HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 22/22

Selecting Valid Instrumental Variables in Linear Models with Multiple Exposure Variables: Adaptive Lasso and the Median-of-Medians Estimator

Xiaoran Liang; Eleanor Sanderson and Frank Windmeijer

July 2022

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Selecting Valid Instrumental Variables in Linear Models with Multiple Exposure Variables: Adaptive Lasso and the Median-of-Medians Estimator

Xiaoran Liang[a], Eleanor Sanderson[b,d], Frank Windmeijer[b,c]

[a]School of Economics, University of Bristol, UK

[b]MRC Integrative Epidemiology Unit, University of Bristol, UK

[c]Dept of Statistics and Nuffield College, University of Oxford, UK

[d]Population Health Science Institute, Bristol Medical School, University of Bristol, UK

May 31, 2022

Abstract

In a linear instrumental variables (IV) setting for estimating the causal effects of multiple confounded exposure/treatment variables on an outcome, we investigate the adaptive Lasso method for selecting valid instrumental variables from a set of available instruments that may contain invalid ones. An instrument is invalid if it fails the exclusion conditions and enters the model as an explanatory variable. We extend the results as developed in Windmeijer et al. (2019) for the single exposure model to the multiple exposures case. In particular we propose a median-of-medians estimator and show that the conditions on the minimum number of valid instruments under which this estimator is consistent for the causal effects are only moderately stronger than the simple majority rule that applies to the median estimator for the single exposure case. The adaptive Lasso method which uses the initial median-of-medians estimator for the penalty weights achieves consistent selection with oracle properties of the resulting IV estimator. This is confirmed by some Monte Carlo simulation results. We apply the method to estimate the causal effects of educational attainment and cognitive ability on body mass index (BMI) in a Mendelian Randomization setting.

Keywords: Causal inference; Adaptive Lasso; Instrumental variables; Invalid instruments; Mendelian randomization; Median-of medians estimator

1 Introduction

Instrumental variable (IV) methods are widely used to determine the causal effect of a treatment/exposure on an outcome when their relationship is potentially confounded by unobserved factors. In IV estimation, it is crucial that instruments are valid. This requires that (a) the instruments must be associated with the exposure variable (the relevance condition), and (b) the only pathway from the instruments to the outcome is through the exposure; the instruments do not have direct effects on the outcome nor affect the outcome through unobservables (the exclusion conditions). In our setting, we are concerned with the situation where we have a fixed, but large number of available instruments that satisfy the relevance condition. However, some of the instruments may violate the exclusion conditions and hence are invalid. If we include these invalid instruments in IV estimation, the resulting estimator will be inconsistent. It is therefore important to have selection methods that consistently selects the valid instruments.

Previous work has addressed the IV selection problem in the case of a single exposure variable. Kang et al. (2016) establish the model setup for this IV selection. They develop the identification conditions and propose a selection method based on the Lasso (Tibshirani, 1996). Windmeijer et al. (2019) propose a method based on the adaptive Lasso (Zou, 2006) under the assumption that more than half of the candidate instruments are valid; the so-called majority rule. The median of the instrument-specific estimates is then a consistent estimator of the causal effect and can be used for the penalisation of the adaptive Lasso, resulting in consistent selection of the valid instruments and oracle properties of the post-selection IV estimator, meaning that the IV estimator behaves in large samples as if the set of valid instruments were known (Fan and Li, 2001). Guo et al. (2018) refine the identification condition proposed by Kang et al. (2016) and establish the sufficient and necessary identification condition which is the plurality rule. It states that the valid instruments form the largest group, where instruments form a group if the instrument-specific estimators for the causal effect converge to the same value, and is hence a relaxation of the majority rule. The Hard Thresholding with Voting method proposed by Guo et al. (2018) can achieve consistent selection under the plurality rule. Also assuming the plurality rule, Windmeijer et al. (2021) propose the Confidence Interval method which result in consistent selection, and has as an advantage over the Hard Thresholding with Voting method that the number of instruments selected as valid in the Confidence Interval method decreases monotonically when decreasing the tuning

parameter.

Unlike the existing literature above, we consider here the case of multiple, potentially confounded exposure variables. This setting can be motivated by recent Mendelian Randomization (MR) studies in epidemiology. In MR studies, genetic variants are used as instruments for estimating the causal effect of an modifiable exposure on a health-related outcome. In many cases, there are additional exposure variables that need to be considered apart from the primary exposure. For example, Sanderson et al. (2019) estimate the effect of educational attainment on body mass index (BMI) conditional on cognitive ability. Both educational attainment and cognitive ability are confounded by unobserved factors that affect both the outcome and the exposure variables. Therefore, a method to select the valid instruments needs to take account of the multiple exposure variables problem.

We contribute to the literature by extending the adaptive Lasso method in Windmeijer et al. (2019) to allow for multiple exposure variables. The main issue for the adaptive Lasso is to have an initial consistent estimator of the causal effects that can be used for the penalisation. For the single exposure case, the median of the instrument-specific estimates of the causal effect is a consistent estimator when more than 50% of the instruments are valid and satisfies the conditions for oracle properties when used in the adaptive Lasso for instrument selection. This could simply be extended for the multiple exposure case to the medians of all just-identified estimates of the causal effects. A just-identified estimator is one where the number of instruments used is equal to the number of exposure variables. Let k_x and k_z denote the number of exposure and instrumental variables respectively. Then there are $\binom{k_z}{k_x}$ just-identified estimators of the causal effects and if more than 50% of these are consistent, then the medians of these p estimators are consistent. Let $k_\mathcal{V}$ denote the number of valid instruments. Under a strong relevance assumption that each set of just-identifying instruments are jointly relevant for all exposure variables, this majority rule then implies that $\binom{k_\mathcal{V}}{k_x} > \frac{1}{2}\binom{k_z}{k_x}$. As an example, with $k_x = 2$ and $k_z = 21$, we have 210 just-identifying pairs of instruments, of which more than 105 need to be pairs of valid instruments. This implies that for this naive median estimator at least 16 instruments need to be valid.

We propose a novel median-of-medians estimator which we show to be a consistent estimator of the causal effects and which utilises the available information better, in the sense that it requires less instruments to be valid for consistency compared to the naive median estimator. We show in Section 3 that for $k_x \geq 2$, the median-of-medians estimator

is consistent if $k_{\mathcal{V}} > \frac{k_z+k_x-1}{2}$. This condition is a (weakly) weaker condition on the number of valid instruments than for the naive median estimator, with the difference increasing in k_z. For the case of $k_x = 2$ this results in the condition that $k_{\mathcal{V}} > \frac{k_z+1}{2}$, and so for $k_z = 21$ this implies that at least 12 instruments need to be valid. In other words, whereas the naive median estimator allows in this case for a maximum of 5 instruments to be invalid, the median-of-medians estimator is still consistent with 9 invalid instruments. The condition for the median-of-medians estimator is a natural progression of the condition for the single-exposure majority rule that $k_{\mathcal{V}} > \frac{k_z}{2}$, and for $k_x = 2$, it is only stricter when k_z is odd. For $k_x > 2$, the estimator is strictly speaking a median-of-medians-of-medians.... estimator, but for brevity, we will call it the median-of-medians estimator for all k_x.

The assumption that all just-identifying sets of instruments are jointly relevant for all exposure variables may not hold in practice. In our application, genetic variants that are candidate instruments for educational attainment and cognitive ability are identified in separate GWAS studies and there is very little overlap of genetic variants between the two traits. We can adjust the median-of-medians estimator for this block structure of the instruments by only considering in this case the just-identifying pairs of instruments, where each pair contains one instrument from each group.

The paper is structured as follows. Section 2 introduces the model, IV estimation and the adaptive Lasso IV selection method for selecting the valid/invalid instruments. Section 3 introduces the median-of-medians estimator and derives its properties. Section 4.1 we discuss the median-of-medians estimator based consistent selection and oracle properties of the adaptive Lasso method, also combining it with the downward testing procedure for model selection proposed by Andrews (1999). In Section 5, we introduce the block structure variation of the method that accounts for violation of the full rank assumption. Section 6 presents some Monte Carlo simulation results. In Section 7, we apply our method to Mendelian randomisation and estimate the causal effects of educational attainment and cognitive ability on BMI. Section 8 concludes.

Notation. In the remainder of the paper, let $\|\{.\}\|_q$ denote the l_q-norm of a vector. For a matrix $\mathbf{X}_{n\times p}$ with full column rank, let $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and $\mathbf{M}_X = \mathbf{I}_n - \mathbf{P}_X$, where $\mathbf{I}_n$ is the n-dimensional identity matrix. For a general matrix $\mathbf{A}$, $r(\mathbf{A})$ denotes its rank. Convergence in probability and distribution are indicated by $\xrightarrow{p}$ and $\xrightarrow{d}$ respectively.

2 Model, IV Estimation and Adaptive Lasso

We have an i.i.d. sample $\{Y_i, \mathbf{X}_i^T, \mathbf{Z}_i^T\}_{i=1}^n$, where Y_i is the outcome of interest for observation i, $\mathbf{X}_i$ is a k_x-vector of exposure variables, $\mathbf{Z}_i$ is a k_z-vector of putative instrumental variables and n is the sample size. As in Guo et al. (2018), Windmeijer et al. (2019) and Windmeijer et al. (2021), we follow Kang et al. (2016) who, starting from the additive linear constant effects model of Holland (1988), arrived at the observed data model for the random sample given by

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \mathbf{Z}_i^T \boldsymbol{\alpha} + U_i, \tag{1}$$

where $\boldsymbol{\beta}$ is the causal parameter vector of interest, and with $\mathbb{E}[U_i|\mathbf{Z}_i] = 0$, but $\mathbf{X}_i$ may be confounded by U_i. The parameter vector $\boldsymbol{\alpha}$ captures the violations of the exclusion restriction. Formally, following the definition of invalid instruments as in Guo et al. (2018, p797), for $j \in 1, ..., k_z$, an instrument Z_j is invalid if $\alpha_j \neq 0$ and valid if $\alpha_j = 0$. Let $\mathcal{V}$ and $\mathcal{A}$ be the sets of indices of the valid and invalid instruments respectively: $\mathcal{V} = \{j : \alpha_j = 0\}$, $\mathcal{A} = \{j : \alpha_j \neq 0\}$, with dimensions $k_\mathcal{V}$ and $k_\mathcal{A}$ respectively, then $k_z = k_\mathcal{V} + k_\mathcal{A}$.

Let $\mathbf{y}$ be the n-vector of n observations on $\{Y_i\}$, and let $\mathbf{X}$ and $\mathbf{Z}$ be the $n \times k_x$ and $n \times k_z$ matrices of the exposure variables and candidate instrumental variables, respectively. Let $\mathbf{Z}_\mathcal{V}$ and $\mathbf{Z}_\mathcal{A}$ denote the $n \times k_\mathcal{V}$ and $n \times k_\mathcal{A}$ matrices of valid and invalid instruments. The oracle model is the model where the set of invalid instruments is known, and is hence given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_\mathcal{A}\boldsymbol{\alpha}_\mathcal{A} + \mathbf{u},$$

where $\mathbf{u}$ is the n-vector with elements $\{U_i\}$. The so-called first-stage regression model of $\mathbf{X}$ on $\mathbf{Z}$ is given by

$$\begin{aligned} \mathbf{X} &= \mathbf{Z}\boldsymbol{\Pi} + \mathbf{E} \\ &= \mathbf{Z}_\mathcal{V}\boldsymbol{\Pi}_\mathcal{V} + \mathbf{Z}_\mathcal{A}\boldsymbol{\Pi}_\mathcal{A} + \mathbf{E}, \end{aligned}$$

where $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_\mathcal{V} \ \boldsymbol{\Pi}_\mathcal{A}] = \left(\mathbb{E}\left[\mathbf{Z}_i\mathbf{Z}_i^T\right]\right)^{-1} \mathbb{E}\left[\mathbf{Z}_i\mathbf{X}_i^T\right]$ and $\mathbb{E}\left[\mathbf{E}_i|\mathbf{Z}_i\right] = 0$.

