Post-selection Problems for Causal Inference with Invalid Instruments: A Solution Using Searching and Sampling

Zijian Guo*

Abstract

Instrumental variable method is among the most commonly used causal inference approaches for analyzing observational studies with unmeasured confounders. Despite its popularity, the instruments' invalidity is a major concern for practical applications and a fast-growing area of research is inference for the causal effect with possibly invalid instruments. In this paper, we construct uniformly valid confidence intervals for the causal effect when the instruments are possibly invalid. We illustrate the post-selection problem of existing inference methods relying on instrument selection. Our proposal is to search for the value of the treatment effect such that a sufficient amount of candidate instruments are taken as valid. We further devise a novel sampling method, which, together with searching, lead to a more precise confidence interval. Our proposed searching and sampling confidence intervals are shown to be uniformly valid under the finite-sample majority and plurality rules. We compare our proposed methods with existing inference methods over a large set of simulation studies and apply them to study the effect of the triglyceride level on the glucose level over a mouse data set.

Key words: unmeasured confounders; uniform inference; post-selection inference; majority rule; plurality rule; mendelian randomization.

1 Introduction

Existence of unmeasured confounders is a major concern for causal inference from observational studies. Instrumental Variable (IV) method is one of the most commonly used causal inference approaches to deal with unmeasured confounders. The validity of IV methods relies on the proposed instruments satisfying three identification conditions: conditioning on the baseline covariates, (A1) the instruments are associated with the treatment; (A2)

*Z. Guo is an assistant professor at Department of Statistics, Rutgers University. The research of Z. Guo was supported in part by the NSF DMS 1811857, 2015373 and NIH R01GM140463-01.

the instruments are independent with the unmeasured confounders; (A3) the instruments have no direct effect on the outcome.

In practice, the main challenge of applying IV-based methods is to identify instruments satisfying (A1)-(A3). Assumptions (A2) and (A3), the so-called "exclusion restriction" assumptions, are crucial for the causal effect identification as they assume that the instruments can only affect the outcome through the treatment and exclude all other pathways. However, assumptions (A2) and (A3) cannot even be tested in a data-dependent way. A fast-growing literature [3, 5, 14, 18–20, 25, 30, 34] is focused on causal inference with invalid instruments which may not satisfy assumptions (A2) and (A3). Many of these works are motivated from applications in Mendelian Randomization (MR), which study causal effects using genetic variants as instruments; see [6] for a review of IV methods in MR. In MR applications, the adopted genetic variants are possibly invalid instruments due to the pleiotropic effects [9,10], that is, a genetic variant may affect both the treatment and the outcome variable.

The existing works [11,14,19,25,33,34] propose to first select valid instruments in a data-dependent way and then make inference for the effect with the selected instruments. With a finite amount of data, the selection step may make mistakes in detecting (and removing) invalid instruments, especially when the assumptions (A2) and (A3) are violated mildly; see Section 3 for an illustration. There is a pressing need to address the post-selection issue of instrument selection and propose uniformly valid confidence intervals.

We shall emphasize the practical importance of developing uniform inference methods which are robust to IV selection mistakes. In applications, there are chances that the invalid instruments are weakly invalid, that is, they only violate assumptions (A2) and (A3) mildly. The weakly invalid instruments are hard to be detected with a finite amount of data. The goal of the current paper is to develop a uniform statistical inference method for the treatment effect, which is valid even in the presence of weakly invalid instruments.

1.1 Results and Contributions

The current paper is focused on linear outcome models with a few candidate instruments (IVs). We illusrate the post-selection problem for the existing inference methods `TSHT` [14] and `CIIV` [34], both of which rely on invalid instruments being correctly detected. Since some weakly invalid instruments are hard to be identified as invalid, the selection errors lead to the post-selection bias and under-coverage; see Figure 2 and Table 1 for an illustration.

Identification conditions are needed for causal inference with invalid instruments, such as majority rule [5,19] and plurality rule [14]. Both rules require that there are enough valid instruments among the candidate instruments, even though the validity of any instrument is not known a priori; see (5) and (6) for the exact definitions. These rules enable us to detect invalid instruments (even weakly invalid) in the setting that we have an infinite amount of data. However, these identification conditions might not copy well with the practical applications with a finite amount of data. To address this, we introduce the finite-sample majority rule (Condition 1) and the finite-sample plurality rule (Condition 2). When there is an infinite amount of data, the finite-sample majority (or plurality) rule is equivalent to the existing majority (or plurality) rule. We shall highlight that, even the data generating mechanism satisfies the finite sample majority or plurality rule, there is still a lack of uniformly valid inference methods in the literature.

Our first proposed confidence interval (CI) is based on the *searching* idea. For every β value, we implement a thresholding step and decide which candidate instruments are valid. If the majority rule (Condition 1) holds, we search for a range of β values such that, with the corresponding β, more than half of candidate instruments are taken as valid. If only the plurality rule (Condition 2) holds, we estimate the set $\mathcal{V}$ of valid IVs by $\widehat{\mathcal{V}}$ and then apply the searching method with the estimated set $\widehat{\mathcal{V}}$. Our proposed searching CI works even if $\widehat{\mathcal{V}}$ is not equal to $\mathcal{V}$. The set $\widehat{\mathcal{V}}$ is allowed to include weakly invalid instruments, which are hard to detect in practice. We modify the `TSHT` method in [14] to construct such an initial set and a general requirement for $\widehat{\mathcal{V}}$ is stated in (32).

We further propose a novel *sampling* method to construct uniform CIs. Conditioning on the observed data, we repeatedly sample the reduced form estimators centering at the reduced form estimates (constructed from the observed data); see equation (21). We show in Proposition 1 that, after sampling sufficiently many times, there exists one sampled reduced form estimator converging to the corresponding true parameters at a rate faster than the regular parametric rate. With this property, we propose to reduce thresholding levels that are used to decide the instruments' validity for a given β value. For each sampled reduced form estimator, we construct a searching interval for β which gives enough number of valid instruments. We then take a union of the intervals across all sampled estimators.

The proposed CIs are computationally efficient as the searching method searches over a subset of the one-dimension space and we sample the reduced form estimators instead of the observed data.

One interesting observation is that our proposed sampling idea is useful in producing shorter CIs than the regular searching idea. This happens due to the fact that the decreased thresholds lead to a large proportions of searching intervals based on the sampled reduced form estimators being empty.

Our proposed searching and sampling CIs are shown to achieve the desired coverage under the finite-sample majority or plurality rule. The CIs are uniformly valid in the sense that the coverage is guaranteed even if invalid instruments only violate (A2) and (A3) mildly. We conduct a large set of simulation studies to compare our proposed methods with existing

CI construction methods: `TSHT` [14], `CIIV` [34] and `Union` method [18].

To sum-up, the contribution of the current paper is two-folded.

1. We illustrate the post-selection problem of inference after instrument selection and introduce finite-sample identifiability conditions.

2. We propose novel searching and sampling methods to construct uniform CIs for the treatment effect when the candidate IVs are possibly invalid. These intervals are robust to the existence of weakly invalid IVs, which are hard to detect in finite-sample.

The current paper is focused on the low-dimension with homoscedastic regression errors. The proposed methods can be generalized to handle summary statistics, heteroscedastic errors and high-dimensional covariates and instruments; see Section 2.3 for further discussion.

1.2 Literature Comparison

The most relevant papers to the current work are [14, 34], who propose data-dependent IV selection methods and construct CIs with the selected IVs. The validity of the CIs in [14, 34] relies on the condition that the invalid instruments are correctly identified. However, it is challenging to have a perfect separation between valid and invalid IVs in practice, especially in the presence of weakly invalid instruments. In Section 7, it is observed that the CIs by [14, 34] are under-coverage when the sample size is not large enough or there exist weakly invalid IVs. In contrast, in these challenging settings, our proposed searching and sampling CIs achieve the desired coverage level; see tables 3 and 5 for details.

Another closely related work [18] propose to constructing uniform CI for the effect by taking a union of CIs by Two Stage Least Squares with a given number of candidate IVs and not rejected by the Sargan test [16, 27]. The number of invalid IVs is required for the construction and is taken as a sensitivity parameter. Our proposed uniformly valid CIs based on searching and sampling are computationally more efficient than [18], which searches over a large number of sub-models; see Tables D.13 and D.14 in the supplementary material. As illustrated in Table 3, the CIs in [18] assuming two valid IVs achieve the desired coverage properties across all settings. However, they are typically much longer than our proposed sampling and searching CIs; see Tables 4 and 5 for details.

When the classical IV assumptions (A2) and (A3) fail to hold, different identifiability conditions have been proposed to identify the causal effect. The paper [4] and [20] consider inference for the treatment effect under the condition that the direct effect of the instruments on the outcome and the association between the treatment and the instruments are orthogonal. The majority rule (more than 50% of the IVs are valid) are applied to identify the treatment effect in both linear outcome model [5, 15, 19, 33] and nonlinear outcome model [25]. [29] propose to identify the treatment effect by requiring the interaction of the candidate instruments and an environmental factor to satisfy the classical IV assumptions (A1)-(A3). In the presence of invalid instruments, [23, 24, 30] leverage the heteroscedastic covariance restriction to identify the treatment effect.

Construction of uniformly valid CIs after model selection has been a major focus in statistics, under the name of post-selection inference. Many useful methods [2, 7, 8, 17, 21, 22,31,36] have been proposed and the focus is on (but not limited to) inference for regression coefficients after some variables or sub-models are selected. In this paper, we consider a different problem, post-selection inference for causal effect with possibly invalid instruments. To our best knowledge, this problem has not been carefully investigated in the post-selection inference literature. Furthermore, our proposed sampling method is different from other existing post-selection inference methods. The sampling method can be of independent interest and find applications in other post-selection inference problems.

Paper Structure. In Section 2, we introduce the model set-up and the reduced form estimators. In Section 3, we illustrate the post-selection problem. In Section 4, we propose the searching and sampling CIs under the majority rule; In Section 5, we extend the methods to the plurality rule. The theoretical justification is provided in Section 6. In Section 7, we conduct a large set of simulation studies. In Section 8, our proposed methods are applied to to a stock mouse data set to study the effect of the triglyceride level on the glucose level. The proofs are presented in Sections B and C in the supplementary material.

Notations. For a set S and a vector $x \in \mathbb{R}^p$, $|S|$ denotes the cardinality of S and x_S is the sub-vector of x with indices in S. The ℓ_q norm of a vector x is defined as $\|x\|_q = \left(\sum_{l=1}^{p} |x_l|^q\right)^{\frac{1}{q}}$ for $q \geq 0$ with $\|x\|_0 = |\{1 \leq l \leq p : x_l \neq 0\}|$ and $\|x\|_\infty = \max_{1 \leq l \leq p} |x_l|$. We use $\mathbf{0}_q$ and $\mathbf{1}_q$ to denote the q-dimension vector with all entries equal to 0 and 1, respectively. For a matrix X, $X_{i\cdot}$, $X_{\cdot j}$ and X_{ij} are used to denote its i-th row, j-th column and (i, j) entry. For a sequence of random variables X_n indexed by n, we use $X_n \xrightarrow{d} X$ to represent that X_n converges to X in distribution. We use c and C to denote generic positive constants that may vary from place to place. For two positive sequences a_n and b_n, $a_n \lesssim b_n$ means that $\exists C > 0$ such that $a_n \leq C b_n$ for all n; $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$, and $a_n \ll b_n$ if $\limsup_{n\to\infty} a_n/b_n = 0$. For a matrix A, we use $\|A\|_2$ and $\|A\|_\infty$ to denote its spectral norm and element-wise maximum norm, respectively. For a symmetric matrix A, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its maximum and minimum eigenvalues, respectively.

2 Statistical Modeling and Reduced Form Estimators

We introduce the model set-up in Section 2.1 and review the existing identifiability conditions in Section 2.2. In Section 2.3, we introduce the reduced form estimators.

2.1 Models with Invalid IVs

We consider the i.i.d. data $\{Y_i, D_i, X_{i\cdot}, Z_{i\cdot}\}_{1\leq i\leq n}$, where $Y_i \in \mathbb{R}$, $D_i \in \mathbb{R}$ and $X_{i\cdot} \in \mathbb{R}^{p_{\mathrm{x}}}$ and $Z_{i\cdot} \in \mathbb{R}^{p_{\mathrm{z}}}$ denote the outcome, the treatment, the baseline covariates, and candidate instruments (or IVs), respectively. For two possible treatment values $d \in \mathbb{R}$ and $d' \in \mathbb{R}$ and two possible realizations of IVs $\mathbf{z} \in \mathbb{R}^{p_{\mathrm{z}}}$ and $\mathbf{z}' \in \mathbb{R}^{p_{\mathrm{z}}}$, we define the following potential outcome model:

$$Y^{(d',\mathbf{z}')} - Y^{(d,\mathbf{z})} = (d'-d)\beta^* + (\mathbf{z}'-\mathbf{z})^{\intercal}\kappa^* \quad \text{and} \quad \mathbf{E}\left(Y^{(0,\mathbf{0})} \mid Z_{i\cdot}, X_{i\cdot}\right) = Z_{i\cdot}^{\intercal}\eta^* + X_{i\cdot}^{\intercal}\phi^* \tag{1}$$

where $\beta^* \in \mathbb{R}$ is the treatment effect, $\kappa^*, \eta^* \in \mathbb{R}^{p_{\mathrm{z}}}$ and $\phi^* \in \mathbb{R}^{p_{\mathrm{x}}}$. Define $e_i = Y^{(0,\mathbf{0})} - \mathbf{E}\left(Y^{(0,\mathbf{0})} \mid Z_{i\cdot}, X_{i\cdot}\right)$. Under the consistency condition $Y_i = Y_i^{D_i, Z_{i\cdot}}$, the potential outcome model (1) implies the following observed outcome model:

$$Y_i = D_i\beta^* + Z_{i\cdot}^{\intercal}\pi^* + X_{i\cdot}^{\intercal}\phi^* + e_i \quad \text{with} \quad \pi^* = \kappa^* + \eta^* \in \mathbb{R}^{p_{\mathrm{z}}} \quad \text{and} \quad \mathbf{E}(e_i \mid Z_{i\cdot}, X_{i\cdot}) = 0. \tag{2}$$

The models (1) and (2) have been considered in [14,18,19,28,34]. As illustrated in Figure 1, $\kappa_j^* \neq 0$ in the model (1) indicates that the j-th candidate IV has a direct effect on outcome, which violates the assumption (A3); $\eta_j \neq 0$ in the model (1) indicates that the j-th candidate IV is associated with the unmeasured confounders, which violates the assumption (A2). The vector $\pi^* = \kappa^* + \eta^*$ characterizes the invalidity of the candidate IVs $Z_{i\cdot} \in \mathbb{R}^{p_{\mathrm{z}}}$.

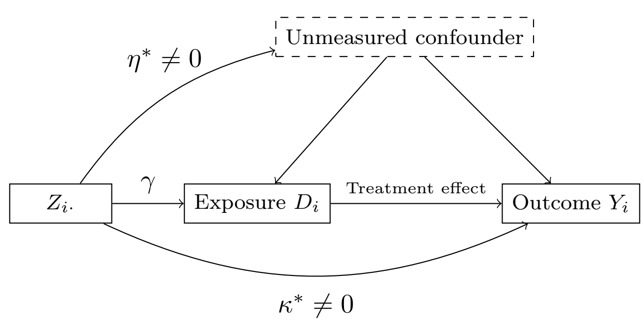

Figure 1: Illustration of violations of (A2) and (A3) in the model (1) or (2).

