E}[\phi(\Delta, V)|Z]$  for any bounded function  $\phi$ . Since  $p_Z/v$  is uniformly bounded away from zero by some constant  $C > 0$  we obtain

$$\begin{aligned}
S_n &= \sum_{j=1}^{m_n} \tau_j \left| n^{-1} \sum_i r(\Delta_i, V_i) f_j(Z_i) \right|^2 + o_p(1) \\
&= \sum_{j=1}^{m_n} \tau_j \left| \mathbb{E} \left[ r(\Delta, V) f_j(Z) \right] \right|^2 + o_p(1) \\
&= \sum_{j=1}^{\infty} \left| \int_{\mathcal{X} \times \mathcal{W}} \sqrt{\tau_j} \mathbb{E}(r(\Delta, V)|Z = z) \frac{p_Z(z)}{v(z)} f_j(z) v(z) dz \right|^2 + o_p(1) \\
&= \int_{\mathcal{X} \times \mathcal{W}} \left| (KTr)(z) \frac{p_Z(z)}{v(z)} \right|^2 v(z) dz + o_p(1) \\
&\geq C \mathbb{E} \left| (KTr)(Z) \right|^2 + o_p(1).
\end{aligned}$$

Since  $K$  is nonsingular by construction it follows from the proof of Theorem 2.1 that  $\mathbb{E} |(KTr)(Z)|^2 > 0$ .  $\square$

## References

- T. B. Armstrong. Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics*, 2015.
- A. Belloni, V. Chernozhukov, D. Chetverikov, and K. Kato. On the asymptotic theory for least squares series: pointwise and uniform results. *arXiv preprint arXiv:1212.0442*, 2012.
- R. Blundell and J. Horowitz. A nonparametric test of exogeneity. *Review of Economic Studies*, 74(4):1035–1058, Oct 2007.
- R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.
- C. Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.
- C. Breunig, E. Mammen, and A. Simoni. Semi/nonparametric estimation in case of endogenous selection. Technical report, University of Mannheim (submitted.), 2014.
- H. Y. Chen and R. Little. A test of missing completely at random for generalised estimating equations with missing data. *Biometrika*, 86(1):1–13, 1999.
- 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. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 2007.

- M. Das, W. K. Newey, and F. Vella. Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1):33–58, 2003.
- 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(03):460–471, 2011.
- F. Fève, J.-P. Florens, I. Van Keilegom, et al. Estimation of conditional ranks and tests of exogeneity in nonparametric nonseparable models. Technical report, UCL, 2012.
- C. Fuchs. Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, 77(378):270–278, 1982.
- T. Hayfield and J. S. Racine. np: Nonparametric kernel smoothing methods for mixed datatypes. *R package version 0.13-1*, 2007.
- J. Heckman. Shadow prices, market wages, and labor supply. *Econometrica: Journal of the Econometric Society*, pages 679–694, 1974.
- J. L. Horowitz. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74(2):521–538, 2006.
- Y. Hu and S. M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008.
- J. G. Ibrahim, M.-H. Chen, S. R. Lipsitz, and A. H. Herring. Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 100(469):332–346, 2005.
- G. W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29, 2004.
- P. Kline and A. Santos. Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics*, 4(2):231–267, 2013.
- K.-Y. Liang and J. Qin. Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):773–786, 2000.
- R. J. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404):1198–1202, 1988.
- R. J. Little and D. B. Rubin. Statistical analysis with missing data. 2002.
- L. Mattner. Some incomplete but boundedly complete location families. *The Annals of Statistics*, 21(4):2158–2162, 1993.
- W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147 – 168, 1997.
- D. Picard and K. Tribouley. Adaptive confidence interval for pointwise curve estimation. *Annals of statistics*, pages 298–335, 2000.

- A. Qu and P. X.-K. Song. Testing ignorable missingness in estimating equation approaches for longitudinal data. *Biometrika*, 89(4):841–850, 2002.
- E. A. Ramalho and R. J. Smith. Discrete choice non-response. *The Review of Economic Studies*, 80(1):343–364, 2013.
- M. Reilly and M. S. Pepe. A mean score method for missing and auxiliary covariate data in regression models. *Biometrika*, 82(2):299–314, 1995.
- J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.
- D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
- D. B. Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons, 2004.
- M. Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.
- R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 1981.
- 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.

**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".