We assume the instrument relevance condition for the oracle model holds:

Assumption 1. *Relevance:* $r(\boldsymbol{\Pi}_\mathcal{V}) = k_x$.

We further make the standard assumptions as in Windmeijer et al. (2019):

Assumption 2. $\mathbb{E}\left[\mathbf{Z}_i\mathbf{Z}_i^T\right] = \mathbf{Q}_{zz}$, *with* $\mathbf{Q}_{zz}$ *a finite and full rank matrix;* $\mathbb{E}\left[\mathbf{Z}_{i.}\mathbf{X}_{i.}^T\right] = \mathbf{Q}_{zx}$, *with* $\mathbf{Q}_{zx}$ *a finite matrix.*

Assumption 3. $\frac{1}{\sqrt{n}}\boldsymbol{Z}^T\boldsymbol{u} \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_{zu}\right)$ *as* $n \to \infty$, *with* $\boldsymbol{\Sigma}_{zu}$ *a finite and full rank matrix.*

We further make a conditional homoskedasticity assumption,

Assumption 4. *Homoskedasticity:* $\mathbb{E}\left[U_i^2|\mathbf{Z}_i\right] = \sigma_u^2$.

It follows that under under the homoskedasticity assumption, $\boldsymbol{\Sigma}_{zu} = \sigma_u^2\boldsymbol{Q}_{zz}$.

2.1 IV Estimation

Let $\boldsymbol{\theta}^{or} = \left(\boldsymbol{\beta}^T \ \boldsymbol{\alpha}_{\mathcal{A}}^T\right)^T$ and $\mathbf{R} = [\mathbf{X} \ \mathbf{Z}_{\mathcal{A}}]$. A standard two-stage least squares (2sls) IV estimator of $\boldsymbol{\theta}^{or}$ is defined as

$$\widehat{\boldsymbol{\theta}}_{2sls}^{or} = \arg\min_{\boldsymbol{\theta}} \left(\mathbf{y} - \mathbf{R}\boldsymbol{\theta}\right)^T \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{y} - \mathbf{R}\boldsymbol{\theta}\right),$$

resulting in

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{2sls}^{or} &= \left(\mathbf{R}^T\mathbf{P}_Z\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{P}_Z\mathbf{y} \\ &= \left(\widehat{\mathbf{R}}^T\widehat{\mathbf{R}}\right)^{-1}\widehat{\mathbf{R}}^T\mathbf{y},\end{aligned}$$

where $\widehat{\mathbf{R}} = \left[\widehat{\mathbf{X}} \ \mathbf{Z}_{\mathcal{A}}\right]$, with $\widehat{\mathbf{X}} = \mathbf{Z}\widehat{\mathbf{\Pi}}$, and $\widehat{\mathbf{\Pi}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{X}$. Under Assumptions 1-4 the 2sls estimator is asymptotically efficient, and its limiting distribution is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{2sls}^{or} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, \sigma_u^2\left(\boldsymbol{Q}_{zx}^T\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{zx}\right)^{-1}\right). \tag{2}$$

From standard partitioned regression results, we can express the 2sls estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_{\mathcal{A}}$ as

$$\widehat{\boldsymbol{\beta}}_{2sls}^{or} = \left(\widehat{\mathbf{X}}^T\mathbf{M}_{Z_{\mathcal{A}}}\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}^T\mathbf{M}_{Z_{\mathcal{A}}}\mathbf{y}, \tag{3}$$

$$\widehat{\boldsymbol{\alpha}}_{\mathcal{A}}^{or} = \left(\mathbf{Z}_{\mathcal{A}}^T\mathbf{M}_{\widehat{X}}\mathbf{Z}_{\mathcal{A}}\right)^{-1}\mathbf{Z}_{\mathcal{A}}^T\mathbf{M}_{\widehat{X}}\mathbf{y}. \tag{4}$$

When $k_{\mathcal{V}} > k_x$, the test for overidentifying restrictions is a test for $H_0 : \mathbb{E}\left[\mathbf{Z}_iU_i\right] = 0$. The Sargan (1958) test statistic is given by

$$S\left(\widehat{\boldsymbol{\theta}}_{2sls}^{or}\right) = \frac{\left(\mathbf{y} - \mathbf{R}\widehat{\boldsymbol{\theta}}_{2sls}^{or}\right)^T\mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{y} - \mathbf{R}\widehat{\boldsymbol{\theta}}_{2sls}^{or}\right)}{\left(\mathbf{y} - \mathbf{R}\widehat{\boldsymbol{\theta}}_{2sls}^{or}\right)^T\left(\mathbf{y} - \mathbf{R}\widehat{\boldsymbol{\theta}}_{2sls}^{or}\right)/n}, \tag{5}$$

which, under the null, converges to a $\chi^2_{k_{\mathcal{V}}-k_x}$ distributed random variable under Assumptions 1-4.

Let a selection of $k_{\mathcal{A}^*}$ instruments classified as invalid be denoted $\mathbf{Z}_{\mathcal{A}^*}$, with $k_z - k_{\mathcal{A}^*} \geq k_x$. The corresponding model is given by

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{\mathcal{A}^*}\boldsymbol{\alpha}_{\mathcal{A}^*} + \mathbf{u}^* \\ &= \mathbf{R}^*\boldsymbol{\theta}^* + \mathbf{u}^*.\end{aligned}$$

Then, if the set contains all invalid instruments, such that $\mathbf{Z}_{\mathcal{A}} \subseteq \mathbf{Z}_{\mathcal{A}^*}$, it follows that the IV estimator for $\boldsymbol{\beta}$ is consistent and normal and, for $k_z - k_{\mathcal{A}^*} > k_x$, $S\left(\widehat{\boldsymbol{\theta}}^*_{2sls}\right) \xrightarrow{d} \chi^2_{k_z-k_x-k_{\mathcal{A}^*}}$. Alternatively, under the plurality rule that the valid instruments form the largest group, it follows that for all sets with $k_{\mathcal{A}^*} = k_{\mathcal{A}}$, if $\mathbf{Z}_{\mathcal{A}^*} \neq \mathbf{Z}_{\mathcal{A}}$, then the IV estimator for $\boldsymbol{\beta}$ is inconsistent and $S\left(\widehat{\boldsymbol{\theta}}^*_{2sls}\right) = O_p(n)$.

2.2 Adaptive Lasso

Based on the definition of a valid instrument, selection of the valid instruments is equivalent to identifying which entries in $\boldsymbol{\alpha}$ are zero. For this purpose, we consider using the adaptive Lasso to estimate $\boldsymbol{\alpha}$, as the Lasso will shrink some entries in $\boldsymbol{\alpha}$ to exactly zero. Hence, we can obtain estimators for $\mathcal{V}$ and $\mathcal{A}$ from the adaptive Lasso estimator for $\boldsymbol{\alpha}$, which we denote by $\widehat{\boldsymbol{\alpha}}_{ad}$. The estimators for $\mathcal{V}$ and $\mathcal{A}$ are then $\widehat{\mathcal{V}} = \{j : \widehat{\alpha}_{ad,j} = 0\}$ and $\widehat{\mathcal{A}} = \{j : \widehat{\alpha}_{ad,j} \neq 0\}$.

Kang et al. (2016) introduced the Lasso method for IV selection for the single exposure case. Windmeijer et al. (2019) showed that the Lasso irrepresentable condition (see Zhao and Yu, 2006, and Zou, 2006) could be violated, depending on the relative strengths of the invalid and valid instruments, leading to inconsistent selection of the valid/invalid instruments. They adopted the adaptive Lasso estimator of Zou (2006). Let now $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T \ \boldsymbol{\alpha}^T\right)^T$, then the penalised objective function is based on the 2sls criterion and the adaptive Lasso estimator is given by

$$\widehat{\boldsymbol{\theta}}_{ad} = \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\arg\min} \frac{1}{2} \left\| \{\mathbf{P}_Z(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})\} \right\|_2^2 + \lambda_n \sum_{j=1}^{k_z} \frac{|\alpha_j|}{|\widehat{\alpha}_j|^{\nu}}, \tag{6}$$

where $\widehat{\boldsymbol{\alpha}}$, with j-th element equal to $\widehat{\alpha}_j$, is an initial estimator of $\boldsymbol{\alpha}$, and $\nu > 0$. As $\boldsymbol{\beta}$ is not penalized, the adaptive Lasso estimator for $\boldsymbol{\alpha}$ can be obtained as

$$\widehat{\boldsymbol{\alpha}}_{ad} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \left\| \mathbf{y} - \widetilde{\mathbf{Z}}\boldsymbol{\alpha} \right\|_2^2 + \lambda_n \sum_{j=1}^{k_z} \frac{|\alpha_j|}{|\widehat{\alpha}_j|^{\nu}}, \tag{7}$$

where $\widetilde{\mathbf{Z}} = \mathbf{M}_{\widehat{X}}\mathbf{Z}$, see Kang et al. (2016) and Windmeijer et al. (2019).

λ_n is the tuning parameter controlling the strength of the penalization. A larger λ_n leads to more entries in $\boldsymbol{\alpha}$ being shrunk to zero, which implies that the adaptive Lasso selects more instruments as valid. From Theorem 2 and Remark 1 in Zou (2006) the adaptive Lasso estimator for $\boldsymbol{\alpha}$, as defined in (7) has oracle properties and hence selects the valid instruments consistently under the following assumptions:

Assumption 5. $\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}$ *and* $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) = O_p\left(1\right)$.

Assumption 6. $\lambda_n = o(\sqrt{n})$, $n^{\frac{\nu-1}{2}}\lambda_n \to \infty$.

The intuition for $\widehat{\alpha}_j$ is clear. As a consistent estimator for α_j, $\widehat{\alpha}_j$ will be close to zero when $\alpha_j = 0$. Since $\widehat{\alpha}_j$ enters in the denominator in (7), a value close to zero will produce a large penalty weight, and make it more likely that $\widehat{\alpha}_{ad,j}$ is equal to zero. The oracle properties then apply to $\widehat{\boldsymbol{\theta}}_{ad}$ and hence the estimator of interest $\widehat{\boldsymbol{\beta}}_{ad}$.

Clearly, a crucial component for the application of the adaptive Lasso estimator is the initial consistent estimator of $\boldsymbol{\alpha}$. We propose an initial consistent estimator of $\boldsymbol{\beta}$, the median-of-medians estimator as described in the next section, from which the required estimator for $\boldsymbol{\alpha}$ can be derived.