We consider the association model between D_i and $Z_{i\cdot}, X_{i\cdot}$

$$D_i = Z_{i\cdot}^{\intercal}\gamma^* + X_{i\cdot}^{\intercal}\phi^* + \delta_i \quad \text{with} \quad \mathbf{E}(\delta_i Z_{i\cdot}) = \mathbf{0} \quad \text{and} \quad \mathbf{E}(\delta_i X_{i\cdot}) = \mathbf{0}. \tag{3}$$

Here, $\gamma_j^* \neq 0$ indicates that the j-th IV is associated with the treatment conditioning on baseline covariates, that is, satisfying assumption (A1). We define the sets of instruments,

$$\mathcal{S} = \{j : \gamma_j^* \neq 0\} \quad \text{and} \quad \mathcal{V} = \{j \in \mathcal{S} : \pi_j^* = 0\}. \tag{4}$$

$\mathcal{S}$ denotes the set of relevant instruments, $\mathcal{V}$ denotes the set of valid instruments and $\mathcal{I} = \{j \in \mathcal{S} : \pi_j^* \neq 0\} = \mathcal{S}\backslash\mathcal{V}$ denotes the set of invalid instruments.

We estimate the set $\mathcal{S}$ by conducting significance tests for $\gamma_j^* = 0$; see (12). Since there is no prior information on the set $\mathcal{V}$, additional identifiability conditions are required to identify $\mathcal{V}$ and hence β^*. The existing conditions are reviewed in the next subsection.

2.2 Population Identifiability Conditions

The identification of β is impossible under models (2) and (3) without further structural assumptions on the invalidity vector π^* and the IV strength vector γ^* [14, 19, 20]. We now review existing identifiability conditions and start with the majority rule.

Population Majority Rule: More than half of the relevant IVs are valid instruments, that is,

$$|\mathcal{V}| > |\mathcal{S}|/2. \tag{5}$$

Majority rule requires that the majority of the relevant IVs are valid but does not directly require the knowledge of the set $\mathcal{V}$. $|\mathcal{V}| > |\mathcal{S}|/2$ is equivalent to $|\mathcal{V}| > |\mathcal{I}|$. The following plurality rule has been proposed in [14] to weaken the majority rule in (5).

Population Plurality Rule: The number of valid IVs is larger than the number of invalid IVs with any given invalidity level $\nu \neq 0$, that is,

$$|\mathcal{V}| > \max_{\nu \neq 0} |\mathcal{I}_\nu| \quad \text{with} \quad \mathcal{I}_\nu = \left\{ j \in \mathcal{S} : \pi_j^*/\gamma_j^* = \nu \right\}. \tag{6}$$

For the j-th IV, π_j^*/γ_j^* represents its invalidity level: $\pi_j^*/\gamma_j^* = 0$ indicates indicates that the j-th IV satisfies assumptions (A2) and (A3); a small but non-zero $|\pi_j^*/\gamma_j^*|$ indicates that the j-th IV weakly violates assumptions (A2) and (A3); a large $|\pi_j^*/\gamma_j^*|$ indicates that the j-th IV strongly violates assumptions (A2) and (A3). For $\nu \in \mathbb{R}$, $\mathcal{I}_\nu$ denotes the set of all IVs with the same invalidity level ν. The plurality rule requires that the number of valid IVs is larger than the number of invalid IVs with any level $\nu \neq 0$.

The majority rule in (5) and the plurality rule in (6) are referred to as *population identifiability conditions* since they are used to identify β with an infinite (or at least a very large) amount of data. These *population identifiability conditions* may not work in practical applications with only a finite amount of data. In Section 3, we demonstrate that, even these population identifiability conditions hold, the existing inference procedures can produce unreliable CIs with a finite amount of data.

2.3 Problem Formulation with Reduced-form Estimators

We combine models (2) and (3) and establish the following reduced form models

$$\begin{aligned} Y_i = &\quad Z_{i\cdot}^\intercal \Gamma^* + X_{i\cdot}^\intercal \Psi^* + \epsilon_i \quad \text{with} \quad \mathbf{E}(Z_{i\cdot}\epsilon_i) = \mathbf{0},\ \mathbf{E}(X_{i\cdot}\epsilon_i) = \mathbf{0}, \\ D_i = &\quad Z_{i\cdot}^\intercal \gamma^* + X_{i\cdot}^\intercal \psi^* + \delta_i \quad \text{with} \quad \mathbf{E}(Z_{i\cdot}\delta_i) = \mathbf{0},\ \mathbf{E}(X_{i\cdot}\delta_i) = \mathbf{0}, \end{aligned} \tag{7}$$

where $\Gamma^* = \beta^*\gamma^* + \pi^* \in \mathbb{R}^{p_z}$, $\Psi^* = \beta^*\phi^* + \psi^* \in \mathbb{R}^{p_x}$ and $\epsilon_i = \beta^*\delta_i + e_i$.

The parameters $\Gamma^* \in \mathbb{R}^{p_z}$ and $\gamma^* \in \mathbb{R}^{p_z}$ in the reduced-form model (7) can be consistently estimated since $Z_{i\cdot}$ and $X_{i\cdot}$ are uncorrrelated with the model errors ϵ_i and δ_i. We construct the reduced form estimator $\widehat{\Gamma}$ and $\widehat{\gamma}$ satisfying

$$\sqrt{n}\begin{pmatrix}\widehat{\Gamma}-\Gamma^* \\ \widehat{\gamma}-\gamma^*\end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \mathrm{Cov}\right) \quad \text{with} \quad \mathrm{Cov} = \begin{pmatrix}\mathbf{V}^{\Gamma} & \mathbf{C} \\ \mathbf{C}^{\intercal} & \mathbf{V}^{\gamma}\end{pmatrix} \tag{8}$$

where $\mathbf{V}^{\Gamma} \in \mathbb{R}^{p_z \times p_z}$, $\mathbf{V}^{\gamma} \in \mathbb{R}^{p_z \times p_z}$ and $\mathbf{C} \in \mathbb{R}^{p_z \times p_z}$. In the low-dimensional setting, we construct $\widehat{\Gamma}$ and $\widehat{\gamma}$ by the Ordinary Least Squares (OLS):

$$(\widehat{\Gamma}, \widehat{\Psi})^{\intercal} = (W^{\intercal}W)^{-1}W^{\intercal}Y \quad \text{and} \quad (\widehat{\gamma}, \widehat{\psi})^{\intercal} = (W^{\intercal}W)^{-1}W^{\intercal}D, \tag{9}$$

where $W = (Z, X) \in \mathbb{R}^{n \times p}$ with $p = p_x + p_z$. We estimate $\mathrm{Var}(\epsilon_i)$, $\mathrm{Var}(\delta_i)$ and $\mathrm{Cov}(\epsilon_i, \delta_i)$ by

$$\begin{aligned} \widehat{\sigma}_{\epsilon}^2 &= \|Y - Z\widehat{\Gamma} - X\widehat{\Psi}\|_2^2/(n-1), \quad \widehat{\sigma}_{\delta}^2 = \|D - Z\widehat{\gamma} - X\widehat{\psi}\|_2^2/(n-1), \\ \widehat{\sigma}_{\epsilon,\delta} &= (Y - Z\widehat{\Gamma} - X\widehat{\Psi})^{\intercal}(D - Z\widehat{\gamma} - X\widehat{\psi})/(n-1). \end{aligned} \tag{10}$$

We further estimate $\mathbf{V}^{\Gamma} \in \mathbb{R}^{p_z \times p_z}$, $\mathbf{V}^{\gamma} \in \mathbb{R}^{p_z \times p_z}$ and $\mathbf{C} \in \mathbb{R}^{p_z \times p_z}$ in (8) respectively by

$$\widehat{\mathbf{V}}^{\Gamma} = \widehat{\sigma}_{\epsilon}^2\widehat{\Omega}/n, \; \widehat{\mathbf{V}}^{\gamma} = \widehat{\sigma}_{\delta}^2\widehat{\Omega}/n, \; \widehat{\mathbf{C}} = \widehat{\sigma}_{\epsilon,\delta}\widehat{\Omega}/n \quad \text{where} \quad \widehat{\Omega} = [(W^{\intercal}W/n)^{-1}]_{1:p_z, 1:p_z}. \tag{11}$$

Several important extensions on constructing $\widehat{\Gamma}$ and $\widehat{\gamma}$ are discussed at the end of this section.

With $\widehat{\gamma}$ defined in (9), we estimate the set $\mathcal{S}$ defined in (4) by

$$\widehat{\mathcal{S}} = \left\{1 \leq j \leq p : |\widehat{\gamma}_j| \geq \sqrt{\log n} \cdot \sqrt{\widehat{\mathbf{V}}^{\gamma}_{jj}/n}\right\} \tag{12}$$

where $\widehat{\mathbf{V}}^{\gamma}_{jj}$ is defined in (11) and the term $\sqrt{\log n}$ is introduced to adjust for multiplicity.

Identification of β relies on the reduced form equation $\Gamma^* = \beta^*\gamma^* + \pi^*$. We have the following data-dependent reduced form equation

$$\beta \cdot \widehat{\gamma}_j + \pi_j \approx \widehat{\Gamma}_j \quad \text{for} \quad j \in \widehat{\mathcal{S}}.$$

Since there are $|\widehat{\mathcal{S}}| + 1$ parameters in $|\widehat{\mathcal{S}}|$ equations, we need the identifiability conditions (e.g. majority or plurality rule in Section 2.2) to identify $\beta \in \mathbb{R}$.

We now discuss a few important extentions about construction of the reduced form estimators satisfying (8). Firstly, $\widehat{\Gamma}$ and $\widehat{\gamma}$ in (9) can be calculated with summary statistics. In medical applications (e.g. mendelian randomization), there are constraints on sharing the raw data. The implementation of (9) and (11) relies on the summary statistics $W^{\intercal}W$,$W^{\intercal}Y$ and $W^{\intercal}D$ together with the noise level estimates $\widehat{\sigma}_{\epsilon}^2$, $\widehat{\sigma}_{\delta}^2$ and $\widehat{\sigma}_{\epsilon,\delta}$.

Secondly, even with OLS reduced form estimators in (9), the corresponding variance covariance matrices $\mathbf{V}^{\Gamma}$ and $\mathbf{V}^{\gamma}$ in (11) only hold by assuming the homoscedastic regression error assumption $\mathbf{E}(e_i^2 \mid Z_{i\cdot}, X_{i\cdot}) = \sigma_e^2$ and $\mathbf{E}(\delta_i^2 \mid Z_{i\cdot}, X_{i\cdot}) = \sigma_\delta^2$. If this assumption does not hold, we can adopt the robust variance and covariance estimator [35]

$$\widehat{\mathbf{V}}^{\Gamma} = \left[(W^{\intercal}W)^{-1} \left(\sum_{i=1}^{n} \widehat{u}_i^2 W_{i\cdot} W_{i\cdot}^{\intercal} \right) (W^{\intercal}W)^{-1} \right]_{1:p_z, 1:p_z} \tag{13}$$

where $\widehat{u}_i = Y_i - Z_{i\cdot}^{\intercal}\widehat{\Gamma} - X_{i\cdot}^{\intercal}\widehat{\Psi}$ for $1 \leq i \leq n$. Similarly, $\widehat{\mathbf{V}}^{\gamma}$ can be computed using the same formula in (13) with $\widehat{u}_i = D_i - Z_{i\cdot}^{\intercal}\widehat{\gamma} - X_{i\cdot}^{\intercal}\widehat{\psi}$ for $1 \leq i \leq n$.

Thirdly, we can construct reduced form estimators $\widehat{\Gamma}$ and $\widehat{\gamma}$ satisfying (8) in high-dimensional setting where $p > n$. We may apply the debiased lasso estimators [17, 31, 36] or the orthogonal estimating equations estimator [8]. A detailed discussion about obtaining such reduced form estimators can be found in Section 4.1 of [13].

3 Post-selection Problem

In this section, we demonstrate the post-selection problem of inference for β^* with invalid IVs. The existing CI construction methods, `TSHT` [14] and `CIIV` [34], first select valid IVs and then construct CIs for β^* with the selected IVs. The validity of these methods relies on the consistency of model (or valid IV) selection, that is, the set $\mathcal{V}$ of valid IVs defined in (4) has been correctly selected. However, in finite samples, it is challenging to achieve this if some invalid instruments only violate assumptions (A2) and (A3) mildly.

We define the index set $\mathcal{I}(0, \tau_n) = \{j \in \mathcal{S} : |\pi_j^*/\gamma_j^*| \leq \tau_n\}$. For $\tau_n \asymp \sqrt{\log n/n}$, the set $\mathcal{I}(0, \tau_n)$ consists of the set $\mathcal{V}$ of valid IVs and the set of weakly invalid IVs:

$$\mathcal{I}(0, \tau_n) \backslash \mathcal{V} = \left\{ j \in \mathcal{S} : 0 < \left| \pi_j^*/\gamma_j^* \right| \leq \tau_n \right\}.$$

The definition of weakly invalid instruments $\mathcal{I}(0, \tau_n)\backslash\mathcal{V}$ depends both on the invalidity level π_j^*/γ_j^* and the sample size n (via $\tau_n \asymp \sqrt{\log n/n}$). In the favorable setting with a very large n, the set $\mathcal{I}(0, \tau_n)\backslash\mathcal{V}$ becomes empty and none of the invalid instruments will be weakly invalid.

The estimated sets $\widetilde{\mathcal{V}}$ of valid IVs by `TSHT` and `CIIV` satisfy $\widetilde{\mathcal{V}} \subset \mathcal{I}(0, C\sqrt{\log n/n})$ for some positive constant $C > 0$. There are chances that the weakly invalid IVs are included in the estimated sets $\widetilde{\mathcal{V}}$. If the set $\widetilde{\mathcal{V}}$ includes the weakly violated IVs, the resulting estimators of β using $\widetilde{\mathcal{V}}$ are biased. The theoretical justifications of `TSHT` [14] or `CIIV` [34] rely on the absence of weakly invalid IVs, that is, $\min_{\pi_j^*/\gamma_j^* \neq 0} |\pi_j^*/\gamma_j^*| \geq C\sqrt{\log n/n}$ for some $C > 0$. With this well-separation condition, the set $\widetilde{\mathcal{V}}$ constructed by `TSHT` and `CIIV` will achieve the selection consistency $\widetilde{\mathcal{V}} = \mathcal{V}$. However, the absence of weakly invalid IVs may not properly accommodate for the finite-sample data analysis. The instruments with a small non-zero value $|\pi_j^*/\gamma_j^*|$ are likely to be taken as a valid IV when n is not sufficiently large.

We illustrate this post-selection problem with a numerical example.

Example 1. (Setting **S2** in Section 7.) For the models (2) and (3), set $\gamma^* \in \mathbb{R}^{10}$ with $\gamma_j^* = 0.5$ for $1 \leq j \leq 10$ and $\pi^* = (\mathbf{0}_4, \tau/2, \tau/2, -1/3, -2/3, -1, -4/3)^\intercal$ and vary τ across $\{0.1, 0.2, 0.4\}$. The plurality rule is satisfied in this setting: $|\mathcal{V}| = 4 > \max\{|\mathcal{I}_\tau|, |\mathcal{I}_{-0.5}|\} = 3$.

In Figure 2, we plot the histogram of the `TSHT` and `CIIV` estimators for $\tau = 0.1$ over 500 simulations. The top panel of the plot corresponds to $n = 500$ and the bottom panel to $n = 2000$. The histogram centers round the sample average of the 500 estimates (the dashed line) and deviates from the true value $\beta^* = 1$ (the solid line). This bias results from the post-selection problem, where weakly invalid IVs are selected as valid IVs.