3 The Median-of-Medians Estimator

For the single exposure case, Windmeijer et al. (2019), following Han (2008), showed that the median of the instrument specific, just-identified estimators for β is a consistent estimator of β when more than 50% of the instruments are valid, i.e. $k_{\mathcal{V}} > \frac{k_z}{2}$. These just-identified estimators are the IV, or 2sls estimators in the model specifications

$$\mathbf{y} = \mathbf{x}\beta_j + \mathbf{Z}_{\{-j\}}\boldsymbol{\alpha}_{\{-j\}} + \mathbf{u}_j,$$

for $j = 1, \ldots, k_z$, and where $\mathbf{Z}_{\{-j\}} = \mathbf{Z} \setminus \{\mathbf{Z}_j\}$ is the full set of instruments with the j-th instrument omitted, which is used as the excluded instrument for $\mathbf{x}$, and $\mathbf{u}_j = \mathbf{z}_j\alpha_j + \mathbf{u}$,

where $\mathbf{z}_j$ is the j-th instrument vector, see also Windmeijer et al. (2021). Let $\widehat{\beta}_j$ denote IV estimator for β, treating the j-th instrument as the valid instrument and all other instruments as invalid. Provided all instruments are relevant, $\pi_j \neq 0$ for $j = 1, \ldots, k_z$ in $\mathbf{x} = \mathbf{Z}\boldsymbol{\pi} + \mathbf{e}$, it follows that $\mathbf{z}_j \alpha_j = \left(\mathbf{x} - \mathbf{Z}_{\{-j\}}\boldsymbol{\pi}_{\{-j\}} - \mathbf{e}\right)\frac{\alpha_j}{\pi_j}$. Then for the valid instruments, $j \in \mathcal{V}$, $\widehat{\beta}_j$ is a consistent and normal estimator of β, whereas for the invalid instruments, $j \in \mathcal{A}$, $\widehat{\beta}_j$ is a consistent and normal estimator of $\beta + \frac{\alpha_j}{\pi_j}$, and hence an inconsistent estimator of β. It then follows that the median estimator, given by

$$\widehat{\beta}_m = \left\{\widehat{\beta}_j\right\}_{j=1}^{k_z} \tag{8}$$

is a consistent estimator of β if $k_{\mathcal{V}} > \frac{k_z}{2}$ and Windmeijer et al. (2019) show that then $\sqrt{n}\left(\widehat{\beta}_m - \beta\right) = O_p(1)$. A consistent estimator for $\boldsymbol{\alpha}$ is then given by

$$\widehat{\boldsymbol{\alpha}}_m = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{y} - \mathbf{x}\widehat{\beta}_m\right),$$

with also $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_m - \boldsymbol{\alpha}\right) = O_p(1)$, satisfying the conditions for oracle properties of the adaptive Lasso estimator, leading to consistent selection of the valid and invalid instruments and oracle properties of $\widehat{\beta}_{ad}$.

We can extend the median estimator to the case where there are multiple exposure variables, $k_x \geq 2$. We initially assume that all $p = \binom{k_z}{k_x}$ just-identifying sets of instruments are jointly relevant for all exposure variables. Denote the just-identifying sets of instruments by $\mathbf{Z}_s$, for $s = 1, \ldots, p$. The just-identified model specifications are then given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_s + \mathbf{Z}_{\{-s\}}\boldsymbol{\alpha}_{\{-s\}} + \mathbf{u}_s \tag{9}$$

$$\mathbf{X} = \mathbf{Z}_s\boldsymbol{\Pi}_s + \mathbf{Z}_{\{-s\}}\boldsymbol{\Pi}_{\{-s\}} + \mathbf{E}, \tag{10}$$

where $\mathbf{Z}_{\{-s\}} = \mathbf{Z} \setminus \{\mathbf{Z}_s\}$ and $\mathbf{u}_s = \mathbf{Z}_s\boldsymbol{\alpha}_s + \mathbf{u}$. These relevance conditions can then be stated as follows,

Assumption 7. *Relevance of just-identifying sets. Let the $p = \binom{k_z}{k_x}$ sets of just-identifying instruments be denoted $\{\mathbf{Z}_s\}_{s=1}^p$ and let $\boldsymbol{\Pi}_s$ be defined as in (10). Then all these sets are jointly relevant for all exposure variables, $r(\boldsymbol{\Pi}_s) = k_x$, for $s = 1, \ldots, p$.*

It then follows that $\mathbf{Z}_s\boldsymbol{\alpha}_s = \left(\mathbf{X} - \mathbf{Z}_{\{-s\}}\boldsymbol{\Pi}_{\{-s\}} - \mathbf{E}\right)\boldsymbol{\Pi}_s^{-1}\boldsymbol{\alpha}_s$ and so the estimands for the just-identified IV estimators are given by $\boldsymbol{\beta}_s = \boldsymbol{\beta} + \boldsymbol{\Pi}_s^{-1}\boldsymbol{\alpha}_s$, resulting in consistent

and normal IV estimators for the sets that contain valid instruments only, with $\boldsymbol{\alpha}_s = \mathbf{0}$, and inconsistent estimators of $\boldsymbol{\beta}$ for all other sets. Let $\widehat{\boldsymbol{\beta}}_m$ denote the medians of the p just-identified estimators, then it follows directly from the results in Windmeijer et al. (2019) that $\widehat{\boldsymbol{\beta}}_m$ is a consistent estimator of $\boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}\right) = O_p(1)$ if more than half of all just-identified sets are sets of valid instruments only, or $\binom{k_\mathcal{V}}{k_x} > \frac{p}{2}$, see also Apfel (2019).

3.1 $k_x = 2$

In comparison to this naive median estimator, we can allow for a weaker condition on the number of valid instruments required to obtain an initial consistent estimator based on the just-identified estimators. To this end we propose the median-of-medians estimator. We first consider the $k_x = 2$ case. Consider for each instrument $j = 1, \ldots, k_z$, all just-identifying sets of instruments that contain instrument j. There are $k_z - 1$ such sets for each j. Denote the just-identified estimators based on the sets that contain j by $\widehat{\boldsymbol{\beta}}^j_\ell$, $\ell = 1, \ldots, k_z, \ell \neq j$. If instrument j is invalid, $j \in \mathcal{A}$, then none of the $\widehat{\boldsymbol{\beta}}^j_\ell$ are consistent estimators of $\boldsymbol{\beta}$. If instrument j is valid, $j \in \mathcal{V}$, then the $k_\mathcal{V} - 1$ sets of instruments containing j and another valid instrument result in consistent and normal IV estimators of $\boldsymbol{\beta}$, whereas the remaining $k_z - k_\mathcal{V}$ sets contain invalid instruments resulting in inconsistent IV estimators. Let

$$\widehat{\boldsymbol{\beta}}^j_m = \text{median}\left\{\widehat{\boldsymbol{\beta}}^j_\ell\right\}^{k_z}_{\ell=1,\ell\neq j}, \tag{11}$$

where the medians are taking element wise, so $\widehat{\beta}^j_{m,q} = \text{median}\left\{\widehat{\beta}^j_{\ell,q}\right\}^{k_z}_{\ell=1,\ell\neq j}$ for $q = 1, 2$.

For $\widehat{\boldsymbol{\beta}}^j_m$ to be a consistent estimator of $\boldsymbol{\beta}$ for a valid instrument $j \in \mathcal{V}$, we need the following further assumption,

Assumption 8. *Condition on number of valid instruments. For $k_x = 2$, the number of valid instruments $k_\mathcal{V}$ satisfies $(k_\mathcal{V} - 1) > \frac{k_z - 1}{2}$, or equivalently $k_\mathcal{V} > \frac{k_z + 1}{2}$.*

Under Assumption 8, it follows that for a valid instrument $j \in \mathcal{V}$ the majority rule is satisfied. This implies that more than half of the just-identifying sets of instruments containing j are sets of valid instruments only, and hence the result follows straightforwardly that $\widehat{\boldsymbol{\beta}}^j_m$ is a consistent estimator of $\boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^j_m - \boldsymbol{\beta}\right) = O_p(1)$, from the results and proof of Theorem 1 in Windmeijer et al. (2019). The result is stated in the following proposition, with the proof given in the Appendix.

Proposition 1. *For $k_x = 2$, let for each instrument $j = 1, \ldots, k_z$, the median estimator $\widehat{\boldsymbol{\beta}}_m^j$ be defined as in (11). Then, under Assumptions 2, 3, 7 and 8, for valid instruments $j \in \mathcal{V}$, $\widehat{\boldsymbol{\beta}}_m^j$ is a consistent estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_m^j \xrightarrow{p} \boldsymbol{\beta}$, and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_m^j - \boldsymbol{\beta}\right) = O_p(1)$.*

Under the conditions of Proposition 1 it follows that $k_{\mathcal{V}}$ out of k_z estimators $\left\{\widehat{\boldsymbol{\beta}}_m^j\right\}_{j=1}^{k_z}$ are consistent estimators of $\boldsymbol{\beta}$. It then follows that the median-of-medians estimator, defined as

$$\widehat{\boldsymbol{\beta}}_{mm} = \text{median}\left\{\widehat{\boldsymbol{\beta}}_m^j\right\}_{j=1}^{k_z} \tag{12}$$

where the medians are taken element wise, so $\widehat{\beta}_{mm,q} = \text{median}\left\{\widehat{\beta}_{m,q}^j\right\}_{j=1}^{k_z}$ for $q = 1, 2$, is a consistent estimator of $\boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$ if $k_{\mathcal{V}} > \frac{k_z}{2}$, but this condition is implied by Assumption 8. We formally state this result in the following proposition, with the proof presented in the Appendix.

Proposition 2. *For $k_x = 2$, let the median-of-medians estimator $\widehat{\boldsymbol{\beta}}_{mm}$ be defined as in (12), then under the conditions of Proposition 1 it follows that $\widehat{\boldsymbol{\beta}}_{mm} \xrightarrow{p} \boldsymbol{\beta}$, and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$.*

As an illustration of the median-of-medians estimator, we consider the case with $k_x = 2$, $k_z = 7$ and $k_{\mathcal{V}} = 5 > \frac{k_z+1}{2} = 4$, and so Assumption 8 is satisfied. Let instruments 1 and 2 be the invalid ones, so $\mathcal{A} = \{1, 2\}$ and $\mathcal{V} = \{3, 4, 5, 6, 7\}$. Table 1 lists the just-identified estimators for β_q, $q = 1, 2$, and they are estimated using each IV pair. The valid instruments and consistent estimators are indicated in boldface.

Table 1: Illustration of the median-of-medians estimator of β_q, $q = 1, 2$.

Instruments	1	2	**3**	**4**	**5**	**6**	**7**	
1		$\widehat{\beta}_{1,q}^2$	$\widehat{\beta}_{1,q}^3$	$\widehat{\beta}_{1,q}^4$	$\widehat{\beta}_{1,q}^5$	$\widehat{\beta}_{1,q}^6$	$\widehat{\beta}_{1,q}^7$	
2	$\widehat{\beta}_{2,q}^1$		$\widehat{\beta}_{2,q}^3$	$\widehat{\beta}_{2,q}^4$	$\widehat{\beta}_{2,q}^5$	$\widehat{\beta}_{2,q}^6$	$\widehat{\beta}_{2,q}^7$	
3	$\widehat{\beta}_{3,q}^1$	$\widehat{\beta}_{3,q}^2$		$\widehat{\boldsymbol{\beta}}_{\boldsymbol{3,q}}^{\boldsymbol{4}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{3,q}}^{\boldsymbol{5}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{3,q}}^{\boldsymbol{6}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{3,q}}^{\boldsymbol{7}}$	
4	$\widehat{\beta}_{4,q}^1$	$\widehat{\beta}_{4,q}^2$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{4,q}}^{\boldsymbol{3}}$		$\widehat{\boldsymbol{\beta}}_{\boldsymbol{4,q}}^{\boldsymbol{5}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{4,q}}^{\boldsymbol{6}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{4,q}}^{\boldsymbol{7}}$	
5	$\widehat{\beta}_{5,q}^1$	$\widehat{\beta}_{5,q}^2$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{5,q}}^{\boldsymbol{3}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{5,q}}^{\boldsymbol{4}}$		$\widehat{\boldsymbol{\beta}}_{\boldsymbol{5,q}}^{\boldsymbol{6}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{5,q}}^{\boldsymbol{7}}$	
6	$\widehat{\beta}_{6,q}^1$	$\widehat{\beta}_{6,q}^2$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{6,q}}^{\boldsymbol{3}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{6,q}}^{\boldsymbol{4}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{6,q}}^{\boldsymbol{5}}$		$\widehat{\boldsymbol{\beta}}_{\boldsymbol{6,q}}^{\boldsymbol{7}}$	
7	$\widehat{\beta}_{7,q}^1$	$\widehat{\beta}_{7,q}^2$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{7,q}}^{\boldsymbol{3}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{7,q}}^{\boldsymbol{4}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{7,q}}^{\boldsymbol{5}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{7,q}}^{\boldsymbol{6}}$		
median	$\widehat{\beta}_{m,q}^1$	$\widehat{\beta}_{m,q}^2$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m,q}}^{\boldsymbol{3}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m,q}}^{\boldsymbol{4}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m,q}}^{\boldsymbol{5}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m,q}}^{\boldsymbol{6}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m,q}}^{\boldsymbol{7}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{mm,q}}$

Notes: $k_x = 2$, $k_z = 7$, $\mathcal{V} = \{3, 4, 5, 6, 7\}$, $\mathcal{A} = \{1, 2\}$. Valid instruments and consistent estimators are displayed in boldface.