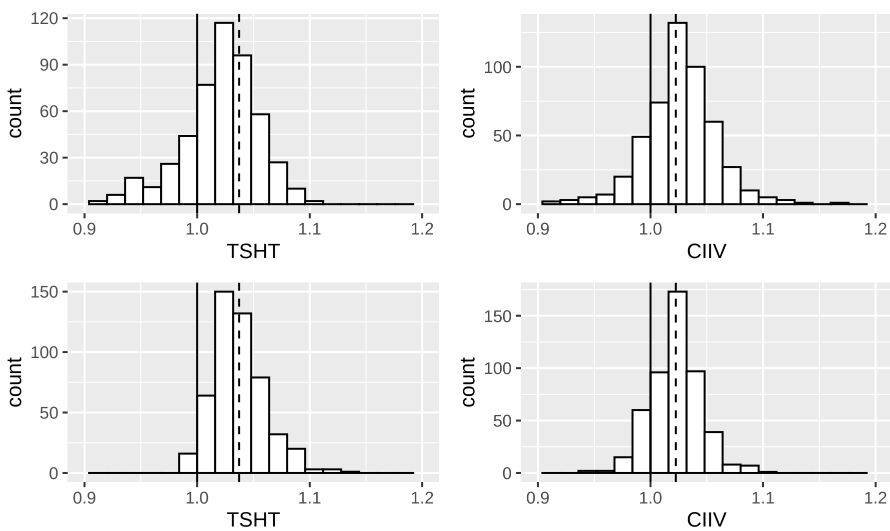

Figure 2: Histogram of 500 `TSHT` and `CIIV` estimates for **Example 1** with $\tau = 0.1$ and $n = 500$ (top panel) and $n = 2000$ (bottom panel).

	$\tau = 0.1$		$\tau = 0.2$		$\tau = 0.5$	
n	`TSHT`	`CIIV`	`TSHT`	`CIIV`	`TSHT`	`CIIV`
500	0.74	0.75	0.41	0.45	0.36	0.74
1000	0.68	0.65	0.24	0.63	0.51	0.90
2000	0.39	0.49	0.18	0.72	0.93	0.95

Table 1: Empirical coverage of `TSHT` and `CIIV` estimators for **Example 1**.

The empirical coverage properties are reported in Table 1. The coverage levels of `TSHT` and `CIIV` are below the nominal level 95% for $\tau = 0.1$ and 0.2, which correspond to the existence of weakly invalid instruments; in contrast, for $\tau = 0.4$, `TSHT` achieves the correct coverage for $n \geq 2000$ and `CIIV` is effective for $n = 1000$ and 2000. Table D.4 in the supplementary material shows that our proposed CIs achieve the desired coverage level.

4 Uniform Inference Methods under Majority Rule

We first introduce the finite-sample majority rule in Section 4.1 and then devise two novel methods to overcome the post-selection issue under the majority rule.

4.1 Finite-sample Majority Rule

Define the set of strongly relevant IVs as

$$\mathcal{S}_{\rm str} = \left\{1 \leq j \leq p_{\rm z} : \left|\gamma_j^*\right| \geq 2\sqrt{\log n} \cdot \sqrt{\mathbf{V}^{\gamma}_{jj}}\right\}. \tag{14}$$

Note that $\sqrt{\mathbf{V}^{\gamma}_{jj}}$ is of order $1/\sqrt{n}$ and $\mathcal{S}_{\rm str}$ is a subset of relevant IVs $\mathcal{S}$ with the individual IV strength above $\sqrt{\log n/n}$. For a sufficiently large sample size, any IV with a fixed individual strength belongs to $\mathcal{S}_{\rm str}$ as $\sqrt{\log n/n}$ diminishes to zero.

Now we introduce the following finite-sample identifiability conditions.

Condition 1 (Finite-sample Majority Rule) *More than half of the relevant IVs are strongly relevant and valid, that is,*

$$|\mathcal{V} \cap \mathcal{S}_{\rm str}| > |\mathcal{S}|/2. \tag{15}$$

where $\mathcal{S}$ and $\mathcal{V}$ are defined in (4) *and $\mathcal{S}_{\rm str}$ is defined in* (14).

For applications with a relatively small n, Condition 1 is more meaningful than the population majority rule in (5). When $n \to \infty$ and the IV strengths $\{\gamma_j^*\}_{1\leq j\leq p_{\rm z}}$ do not grow with n, Condition 1 is reduced to the population majority rule in (5) since $\mathcal{S}_{\rm str}$ converges to $\mathcal{S}$.

4.2 The Searching Method

We propose a searching method to make inference for β^* under Condition 1. Define $\mathcal{B} = \{\beta_1, \beta_2, \cdots, \beta_K\} \subset \mathbb{R}$ as a candidate set for β^*. We construct $\mathcal{B}$ as a grid set between two constants A_1 and A_2 with the grid size $1/n^a$ for some constant $a > 1/2$. The true β^* might not be contained in the set of values $\mathcal{B}$ but our construction guarantees that there exists $\beta_k \in \mathcal{B}$ such that $|\beta_k - \beta^*| < 1/n^a$. The value a can be set as any value larger than 0.5 and our default choice is $a = 0.6$. In Section 4.4, we present the construction of A_1 and A_2.

Let $\widehat{\Gamma}$ and $\widehat{\gamma}$ denote the reduced form estimators satisfying (8), e.g., OLS estimators in (9). For any $\beta \in \mathbb{R}$ and $1 \leq j \leq p_{\mathrm{z}}$, we apply the reduced-form formula $\beta \cdot \widehat{\gamma}_j + \pi_j \approx \widehat{\Gamma}_j$ for $j \in \widehat{\mathcal{S}}$ and use $\widehat{\Gamma}_j - \beta\widehat{\gamma}_j$ as an estimate of π_j^* :

$$\widehat{\Gamma}_j - \beta\widehat{\gamma}_j = \widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*) + (\beta^* - \beta)\gamma_j^* + \pi_j^*. \tag{16}$$

With $\widehat{\mathcal{S}}$ defined in (4), we quantify the uncertainty of $\{\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)\}_{j \in \widehat{\mathcal{S}}, \beta \in \mathcal{B}}$ in (16) through choosing a threshold $\widehat{\rho}(\alpha) > 0$ satisfying

$$\mathbb{P}\left(\max_{\beta \in \mathcal{B}} \max_{j \in \widehat{\mathcal{S}}} \frac{|\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2\widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta\widehat{\mathbf{C}}_{jj})/n}} \leq \widehat{\rho}(\alpha)\right) \geq 1 - \alpha, \tag{17}$$

where $\widehat{\mathbf{V}}^{\Gamma}$, $\widehat{\mathbf{V}}^{\gamma}$ and $\widehat{\mathbf{C}}$ are consistent estimators of $\mathbf{V}^{\Gamma}, \mathbf{V}^{\gamma}$ and $\mathbf{C}$, respectively. For the OLS estimators in low dimensions, $\widehat{\mathbf{V}}^{\Gamma}$, $\widehat{\mathbf{V}}^{\gamma}$ and $\widehat{\mathbf{C}}$ are defined in (11). For theoretical purpose, we may choose $\widehat{\rho}(\alpha) \asymp \sqrt{\log|\mathcal{B}|}$ or $\widehat{\rho}(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{|\mathcal{B}| \cdot p_{\mathrm{z}}}\right)$, where Φ^{-1} is the inverse of the cumulative distribution function of the standard normal. In Section 4.5, we provide an alternative bootstrap method to choose $\widehat{\rho}(\alpha)$.

For any given $\beta \in \mathbb{R}$ and $\widehat{\rho}(\alpha)$ satisfying (17), we define the re-scaled threshold,

$$\widehat{\rho}_j(\beta, \alpha) = \widehat{\rho}(\alpha) \cdot \sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2\widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta\widehat{\mathbf{C}}_{jj})/n}, \tag{18}$$

and estimate π^* by applying hard-thresholding to $\widehat{\Gamma}_{\widehat{\mathcal{S}}} - \beta\widehat{\gamma}_{\widehat{\mathcal{S}}}$:

$$\widehat{\pi}_j(\beta) = \left(\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right) \cdot \mathbf{1}\left(\left|\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right| \geq \widehat{\rho}_j(\beta, \alpha)\right) \quad \text{for} \quad j \in \widehat{\mathcal{S}}. \tag{19}$$

For a specific value β, we use the sparsity of $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta) = (\widehat{\pi}_j(\beta))_{j \in \widehat{\mathcal{S}}}$ as our confidence about this β. If more than $|\widehat{\mathcal{S}}|/2$ entries of $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)$ are zero, the corresponding β is believed to pass the majority rule and is included in our constructed interval. Specifically, we construct the searching CI for β as

$$\mathrm{CI}^{\mathrm{search}} = \left(\min_{\{\beta \in \mathcal{B}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2\}} \beta, \quad \max_{\{\beta \in \mathcal{B}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2\}} \beta\right). \tag{20}$$

where $\|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0$ denotes the number of non-zeros in $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)$.

In construction of $\mathrm{CI}^{\mathrm{search}}$ in (20), we search for the smallest $\beta \in \mathcal{B}$ and largest $\beta \in \mathcal{B}$ such that $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)$ is sparse enough. When the majority is violated, then, with a high probability, there is no β such that $\|\widehat{\pi}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2$ and hence $\mathrm{CI}^{\mathrm{search}}$ is empty. This indicates that the majority rule is violated, which can be used as a partial check of the majority rule.

We illustrate the construction of $\mathrm{CI}^{\mathrm{search}}$ with the following example.

Example 2. Generate the models (2) and (3) with no baseline covariates, set $\beta^* = 1$,

$n = 2000$, $\gamma^* \in \mathbb{R}^{10}$ with $\gamma_j^* = 0.5$ for $1 \leq j \leq 10$ and $\pi^* = (\mathbf{0}_6, 0.05, 0.05, -0.5, -1)^\intercal$. The majority rule is satisfied in this setting with $|\mathcal{V}| = 6 > 5$.

In Figure 3, over all $\beta \in \mathcal{B}$, we plot the number of zero entries of the vector $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)$ defined in (19). The constructed $\mathrm{CI}^{\mathrm{search}}$ is $(0.911, 1.118)$, which contains the true value 1.

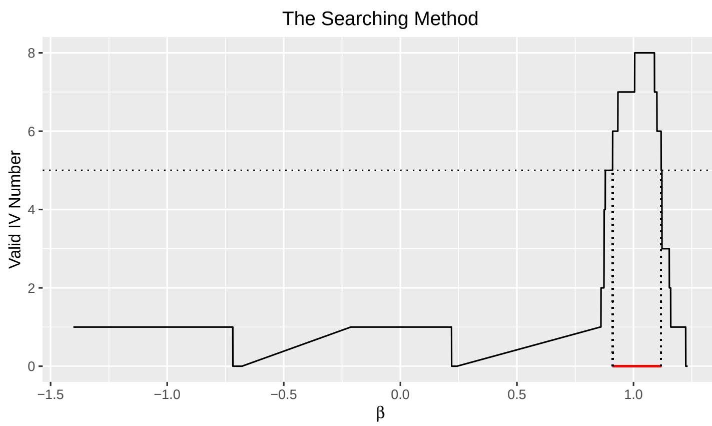

Figure 3: The x-axis are the values of β belonging to the set $\mathcal{B}$, and the y-axis represents the number of zero entries of the thresholded estimator $\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)$ in (19). The red interval $(0.911, 1.118)$ is the interval constructed by the searching method.

As a remark, the proposed searching idea is related to the Anderson-Rubin test [1] for weak IV problem. The key idea of Anderson-Rubin test is to search for β by inverting a χ^2 test statistic [26]. Our proposed method in (20) uses the sparsity as the test statistic and is proposed to address the invalid IV problem, which is different from the weak IV problem.

4.3 The Sampling Method

We now propose another sampling idea and together with the searching method, this can lead to more precision CIs. Conditioning on the reduced form estimators $\widehat{\gamma}$ and $\widehat{\Gamma}$, we sample $\{\widehat{\Gamma}^{[m]}, \widehat{\gamma}^{[m]}\}_{1 \leq m \leq M}$ following

$$\begin{pmatrix} \widehat{\Gamma}^{[m]} \\ \widehat{\gamma}^{[m]} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N\left[\begin{pmatrix} \widehat{\Gamma} \\ \widehat{\gamma} \end{pmatrix}, \begin{pmatrix} \widehat{\mathbf{V}}^{\Gamma}/n & \widehat{\mathbf{C}}/n \\ \widehat{\mathbf{C}}^\intercal/n & \widehat{\mathbf{V}}^{\gamma}/n \end{pmatrix}\right] \quad \text{for} \quad 1 \leq m \leq M, \tag{21}$$

where the sampling size M is a positive integer, $\widehat{\Gamma}$ and $\widehat{\gamma}$ are defined in (9) and $\widehat{\mathbf{V}}^{\Gamma}$, $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{V}}^{\gamma}$ are defined in (11). Our proposed method depends on the following sampling property (see the exact statement in Proposition 1): with a high probability, for a sufficiently large M, there exists $1 \leq m^* \leq M$ such that

$$\max\left\{\left\|\widehat{\gamma}^{[m^*]} - \gamma^*\right\|_\infty, \left\|\widehat{\Gamma}^{[m^*]} - \Gamma^*\right\|_\infty\right\} \lesssim \lambda \cdot \frac{1}{\sqrt{n}} \quad \text{where} \quad \lambda \asymp \left(\frac{\log n}{M}\right)^{\frac{1}{2p_z}}. \tag{22}$$

This sampling property states that, with a good chance, after sampling many times, one of the sampled estimators $\widehat{\gamma}^{[m^*]}$ (or $\widehat{\Gamma}^{[m^*]}$) converges to the truth γ^* (or Γ^*) at a rate faster than $1/\sqrt{n}$. The sampling property in (22), together with the searching method, can be used to overcome the post-selection issue. For each $1 \leq m \leq M$, we define the sampled version of the thresholding step in (19),

$$\widehat{\pi}_j^{[m]}(\beta, \lambda) = \left(\widehat{\Gamma}_j^{[m]} - \beta \widehat{\gamma}_j^{[m]}\right) \cdot \mathbf{1}\left(\left|\widehat{\Gamma}_j^{[m]} - \beta \widehat{\gamma}_j^{[m]}\right| \geq \lambda \cdot \widehat{\rho}_j(\beta, \alpha)\right) \quad \text{for} \quad 1 \leq j \leq |\widehat{\mathcal{S}}| \tag{23}$$

where λ is defined in (22) and $\widehat{\rho}_j(\beta, \alpha)$ is defined in (18). In contrast to (19), (23) shrinks the thresholding level by $\lambda \asymp (\log n/M)^{\frac{1}{2p_z}}$ used in the sampling property.

For $1 \leq m \leq M$, we use $\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m]}(\beta, \lambda) \in \mathbb{R}^{|\widehat{\mathcal{S}}|}$ defined in (23) to search for β:

$$\beta_{\min}^{[m]}(\lambda) = \min_{\left\{\beta \in \mathcal{B}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m]}(\beta, \lambda)\|_0 < |\widehat{\mathcal{S}}|/2\right\}} \beta \quad \text{and} \quad \beta_{\max}^{[m]}(\lambda) = \max_{\left\{\beta \in \mathcal{B}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m]}(\beta, \lambda)\|_0 < |\widehat{\mathcal{S}}|/2\right\}} \beta. \tag{24}$$

If there is no β such that $\|\widehat{\pi}^{[m]}(\beta, \lambda)\|_0 < |\widehat{\mathcal{S}}|/2$ for a given sampled reduced form estimators $\widehat{\Gamma}^{[m]}$ and $\widehat{\gamma}^{[m]}$, we set $(\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda)) = \varnothing$. Define

$$\mathcal{M} = \{1 \leq m \leq M : (\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda)) \neq \varnothing\} \tag{25}$$

and take a union of all sampled CIs with indexes in $\mathcal{M}$:

$$\mathrm{CI}^{\text{sample}} = \left(\min_{m \in \mathcal{M}} \beta_{\min}^{[m]}(\lambda), \max_{m \in \mathcal{M}} \beta_{\max}^{[m]}(\lambda)\right). \tag{26}$$

In Figure 4, we demonstrate the sampling method using Example 2. 52 of $M = 1000$ intervals are non-empty and 8 of them contain $\beta^* = 1$. The red interval $\mathrm{CI}^{\text{sample}} = (0.929, 1.117)$ contains $\beta^* = 1$.

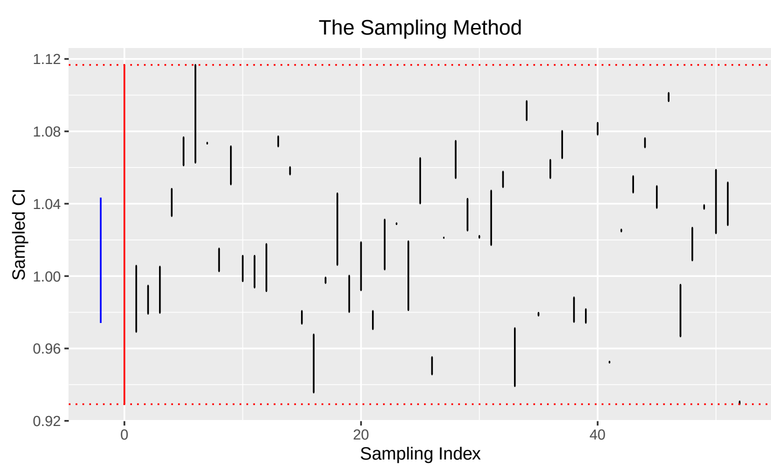

Figure 4: The axis corresponds to different sampling indexes $\{1, 2, \cdots, 52\}$ (after re-ordering) and the y-axis reports the sampled CIs. Along the y-axis, the red interval is $\mathrm{CI}^{\text{sample}} = (0.929, 1.117)$ and the blue interval is the oracle CI $(0.974, 1.043)$ with prior information on valid IVs.

A few remarks are in order for the sampling method. First, even though both $\mathrm{CI}^{\text{search}}$ and $\mathrm{CI}^{\text{sample}}$ achieve the desired coverage level, $\mathrm{CI}^{\text{sample}}$ can further reduce the length of $\mathrm{CI}^{\text{search}}$; see Tables 4 and 5 for details. This happens due to the fact that many of $\{(\beta^{[m]}_{\min}(\lambda), \beta^{[m]}_{\max}(\lambda))\}_{1\leq m\leq M}$ are empty and the non-empty ones are shorter in comparison to the searching interval $\mathrm{CI}^{\text{search}}$.

Second, we may combined the sampled intervals as $\mathrm{CI}^{\text{sampling}}_0 = \cup_{m=1}^{M}(\beta^{[m]}_{\min}(\lambda), \beta^{[m]}_{\max}(\lambda))$. In the current paper, we shall focus on $\mathrm{CI}^{\text{sample}}$ defined in (26) since $\mathrm{CI}^{\text{sampling}}_0$ may lead to a union of disjoint intervals due to its definition. Third, the sampling step results in randomness of the sampling CI in (26). However, any sampling CI from (26) will cover the true β^* with a high probability; see Theorem 2 for the theoretical justification. We can take an intersection of a finite sampling CIs and the coverage property still holds.

4.4 Construction of the Initial Grid Set $\mathcal{B}$

We now discuss how to choose the initial grid set $\mathcal{B} \subset [A_1, A_2]$. For each pair $(\widehat{\gamma}_j, \widehat{\Gamma}_j)$ with $j \in \widehat{\mathcal{S}}$, the ratio $\widehat{\Gamma}_j/\widehat{\gamma}_j$ can be used to estimate β^* with the corresponding variance $\mathrm{Var}(\widehat{\Gamma}_j/\widehat{\gamma}_j)$. For the OLS setting, we have $\mathrm{Var}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j\right) = \left(\frac{\sigma_\epsilon^2}{\gamma_j^2} + \frac{\sigma_\delta^2\gamma_j^2}{\gamma_j^4} - 2\frac{\sigma_{\epsilon,\delta}}{\gamma_j^3}\right)\cdot [(W^\intercal W)^{-1}]_{j,j}$ for $1 \leq j \leq p_{\mathrm{z}}$. We then construct A_1 and A_2 as

$$A_1 = \min_{j\in\widehat{\mathcal{S}}}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j - (\log(n))^{1/4}\sqrt{\cdot\widehat{\mathrm{Var}}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j\right)}\right),\; A_2 = \max_{j\in\widehat{\mathcal{S}}}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j + (\log(n))^{1/4}\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j\right)}\right) \tag{27}$$

where $(\log(n))^{1/4}$ is used to adjust for multiplicity and

$$\widehat{\mathrm{Var}}\left(\widehat{\Gamma}_j/\widehat{\gamma}_j\right) = \left(\frac{\widehat{\sigma}_\epsilon^2}{\widehat{\gamma}_j^2} + \frac{\widehat{\sigma}_\delta^2\widehat{\gamma}_j^2}{\widehat{\gamma}_j^4} - 2\frac{\widehat{\sigma}_{\epsilon,\delta}}{\widehat{\gamma}_j^3}\right)\cdot [(W^\intercal W)^{-1}]_{j,j}.$$

4.5 Bootstrap Approximations of $\widehat{\rho}(\alpha)$

We can approximate $\widehat{\rho}(\alpha)$ defined in (17) through adopting a bootstrap method. Conditioning on the observed data, for $1 \leq l \leq L$ with L being a large integer, we generate

$$\begin{pmatrix} Z^{\Gamma,l} \\ Z^{\gamma,l} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \widehat{\mathbf{V}}^{\Gamma}/n & \widehat{\mathbf{C}}/n \\ \widehat{\mathbf{C}}^\intercal/n & \widehat{\mathbf{V}}^{\gamma}/n \end{pmatrix}\right],$$

where $\widehat{\mathbf{V}}^{\Gamma}$, $\widehat{\mathbf{V}}^{\gamma}$ and $\widehat{\mathbf{C}}$ are defined in (11). For a large integer L, we compute

$$T_l = \max_{\beta\in\mathcal{B}}\max_{j\in\widehat{\mathcal{S}}} \frac{|Z^{\Gamma,l} - \beta Z^{\gamma,l}|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj} + \beta^2\widehat{\mathbf{V}}^{\gamma}_{jj} - 2\beta\widehat{\mathbf{C}}_{jj})/n}} \quad \text{for} \quad 1 \leq l \leq L \tag{28}$$

and choose the empirical upper α quantile of $\{T_l\}_{1\leq l\leq L}$ as an approximation to $\widehat{\rho}(\alpha)$.

5 Uniform Inference Methods under Plurality Rule

We now consider the more challenging setting where the majority rule does not hold. In Section 5.1, we propose the finite-sample plurality rule. With this proposed plurality rule, we propose uniform inference procedures in Sections 5.2 and 5.3.

5.1 Finite-sample Plurality Rule

We introduce the following notations. For $a \in \mathbb{R}$ and $\tau_n \in \mathbb{R}$, define

$$\mathcal{I}(v, \tau_n) = \left\{ j \in \mathcal{S} : \left| \pi_j^*/\gamma_j^* - v \right| \leq \tau_n \right\}. \tag{29}$$

If $\tau_n \to 0$, then $\mathcal{I}(v, \tau_n)$ denotes the set of IVs with the invalidity level near v. Define

$$\mathrm{sep}(n) = \frac{2.1}{\min_{j \in \widehat{\mathcal{S}}} |\gamma_j^*|} \sqrt{\frac{C_1 \log n}{c_0 n}} \sqrt{1 + \max_{l \in \widehat{\mathcal{S}}} \left(\frac{\Gamma_l^*}{\gamma_l^*} \right)^2} \sqrt{1 + \max_{j,l \in \widehat{\mathcal{S}}} \left(\frac{\gamma_j^*}{\gamma_l^*} \right)^2}, \tag{30}$$

where c_0 and C_1 are constants specified in Condition (C1) in Section 6. When γ_j^*'s are of constant orders, $\mathrm{sep}(n)$ is of order $\sqrt{\log n / n}$. Intuitively speaking, $\mathrm{sep}(n)$ denotes the accuracy that we can estimate $\pi_j^*/\gamma_j^* - \pi_k^*/\gamma_k^*$ for any pair (j, k) with $j, k \in \widehat{\mathcal{S}}$; see more discussion in Proposition 2 and its proof.

Condition 2 (Finite-sample Plurality Rule) *For any* $\tau_n \geq 3\mathrm{sep}(n)$ *with* $\mathrm{sep}(n)$ *defined in* (30),

$$|\mathcal{V} \cap \mathcal{S}_{\mathrm{str}}| > \max_{v \in \mathbb{R}} |\mathcal{I}(v, \tau_n)/\mathcal{V}| \tag{31}$$

where $\mathcal{S}_{\mathrm{str}}$ *and* $\mathcal{I}(v, \tau_n)$ *are defined in* (14) *and* (29)*, respectively.*

Here, the set $\mathcal{V} \cap \mathcal{S}_{\mathrm{str}}$ denotes the strongly relevant and valid instruments, which we can rely on to make inference for β^*. The set $\mathcal{I}(v, \tau_n)/\mathcal{V}$ contains all invalid instruments with invalidity levels $\pi_j^*/\gamma_j^* \approx v$. Condition 2 states that the number of valid and strongly relevant IVs is larger than the number of invalid IVs with $\pi_j^*/\gamma_j^* \approx v$ for $v \neq 0$. When $v = 0$, the set $\mathcal{I}(0, \tau_n)/\mathcal{V}$ is the set of weakly invalid IVs. Condition 2 also requires that the number of valid and strongly relevant IVs is larger than that of weakly invalid IVs. In contrast to the population plurality rule, Condition 2 accommodates for the finite-sample approximation error by grouping together invalid IVs with similar invalidity levels.

As a remark, (31) with a smaller τ_n will become a weaker condition. The weakest form of (31) is $|\mathcal{V} \cap \mathcal{S}_{\mathrm{str}}| > \max_{v \in \mathbb{R}} |\mathcal{I}(v, \mathrm{sep}(n))/\mathcal{V}|$ and $|\mathcal{V} \cap \mathcal{S}_{\mathrm{str}}| > |\mathcal{I}(0, 3\mathrm{sep}(n))/\mathcal{V}|$.

For a large sample size, Condition 2 is reduced to the population plurality rule (6). Specifically, if $n \to \infty$ and $\{\pi_j^*\}_{1 \leq j \leq p_{\mathrm{z}}}$ and $\{\gamma_j^*\}_{1 \leq j \leq p_{\mathrm{z}}}$ do not grow with the sample size n, then

$$\lim_{n\to\infty} |\mathcal{V} \cap \mathcal{S}_{\rm str}| = |\mathcal{V}|, \quad \lim_{n\to\infty} |\mathcal{I}(v,\tau_n)/\mathcal{V}| = |\mathcal{I}_\nu/\mathcal{V}| = \begin{cases} 0 & \text{if } \nu = 0 \\ |\mathcal{I}_\nu| & \text{if } \nu \neq 0 \end{cases}$$

with $\mathcal{I}_\nu$ defined in (6). Hence, as $n \to \infty$, (31) implies the population plurality rule (6).

We will propose an inference procedure for β^* under the finite-sample plurality rule. The procedure is two-step: (1) we construct an initial valid IV set $\widehat{\mathcal{V}}$ satisfying

$$\mathcal{V} \cap \mathcal{S}_{\rm str} \subset \widehat{\mathcal{V}} \subset \mathcal{I}(0, 3{\rm sep}(n)). \tag{32}$$

(2) Together with $\widehat{\mathcal{V}}$, we apply the searching and sampling methods to construct CIs for β^*.

5.2 Initial Estimate of $\mathcal{V}$ via TSHT

In the following, we construct $\widehat{\mathcal{V}}$ satisfying (32) through modifying the two-stage hard thresholding (TSHT). The first step of two-stage hard thresholding is to select the set of relevant IVs as in (12). Without loss of generality, we set $\widehat{\mathcal{S}} = \{1, 2, \cdots, |\widehat{\mathcal{S}}|\}$. Next, we apply the reduced-form formula $\beta \cdot \widehat{\gamma}_j + \pi_j \approx \widehat{\Gamma}_j$ for $j \in \widehat{\mathcal{S}}$. For any $j \in \widehat{\mathcal{S}}$, we propose a plug-in estimate of β^* and π^* as

$$\widehat{\beta}^{[j]} = \widehat{\Gamma}_j/\widehat{\gamma}_j \quad \text{and} \quad \widehat{\pi}_k^{[j]} = \widehat{\Gamma}_k - \widehat{\beta}^{[j]}\widehat{\gamma}_k \quad \text{for} \quad k \in \widehat{\mathcal{S}}. \tag{33}$$

We further estimate the standard error of $\widehat{\pi}_k^{[j]}$ for $k \in \widehat{\mathcal{S}}$ by

$$\widehat{\rm SE}(\widehat{\pi}_k^{[j]}) = \sqrt{\frac{\widehat{\sigma}_\epsilon^2 + (\widehat{\beta}^{[j]})^2\widehat{\sigma}_\delta^2 - 2\widehat{\beta}^{[j]}\widehat{\sigma}_{\epsilon,\delta}}{n}} \sqrt{\widehat{\Omega}_{kk} - 2\frac{\widehat{\gamma}_k}{\widehat{\gamma}_j}\widehat{\Omega}_{jk} + (\frac{\widehat{\gamma}_k}{\widehat{\gamma}_j})^2\widehat{\Omega}_{jj}}, \tag{34}$$

where $\widehat{\gamma}$ is defined in (9), $\widehat{\sigma}_\epsilon, \widehat{\sigma}_\delta$ and $\widehat{\sigma}_{\epsilon,\delta}$ are defined in (10) and $\widehat{\Omega}$ is defined in (11).

Construction of a Symmetric Voting Matrix

We construct a voting matrix $\widehat{\Pi} \in \mathbb{R}^{|\widehat{\mathcal{S}}|\times|\widehat{\mathcal{S}}|}$ using $\{\widehat{\pi}^{[j]}\}_{j\in\widehat{\mathcal{S}}}$ and $\{\widehat{\rm SE}(\widehat{\pi}_k^{[j]})\}_{j,k\in\widehat{\mathcal{S}}}$. Specifically, the (k,j) entry of $\widehat{\Pi} \in \mathbb{R}^{|\widehat{\mathcal{S}}|\times|\widehat{\mathcal{S}}|}$ represents whether k-th IV and j-th IV vote for each other to be valid IVs. For $1 \leq k, j \leq |\widehat{\mathcal{S}}|$, we define the (k,j) element of $\widehat{\Pi}$ as

$$\widehat{\Pi}_{k,j} = \mathbf{1}\left(|\widehat{\pi}_k^{[j]}| \leq \widehat{\rm SE}(\widehat{\pi}_k^{[j]}) \cdot \sqrt{\log n} \quad \text{and} \quad |\widehat{\pi}_j^{[k]}| \leq \widehat{\rm SE}(\widehat{\pi}_j^{[k]}) \cdot \sqrt{\log n}\right) \tag{35}$$

where $\widehat{\pi}_k^{[j]}$ and $\widehat{\pi}_j^{[k]}$ are defined in (33) and $\widehat{\rm SE}(\widehat{\pi}_k^{[j]})$ and $\widehat{\rm SE}(\widehat{\pi}_j^{[k]})$ are defined in (34) and $\sqrt{\log n}$ is used to adjust for multiplicity. Based on the definition (35), $\widehat{\Pi}_{k,j}$ denotes the vote between the k-th and j-th instrument: $\widehat{\Pi}_{k,j} = 1$ represents that they support each other to be valid; $\widehat{\Pi}_{k,j} = 0$ represents that they do not.

Example 3. We consider an example with $p_z = 8$ candidate IVs, where $\{Z_1, Z_2, Z_3, Z_4\}$ are valid IVs, $\{Z_5, Z_6, Z_7\}$ are invalid IVs sharing the same invalidity level and Z_8 is invalid with a different invalidity level. The left panel of Table 2 corresponds to a favorable scenario where the valid IVs $\{Z_1, Z_2, Z_3, Z_4\}$ only vote for each other. On the right panel of Table 2, the candidate IV Z_5 receives the votes (by mistake) from three valid IVs $\{Z_2, Z_3, Z_4\}$. This might happen when the IV Z_5 is a weakly invalid IV.