For the general case, the just-identified estimator $\widehat{\boldsymbol{\beta}}_{\ell}^{j}$ is a consistent estimator of $\boldsymbol{\beta}$ if and only if both instruments j and ℓ are valid. Hence, all the estimators of β_q, $q = 1, 2$, in the columns for instruments 1 and 2 are inconsistent as at least one of the invalid instruments is involved in the estimation and the resulting median estimators $\widehat{\beta}_{m,q}^{1}$ and $\widehat{\beta}_{m,q}^{2}$ are inconsistent. For instruments 3-6 more than half of the $k_z - 1$ estimators in each column are consistent as here we have $k_{\mathcal{V}} - 1 > \frac{k_z - 1}{2}$. Hence, the median estimators $\widehat{\beta}_{m,q}^{3}$ to $\widehat{\beta}_{m,q}^{6}$ are all consistent. Now, we take the median of all these column median estimators (as shown in the last row of Table 1), i.e. $\widehat{\beta}_{mm,q} = \text{median}\left(\widehat{\beta}_{m,q}^{1}, ..., \widehat{\beta}_{m,q}^{6}\right)$. The assumption $k_{\mathcal{V}} > \frac{k_z+1}{2}$ implies $k_{\mathcal{V}} > \frac{k_z}{2}$. Thus, more than half of the column median estimators $\widehat{\beta}_{m,q}^{1}, ..., \widehat{\beta}_{m,q}^{6}$ are consistent and therefore the median of these median estimators $\widehat{\beta}_{mm,q}$ is also consistent. Therefore, for $k_x = 2$, under the assumption $k_{\mathcal{V}} > \frac{k_z+1}{2}$ the median-of-medians estimator is consistent even if we have no knowledge about which of the instruments are valid.

For comparison, for the naive median estimator to be consistent, the condition $\binom{k_{\mathcal{V}}}{k_x} > \frac{1}{2}\binom{k_z}{k_x}$ implies here that $k_{\mathcal{V}} > 4$, so only one instrument is allowed to be invalid. Increasing k_z to $k_z = 100$, the condition for the median-of-medians estimator is that $k_{\mathcal{V}} > 50.5$, whereas for the naive median estimator this is $k_{\mathcal{V}} > 70$. For $k_x = 2$, Assumption 8 for the median-of-medians estimator is only stronger for the minimum number of valid instruments required than the simple majority rule for the single exposure model when k_z is odd with the difference then equal to 1.

3.2 $k_x > 2$

We can extend the results for the median-of-median estimator for the $k_x = 2$ case to the $k_x > 2$ case, where the estimator becomes a median-of-medians-of-medians.... estimator, but we simply refer to it as the median-of-medians estimator for brevity.

Let the instrument set be denoted $S = \{1, \ldots, k_z\}$. For $k_x > 2$, consider the $l = \binom{k_z}{k_x - 2}$ sets of $k_x - 2$ instruments. For each of these sets $L = 1, \ldots, l$, and for each instrument $j \in S \setminus L$, let $\widehat{\boldsymbol{\beta}}_{j,k}^{L}$ denote the just-identified IV estimator of $\boldsymbol{\beta}$, using the k_x instruments $\{L, j, k\}$, $k \in S \setminus \{L, j\}$. For brevity, denote $S_L := S \setminus L$ and $S_{L,j} := S \setminus \{L, j\}$. Given L, we have thus for each j, $k_z - k_x + 1$ just identified estimators. If the set L is a set of $k_x - 2$ valid instruments and j is a valid instrument, then the majority of $\left\{\widehat{\boldsymbol{\beta}}_{j,k}^{L}\right\}_{k \in S_{L,j}}$ are consistent and normal if there are additionally more than $\frac{k_z - k_x + 1}{2}$ valid instruments.

Therefore, if $k_{\mathcal{V}} > \frac{k_z+k_x-1}{2}$, L and j are valid instruments, then

$$\widehat{\boldsymbol{\beta}}^{L}_{j,m} = \text{median}\left\{\widehat{\boldsymbol{\beta}}^{L}_{j,k}\right\}_{k\in S_{L,j}}$$

is a consistent estimator of $\boldsymbol{\beta}$.

Given L, we have a total of $k_z - k_x + 2$ estimators $\widehat{\boldsymbol{\beta}}^{L}_{j,m}$. For $k_{\mathcal{V}} > \frac{k_z+k_x-1}{2}$ it follows that $k_{\mathcal{V}_{S_L}} > \frac{k_z-k_x+3}{2} > \frac{k_z-k_x+2}{2}$ for a set L that contains valid instruments only, where $k_{\mathcal{V}_{S_L}}$ denote the number of valid instruments in S_L. This then implies that more than half of the $\left\{\widehat{\boldsymbol{\beta}}^{L}_{j,m}\right\}_{j\in S_L}$ are consistent, and hence the median of the medians

$$\widehat{\boldsymbol{\beta}}^{L}_{mm} = \text{median}\left\{\widehat{\boldsymbol{\beta}}^{L}_{j,m}\right\}_{j\in S_L}$$

is again consistent for all L with valid instruments only. From the proofs of Propositions 1 and 2 it follows that $\widehat{\boldsymbol{\beta}}^{L}_{mm} \xrightarrow{p} \boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{L}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$.

If $k_x = 3$, then $|L| = 1$, and we thus have k_z estimators $\left\{\widehat{\boldsymbol{\beta}}^{\ell}_{mm}\right\}_{\ell=1}^{k_z}$. For $k_{\mathcal{V}} > \frac{k_z+k_x-1}{2}$ it follows that $k_{\mathcal{V}} > \frac{k_z}{2}$ and so more than half of the instruments are valid and so more than half of the $\left\{\widehat{\boldsymbol{\beta}}^{\ell}_{mm}\right\}_{\ell=1}^{k_z}$ are consistent. Hence it follows that the median of the medians of the medians,

$$\widehat{\boldsymbol{\beta}}_{mm} = \text{median}\left\{\widehat{\boldsymbol{\beta}}^{\ell}_{mm}\right\}_{\ell=1}^{k_z} \tag{13}$$

is a consistent estimator of $\boldsymbol{\beta}$, and again, from the proof of Proposition 2 it follows that $\widehat{\boldsymbol{\beta}}_{mm} \xrightarrow{p} \boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$.

For $k_x > 3$, consider the $l_{-1} = \binom{k_z}{k_x-3}$ sets of $k_x - 3$ instruments, and denote these sets L_{-1}. For each set L_{-1} consider the sets $L_{-1,j} = \{L_{-1}, j\}$ for $j \in S_{L_{-1}}$, where $S_{L_{-1}} = S \setminus L_{-1}$. Then, if L_{-1} is a set containing valid instruments only, it follows that $\widehat{\boldsymbol{\beta}}^{L_{-1,j}}_{mm}$ is consistent iff j is a valid instrument. Given L_{-1}, we have a total of $k_z - k_x + 3$ estimators $\widehat{\boldsymbol{\beta}}^{L_{-1,j}}_{mm}$. For $k_{\mathcal{V}} > \frac{k_z+k_x-1}{2}$ it follows that $k_{\mathcal{V}_{S_{L_{-1}}}} > \frac{k_z-k_x+5}{2} > \frac{k_z-k_x+3}{2}$ for a set L_{-1} that contains valid instruments only, where $k_{\mathcal{V}_{S_{L_{-1}}}}$ denotes the number of valid instruments in $S_{L_{-1}}$. It then follows that

$$\widehat{\boldsymbol{\beta}}^{L_{-1}}_{mm_{-1}} = \text{median}\left\{\widehat{\boldsymbol{\beta}}^{\{L_{-1},j\}}_{mm}\right\}_{j\in S_{L_{-1}}}$$

is consistent for all L_{-1} that contain valid instruments only. Cascading, repeat this

exercise for $L_{-2}, \ldots, L_{-(k_x-3)}$. As $\left|L_{-(k_x-3)}\right| = 1$, we get the final result that

$$\widehat{\boldsymbol{\beta}}_{mm} = \text{median}\left\{\widehat{\boldsymbol{\beta}}^{\ell}_{mm_{-(k_x-3)}}\right\}_{\ell=1}^{k_z} \tag{14}$$

is a consistent estimator of $\boldsymbol{\beta}$. From the proof of Proposition 2 it follows that $\widehat{\boldsymbol{\beta}}_{mm} \xrightarrow{p} \boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$.

We can now summarise the results obtained for general k_x in the following Proposition.

Proposition 3. *Under Assumptions 2, 3, 7 and the generalized majority rule*

$$k_{\mathcal{V}} > \frac{k_z + k_x - 1}{2},$$

consider the median estimator $\widehat{\beta}_m$ as defined in (8) for $k_x = 1$, the median-of-medians estimator as defined in (12) for $k_x = 2$ and the generalized median-of-medians estimators for $k_x = 3$ and $k_x > 3$ as defined in (13) and (14). The latter three denoted generically by $\widehat{\boldsymbol{\beta}}_{mm}$. Then $\widehat{\beta}_m \xrightarrow{p} \beta$, $\sqrt{n}\left(\widehat{\beta}_m - \beta\right) = O_p(1)$, $\widehat{\boldsymbol{\beta}}_{mm} \xrightarrow{p} \boldsymbol{\beta}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$.

4 Consistent Selection and Oracle Estimator

Given the consistent estimator $\widehat{\boldsymbol{\beta}}_{mm}$ as defined in (12), (13) or (14) we obtain a consistent estimator for $\boldsymbol{\alpha}$ as

$$\widehat{\boldsymbol{\alpha}}_{mm} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{mm}\right). \tag{15}$$

From the properties of $\widehat{\boldsymbol{\beta}}_{mm}$ as given in Proposition 3, $\widehat{\boldsymbol{\alpha}}_{mm}$ satisfies the conditions of Assumption 5, $\widehat{\boldsymbol{\alpha}}_{mm} \xrightarrow{p} \boldsymbol{\alpha}$ and $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_{mm} - \boldsymbol{\alpha}\right) = O_p(1)$. Therefore, from Theorem 2 and Remark 1 in Zou (2006), it follows that the adaptive Lasso estimator $\widehat{\boldsymbol{\alpha}}_{ad}$ that uses $\widehat{\boldsymbol{\alpha}}_{mm}$ as the initial consistent estimator satisfies consistency of selection and oracle properties as stated in the following proposition.