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Z_1	✓	✓	✓	✓	X	X	X	X
Z_2	✓	✓	✓	✓	X	X	X	X
Z_3	✓	✓	✓	✓	X	X	X	X
Z_4	✓	✓	✓	✓	X	X	X	X
Z_5	X	X	X	X	✓	✓	✓	X
Z_6	X	X	X	X	✓	✓	✓	X
Z_7	X	X	X	X	✓	✓	✓	X
Z_8	X	X	X	X	X	X	X	✓
Votes	4	4	4	4	3	3	3	1

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Z_1	✓	✓	✓	✓	X	X	X	X
Z_2	✓	✓	✓	✓	✓	X	X	X
Z_3	✓	✓	✓	✓	✓	X	X	X
Z_4	✓	✓	✓	✓	✓	X	X	X
Z_5	X	✓	✓	✓	✓	✓	✓	X
Z_6	X	X	X	X	✓	✓	✓	X
Z_7	X	X	X	X	✓	✓	✓	X
Z_8	X	X	X	X	X	X	X	✓
Votes	4	5	5	5	6	3	3	1

Table 2: The left voting matrix $\widehat{\Pi}$ denotes that all valid IVs $\{Z_1, Z_2, Z_3, Z_4\}$ support each other but not any other invalid IV. The right voting matrix $\widehat{\Pi}$ denotes that the (weakly) invalid IV Z_5 receives support from valid IVs $\{Z_2, Z_3, Z_4\}$ and invalid IVs $\{Z_6, Z_7\}$.

Construction of $\widehat{\mathcal{V}}^{\texttt{TSHT}}$

We define the winner set $\widehat{\mathcal{W}} \subset \{1, 2, \cdots, |\widehat{\mathcal{S}}|\}$ as the set of IVs receiving the largest number of votes:

$$\widehat{\mathcal{W}} = \arg\max_{1 \leq k \leq |\widehat{\mathcal{S}}|} \|\widehat{\Pi}_{k\cdot}\|_0.$$

Based on $\widehat{\mathcal{W}}$, we construct

$$\widehat{\mathcal{V}}^{\texttt{TSHT}} = \{1 \leq l \leq |\widehat{\mathcal{S}}| : \text{there exists } 1 \leq j \leq |\widehat{\mathcal{S}}| \text{ such that } \widehat{\Pi}_{k,j}\widehat{\Pi}_{j,l} = 1 \text{ for } k \in \widehat{\mathcal{W}}\}. \tag{36}$$

If the k-th IV from the winner set $\widehat{\mathcal{W}}$ and the l-th candidate IV are claimed to be valid by the same IV, then the k-th candidate IV is also included in our initial construction $\widehat{\mathcal{V}}^{\texttt{TSHT}}$. In general, we have $\widehat{\mathcal{W}} \subset \widehat{\mathcal{V}}^{\texttt{TSHT}} \subset \mathcal{I}(0, 3\text{sep}(n))$ with $\text{sep}(n)$ defined in (30). If there are no weakly invalid IVs (that is, $\mathcal{I}(0, 3\text{sep}(n)) = \mathcal{V}$), then we have $\widehat{\mathcal{W}} = \widehat{\mathcal{V}}^{\texttt{TSHT}} = \mathcal{V}$. However, in practice, if there exists weakly invalid IVs, the winner set $\widehat{\mathcal{W}}$ and $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ can be different and only $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ is guaranteed to satisfy (32).

We illustrate the definitions of $\widehat{\mathcal{W}}$ and $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ using the examples in Table 2.

Continuing Example 3. For the left panel of Table 2, we have $\widehat{\mathcal{V}}^{\texttt{TSHT}} = \widehat{\mathcal{W}} = \{1, 2, 3, 4\} = \mathcal{V}$ and the property (32) is satisfied. For the right panel of Table 2, we have $\widehat{\mathcal{W}} = \{5\}$ and $\widehat{\mathcal{V}}^{\texttt{TSHT}} = \{1, 2, 3, 4, 5, 6, 7\}$. Only $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ satisfies (32) but not $\widehat{\mathcal{W}}$.

5.3 Uniform Confidence Intervals by Searching and Sampling

After constructing $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ in (36), we generalize the methods proposed in Section 4 since Condition 2 implies that $\mathcal{V} \cap \mathcal{S}_{\rm str}$ is the majority of the initial set $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$.

We modify the definition of $\widehat{\rho}(\alpha)$ in (17) as

$$\mathbb{P}\left(\max_{\beta\in\mathcal{B}}\max_{j\in\widehat{\mathcal{V}}}\frac{|\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj} + \beta^2\widehat{\mathbf{V}}^{\gamma}_{jj} - 2\beta\widehat{\mathbf{C}}_{jj})/n}} \leq \widehat{\rho}(\alpha)\right) \geq 1-\alpha. \tag{37}$$

Similar to (17), we can choose $\widehat{\rho}(\alpha) \asymp \sqrt{\log|\mathcal{B}|}$ or $\widehat{\rho}(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{|\mathcal{B}|\cdot p_{\rm z}}\right)$. By replacing $\widehat{\mathcal{S}}$ with $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$, we can also implement the bootstrap method in Section 4.5 to choose $\widehat{\rho}(\alpha)$.

For $j \in \widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\beta \in \mathcal{B}$, define as in (18)

$$\widehat{\rho}_j(\beta,\alpha) = \widehat{\rho}(\alpha)\cdot\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj} + \beta^2\widehat{\mathbf{V}}^{\gamma}_{jj} - 2\beta\widehat{\mathbf{C}}_{jj})/n}$$

and

$$\widehat{\pi}_j(\beta) = \left(\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right)\cdot\mathbf{1}\left(\left|\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right| \geq \widehat{\rho}_j(\beta,\alpha)\right).$$

With $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$, we construct the confidence interval for β^* as

$$\mathrm{CI}^{\rm search} = \left(\min_{\left\{\beta\in\mathcal{B}:\|\widehat{\pi}_{\widehat{\mathcal{V}}}(\beta)\|_0 < \frac{|\widehat{\mathcal{V}}|}{2}\right\}}\beta,\ \max_{\left\{\beta\in\mathcal{B}:\|\widehat{\pi}_{\widehat{\mathcal{V}}}(\beta)\|_0 < \frac{|\widehat{\mathcal{V}}|}{2}\right\}}\beta\right). \tag{38}$$

In comparison to (20), the main difference is that the initial set $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ is used in (38), instead of $\widehat{\mathcal{S}}$. It is possible that there is no β such that $\|\widehat{\pi}_{\widehat{\mathcal{V}}}(\beta)\|_0 < |\widehat{\mathcal{V}}|/2$ and in this case, $\mathrm{CI}^{\rm search}$ is empty. This is used as a partial check for this finite-sample plurality rule.

We also utilize the initial estimator $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ and the sampling idea detailed in Section 4.3. We sample $\{\widehat{\Gamma}^{[m]}, \widehat{\gamma}^{[m]}\}_{1\leq m\leq M}$ as in (21). For $\lambda \asymp (\log n/M)^{\frac{1}{2p_{\rm z}}}$ and $\widehat{\rho}_j(\beta,\alpha)$ in (18), we implement the sampled thresholding step,

$$\widehat{\pi}_j^{[m]}(\beta,\lambda) = \left(\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right)\cdot\mathbf{1}\left(\left|\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right| \geq \lambda\widehat{\rho}_j(\beta,\alpha)\right) \quad \text{for} \quad j \in \widehat{\mathcal{V}}^{\texttt{TSHT}}. \tag{39}$$

For $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $1 \leq m \leq M$, we use $\widehat{\pi}^{[m]}_{\widehat{\mathcal{V}}}(\beta,\lambda) \in \mathbb{R}^{|\widehat{\mathcal{V}}|}$ defined in (39) to search for β:

$$\beta^{[m]}_{\min}(\lambda) = \min_{\left\{\beta\in\mathcal{B}:\|\widehat{\pi}^{[m]}_{\widehat{\mathcal{V}}}(\beta,\lambda)\|_0 < |\widehat{\mathcal{V}}|/2\right\}}\beta \quad \text{and} \quad \beta^{[m]}_{\max}(\lambda) = \max_{\left\{\beta\in\mathcal{B}:\|\widehat{\pi}^{[m]}_{\widehat{\mathcal{V}}}(\beta,\lambda)\|_0 < |\widehat{\mathcal{V}}|/2\right\}}\beta. \tag{40}$$

For a given sample $1 \leq m \leq M$, if there is no β such that $\|\widehat{\pi}^{[m]}(\beta,\lambda)\|_0 < |\widehat{\mathcal{V}}|/2$, we simply set $(\beta^{[m]}_{\min}(\lambda), \beta^{[m]}_{\max}(\lambda)) = \varnothing$. We take a union of all non-empty sampled CIs :

$$\mathrm{CI}^{\rm sample} = \left(\min_{m\in\mathcal{M}}\beta^{[m]}_{\min}(\lambda), \max_{m\in\mathcal{M}}\beta^{[m]}_{\max}(\lambda)\right), \tag{41}$$

with $\mathcal{M} = \{1 \leq m \leq M : (\beta^{[m]}_{\min}(\lambda), \beta^{[m]}_{\max}(\lambda)) \neq \varnothing\}$.

Remark 1 We have demonstrated our method by constructing $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ as in (36). However, as long as the constructed $\widehat{\mathcal{V}}$ satisfy (32), our proposed CIs in (38) and (41) are effective under the finite-sample plurality rule (Condition 2). In Simulation studies, we also investigate the finite-sample performance using $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, the set of valid IVs selected by the `CIIV` method proposed in [34]. We can combine the interval by taking a union of the CI by $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and the corresponding CI by $\widehat{\mathcal{V}}^{\texttt{CIIV}}$. In terms of the coverage property, the validity of this combined interval follows from that of $\mathrm{CI}^{\mathrm{search}}$ or $\mathrm{CI}^{\mathrm{sample}}$.

5.4 Algorithms

We summarize our proposed method with searching and sampling in Algorithm 1. A simplified algorithm by assuming the majority rule is presented in Section A.2 in the supplementary material. In Section 7.1, we provide further discussion about implementation of Algorithm 1, including how to choose the tuning parameter λ for the sampling CI.

Algorithm 1 Uniform inference with Searching and Sampling (Plurality Rule)

Input: Outcome $Y \in \mathbb{R}^n$; Treatment $D \in \mathbb{R}^n$; Candidate IVs $Z \in \mathbb{R}^{n \times p_{\mathrm{z}}}$; Baseline Covariates $X \in \mathbb{R}^{n \times p_{\mathrm{x}}}$; significance level $\alpha \in (0,1)$; $M > 0$; $\lambda \asymp (\log n/M)^{1/(2p_{\mathrm{z}})}$.
Output: Confidence intervals $\mathrm{CI}^{\mathrm{search}}$ and $\mathrm{CI}^{\mathrm{sample}}$

Construct reduced form estimators $\widehat{\Gamma} \in \mathbb{R}^{p_{\mathrm{z}}}$ and $\widehat{\gamma} \in \mathbb{R}^{p_{\mathrm{z}}}$ as in (9);
Construct the noise level estimators $\widehat{\sigma}_{\epsilon}^2, \widehat{\sigma}_{\delta}^2$ and $\widehat{\sigma}_{\epsilon,\delta}$ as in (10);
Construct A_1 and A_2 as in (27);
Construct the grid set $\mathcal{B} \subset [A_1, A_2]$ with the grid size $n^{-0.6}$;
Select the set of relevant IVs $\widehat{\mathcal{S}}$ as in (12);
Construct the voting matrix $\widehat{\Pi} \in \mathbb{R}^{|\widehat{\mathcal{S}}| \times |\widehat{\mathcal{S}}|}$ as in (35);
Construct $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ as in (36); ▷ Construction of the initial set $\widehat{\mathcal{V}}$
Compute $\{T_l\}_{1 \leq l \leq L}$ where T_l is defined in (28) with $\widehat{\mathcal{S}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$;
Compute $\widehat{\rho}(\alpha)$ using the upper α quantile of $\{T_l\}_{1 \leq l \leq L}$;
Construct $\mathrm{CI}^{\mathrm{search}}$ in (38); ▷ Construction of Searching CI
for $m \leftarrow 1$ to M **do**
    Sample $\widehat{\Gamma}^{[m]}$ and $\widehat{\gamma}^{[m]}$ as in (21);
    Compute $\{\widehat{\pi}_j^{[m]}(\beta, \lambda)\}_{j \in \widehat{\mathcal{V}}, \beta \in \mathcal{B}}$ as in (39);
    Construct $(\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda))$ with $\beta_{\min}^{[m]}(\lambda)$ and $\beta_{\max}^{[m]}(\lambda)$ in (40);
end for
Construct $\mathrm{CI}^{\mathrm{sample}}$ as in (41) ▷ Construction of Sampling CI.

---

6 Theoretical Justification

We present the theory under the finite-sample majority rule in Section 6.1 and that under the finite-sample plurality rule in Section 6.2. We focus on the homoscedastic setting with a fixed dimension p and introduce the following regularity conditions.

(C1) Consider the model (7). For $1 \leq i \leq n$, $W_{i\cdot} = (X_{i\cdot}^{\intercal}, Z_{i\cdot}^{\intercal})^{\intercal} \in \mathbb{R}^p$ are i.i.d. Sub-gaussian random vectors with $\Sigma = \mathbf{E}(W_{i\cdot} W_{i,\cdot}^{\intercal})$ satisfying $c_0 \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq C_0$ for some positive constants $C_0 \geq c_0 > 0$; the errors $(\epsilon_i, \delta_i)^{\intercal}$ in (7) are i.i.d Sub-gaussian random vectors with its covariance matrix satisfying $c_1 \leq \lambda_{\min}(\Sigma)\left[\mathrm{Cov}((\epsilon_i, \delta_i)^{\intercal})\right] \leq \lambda_{\max}\left[\mathrm{Cov}((\epsilon_i, \delta_i)^{\intercal})\right] \leq C_1$ for some positive constants $C_1 \geq c_1 > 0$.

(C2) The errors ϵ_i and δ_i satisfy $\mathbf{E}(\epsilon_i^2 \mid W_{i\cdot}) = \sigma_{\epsilon}^2, \mathbf{E}(\delta_i^2 \mid W_{i\cdot}) = \sigma_{\delta}^2$, and $\mathbf{E}(\epsilon_i \cdot \delta_i \mid W_{i\cdot}) = \sigma_{\epsilon,\delta}$.

Both conditions (C1) and (C2) are mild and standard for theoretical justification of linear models with instrumental variables [35]. Condition (C1) is imposed on the reduced form model (7), which includes the outcome model (2) and the treatment model (3) as a special case. We assume that the covariance matrix of $W_{i\cdot}$ is well conditioned and also the covariance matrix of the errors is well conditioned. The later condition holds as long as e_i in (2) and δ_i in (3) are not perfectly correlated. Condition (C2) assumes the homoscedastic error, which can be further relaxed with the robust covariance matrix estimator in (13).

6.1 Majority Rule

The following theorem justifies the searching CI under the majority rule.

Theorem 1 *Suppose that Condition 1, Conditions* (C1) *and* (C2) *hold and* $\widehat{\rho}(\alpha)$ *satisfies* (17)*, then the searching confidence interval* $\mathrm{CI}^{\mathrm{search}}$ *defined in* (20) *satisfies*

$$\liminf_{n \to \infty} \mathbb{P}\left(\beta^* \in \mathrm{CI}^{\mathrm{search}}\right) \geq 1 - \alpha,$$

where α is the pre-specified significance level. For a sufficiently large n, the length $\mathbf{L}(\mathrm{CI}^{\mathrm{search}})$ *of the interval* $\mathrm{CI}^{\mathrm{search}}$ *satisfies*

$$\mathbb{P}\left(\mathbf{L}(\mathrm{CI}^{\mathrm{search}}) \leq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{4\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|} \leq C \frac{\sqrt{\log n / n}}{\min_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} |\gamma_j^*|}\right) \geq 1 - \alpha - \exp(-c\sqrt{\log n})$$

where $c > 0$ and $C > 0$ are positive constants independent of n.

Theorem 1 only requires the finite-sample majority rule (Condition 1) and the regularity conditions (C1) and (C2). Lemma 2 in the supplementary material shows that the choices

$\widehat{\rho}(\alpha) = C\sqrt{\log|\mathcal{B}|}$ or $\widehat{\rho}(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{|\mathcal{B}|\cdot p_{\rm z}}\right)$ will guarantee (17). If the non-zero individual IV strength γ_j^* is of a constant order for $j \in \mathcal{S}$, then (1) implies that $\mathbf{L}(\mathrm{CI}^{\rm search})$ is of order $\sqrt{\log n/n}$ and is at most worse off than the regular parametric rate by $\sqrt{\log n}$.

Importantly, Theorem 1 is valid without requiring invalid IVs to have a sufficiently large violation level $|\pi_j^*/\gamma_j^*|$, which is a key assumption for the theoretical justification of `TSHT` [14] and `CIIV` [34]. Both `TSHT` and `CIIV` rely on the assumption that the invalid IVs (even if weakly invalid ones) are correctly identified. As shown in [14], the invalid IVs can be correctly identified if all invalid IVs are well-separated from the valid instruments, that is, invalid IVs have a sufficiently large violation level $|\pi_j^*/\gamma_j^*|$ (Assumption 8 in [14]). Without this well-separation condition, we demonstrate in numerical studies that the CIs by `TSHT` and `CIIV` are under-coverage even if the majority rule holds; see the details in Tables D.1 and D.2 in the supplemental materials.

With imposing the well-separation condition, the CIs based on `TSHT` and `CIIV` can be as efficient as an oracle setting of knowing which IVs are valid a priori. Our proposed $\mathrm{CI}^{\rm search}$ tends to be longer than those based on IV selection, which can be a price to pay for constructing uniformly valid CIs.

Now we provide justification for the sampling CI. For $\alpha_0 \in (0, 1/4)$, define the constant,

$$c^*(\alpha_0) = \frac{1}{(2\pi)^{p_{\rm z}}} \prod_{i=1}^{2p_{\rm z}} \left[\lambda_i(\mathrm{Cov}) + \frac{1}{2}\lambda_{\min}(\mathrm{Cov})\right]^{-\frac{1}{2}} \exp\left(-F^{-1}_{\chi^2_{2p_{\rm z}}}(1-\alpha_0)\right), \tag{42}$$

where Cov is defined in (8) and $F^{-1}_{\chi^2_{2p_{\rm z}}}(1-\alpha_0)$ denotes $1-\alpha_0$ quantile of the χ^2 distribution with degree of freedom $2p_{\rm z}$. Note that, for a fixed $p_{\rm z}$ and $\alpha_0 \in (0,1)$, $c^*(\alpha_0)$ is a constant independent of n. Condition (C1) implies a constant lower bound for $\lambda_{\min}(\mathrm{Cov})$, which is established in (47) in the supplementary material. We state the following sampling property, which motivates our procedure.

Proposition 1 *Suppose Conditions* (C1) *and* (C2) *hold and* $\alpha_0 \in (0, 1/4)$ *is a small positive constant. If* $\mathrm{err}_n(M, \alpha_0) = \frac{1}{2}\left[\frac{2\log n}{c^*(\alpha_0) M}\right]^{\frac{1}{2p_{\rm z}}} \lesssim c^*(\alpha_0)$ *with* $c^*(\alpha_0)$ *defined in* (42)*, then*

$$\liminf_{n\to\infty} \mathbb{P}\left(\min_{1\le m\le M} \max\left\{\left\|\widehat{\gamma}^{[m]} - \gamma^*\right\|_\infty, \left\|\widehat{\Gamma}^{[m]} - \Gamma^*\right\|_\infty\right\} \le \frac{\mathrm{err}_n(M,\alpha_0)}{\sqrt{n}}\right) \ge 1 - \alpha_0 - M^{-c}$$

where $c > 0$ *is a positive constant.*

Since $c^*(\alpha_0)$ is of a constant order, a large sampling size M guarantees the condition $\mathrm{err}_n(M, \alpha_0) \lesssim c^*(\alpha_0)$. The above proposition states that, after sampling M times, there exists one good sampled estimator $\widehat{\gamma}^{[m]}, \widehat{\Gamma}^{[m]}$ converging to γ^*, Γ^* at a rate faster than $1/\sqrt{n}$. Such a sampling property has been first proved in [12] to address a different problem.

Now we apply Proposition 1 to justify the sampling CI under the majority rule.

Theorem 2 *Suppose that Condition 1, Conditions* (C1) *and* (C2) *hold. If* $\text{err}_n(M, \alpha_0) = \frac{1}{2}\left[\frac{2\log n}{c^*(\alpha_0)M}\right]^{\frac{1}{2p_z}} \lesssim c^*(\alpha_0)$ *with* $c^*(\alpha_0)$ *defined in* (42) *and* λ *and* $\widehat{\rho}(\alpha)$ *satisfy*

$$\lambda \geq \max_{j\in\widehat{\mathcal{S}}} \frac{(1+|\beta^*|)\frac{\text{err}_n(M,\alpha_0)}{\sqrt{n}}}{\widehat{\rho}(\alpha)\cdot\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj}+(\beta^*)^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta^*\widehat{\mathbf{C}}_{jj})/n}}, \tag{43}$$

then $\text{CI}^{\text{sample}}$ *defined in* (26) *satisfies*

$$\liminf_{n\to\infty} \mathbb{P}\left(\beta^* \in \text{CI}^{\text{sample}}\right) \geq 1-2\alpha_0,$$

where $\alpha_0 \in (0, 1/4)$ *is a small constant used in the definition of* (42)*. For a sufficiently large* n*, with probability larger than* $1-|\mathcal{B}|^{-c}-\exp(-c\sqrt{\log n})$*, the length* $\mathbf{L}(\text{CI}^{\text{sample}})$ *satisfies*

$$\mathbf{L}(\text{CI}^{\text{sample}}) \lesssim \frac{\sqrt{\frac{\log|\mathcal{B}|+\log|\mathcal{M}|}{n}}+\lambda\sqrt{\frac{\log n}{n}}}{\min_{j\in\widehat{\mathcal{S}}\cap\mathcal{V}}|\gamma^*_j|}$$

where $\mathcal{M}$ *is defined in* (25) *and* $c > 0$ *is a constant.*

We can choose $\widehat{\rho}(\alpha) \asymp \sqrt{\log|\mathcal{B}|}$ or $\widehat{\rho}(\alpha) = \Phi^{-1}\left(1-\frac{\alpha}{|\mathcal{B}|\cdot p_z}\right)$ or via the bootstrap method. For all cases, $\widehat{\rho}(\alpha)$ is at least of a constant order. The choice $\lambda = c_*(\log n/M)^{1/(2p_z)}$ will guarantee the condition (43) to hold. We will propose a data-dependent way of choosing the constant c_* in Section 7.1.

Similar to Theorem 1, Theorem 2 shows that our proposed searching CI does not require the well-separation condition on invalidity levels of the invalid instruments. If the instrument strengths $\{\gamma^*_j\}_{j\in\mathcal{S}}$ are assumed to be of a constant order, then the interval length is upper bounded by $\sqrt{\log|\mathcal{B}|\cdot|\mathcal{M}|/n}$. With only upper bounds for $\mathbf{L}(\text{CI}^{\text{search}})$ and $\mathbf{L}(\text{CI}^{\text{sample}})$, we cannot compare their exact lengths. However, the component $\lambda\sqrt{\log n/n}$ of the upper bound for $\mathbf{L}(\text{CI}^{\text{sample}})$ indicates why the sampling CIs tend to be (much) shorter than the searching CIs. The corresponding component for the searching CI is $\sqrt{\log n/n}$.

6.2 Plurality Rule

We now consider the more challenging setting which only assumes the finite-sample plurality rule (Condition 2). The following proposition quantifies when the j-th and k-th candidate instruments vote for each other, that is, $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 1$.

Proposition 2 *Suppose that Conditions* (C1) *and* (C2) *hold. Consider the indexes* $j \in \widehat{\mathcal{S}}$ *and* $k \in \widehat{\mathcal{S}}$*. (1) If* $\pi^*_k/\gamma^*_k = \pi^*_j/\gamma^*_j$*, then with probability larger than* $1-\exp(-c\sqrt{\log n})$ *for some positive constant* $c > 0$*,* $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 1$*. (2) If* $\left|\pi^*_k/\gamma^*_k - \pi^*_j/\gamma^*_j\right| \geq \text{sep}(n)$ *with* $\text{sep}(n)$ *defined in* (30)*, then with probability larger than* $1-\exp(-c\sqrt{\log n})$ *for some positive constant* $c > 0$*,* $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 0$.

The above proposition shows that if two candidate IVs are of the same invalidity level, then they vote for each other with a high probability. If two candidate IVs are well-separated (i.e. $|\pi_k^*/\gamma_k^* - \pi_j^*/\gamma_j^*| \geq \mathrm{sep}(n)$), then they vote against each other with a high probability. If $0 < |\pi_k^*/\gamma_k^* - \pi_j^*/\gamma_j^*| < \mathrm{sep}(n)$, there is no theoretical guarantee on how the two candidate IVs will vote. The above proposition also reveals that we are likely to make a mistake in selecting valid IVs if the invalidity levels of some IVs are below $\mathrm{sep}(n)$.

The following proposition shows that $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ satisfies (32).

Proposition 3 *Suppose that Condition 2 and Conditions* (C1) *and* (C2) *hold. Then with probability larger than* $1 - \exp(-c\sqrt{\log n})$ *for some positive constant* $c > 0$, *the constructed* $\widehat{\mathcal{V}}$ *in* (45) *satisfies* (32).

With Proposition 3, we connect the plurality rule to the majority rule. In our construction of $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$, we are able to remove all strongly invalid IVs and the set $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ consists of valid IVs and the weakly invalid IVs. The finite-sample plurality condition (Condition 2) assumes that the number of valid IVs is more than that of the weakly invalid IVs. Then it is sufficient to apply the theoretical analysis of the majority rule by replacing $\widehat{\mathcal{S}}$ with $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ or any $\widehat{\mathcal{V}}$ satisfying (32).

The following theorem justifies the searching CI for the setting with plurality rule.

Theorem 3 *Suppose that Condition 2 and Conditions* (C1) *and* (C2) *hold and* $\widehat{\rho}(\alpha)$ *satisfies* (37), *then* $\mathrm{CI}^{\mathrm{search}}$ *defined in* (20) *satisfies*

$$\liminf_{n\rightarrow\infty} \mathbb{P}\left(\beta^* \in \mathrm{CI}^{\mathrm{search}}\right) \geq 1-\alpha.$$

where α *is the pre-specified significance level. For a sufficiently large* n, *the length* $\mathbf{L}(\mathrm{CI}^{\mathrm{search}})$ *of the interval* $\mathrm{CI}^{\mathrm{search}}$ *satisfies*

$$\mathbb{P}\left(\mathbf{L}(\mathrm{CI}^{\mathrm{search}}) \leq C\frac{\sqrt{\log|\mathcal{B}|/n}}{\max_{j\in\mathcal{V}}|\gamma_j^*|}\right) \geq 1-\alpha-\exp(-c\sqrt{\log n})$$

where $c > 0$ *is a positive constant.*

The following theorem justifies the sampling CI for the setting with plurality rule.

Theorem 4 *Suppose that Condition 1 and Conditions* (C1) *and* (C2) *hold. If* $\mathrm{err}_n(M,\alpha_0) = \frac{1}{2}\left[\frac{2\log n}{c^*(\alpha_0)M}\right]^{\frac{1}{2p_z}} \lesssim c^*(\alpha_0)$ *with* $c^*(\alpha_0)$ *defined in* (42) *and* λ *and* $\widehat{\rho}(\alpha)$ *satisfies*

$$\lambda \geq \max_{j\in\widehat{\mathcal{V}}} \frac{(1+|\beta^*|)\frac{\mathrm{err}_n(M,\alpha_0)}{\sqrt{n}}}{\widehat{\rho}(\alpha)\cdot\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj}+(\beta^*)^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta^*\widehat{\mathbf{C}}_{jj})/n}},$$

then $\mathrm{CI}^{\mathrm{sample}}$ *defined in* (26) *satisfies* $\liminf_{n\to\infty} \mathbb{P}\left(\beta^* \in \mathrm{CI}^{\mathrm{sample}}\right) \geq 1 - 2\alpha_0$ *where* $\alpha_0 \in (0, 1/4)$ *is a small constant used in the definition of* (42). *For a sufficiently large* n, *with probability larger than* $1 - |B|^{-c} - \exp(-c\sqrt{\log n})$, *the length* $\mathbf{L}(\mathrm{CI}^{\mathrm{sample}})$ *satisfies*

$$\mathbf{L}(\mathrm{CI}^{\mathrm{sample}}) \lesssim \frac{\sqrt{\frac{\log|\mathcal{B}|+\log|\mathcal{M}|}{n}} + \lambda\sqrt{\frac{\log n}{n}}}{\min_{j\in\mathcal{V}} |\gamma_j^*|}$$

where $\mathcal{M}$ *is defined in* (25) *and* $c > 0$ *is a constant.*

7 Simulation Studies

7.1 Algorithm Implementation

We implement our proposed Algorithm 1 throughout the numerical studies and discuss two practical issues about the implementation. First, we can shorten the initial interval $[A_1, A_2]$ in (27) and obtain a more refined interval $[A_1^{\mathrm{ref}}, A_2^{\mathrm{ref}}] \subset [A_1, A_2]$. Specifically, we conduct an initial searching over $[A_1, A_2]$ with grid size n^{-1} by applying Algorithm 1 with $\widehat{\rho}(\alpha) = \sqrt{2.005 \cdot \log|\mathcal{B}|}$ where $|\mathcal{B}| = (A_2 - A_1)\cdot n$ denotes the cardinality of the set $\mathcal{B}$. The output interval $[A_1^{\mathrm{ref}}, A_2^{\mathrm{ref}}]$ of this initial searching is a subset of $[A_1, A_2]$. We then construct the grid set $\mathcal{B}^{\mathrm{ref}} \subset [A_1^{\mathrm{ref}}, A_2^{\mathrm{ref}}]$ with the grid size $n^{-0.6}$ and construct $\widehat{\rho}(\alpha)$ using the bootstrap method detailed in Section 4.5. The pre-searching does not change the theoretical justification as the proof of Theorem 1 implies that $\beta^* \in [A_1^{\mathrm{ref}}, A_2^{\mathrm{ref}}]$ with a high probability. The main purpose of this pre-searching step is to reduce the computational time for the following-up bootstrap procedure; see Section D.3 in the supplementary material.

For the sampling part, the size M is set as 1000 by default. The main step is to choose the tuning parameter $\lambda > 0$. To start with, we construct the sampling CI with a small tuning parameter $\lambda = 1/6 \cdot (\log n/M)^{1/(2p_z)}$. If the value of λ is too small, then most of the $M = 1000$ intervals (based on the sampled reduced form estimators) will be empty. We then increase the value of λ until more than 5% of the M intervals are non-empty. The smallest value of λ achieving this will be used in 1.

Compared Methods. We shall implement our proposed $\mathrm{CI}^{\mathrm{search}}$ and $\mathrm{CI}^{\mathrm{sample}}$ in Algorithm 1. For both methods, we try two different initial estimators of $\mathcal{V}$: $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ defined in (36) and the set of valid IVs $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ output by `CIIV` [34]. After obtaining these two CIs, we combine them by taking the union. As a benchmark, we implement the `oracle` TSLS estimator assuming the prior knowledge of valid IVs.

We also compare with three existing CI construction methods allowing for invalid IVs: `TSHT` [14], `CIIV` [34] and `Union` method [18]. `TSHT` and `CIIV` are implemented with the codes on the Github websites [1] while `Union` method is implemented with the code shared by the authors of [18]. Both `TSHT` and `CIIV` select valid IVs from a set of candidate IVs and then make inference for β^* using the selected IVs. The `Union` method takes a union of intervals which are constructed by a given number of candidate IVs and are not rejected by the Sargan test. An upper bound $\bar{s}$ for the number of invalid IVs is required for the construction. We consider two specific upper bounds: $\bar{s} = p_{\mathrm{z}} - 1$ corresponds to two valid instruments and $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ corresponds to the majority rule (that is, more than half instruments are valid).