Proposition 4. *Under the conditions of Proposition 3 and Assumption 6 for λ_n, let $\widehat{\boldsymbol{\alpha}}_{mm}$ as defined in (15) be the initial consistent estimator in the adaptive Lasso criterion (7). Let $\widehat{\mathcal{A}}_{ad} = \{j : \widehat{\alpha}_{ad,j} \neq 0\}$, then the adaptive Lasso estimator $\widehat{\boldsymbol{\alpha}}_{ad}$ satisfies $\lim_{n\to\infty} P\left(\widehat{\mathcal{A}}_{ad} = \mathcal{A}\right) = 1$ and the limiting normal distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_{ad,\mathcal{A}} - \boldsymbol{\alpha}_{\mathcal{A}}\right)$ is that of the oracle 2sls estimator $\widehat{\boldsymbol{\alpha}}^{or}_{2sls}$ as defined in (4) with the limiting distribution as given in 2.*

Similar to Kang et al. (2016) and Windmeijer et al. (2019), the adaptive Lasso esti-

mator for $\boldsymbol{\beta}$ is obtained as

$$\widehat{\boldsymbol{\beta}}_{ad} = \left(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}^T\left(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\alpha}}_{ad}\right). \tag{16}$$

From the results of Proposition 4 it follows that the limiting distribution of $\widehat{\boldsymbol{\beta}}_{ad}$ is that of the oracle 2sls estimator, as stated in the next corollary.

Corollary 1. *Let $\widehat{\boldsymbol{\beta}}_{ad}$ as defined in (16). Under the conditions of Proposition 4 the limiting normal distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{ad} - \boldsymbol{\beta}\right)$ is that of the oracle 2sls estimator $\widehat{\boldsymbol{\beta}}^{or}_{2sls}$ as defined in (3), with the limiting distribution as given in (2).*

As an alternative to obtaining the causal estimator directly from the adaptive Lasso as in (16), we can also estimate $\boldsymbol{\beta}$ by post-selection 2sls using the estimated set of invalid instruments $\widehat{\mathcal{A}}_{ad}$ in the following specification:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{\widehat{\mathcal{A}}_{ad}}\boldsymbol{\alpha}_{\widehat{\mathcal{A}}_{ad}} + \mathbf{u}, \tag{17}$$

using $\mathbf{Z}_{\widehat{\mathcal{V}}_{ad}}$ as the set of valid instruments, where $\widehat{\mathcal{V}}_{ad} = \{j : \widehat{\alpha}_{ad,j} = 0\} = \mathcal{V} \setminus \widehat{\mathcal{A}}_{ad}$. The next proposition states the oracle properties of the post-selection 2sls estimator in model specification (17). The proof follows directly from Theorem 2 in Guo et al. (2018) as, under the stated conditions, $\lim_{n\to\infty} P(\widehat{\mathcal{V}}_{ad} = \mathcal{V}) = 1$.

Proposition 5. *Let $\widehat{\boldsymbol{\beta}}_{2sls,p}$ be the post-selection 2sls estimator of $\boldsymbol{\beta}$ in model (17), which is given by*

$$\widehat{\boldsymbol{\beta}}_{2sls,p} = \left(\widehat{\mathbf{X}}'\mathbf{M}_{\mathbf{Z}_{\widehat{\mathcal{A}}_{ad}}}\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{M}_{\mathbf{Z}_{\widehat{\mathcal{A}}_{ad}}}\mathbf{y}.$$

Under the conditions of Corollary 1, it follows that the limiting normal distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{2sls,p} - \boldsymbol{\beta}\right)$ is that of the of the oracle 2sls estimator $\widehat{\boldsymbol{\beta}}^{or}_{2sls}$ as defined in (3), with the limiting distribution as given in (2).

4.1 Downward Testing Procedure

Consistent IV selection using the adaptive Lasso depends on the choice of the tuning parameter λ_n which controls the strength of penalization. While λ_n needs to satisfy the theoretical conditions of Assumption 6, $n^{\frac{1-\nu}{2}}\lambda_n \to \infty$, $\lambda_n = o(\sqrt{n})$, it can be challenging to pick a specific value of λ_n for a given sample. A common practice of choosing the tuning parameter is k-fold cross-validation. However, it is well known that cross-validation works better for prediction rather than model selection (Bühlmann and Van De Geer, 2011), and

cross-validation almost always results in inconsistent variable selection, as stated in Chand (2012).

As an alternative, and similar to Windmeijer et al. (2019) and Windmeijer et al. (2021), we combine the adaptive Lasso with the downward testing procedure for moment selection as proposed by Andrews (1999), which uses the Sargan test statistic as the selection criterion, as defined in (5). A crude downward testing procedure starts with the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, treating all k_z instruments as valid. If the Sargan test rejects the model, then the procedure moves to models with $k_z - 1$ treated as valid instruments and tests all such models $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_j\alpha_j + \mathbf{u}_j$, $j = 1, \ldots, k_z$. If the Sargan test rejects them all, then it moves to evaluate all $\binom{k_z}{2}$ models with $k_z - 2$ instruments treated as valid, and so on, until it finds a model that is not rejected by the Sargan test. This procedure can become computationally infeasible since for each number of instruments, $k_z, k_z - 1, ...$, we need to exhaustively test models corresponding to all possible combinations of instruments.

The adaptive Lasso can mitigate the computational challenges in the downward testing procedure. When the adaptive Lasso is implemented using the Least-Angle Regression (LARS) algorithm (Efron et al., 2004), it generates a selection path starting with a model with k_z valid instruments, and, for each LARS step, the number of instruments treated as valid decreases by one. This means that, for each number of instruments treated as valid, $k_z, k_z - 1, ...$, we only need to evaluate a single model, i.e. the one on the LARS selection path. Given the consistency of selection and oracle results of the adaptive Lasso estimator as given in Proposition 4, the oracle model lies on this path in large samples. Given the properties of the Sargan/Hansen test as described in Section 2.1 and the adjusted majority rule requirement as given in Assumption 8, it follows that selecting the first model on this LARS path that does not reject the Sargan test is a consistent selection rule, when for a model with k_{inv} instruments selected as invalid, the critical value $\zeta_{n,k_z-k_x-k_{inv}}$ used for the $\chi^2_{k_z-k_x-k_{inv}}$ distribution satisfies

$$\zeta_{n,k_z-k_x-k_{inv}} \to \infty \text{ for } n \to \infty, \text{ and } \zeta_{n,k_z-k_x-k_{inv}} = o(n), \tag{18}$$

see Andrews (1999). In practice, following Windmeijer et al. (2019) and Windmeijer et al. (2021), instead of a critical value $\zeta_{n,k_z-k_x-k_{inv}}$ for the Sargan test, we use a p-value p_n. If p_n satisfies $\lim_{n\to\infty} p_n = 0$ and $\log(p_n) = o(n)$, then condition (18) is satisfied. As in Windmeijer et al. (2019) and Windmeijer et al. (2021), for a given sample, we set $p_n = 0.1/\log(n)$, as suggested by Belloni et al. (2012). This procedure leads to

consistent selection and oracle properties of the post-selection 2sls estimator as detailed in Proposition 5.

5 Instrument Relevance

In the previous sections, we maintained Assumption 7, requiring that all just identified models identify all parameters in $\boldsymbol{\beta}$. However, in practical applications, it may well be the case that a given instrument is not relevant for all endogenous exposure variables. In this case, some just-identifying combination of instruments would violate the full rank assumption. In practice, one could test for underidentification, as described in e.g. Windmeijer (2021), and discard just identified estimates where the test for underidentification fails to reject, similar to the first-stage hard-thresholding method of Guo et al. (2018). However, it may be difficult then to establish an adjusted majority rule as in Assumption 8 and is the subject of future research.

Instead, we consider here the case, as in our application, where the the relevance of the instruments with respect to each endogenous exposure variable is known. In our application of Mendelian randomisation, the potential instruments are genetic markers, which are identified from GWAS studies, and hence it is known from these studies which genetic marker is relevant for which exposure variable. In Mendelian randomisation studies, the genetic markers are also independently distributed. We show here how to obtain the consistent median-of-medians estimator that incorporates this information. For ease of exposition and in line with our application, we focus here on the $k_x = 2$ case.

We first consider the case where each instrument is relevant only for one of the exposure variables. For the $k_x = 2$ case, the first-stage linear specification can then be written as

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2] = [\mathbf{Z}_1 \ \mathbf{Z}_2] \begin{bmatrix} \boldsymbol{\pi}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\pi}_2 \end{bmatrix} + \mathbf{E},$$

where $\mathbf{Z}_1$ is the $n \times k_1$ matrix of instruments relevant for $\mathbf{x}_1$ and $\mathbf{Z}_2$ is the $n \times k_2$ matrix of instruments relevant for $\mathbf{x}_2$. When instruments are independent, as often the case in Mendelian randomisation studies, any just-identifying pair of them can identify the parameter vector $\boldsymbol{\beta} = (\beta_1 \ \beta_2)^T$only if it combines one instrument from $\mathbf{Z}_1$ with one instrument from $\mathbf{Z}_2$. Hence there are now $k_1 \times k_2$ sets of just-identifying instruments that are relevant for both exposure variables. Let $k_{\mathcal{V}_1}$ and $k_{\mathcal{V}_2}$ denote the number of valid instruments in $\mathbf{Z}_1$ and $\mathbf{Z}_2$ respectively. Then there are $k_{\mathcal{V}_1} \times k_{\mathcal{V}_2}$ pairs of instruments where

the instruments are both valid and hence for the naive median estimator to be consistent, the condition that $k_{\mathcal{V}_1} \times k_{\mathcal{V}_2} > k_1 \times k_2/2$ needs to hold.

For the median-of medians estimator to be consistent, we now require that $k_{\mathcal{V}_1} > \frac{k_1}{2}$ and $k_{\mathcal{V}_2} > \frac{k_2}{2}$, or the standard majority rule holds for each set. This can be shown as follows. Let the indices of the instruments relevant for $\mathbf{x}_1$ be $S_1 = \{1, 2, \ldots, k_1\}$, and those for $\mathbf{x}_2$ be $S_2 = \{k_1 + 1, k_1 + 2, \ldots, k_z\}$, where $k_z = k_1 + k_2$. We then have the just identified IV estimators $\widehat{\boldsymbol{\beta}}_s^j$, with, when $j \in S_1$, $s \in S_2$ and vice versa. For each element β_q in $\boldsymbol{\beta}$, $q = 1, 2$, we have for each instrument j a vector of estimators $\widehat{\boldsymbol{\beta}}_q^j = \left(\widehat{\beta}_{s,q}^j\right)$. This is a k_2-vector if $j \in S_1$ and a k_1-vector is $j \in S_2$. Let $\widehat{\beta}_{m,q}^j = \text{median}\left(\widehat{\boldsymbol{\beta}}_q^j\right)$. Then $\widehat{\beta}_{m,q}^j$ is consistent if j is a valid instrument and if $j \in S_1$, $k_{\mathcal{V}_2} > \frac{k_2}{2}$, or if $j \in S_2$, $k_{\mathcal{V}_1} > \frac{k_1}{2}$. There are then $k_{\mathcal{V}_1} + k_{\mathcal{V}_2} > \frac{k_z}{2}$ consistent estimators in $\left\{\widehat{\beta}_{m,q}^j\right\}_{j=1}^{k_z}$ and hence $\widehat{\beta}_{mm,q} = \text{median}\left\{\widehat{\beta}_{m,q}^j\right\}_{j=1}^{k_z}$ is a consistent estimator of β_q, for $q = 1, 2$.