With 500 replications of simulations, we compare different CIs in terms of empirical coverage and average lengths.

7.2 Simulation Settings and Numerical Results

We generate the i.i.d. data $\{Y_i, D_i, Z_{i\cdot}, X_{i\cdot}\}_{1 \le i \le n}$ using the outcome model (2) and treatment model (3). For $1 \le i \le n$, generate the covariates $W_{i\cdot} = (Z_{i\cdot}^{\intercal}, X_{i\cdot}^{\intercal})^{\intercal} \in \mathbb{R}^p$ following a multivariate normal distribution with zero mean and covariance $\Sigma \in \mathbb{R}^{p \times p}$ where $\Sigma_{j,l} = 0.5^{|j-l|}$ for $1 \le j, l \le p$; generate the errors $(e_i, \delta_i)^{\intercal}$ following bivariate normal with zero mean, unit variance and $\mathrm{Cov}(\epsilon_i, \delta_i) = 0.8$. Set $p_{\mathrm{x}} = 10$, $\psi^* = (0.6, 0.7, \cdots, 1.5)^{\intercal} \in \mathbb{R}^{10}$ in (3) and $\Psi^* = (1.1, 1.2, \cdots, 2)^{\intercal} \in \mathbb{R}^{10}$ in (2). We vary n across $\{500, 1000, 2000, 5000\}$. We generate the IV strength vector $\gamma^* \in \mathbb{R}^{p_{\mathrm{z}}}$ and the violation vector $\pi^* \in \mathbb{R}^{p_{\mathrm{z}}}$ as follows,

S1 (Majority rule): set $\gamma^* = \gamma_0 \cdot \mathbf{1}_{10}$ and $\pi^* = (\mathbf{0}_6, \tau \cdot \gamma_0, \tau \cdot \gamma_0, -0.5, -1)^{\intercal}$;

S2 (Plurality rule): set $\gamma^* = \gamma_0 \cdot \mathbf{1}_{10}$ and $\pi^* = (\mathbf{0}_4, \tau \cdot \gamma_0, \tau \cdot \gamma_0, -\frac{1}{3}, -\frac{2}{3}, -1, -\frac{4}{3})^{\intercal}$;

S3 (Plurality rule): set $\gamma^* = \gamma_0 \cdot \mathbf{1}_{10}$ and $\pi^* = (\mathbf{0}_4, \tau \cdot \gamma_0, \tau \cdot \gamma_0, -\frac{1}{6}, -\frac{1}{3}, -\frac{1}{2}, -\frac{2}{3})^{\intercal}$;

S4 (Plurality rule): set $\gamma^* = \gamma_0 \cdot \mathbf{1}_6$ and $\pi^* = (\mathbf{0}_2, -0.8, -0.4, \tau \cdot \gamma_0, 0.6)^{\intercal}$;

S5 (Plurality rule): set $\gamma^* = \gamma_0 \cdot \mathbf{1}_6$ and $\pi^* = (\mathbf{0}_2, -0.8, -0.4, \tau \cdot \gamma_0, \tau \cdot \gamma_0 + 0.1)^{\intercal}$;

The parameter γ_0 denotes the IV strength and is varied across $\{0.25, 0.5\}$. The parameter τ denotes the invalidity level of the invalid IV and is varied across $\{0.1, 0.2, 0.3, 0.4\}$, where the smaller values indicate the existence of weakly invalid IVs. Only setting **S1** satisfies the (population) majority rule while the other settings only satisfy the (population) plurality rule. Settings **S4** and **S5** represent the challenging settings where there are only two valid IVs. We introduce settings **S3** and **S5** to test the robustness of our proposed method when the finite-sample plurality rule might be violated. For example, for the setting **S3** with small n (e.g. $n = 500$), the invalid IVs with π_j^* values $\tau \cdot \gamma_0, \tau \cdot \gamma_0, -\frac{1}{6}, -\frac{1}{3}$ may be weakly invalid

[1] The code for `TSHT` is obtained from `https://github.com/hyunseungkang/invalidIV` and for `CIIV` is obtained from `https://github.com/xlbristol/CIIV`.

and hence such a setting may violate the finite-sample plurality rule (Condition 2); for the setting **S5** with small n (e.g. $n = 500$), the invalid IVs with π_j^* values $\tau \cdot \gamma_0, \tau \cdot \gamma_0 + 0.1$ have similar invalidity levels and may violate Condition 2.

In Table 3, we report the empirical coverage for settings **S2, S3, S4, S5** with $\gamma_0 = 0.5$ and $\tau = 0.2$ or 0.4. Our proposed searching and sampling CIs achieve the desired coverage levels in most settings. For settings **S2** and **S4**, both initial estimates of set of valid IVs $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ lead to CIs achieving the 95% coverage level; so does the combined intervals. For the more challenging settings **S3** and **S5**, the empirical coverage level of the combined interval achieves the desired coverage level, except for **S5** with $\tau = 0.4$ and $n = 500$. For $n = 500, 1000$, the empirical coverage for CIs with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ can be under-coverage while that with $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ is closer to the desired coverage level. For $n = 2000, 5000$, the empirical coverage levels reach the desired 95%. This happens mainly due to the fact that the finite-sample plurality rule tends to fail for settings **S3** and **S5** with $n = 500$ and $n = 1000$.

As observed in Table 3, the CIs by `TSHT` [14] and `CIIV` [34] achieve the 95% coverage level for a large sample size and a relatively large violation level, such as $n = 5000$ and $\tau = 0.4$. The CI by `CIIV` is more robust in the sense that its validity may require a smaller sample size than `TSHT`. The CIs by the `Union` method [18] with $\bar{s} = p_{\rm z} - 1$ (assuming there are two valid IVs) achieve the desired coverage levels while those with $\bar{s} = \lceil p_{\rm z}/2 \rceil$ (assuming the majority rule) do not achieve the desired coverage levels.

We compare the lengths of different CIs in Table 4. For settings **S2** and **S3**, the proposed sampling CI is shorter than the searching CI and the CIs by the `Union` method. The CIs by the `Union` method can be three to six times longer than that of the (combined) sampling CI. For settings **S4** and **S5** with $n = 500, 1000$, the sampling CI is still shorter than the searching CI and the CIs by the `Union` method. However, when the sample size is relatively large, the searching CI and the CIs by the `Union` method can be better than the sampling method. When the CIs by `TSHT` [14] and `CIIV` [34] are valid, their lengths are similar to the length of the CI by `oracle` TSLS, which has been justified in [14, 34]. The sampling CI, searching CI and CI by the `Union` are in general longer than the CI by the `oracle` TSLS, which is a price to pay for constructing uniformly valid CIs. The full details of settings **S1** to **S5** are reported in Section D.1 in the supplementary material.

Similar to the setting in CIIV paper [34], we further consider the following settings.

CIIV-1 (Plurality rule): set $\gamma^* = 0.4 \cdot \mathbf{1}_{21}$ and $\pi^* = (\mathbf{0}_9, \tau \cdot \mathbf{1}_6, \frac{\tau}{2} \cdot \mathbf{1}_6)^{\intercal}$.

CIIV-2 (Plurality rule): set $\gamma^* = 0.4 \cdot \mathbf{1}_{21}$ and $\pi^* = (\mathbf{0}_9, \tau \cdot \mathbf{1}_3, -\tau \cdot \mathbf{1}_3, \frac{\tau}{2} \cdot \mathbf{1}_3, -\frac{\tau}{2} \cdot \mathbf{1}_3)^{\intercal}$.

We vary τ across $\{0.1, 0.2, 0.3, 0.4\}$ where τ represents the invalidity level. The setting **CIIV-1** with $\tau = 0.4$ corresponds to the exact setting considered in [34]. For a small τ and

Empirical Coverage for $\gamma_0 = 0.5$

						Proposed Searching			Proposed Sampling			Union	
Set	τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$p_z - 1$	$\lceil p_z/2 \rceil$
S2	0.2	500	0.95	0.41	0.45	0.99	1.00	1.00	0.99	0.99	1.00	1.00	0.26
		1000	0.95	0.24	0.63	1.00	1.00	1.00	0.99	0.98	1.00	1.00	0.04
		2000	0.94	0.18	0.72	0.98	0.99	0.99	1.00	0.97	1.00	1.00	0.00
		5000	0.95	0.70	0.92	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.00
S2	0.4	500	0.95	0.36	0.74	0.90	0.99	0.99	0.97	0.97	1.00	1.00	0.01
		1000	0.95	0.51	0.90	0.97	1.00	1.00	0.99	0.99	1.00	1.00	0.00
		2000	0.96	0.93	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.00
		5000	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
S3	0.2	500	0.97	0.63	0.64	0.90	0.99	1.00	1.00	0.98	1.00	1.00	0.63
		1000	0.95	0.40	0.63	0.92	0.99	0.99	0.99	0.96	1.00	1.00	0.17
		2000	0.95	0.38	0.73	0.96	0.98	0.99	0.98	0.98	1.00	1.00	0.00
		5000	0.96	0.72	0.93	0.99	1.00	1.00	1.00	0.99	1.00	1.00	0.00
S3	0.4	500	0.93	0.45	0.73	0.60	0.96	0.97	0.93	0.94	0.98	1.00	0.22
		1000	0.95	0.66	0.87	0.71	0.99	1.00	0.92	0.98	1.00	1.00	0.01
		2000	0.94	0.86	0.93	0.98	1.00	1.00	0.99	0.99	1.00	1.00	0.00
		5000	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
S4	0.2	500	0.93	0.75	0.64	0.94	0.96	0.98	0.96	0.94	0.98	0.97	0.00
		1000	0.93	0.49	0.56	0.96	0.95	0.98	0.97	0.93	0.98	0.97	0.00
		2000	0.95	0.41	0.60	0.97	0.92	0.97	0.96	0.92	0.98	0.94	0.00
		5000	0.93	0.77	0.88	0.95	0.93	0.95	0.96	0.94	0.98	0.94	0.00
S4	0.4	500	0.94	0.48	0.54	0.69	0.84	0.91	0.73	0.84	0.92	0.94	0.00
		1000	0.94	0.33	0.81	0.92	0.90	0.96	0.92	0.88	0.96	0.93	0.00
		2000	0.93	0.73	0.91	0.95	0.95	0.95	0.96	0.94	0.97	0.93	0.00
		5000	0.98	0.97	0.97	0.98	0.98	0.98	0.98	0.97	0.99	0.97	0.00
S5	0.2	500	0.94	0.45	0.50	0.83	0.91	0.93	0.87	0.84	0.93	0.98	0.12
		1000	0.94	0.29	0.54	0.68	0.90	0.93	0.84	0.88	0.95	0.97	0.00
		2000	0.96	0.28	0.57	0.75	0.91	0.95	0.81	0.89	0.96	0.96	0.00
		5000	0.95	0.79	0.89	0.96	0.93	0.96	0.96	0.94	0.97	0.95	0.00
S5	0.4	500	0.95	0.26	0.43	0.38	0.73	0.81	0.63	0.70	0.86	0.93	0.00
		1000	0.95	0.25	0.77	0.67	0.86	0.93	0.66	0.86	0.93	0.93	0.00
		2000	0.95	0.58	0.92	0.97	0.94	0.97	0.99	0.95	0.99	0.96	0.00
		5000	0.94	0.92	0.93	0.97	0.96	0.97	0.97	0.96	0.99	0.96	0.00

Table 3: The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of $\mathcal{V}$, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching (or sampling) CI with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. The columns indexed with `Union` represent the union of TSLS estimators, which pass the Sargan test. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the `Union` methods assuming two valid IVs and the majority rule, respectively.

Average Length of Confidence Intervals for $\gamma_0 = 0.5$

						Proposed Searching			Proposed Sampling			Union	
Set	τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$p_z - 1$	$\lceil p_z/2 \rceil$
S2	0.2	500	0.13	0.10	0.10	0.58	0.61	0.65	0.32	0.34	0.38	2.45	0.07
		1000	0.09	0.11	0.08	0.38	0.42	0.43	0.24	0.26	0.29	1.45	0.01
		2000	0.06	0.13	0.06	0.25	0.28	0.29	0.16	0.18	0.20	0.75	0.00
		5000	0.04	0.08	0.04	0.13	0.16	0.16	0.09	0.10	0.11	0.28	0.00
S2	0.4	500	0.13	0.18	0.12	0.51	0.60	0.64	0.37	0.36	0.44	2.57	0.00
		1000	0.09	0.21	0.09	0.34	0.38	0.39	0.23	0.22	0.27	1.49	0.00
		2000	0.06	0.07	0.06	0.25	0.25	0.26	0.15	0.15	0.17	0.73	0.00
		5000	0.04	0.04	0.04	0.16	0.16	0.16	0.09	0.09	0.10	0.27	0.00
S3	0.2	500	0.13	0.09	0.10	0.57	0.66	0.72	0.45	0.36	0.51	1.77	0.13
		1000	0.09	0.08	0.08	0.35	0.42	0.43	0.26	0.26	0.30	1.36	0.03
		2000	0.06	0.11	0.06	0.25	0.28	0.29	0.17	0.18	0.21	0.87	0.00
		5000	0.04	0.08	0.04	0.14	0.16	0.16	0.09	0.10	0.11	0.33	0.00
S3	0.4	500	0.13	0.10	0.13	0.43	0.66	0.72	0.57	0.40	0.64	1.90	0.03
		1000	0.09	0.23	0.09	0.30	0.39	0.42	0.26	0.24	0.31	1.42	0.00
		2000	0.06	0.15	0.06	0.25	0.26	0.28	0.16	0.15	0.18	0.82	0.00
		5000	0.04	0.05	0.04	0.16	0.16	0.16	0.09	0.09	0.10	0.32	0.00
S4	0.2	500	0.23	0.34	0.17	0.53	0.55	0.59	0.51	0.45	0.57	0.88	0.00
		1000	0.16	0.15	0.13	0.38	0.35	0.39	0.37	0.29	0.39	0.42	0.00
		2000	0.11	0.12	0.10	0.23	0.21	0.24	0.23	0.18	0.25	0.20	0.00
		5000	0.07	0.08	0.07	0.12	0.11	0.12	0.10	0.09	0.11	0.09	0.00
S4	0.4	500	0.23	0.30	0.23	0.45	0.49	0.60	0.51	0.39	0.63	0.80	0.00
		1000	0.16	0.19	0.16	0.39	0.28	0.41	0.61	0.22	0.64	0.33	0.00
		2000	0.11	0.12	0.11	0.20	0.18	0.20	0.26	0.14	0.27	0.14	0.00
		5000	0.07	0.08	0.07	0.11	0.11	0.11	0.09	0.09	0.10	0.08	0.00
S5	0.2	500	0.23	0.25	0.17	0.43	0.52	0.55	0.39	0.41	0.49	1.00	0.05
		1000	0.16	0.19	0.12	0.26	0.35	0.37	0.29	0.29	0.36	0.50	0.00
		2000	0.11	0.13	0.10	0.20	0.22	0.25	0.21	0.18	0.24	0.23	0.00
		5000	0.07	0.08	0.07	0.12	0.11	0.12	0.10	0.09	0.11	0.09	0.00
S5	0.4	500	0.23	0.31	0.22	0.33	0.48	0.60	0.51	0.38	0.67	0.97	0.00
		1000	0.16	0.15	0.17	0.35	0.29	0.44	0.60	0.22	0.67	0.40	0.00
		2000	0.11	0.11	0.11	0.22	0.18	0.22	0.34	0.14	0.35	0.15	0.00
		5000	0.07	0.08	0.07	0.11	0.11	0.11	0.09	0.09	0.10	0.08	0.00

Table 4: The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of $\mathcal{V}$, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching (or sampling) CI with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. The columns indexed with `Union` represent the union of TSLS estimators, which pass the Sargan test. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the `Union` methods assuming two valid IVs and the majority rule, respectively.

sample size n, the setting **CIIV-1** does not necessarily satisfy the finite-sample plurality rule (Condition 2) since τ and $\tau/2$ are close to each other. For the setting **CIIV-2**, the invalid levels are more spread out and the finite-sample plurality rule may hold more plausibly.