This result is illustrated in Table 2 with an example where $S_1 = \{1, 2, 3, 4\}$, and $S_2 = \{5, 6, 7\}$. Valid instruments are $\mathcal{V}_1 = \{2, 3, 4\}$, and $\mathcal{V}_2 = \{6, 7\}$, and so $\mathcal{A}_1 = \{1\}$ and $\mathcal{A}_2 = \{5\}$. Therefore the individual majority rule for each set holds, and no more instruments can be invalid. Although the total number of two instruments allowed to be invalid is the same here as in the example of Table 1, it is clear that they cannot be both in the set that is relevant for one of the exposure variables. This conditions can change if there is some overlap between the two groups, for example if $S_1 = \{1, 2, 3, 4, 5\}$ and $S_2 = \{5, 6, 7\}$, both instruments 1 and 2 can be invalid, as illustrated in Table 3, as the majority in S_1 is valid. This is not the case if instead $S_1 = \{1, 2, 3, 4\}$ and $S_2 = \{4, 5, 6, 7\}$.

Table 2: Illustration of median-of-medians estimator with block structure relevance of instruments.

Instruments	1	**2**	**3**	**4**	5	**6**	**7**	
1					$\widehat{\beta}_{1,q}^5$	$\widehat{\beta}_{1,q}^6$	$\widehat{\beta}_{1,q}^7$	
2					$\widehat{\beta}_{2,q}^5$	$\widehat{\boldsymbol{\beta}}_{\mathbf{2},\boldsymbol{q}}^{\mathbf{6}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{2},\boldsymbol{q}}^{\mathbf{7}}$	
3					$\widehat{\beta}_{3,q}^5$	$\widehat{\boldsymbol{\beta}}_{\mathbf{3},\boldsymbol{q}}^{\mathbf{6}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{3},\boldsymbol{q}}^{\mathbf{7}}$	
4					$\widehat{\beta}_{4,q}^5$	$\widehat{\boldsymbol{\beta}}_{\mathbf{4},\boldsymbol{q}}^{\mathbf{6}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{4},\boldsymbol{q}}^{\mathbf{7}}$	
5	$\widehat{\beta}_{5,q}^1$	$\widehat{\beta}_{5,q}^2$	$\widehat{\beta}_{5,q}^3$	$\widehat{\beta}_{5,q}^2$				
6	$\widehat{\beta}_{6,q}^1$	$\widehat{\boldsymbol{\beta}}_{\mathbf{6},\boldsymbol{q}}^{\mathbf{2}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{6},\boldsymbol{q}}^{\mathbf{3}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{6},\boldsymbol{q}}^{\mathbf{4}}$				
7	$\widehat{\beta}_{7,q}^1$	$\widehat{\boldsymbol{\beta}}_{\mathbf{7},\boldsymbol{q}}^{\mathbf{2}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{7},\boldsymbol{q}}^{\mathbf{3}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{7},\boldsymbol{q}}^{\mathbf{4}}$				
median	$\widehat{\beta}_{m,q}^1$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m},\boldsymbol{q}}^{\mathbf{2}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m},\boldsymbol{q}}^{\mathbf{3}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m},\boldsymbol{q}}^{\mathbf{4}}$	$\widehat{\beta}_{m,q}^5$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m},\boldsymbol{q}}^{\mathbf{6}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{m},\boldsymbol{q}}^{\mathbf{7}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{mm},\boldsymbol{q}}$

Notes: $k_x = 2$, $S_1 = \{1, 2, 3, 4\}$, $S_2 = \{5, 6, 7\}$,$\mathcal{V} = \{2, 3, 4, 6, 7\}$, $\mathcal{A} = \{1, 6\}$. Valid instruments and consistent estimators are displayed in boldface.

Table 3: Illustration of median-of-medians estimator with block structure relevance of instruments with overlap.

Instruments	1	2	**3**	**4**	**5**	**6**	**7**	
1					$\widehat{\beta}^{5}_{1,q}$	$\widehat{\beta}^{6}_{1,q}$	$\widehat{\beta}^{7}_{1,q}$	
2					$\widehat{\beta}^{5}_{2,q}$	$\widehat{\beta}^{6}_{2,q}$	$\widehat{\beta}^{7}_{2,q}$	
3					$\widehat{\boldsymbol{\beta}}^{\mathbf{5}}_{\mathbf{3},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{7}}_{\mathbf{4},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{7}}_{\mathbf{3},\boldsymbol{q}}$	
4					$\widehat{\boldsymbol{\beta}}^{\mathbf{5}}_{\mathbf{4},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{7}}_{\mathbf{3},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{7}}_{\mathbf{4},\boldsymbol{q}}$	
5	$\widehat{\beta}^{1}_{5,q}$	$\widehat{\beta}^{2}_{5,q}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{3}}_{\mathbf{5},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{4}}_{\mathbf{5},\boldsymbol{q}}$		$\widehat{\boldsymbol{\beta}}^{\mathbf{6}}_{\mathbf{5},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{6}}_{\mathbf{5},\boldsymbol{q}}$	
6	$\widehat{\beta}^{1}_{6,q}$	$\widehat{\beta}^{2}_{5,q}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{3}}_{\mathbf{4},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{4}}_{\mathbf{6},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{5}}_{\mathbf{6},\boldsymbol{q}}$			
7	$\widehat{\beta}^{1}_{7,q}$	$\widehat{\beta}^{2}_{5,q}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{3}}_{\mathbf{7},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{4}}_{\mathbf{7},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{5}}_{\mathbf{7},\boldsymbol{q}}$			
median	$\widehat{\beta}^{1}_{m,q}$	$\widehat{\beta}^{2}_{m,q}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{3}}_{\boldsymbol{m},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{4}}_{\boldsymbol{m},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{5}}_{\boldsymbol{m},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{6}}_{\boldsymbol{m},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}^{\mathbf{7}}_{\boldsymbol{m},\boldsymbol{q}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{mm},\boldsymbol{q}}$

Notes: $k_x = 2$, $S_1 = \{1,2,3,4,5\}$, $S_2 = \{5,6,7\}$,$\mathcal{V} = \{3,4,5,6,7\}$, $\mathcal{A} = \{1,2\}$. Valid instruments and consistent estimators are displayed in boldface.

6 Monte Carlo Simulations

We conduct Monte Carlo simulations to evaluate the performance of our method in three settings. In the first setting, all instruments are relevant for both of the endogenous variables, while in the other two settings, some of the instruments are relevant for only one of the variables. We run the simulations for $1,000$ replications, and we implement the adaptive Lasso using the `Lars` package (Hastie and Efron, 2013) in `R`. We set $k_x = 2$ and generate the data from

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}\\ \mathbf{x_1} &= \mathbf{Z}\boldsymbol{\pi}_1 + \mathbf{e}_1\\ \mathbf{x_2} &= \mathbf{Z}\boldsymbol{\pi}_2 + \mathbf{e}_2\end{aligned}$$

where

$$\begin{pmatrix} U_i \\ E_{1i} \\ E_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & 0 \\ \rho_2 & 0 & 1 \end{pmatrix}\right);$$

$$\mathbf{Z}_i \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_z\right);$$

with $\boldsymbol{\beta} = (0.3, 0.6)^T$; $k_z = 21$; $\rho_1 = 0.25$, $\rho_2 = 0.3$; $k_{\mathcal{V}} = 12$, $k_{\mathcal{A}} = 9$, $\boldsymbol{\alpha} = 0.4\left(\boldsymbol{\iota}_9^T, \mathbf{0}_{12}^T\right)^T$, similar to the simulation setup in Windmeijer et al. (2021). We generate the elements of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ from a uniform distribution on the interval $[1.5, 2.5]$, and we set the elements

of $\mathbf{\Sigma}_z$ to $\Sigma_{z,jk} = 0.5^{|j-k|}$, $j, k = 1, \ldots, k_z$. In this setup, all the instruments are relevant for both endogenous variables, and both the pair-wise full rank assumption 7 and the majority assumption 8 are satisfied.

Table 4 presents the IV selection and estimation results of the adaptive Lasso method. with penalty parameters chosen by 10-fold cross-validation. The first two columns of Table 4 report statistics related to estimation, and in both of these columns, we average the statistics over the two entries in $\boldsymbol{\beta}$. Column 1 presents the averaged median absolute error (MAE), while Column 2 shows the averaged standard deviation (SD). The remaining three columns in Table 4 report statistics related to IV selection. Column 3 reports the number of instruments selected as invalid, Column 4 the frequency with which all invalid instruments have been selected as invalid, and Column 5 is the frequency with which the oracle model has been selected. The three panels in Table 4 correspond to the sample sizes $n = 500, 1000, 2000$. In each panel, the first row, denoted "Oracle 2SLS", shows the results for the oracle 2SLS estimator, which is the 2SLS estimator that uses the true set of valid instruments, while it controls for the remaining invalid ones. The second row, denoted "Naive 2SLS", reports the results for the naive 2SLS estimates, which is the 2SLS that considers all candidate instruments to be valid. The third row, denoted $\widehat{\boldsymbol{\beta}}_{mm}$, shows the results for the median-in-medians estimator, as defined in (12). The fourth and fifth rows, denoted "Post-ALasso", report the results for the post-selection 2sls estimators, which are the 2sls estimators that use the instruments selected as valid, and include the invalid instruments as control variables. We present results for the adaptive Lasso and Post-Selection 2SLS estimators using two different types of cross-validation. First, as denoted with the "*cv*" subscript, we show cross-validation using the tuning parameter that gives the minimum cross-validation Sargan statistics. Second, as denoted with the "*cvse*" subscript, we show cross-validation using the tuning parameter chosen by the one-standard-error rule.

In terms of IV selection, in all three sample sizes, the *cv*-procedure dominates the *cvse*-procedure, especially for the smallest sample with $n = 500$. Both methods improve as the sample size increases. The frequencies of selecting the oracle model are both almost equal to 1 at $n = 2,000$ with 0.992 for CV and 0.956 for CVSE. In line with the selection performance, the post-selection 2SLS estimates are close to the oracle model at $n = 2,000$. In all three sample sizes, the post-selection 2SLS estimates outperform the adaptive Lasso estimates.

Table 4: Simulation Results

	MAE (1)	SD (2)	# invalid (3)	p allinv (4)	p oracle (5)
Panel (a), $n = 500$					
Oracle 2SLS	0.0624	0.0912	9	1	1
Naive 2SLS	0.2856	0.2685	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.1263	0.1517			
ALasso$_{cv}$	0.3460	0.4212	8.095	0.440	0.440
Post-ALasso$_{cv}$	0.1295	0.3060			
ALasso$_{cvse}$	0.4393	0.4542	6.908	0.115	0.115
Post-ALasso$_{cvse}$	0.2912	0.4368			
Post-ALasso$_{Sar}$	0.0853	0.1446	9.09	0.987	0.947
Panel (b), $n = 1,000$					
Oracle 2SLS	0.0439	0.0681	9	1	1
Naive 2SLS	0.2889	0.2037	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.0892	0.1268			
ALasso$_{cv}$	0.2047	0.2843	8.857	0.882	0.882
Post-ALasso$_{cv}$	0.0513	0.1627			
ALasso$_{cvse}$	0.2895	0.3446	8.509	0.634	0.634
Post-ALasso$_{cvse}$	0.0716	0.2332			
Post-ALasso$_{Sar}$	0.0544	0.0690	9.02	1	0.983
Panel (c), $n = 2,000$					
Oracle 2SLS	0.0305	0.0473	9	1	1
Naive 2SLS	0.2796	0.1448	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.0618	0.0941			
ALasso$_{cv}$	0.1341	0.1795	8.991	0.992	0.992
Post-ALasso$_{cv}$	0.0307	0.0574			
ALasso$_{cvse}$	0.1753	0.2244	8.949	0.956	0.956
Post-ALasso$_{cvse}$	0.0319	0.0881			
Post-ALasso$_{Sar}$	0.0375	0.0478	9.018	1	0.984

Notes: The reported statistics include median absolute error (Column 1), standard deviation (Column 2), number of IVs selected as invalid (Column 3), frequency with which all invalid IVs have been selected as invalid (Column 4), and frequency with which oracle model has been selected (Column 5). The simulations are based on 1,000 repetitions.

Next, we consider the case where the sets of instruments for $\mathbf{x}_1$ and $\mathbf{x}_2$ are separate, such that no instrument is relevant for both endogenous variables. We set $\boldsymbol{\pi}_1 = (\boldsymbol{\gamma}_1^T, \mathbf{0}_{11}^T)^T$ and $\boldsymbol{\pi}_2 = (\mathbf{0}_{10}^T, \boldsymbol{\gamma}_2^T)^T$, where $\boldsymbol{\gamma}_1$ has length $k_1 = 10$ and $\boldsymbol{\gamma}_2$ has length $k_2 = 11$. We let $\boldsymbol{\alpha} = (\boldsymbol{\iota}_4^T, \mathbf{0}_6^T, \boldsymbol{\iota}_5^T, \mathbf{0}_6^T)^T$ such that $k_{\mathcal{V}_1} = 6$, $k_{\mathcal{A}_1} = 4$ and $k_{\mathcal{V}_2} = 6$, $k_{\mathcal{A}_1} = 5$. All the other parameters are identical to the previous simulation design. Again, the necessary and sufficient majority assumption for consistency of the median-of-medians estimator,

$k_{\mathcal{V}_1} > \frac{k_{z_1}}{2}$ and $k_{\mathcal{V}_2} > \frac{k_{z_2}}{2}$, is satisfied.

Table 5: Simulation results, separate sets of instruments for each exposure variable

	MAE (1)	SD (2)	# invalid (3)	p allinv (4)	p oracle (5)
Panel (a), $n = 500$					
Oracle 2SLS	0.0109	0.0161	9	1	1
Naive 2SLS	0.3285	0.0209	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.1124	0.1472			
Post-ALasso$_{Sar}$	0.0159	0.1779	10.241	0.791	0.515
$\widehat{\boldsymbol{\beta}}_{mm,block}$	0.0839	0.0394			
Post-ALasso$_{block}$	0.0111	0.0192	9.044	0.999	0.971
Panel (b), $N = 1,000$					
Oracle 2SLS	0.0075	0.0111	9	1	1
Naive 2SLS	0.3288	0.0149	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.0892	0.1629			
Post-ALasso$_{Sar}$	0.0102	0.2413	10.052	0.786	0.565
$\widehat{\boldsymbol{\beta}}_{mm,block}$	0.0599	0.0283			
Post-ALasso$_{Sar,block}$	0.0076	0.0115	9.019	1.000	0.987
Panel (c), $N = 2,000$					
Oracle 2SLS	0.0054	0.0080	9	1	1
Naive 2SLS	0.3286	0.0107	0	0	0
$\widehat{\boldsymbol{\beta}}_{mm}$	0.0703	0.1667			
Post-ALasso$_{Sar}$	0.0071	0.1847	10.086	0.803	0.544
$\widehat{\boldsymbol{\beta}}_{mm,block}$	0.0411	0.0196			
Post-ALasso$_{Sar,block}$	0.0054	0.0084	9.020	1.000	0.987

Notes: This table reports IV selection and estimation results of the adaptive Lasso method with the block structure in simulation design (2) with no overlap. The reported statistics include median absolute error (column 1), standard deviation (column 2), number of IVs selected as invalid (column 3), frequency with which all invalid IVs have been selected as invalid (column 4), and frequency with which oracle model has been selected (column 5). The simulations are based on $1,000$ repetitions.

7 Application: The Effects of Educational Attainment and Cognitive Ability on BMI

We apply our IV selection method to a multivariable Mendelian randomisation (MVMR) study. We estimate the effects of educational attainment and cognitive ability on Body Mass Index (BMI), as in Sanderson et al. (2019). Both educational attainment and cognitive ability have been found to be negatively correlated with BMI (Sanderson et al., 2019). However, as educational attainment and cognitive ability are highly correlated, it

is unclear to what extent each of them have a direct effect on BMI. In this application, we account for both variables in order to disentangle their direct effects. We use 74 SNPs as instruments for educational attainment and 18 SNPs for cognitive ability, and one SNPs overlaps between the two sets of candidate instruments. These SNPs have previously been identified in independent Genome-Wide Association Studies (GWAS), see Okbay et al. (2016) for educational attainment, and Sniekers et al. (2017) for cognitive ability. We use data on 107,371 individuals from the UK Biobank. Educational attainment is measured in years of completed education, and it is imputed based on the individuals' qualifications, which is standard in the literature, see, e.g., Okbay et al. (2016). Cognitive ability is measured as a unitless fluid intelligence score that the UK biobank constructs from a series of tests completed by the individuals during assessment. We standardise the cognitive ability to mean zero and variance one. BMI is the ratio of weight to height, both of which were measured for all individuals during assessment, and we log-transform it due to skewness. Hence, we interpret our estimates as the percentage change in BMI that is associated with a one unit increase in the relevant explanatory variable. We also include additional covariates that control for age at assessment, sex, and the first 10 genetic principal components, all of which are available from the UK biobank. See Sanderson et al. (2019) for a detailed definition of the variables and presentation of the data.

Table 6: The impacts of educational attainment and cognitive ability on $\log(BMI)$

	Estimate (1)	Std. error (2)	# Invalid (3)	p-value, Sargan (4)
Panel (a) – 2SLS				
Educational attainment	-0.035	0.004	0	1.69e-13
Cognitive ability	0.031	0.011		
Panel (b) – Post-A Lasso$_{Sar}$				
Educational attainment	-0.029	0.005	10	0.011
Cognitive ability	0.021	0.012		
$\widehat{\beta}_{mm,edu}$	-0.031			
$\widehat{\beta}_{mm,cog}$	0.017			

Notes: The sample size is $n = 107,371$. The number of instruments for educational attainment is $k_{edu} = 74$. The number of instruments for cognitive ability is $k_{cog} = 18$. There is one instrument identified for both educational attainment and cognitive ability.

Table 6 reports the results of our analysis. Columns (1) and (2) show, respectively, the point estimates and their standard errors. Column (3) is the number of instruments selected as invalid, and column (4) shows the p-value of the Sargan test. Panel (a) presents

the estimates from a naive 2SLS regression where we treat all the candidate instruments as valid. Both estimates are statistically significant at the 1% level. However, these results are from the naive 2SLS regression, and they might be biased due to the presence of invalid instruments. This is supported by the small p-value of the Sargan test (1.69e-13). In practice, SNPs can exhibit so-called pleiotropic effects, which would make them invalid instruments. In our setting, pleiotropy would mean that some of the SNPs, either for educational attainment or cognitive ability, have direct effects on BMI.

Instead of the naive 2sls, we now conduct IV selection using the adaptive Lasso with the downward testing procedure, as described in Section 4.1, and we obtain post-selection 2sls estimates. In Panel (b) of Table 6, we report the results for the direct effects of educational attainment and cognitive ability using our method, and we show the estimates taking the block structure into account. We also present the associated median-of-medians estimates, denoted $\widehat{\beta}_{mm,edu}$ and $\widehat{\beta}_{mm,cog}$ for, respectively, educational attainment and cognitive ability. The threshold p-value for the Sargan test is $0.1/\log(n) = 0.0086$.

For educational attainment we find that $\widehat{\beta}_{mm,edu} = -0.0314$. For cognitive ability, the estimate is $\widehat{\beta}_{mm,cog} = 0.017$. We find that our method selects 10 instruments as invalid. Six of these are for educational attainment, three are for cognitive ability, and one is for both variables. As seen in Column (4), the p-value of the Sargan statistic for the selected model is 0.011. We find that the post-selection 2sls estimates are closer to zero compared to the estimates for the naive 2sls, especially for cognitive ability. The post-selection estimate for educational attainment is -0.029, while, for cognitive ability, it is 0.021. The effect of educational attainment on $\log(BMI)$ is still significant at the 1% level, while the effect of cognitive ability is insignificant at the 5% level. We therefore find limited evidence of a direct effect of cognitive ability on BMI.

For the results in Table 6, we assume conditional homoskedasticity. However, a robust version of our method, i.e. using the two-step Hansen J-test and the post-selection two-step GMM estimator, produces almost identical results. We use the Sanderson-Windmeijer conditional F-statistic (Sanderson and Windmeijer, 2016) to evaluate the power of the instruments in predicting educational attainment and cognitive ability jointly. When we include the instruments in the naive 2SLS, the conditional F statistics are 2.57 for educational attainment and 2.65 for cognitive ability. Both of them are significantly lower than the rule-of-thumb value of 10, showing that the joint prediction power of the instruments is relatively weak. One way to deal with this weak IV problem to create a weighted score of all the instruments, that is, one score for each of educational attainment

and cognitive ability, and then use these two scores as the instruments in the regression. In the naive 2SLS, when we use the weighted scores, the conditional F statistics are 67.73 for educational attainment and 68.65 for cognitive ability, which are much larger than the rule-of-thumb value of 10. For the post-selection 2sls, we create the weighted scores using only the selected valid instruments. The estimate for educational attainment is -0.042 (se 0.009) and for cognitive ability it is 0.041 (se 0.024). This maintains the conclusion that educational attainment has a significant negative effect on BMI, while the direct effect of cognitive ability is insignificant.

8 Conclusion

We investigate the use of the adaptive Lasso method for selecting valid instrumental variables from a set of candidate instruments when some of the instruments may be invalid. While existing work has focused on a single endogenous variable, our method contributes to the literature by allowing for multiple endogenous exposure variables. Under a modified majority rule, we show that the adaptive Lasso method can achieve consistent selection and oracle estimation. In this work, we consider the number of candidate instruments to be fixed, but in some settings it may grow with the sample size (or even at a quicker rate), and, therefore, future research will focus on extending the method to handle such cases.

Appendix, Proofs

Proposition 1

Proof. The proof of Proposition 1, with $k_x = 2$, follows the arguments of the proof of Theorem 1 in Windmeijer et al. (2019). The estimands for the $k_z - 1$ just-identified IV estimators in the model specification (9)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_s + \mathbf{Z}_{\{-s\}}\boldsymbol{\alpha}_{\{-s\}} + \mathbf{u}_s,$$

are given by $\boldsymbol{\beta} + \boldsymbol{\Pi}_s^{-1}\boldsymbol{\alpha}_s$, with $\boldsymbol{\Pi}_s$ as defined in (10). It then follows, under Assumptions 2, 3 and 7 that $\widehat{\boldsymbol{\beta}}_\ell^j$ as defined in (11) is a consistent and normal estimator of $\boldsymbol{\beta}$ for a valid instrument $j \in \mathcal{V}$ and when the other instrument ℓ is also valid. For each element β_q in $\boldsymbol{\beta}$, $q = 1, 2$, we have for instrument j a vector of $k_z - 1$ estimators $\widehat{\boldsymbol{\beta}}_q^j$. Let $\boldsymbol{\delta}_\ell^j$ be the

k_x-vector $\mathbf{\Pi}_{\{j,\ell\}}^{-1}\boldsymbol{\alpha}_{\{j,\ell\}}$, with elements $\boldsymbol{\delta}_{\{j,\ell\},q}^{j}$, for $q = 1, \ldots, k_x$. For each element we have the $(k_z - 1)$-vector $\boldsymbol{\delta}_q^j$. It follows that

$$\widehat{\boldsymbol{\beta}}_q^j \xrightarrow{p} \beta_q \boldsymbol{\iota}_{k_z-1} + \boldsymbol{\delta}_q^j,$$

where $\boldsymbol{\iota}_{k_z-1}$ is a $(k_z - 1)$-vector of ones. For a valid instrument $j \in \mathcal{V}$ there are $k_\mathcal{V} - 1$ sets with valid instruments only. By Assumption 8 $(k_\mathcal{V} - 1) > \frac{k_z-1}{2}$, it follows that the majority rule is satisfied and more than 50% of the elements of $\boldsymbol{\delta}_q^j$ are equal to zero. Using a continuity theorem, it then follows that, for a valid instrument $j \in \mathcal{V}$,

$$\text{median}\left(\widehat{\boldsymbol{\beta}}_q^j\right) \xrightarrow{p} \beta_q + \text{median}\left(\boldsymbol{\delta}_q^j\right) = \beta_q, \tag{19}$$

for $q = 1, \ldots, k_x$, and hence the first result of Proposition 1 therefore follows, $\widehat{\boldsymbol{\beta}}_m^j \xrightarrow{p} \boldsymbol{\beta}$.

Under Assumptions 2, 3 and 7 the limiting distribution of $\widehat{\boldsymbol{\beta}}_q^j$, for $q = 1, \ldots, k_x$, is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_q^j - \left(\beta_q \boldsymbol{\iota}_c + \boldsymbol{\delta}_q^j\right)\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_q^j}\right).$$

For $\widehat{\beta}_{m,q}^j = \text{median}\left(\widehat{\boldsymbol{\beta}}_q^j\right)$ we have that

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}_{m,q}^j - \beta_q\right) &= \sqrt{n}\left(\text{median}\left(\widehat{\boldsymbol{\beta}}_q^j\right) - \beta_q\right) \\ &= \text{median}\left(\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_q^j - \beta_q \boldsymbol{\iota}_{k_z-1}\right)\right).\end{aligned}$$

For a valid instrument $j \in \mathcal{V}$, let $\boldsymbol{\delta}_{\mathcal{A},q}^j$ denote the $k_z - k_\mathcal{V}$ values of $\boldsymbol{\delta}_q^j$ for the sets that include invalid instruments and $\boldsymbol{\delta}_{\mathcal{V},q}^j = \mathbf{0}_{k_\mathcal{V}-1}$ the $k_\mathcal{V} - 1$ values of $\boldsymbol{\delta}_q^j$ for the sets that only contain valid instruments. Partition $\boldsymbol{\delta}_q^j = \left(\left(\boldsymbol{\delta}_{\mathcal{A},q}^j\right)^T \ \mathbf{0}_{k_\mathcal{V}-1}^T\right)^T$ and equivalently $\widehat{\boldsymbol{\beta}}_q^j = \left(\left(\widehat{\boldsymbol{\beta}}_{\mathcal{A},q}^j\right)^T \ \left(\widehat{\boldsymbol{\beta}}_{\mathcal{V},q}^j\right)^T\right)^T$. Then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_q^j - \beta_q \boldsymbol{\iota}_c\right) = \begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathcal{A},q}^j - \left(\beta_q \boldsymbol{\iota}_{k_z-k_\mathcal{V}} + \boldsymbol{\delta}_{\mathcal{A},q}^j\right)\right) + \sqrt{n}\boldsymbol{\delta}_{\mathcal{A},q}^j \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathcal{V},q}^j - \beta_q \boldsymbol{\iota}_{k_\mathcal{V}-1}\right) \end{pmatrix},$$

and it follows that

$$\sqrt{n}\left(\widehat{\beta}_{m,q}^j - \beta_q\right) = \text{median}\left(\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_q^j - \beta_q \boldsymbol{\iota}_{k_z-1}\right)\right) \xrightarrow{d} H_{[r_j],k_z-1,q}^j, \tag{20}$$

for $q = 1, \ldots, k_x$, and where, for $k_z - 1$ odd, $H_{[r_j],k_\mathcal{V}-1,q}^j$ is the r_jth-order statistic of the

limiting normal distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{j}_{\mathcal{V},q} - \beta_q \boldsymbol{\iota}_{k_\mathcal{V}-1}\right)$, where r_j is determined by k_z, $k_\mathcal{V}$ and the signs of the elements of $\boldsymbol{\delta}^{j}_{\mathcal{A},q}$. For $k_z - 1$ even, $H^{j}_{[r_j],k_\mathcal{V}-1,q}$ is defined as the average of either the $[r_j]$ and $[r_j - 1]$-order statistics, or the $[r_j]$ and $[r_j + 1]$-order statistics, see Windmeijer et al. (2019, p 1343). From (20) it follows that $\widehat{\beta}^{j}_{m,q}$ converges at the $\sqrt{n}$ rate. It has an asymptotic bias, but $\sqrt{n}\left(\widehat{\beta}^{j}_{m,q} - \beta_q\right) = O_p(1)$ for $q = 1, 2$, and so the second result of Proposition 1 holds, $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{j}_{m} - \boldsymbol{\beta}\right) = O_p(1)$. □

Proposition 2

Proof. For $k_x = 2$, and for $q = 1, 2$ we have the k_z median estimators $\widehat{\beta}^{j}_{m,q}$, $j = 1, \ldots, k_z$, of β_q. Denote the k_z-vector of estimators by $\widehat{\boldsymbol{\beta}}_{m,q}$. Let $\widehat{\boldsymbol{\beta}}^{\mathcal{V}}_{m,q}$ denote the $k_\mathcal{V}$-vector $\left(\widehat{\beta}^{j}_{m,q}\right)_{j\in\mathcal{V}}$ and $\widehat{\boldsymbol{\beta}}^{\mathcal{A}}_{m,q}$ the $(k_z - k_\mathcal{V})$-vector $\left(\widehat{\beta}^{j}_{m,q}\right)_{j\in\mathcal{A}}$. Partition $\widehat{\boldsymbol{\beta}}_{m,q} = \left(\left(\widehat{\boldsymbol{\beta}}^{\mathcal{A}}_{m,q}\right)^T \left(\widehat{\boldsymbol{\beta}}^{\mathcal{V}}_{m,q}\right)^T\right)^T$. Then under the assumptions and from the results of Proposition 1 it follows that

$$\widehat{\boldsymbol{\beta}}_{m,q} \xrightarrow{p} \begin{pmatrix} \beta_q \boldsymbol{\iota}_{k_z - k_\mathcal{V}} + \boldsymbol{\gamma}_q \\ \beta_q \boldsymbol{\iota}_{k_\mathcal{V}} \end{pmatrix},$$

where $\boldsymbol{\gamma}_q$ is the $(k_z - k_\mathcal{V})$-vector with elements median $\left(\boldsymbol{\delta}^{j}_{q}\right)_{j\in\mathcal{A}}$, with $\boldsymbol{\delta}^{j}_{q}$ as defined in the proof of Proposition 1. Therefore, if the majority rule holds that $k_\mathcal{V} > \frac{k_z}{2}$, it follows that

$$\beta_{mm,q} = \text{median}\left(\widehat{\boldsymbol{\beta}}_{m,q}\right) \xrightarrow{p} \beta_q,$$

for $q = 1, 2$. From Assumption 8 it follows $k_\mathcal{V} > \frac{k_z+1}{2} > \frac{k_z}{2}$, and so it follows that the first result of Proposition 2 holds, $\boldsymbol{\beta}_{mm} \xrightarrow{p} \boldsymbol{\beta}$.

For the limiting distribution, we have, for $q = 1, 2$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{m,q} - \beta_q \boldsymbol{\iota}_{k_z}\right) = \begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\mathcal{A}}_{m,q} - \left(\beta_q \boldsymbol{\iota}_{k_z - k_\mathcal{V}} + \boldsymbol{\gamma}_q\right)\right) + \sqrt{n}\boldsymbol{\gamma}_q \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\mathcal{V}}_{m,q} - \beta_q \boldsymbol{\iota}_{k_\mathcal{V}}\right) \end{pmatrix}$$

and, as $k_\mathcal{V} > \frac{k_z}{2}$,

$$\sqrt{n}\left(\widehat{\beta}_{mm,q} - \beta_q\right) = \text{median}\left(\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{m,q} - \beta_q \boldsymbol{\iota}_{k_z}\right)\right)$$

converges in distribution to the distribution of an order statistic of the distribution of the order statistics $\left(H^{j}_{[r_j],k_\mathcal{V}-1,q}\right)_{j\in\mathcal{V}}$, which is again $O_p(1)$. From this, the second result of Proposition 2 holds, $\sqrt{n}\left(\boldsymbol{\beta}_{mm} - \boldsymbol{\beta}\right) = O_p(1)$. □

References

Andrews, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563.

Apfel, N. (2019): "Relaxing the Exclusion Restriction in Shift-Share Instrumental Variable Estimation," *SSRN Electronic Journal, https://dx.doi.org/10.2139/ssrn.3408682*.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80, 2369–2429.

Bühlmann, P. and S. Van De Geer (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science & Business Media.

Chand, S. (2012): "On Tuning Parameter Selection of Lasso-Type Methods: A Monte Carlo Study," in *Proceedings of the 9th International Bhurban Conference on Applied Sciences & Technology (IBCAST)*, IEEE, 120–129.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004): "Least Angle Regression," *The Annals of Statistics*, 32, 407–499.

Fan, J. and R. Li (2001): "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Guo, Z., H. Kang, T. T. Cai, and D. S. Small (2018): "Confidence Intervals for Causal Effects with Invalid Instruments by Using Two-Stage Hard Thresholding with Voting," *Journal of the Royal Statistical Society: Series B*, 80, 793–815.

Han, C. (2008): "Detecting Invalid Instruments Using L1-GMM," *Economics Letters*, 101, 285–287.

Hastie, T. and B. Efron (2013): *lars: Least Angle Regression, Lasso and Forward Stagewise*, r package version 1.2.

Holland, P. W. (1988): "Causal Inference, Path Analysis, and Recursive Structural Equations Models," *Sociological Methodology*, 18, 449–484.

Kang, H., A. Zhang, T. T. Cai, and D. S. Small (2016): "Instrumental Variables Estimation With Some Invalid Instruments and Its Application to Mendelian Randomization," *Journal of the American Statistical Association*, 111, 132–144.

Okbay, A., J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, P. Turley, G.-B. Chen, V. Emilsson, S. F. W. Meddens, et al. (2016): "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment," *Nature*, 533, 539–542.

Sanderson, E., G. D. Smith, F. Windmeijer, and J. Bowden (2019): "An Examination of Multivariable Mendelian Randomization in the Single-Sample and Two-Sample Summary Data Settings," *International Journal of Epidemiology*, 48, 713–727.

Sanderson, E. and F. Windmeijer (2016): "A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics*, 190, 212–221.

Sargan, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Sniekers, S., S. Stringer, K. Watanabe, P. R. Jansen, J. R. Coleman, E. Krapohl, E. Taskesen, A. R. Hammerschlag, A. Okbay, D. Zabaneh, et al. (2017): "Genome-Wide Association Meta-Analysis of 78,308 Individuals Identifies New Loci and Genes Influencing Human Intelligence," *Nature Genetics*, 49, 1107–1112.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

Windmeijer, F. (2021): "Testing underidentification in linear models, with applications to dynamic panel and asset pricing models," *Journal of Econometrics*, https://doi.org/10.1016/j.jeconom.2021.03.007.

Windmeijer, F., H. Farbmacher, N. Davies, and G. D. Smith (2019): "On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments," *Journal of the American Statistical Association*, 114, 1339–1350.

Windmeijer, F., X. Liang, F. P. Hartwig, and J. Bowden (2021): "The confidence interval method for selecting valid instrumental variables," *Journal of the Royal Statistical Society: Series B*, 83, 752–776.

Zhao, P. and B. Yu (2006): "On Model Selection Consistency of LASSO," *Journal of Machine learning research*, 7, 2541–2563.

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.