In Table 5, we consider the setting **CIIV-1** and compare different CIs in terms of empirical coverage and average lengths. In terms of coverage, our proposed (combined) searching and sampling CIs attain the desired coverage level (95%); CIs by the `Union` method achieve the desired coverage level; `CIIV` achieves the desired 95% coverage level for $\tau = 0.2$ with $n = 5000$ and $\tau = 0.4$ with $n = 2000, 5000$; and `TSHT` achieves the desired 95% coverage level only for $\tau = 0.4$ with $n = 5000$. In terms of interval lengths, the sampling and searching CIs are much shorter than the CIs by the `Union` method.

Empirical Coverage for **CIIV-1**

					Proposed Searching			Proposed Sampling			`Union`
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$p_z - 1$
0.2	500	0.94	0.00	0.13	1.00	1.00	1.00	0.84	0.55	0.88	1.00
	1000	0.95	0.00	0.44	1.00	0.94	1.00	0.92	0.73	0.94	1.00
	2000	0.96	0.00	0.76	0.73	0.95	0.98	0.92	0.92	0.97	1.00
	5000	0.96	0.01	0.93	0.06	1.00	1.00	0.11	1.00	1.00	1.00
0.4	500	0.94	0.00	0.65	0.85	0.89	0.96	0.94	0.85	0.96	1.00
	1000	0.94	0.00	0.89	0.02	0.99	0.99	0.12	0.99	0.99	1.00
	2000	0.94	0.13	0.94	0.58	0.92	0.92	0.59	0.92	0.92	1.00
	5000	0.95	0.91	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Average Length of Confidence Intervals for **CIIV-1**

					Proposed Searching			Proposed Sampling			`Union`
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$p_z - 1$
0.2	500	0.09	0.06	0.09	1.07	1.01	1.12	0.48	0.36	0.51	1.40
	1000	0.07	0.04	0.07	0.68	0.62	0.77	0.42	0.26	0.45	1.09
	2000	0.05	0.03	0.05	0.40	0.42	0.57	0.34	0.19	0.38	0.91
	5000	0.03	0.05	0.03	0.05	0.26	0.27	0.26	0.12	0.35	0.72
0.4	500	0.09	0.06	0.10	1.04	0.96	1.39	0.89	0.39	0.95	2.04
	1000	0.07	0.06	0.07	0.22	0.63	0.70	0.48	0.27	0.67	1.66
	2000	0.05	0.22	0.05	0.20	0.39	0.41	0.19	0.18	0.29	1.27
	5000	0.03	0.04	0.03	0.26	0.26	0.26	0.12	0.12	0.13	0.77

Table 5: The columns indexed with `oracle`, `TSHT`, `CIIV` and represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a combination of the corresponding two intervals. The columns indexed with `Union` represent the union of TSLS estimators using two candidate IVs, which pass the Sargan test.

Additional simulation results for settings **CIIV-1** and **CIIV-2** are reported in Section

D.2 in the supplementary material.

8 Real Data Analysis

We apply our proposed method to a stock mouse data set [2] and study the effect of the triglyceride level on the glucose level. After removing the missing values, the data consists of $1,269$ subjects, where for each subject, $10,346$ polymorphic genetic markers, the triglyceride level, the glucose level and baseline covariates (age and sex) are measured. The polymorphic genetic markers and baseline covariates are standardized before analysis.

We follow [25] and construct factor IVs using the polymorphic genetic markers. We sketch the two-step construction in the following and the exact details about the construction of factor IVs can be found in Section 7.1 of [25]. We first run marginal regressions of the triglyceride level (treatment) over $10,346$ polymorphic genetic markers and select the markers with p-values below 10^{-3}. We then conduct the PCA over the selected markers and output the leading principle components as the constructed IVs. For the the triglyceride level, 14 factor IVs have been constructed.

We plot the searching CI (in blue) and the sampling CI (in red) in Figure 5. Out of the 1000 sampled intervals, 83 of them are non-empty and the union of these 83 intervals is shorter than the searching CI.

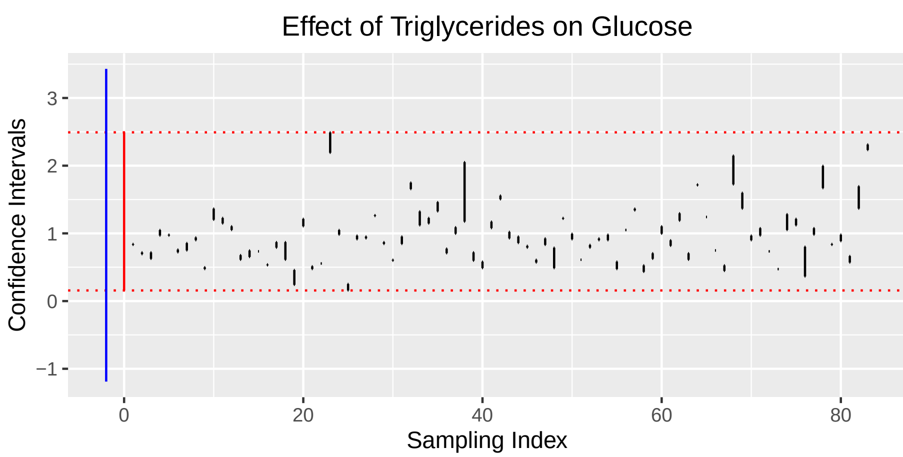

Figure 5: The axis corresponds to sampling indexes $\{1, 2, \cdots, 83\}$ (after re-ordering) and the y-axis reports the sampled CIs. Along the y-axis, the red interval is $\mathrm{CI}^{\mathrm{sample}} = (0.1463, 2.4989)$ and the blue interval is $\mathrm{CI}^{\mathrm{search}} = (-1.1895, 3.4300)$.

In Table 6, we compare our proposed searching and sampling CIs with existing methods.

[2] The data set is available at `https://wp.cs.ucl.ac.uk/outbredmice/heterogeneous-stock-mice/`

Method	CI	Method	CI
OLS	(0.5026, 0.7982)	Searching CI	(-1.1895, 3.4300)
TSLS	(0.5458, 1.4239)	Sampling CI	(0.1463, 2.4989)
TSHT	(0.4204, 1.3513)	Union ($\bar{s} = p_{\mathrm{z}} - 1$)	(-22.271, 25.785)
CIIV	(0.5465, 1.4232)	Union ($\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$)	(-0.9421, 4.3087)

Table 6: Confidence intervals for the effect of the triglyceride level on the glucose level.

TSHT selects four IVs as valid and the CIIV returns the same set of valid IVs [3]. CIs by TSHT and CIIV are relatively short but they may be under-coverage due to the post-selection problem. Regarding the Union method, we report the Sargan TSLS estimator from [18]. If the majority rule is satisfied (at least half of candidate IVs are valid), then the CI by Union with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ is valid; if we can only assume that two of candidate IVs are valid, then the CI by Union with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ may not be valid but the CI by Union with $\bar{s} = p_{\mathrm{z}} - 1$ is valid. The CI by Union with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ is shorter than that with $\bar{s} = p_{\mathrm{z}} - 1$. The searching CI is of a similar length with the Union with with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$. However, the validity of the searching CI only relies on the finite-sample plurality rule instead of the stronger majority rule. The length of the sampling CI is much shorter than other uniform inference methods, including the sampling CI, Union ($\bar{s} = p_{\mathrm{z}} - 1$) and Union ($\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$).

References

[1] Theodore W Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical statistics*, 20(1):46–63, 1949.

[2] Richard Berk, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

[3] Jack Bowden, George Davey Smith, and Stephen Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology*, 44(2):512–525, 2015.

[4] Jack Bowden, George Davey Smith, and Stephen Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology*, 44(2):512–525, 2015.

[3] CIIV reports invalid IVs and we estimate the set of valid IVs by removing the invalid IVs from $\widehat{\mathcal{S}}$.

[5] Jack Bowden, George Davey Smith, Philip C Haycock, and Stephen Burgess. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology*, 40(4):304–314, 2016.

[6] Stephen Burgess, Dylan S Small, and Simon G Thompson. A review of instrumental variable estimators for mendelian randomization. *Statistical methods in medical research*, 26(5):2333–2355, 2017.

[7] T Tony Cai and Zijian Guo. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646, 2017.

[8] Victor Chernozhukov, Christian Hansen, and Martin Spindler. Post-selection and post-regularization inference in linear models with many controls and instruments. *The American Economic Review*, 105(5):486–490, 2015.

[9] George Davey Smith and Shah Ebrahim. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology*, 32(1):1–22, 2003.

[10] George Davey Smith and Gibran Hemani. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*, 23(R1):R89–R98, 2014.

[11] Qingliang Fan and Yaqian Wu. Endogenous treatment effect estimation with some invalid and irrelevant instruments. *arXiv preprint arXiv:2006.14998*, 2020.

[12] Zijian Guo. Inference for high-dimensional maximin effects in heterogeneous regression models using a sampling approach. *arXiv preprint arXiv:2011.07568*, 2020.

[13] Zijian Guo, Hyunseung Kang, T Tony Cai, and Dylan S Small. Testing endogeneity with high dimensional covariates. *Journal of Econometrics*, 207(1):175–187, 2018.

[14] Zijian Guo, Hyunseung Kang, T Tony Cai, and Dylan S Small. Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):793–815, 2018.

[15] Chirok Han. Detecting invalid instruments using l1-gmm. *Economics Letters*, 101(3):285–287, 2008.

[16] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

[17] Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

[18] Hyunseung Kang, Youjin Lee, T Tony Cai, and Dylan S Small. Two robust tools for inference about causal effects with invalid instruments. *Biometrics*, 2020.

[19] Hyunseung Kang, Anru Zhang, T Tony Cai, and Dylan S Small. Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American Statistical Association*, 111(513):132–144, 2016.

[20] Michal Kolesár, Raj Chetty, John Friedman, Edward Glaeser, and Guido W Imbens. Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics*, 33(4):474–484, 2015.

[21] Jason D Lee, Dennis L Sun, Yuekai Sun, and Jonathan E Taylor. Exact post-selection inference, with application to the lasso. *Annals of Statistics*, 44(3):907–927, 2016.

[22] Hannes Leeb and Benedikt M Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, pages 21–59, 2005.

[23] Arthur Lewbel. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80, 2012.

[24] Arthur Lewbel. Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Economics Letters*, 165:10–12, 2018.

[25] Sai Li and Zijian Guo. Causal inference for nonlinear outcome models with possibly invalid instrumental variables. *arXiv preprint arXiv:2010.09922*, 2020.

[26] Anna Mikusheva. Survey on statistical inferences in weakly-identified instrumental variable models. *Applied Econometrics*, 29(1):117–131, 2013.

[27] John D Sargan. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, pages 393–415, 1958.

[28] Dylan S Small. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association*, 102(479):1049–1058, 2007.

[29] Wes Spiller, David Slichter, Jack Bowden, and George Davey Smith. Detecting and correcting for bias in mendelian randomization analyses using gene-by-environment interactions. *International journal of epidemiology*, 48(3):702–712, 2019.

[30] Eric J Tchetgen Tchetgen, BaoLuo Sun, and Stefan Walter. The genius approach to robust mendelian randomization inference. *arXiv preprint arXiv:1709.07779*, 2017.

[31] Sara van de Geer, Peter Bühlmann, Yaacov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

[32] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[33] Frank Windmeijer, Helmut Farbmacher, Neil Davies, and George Davey Smith. On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association*, 114(527):1339–1350, 2019.

[34] Frank AG Windmeijer, Xiaoran Liang, Fernando P Hartwig, and Jack Bowden. The confidence interval method for selecting valid instrumental variables. 2019.

[35] Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

[36] Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.

A Additional Method and Theory

A.1 Equivalent Definition of $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ in (36)

Recall that the winner set is defined as

$$\widehat{\mathcal{W}} = \arg\max_{1 \leq k \leq |\widehat{\mathcal{S}}|} \|\widehat{\Pi}_{k\cdot}\|_0.$$

With $\widehat{\mathcal{W}}$, we further define the index set $\widetilde{\mathcal{V}}$ as

$$\widetilde{\mathcal{V}} = \cup_{k \in \widehat{\mathcal{W}}} \left\{ 1 \leq j \leq |\widehat{\mathcal{S}}| : \widehat{\Pi}_{j,k} = 1 \right\} \tag{44}$$

The set $\widetilde{\mathcal{V}}$ denotes the set of IVs who support (and are also supported by) at least one element in $\widehat{\mathcal{W}}$. We finally construct the index set $\widehat{\mathcal{V}} \subset \{1, 2, \cdots, |\widehat{\mathcal{S}}|\}$ as

$$\widehat{\mathcal{V}}^{\texttt{TSHT}} = \cup_{j \in \widetilde{\mathcal{V}}} \left\{ 1 \leq l \leq |\widehat{\mathcal{S}}| : \widehat{\Pi}_{j,l} = 1 \right\}. \tag{45}$$

This set $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ contains all candidate IVs that are claimed to be valid by at least one element from $\widetilde{\mathcal{V}}$. The set defined in (45) is equivalent to that in (36).

A.2 Algorithm under the finite-sample Majority Rule

In Algorithm 2, we summarize a simplified uniform inference procedure by assuming the finite sample majority rule.

Algorithm 2 Uniform inference with Searching and Sampling (Majority Rule)

Input: Outcome $Y \in \mathbb{R}^n$; Treatment $D \in \mathbb{R}^n$; Candidate IVs $Z \in \mathbb{R}^{n \times p_{\mathrm{z}}}$; Baseline Covariates $X \in \mathbb{R}^{n \times p_{\mathrm{x}}}$; sampling number M; significance level $\alpha \in (0,1)$

Output: Confidence intervals $\mathrm{CI}^{\mathrm{search}}$ and $\mathrm{CI}^{\mathrm{sample}}$

```
Construct reduced form estimators $\widehat{\Gamma} \in \mathbb{R}^{p_{\mathrm{z}}}$ and $\widehat{\gamma} \in \mathbb{R}^{p_{\mathrm{z}}}$ as in (9)
Construct the noise level estimator $\widehat{\sigma}_{\epsilon}^2, \widehat{\sigma}_{\delta}^2$ and $\widehat{\sigma}_{\epsilon,\delta}$ as in (10).
Construct $\mathcal{B} \subset [A_1, A_2]$ with the grid size $n^{-0.6}$ and $A_1$ and $A_2$ defined in (27)
Select the set of relevant IVs $\widehat{\mathcal{S}}$ as in (12)                ▷ Construction of $\widehat{\mathcal{S}}$ and $\mathcal{B}$

Compute $\widehat{\rho}(\alpha) = \sqrt{2.01 \log |\mathcal{B}|}$ or the $\alpha$ quantile of $\{T_l\}_{1 \leq l \leq L}$ with $T_l$ in (28)
Construct $\mathrm{CI}^{\mathrm{search}}$ in (20) with $\widehat{\pi}_j(\beta)$ defined in (19)    ▷ Construction of Searching CI

Compute $\lambda = (\log n / M)^{\frac{1}{2p_{\mathrm{z}}}}$
for $m \leftarrow 1$ to $M$ do
    Sample $\widehat{\Gamma}^{[m]}$ and $\widehat{\gamma}^{[m]}$ as in (21)
    Compute $\{\widehat{\pi}_j^{[m]}(\beta, \lambda)\}_{j \in \widehat{\mathcal{S}}, \beta \in \mathcal{B}}$ as in (23);
    Construct the interval $(\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda))$ with $\beta_{\min}^{[m]}(\lambda)$ and $\beta_{\max}^{[m]}(\lambda)$ in (24)
end for
Construct $\mathrm{CI}^{\mathrm{sample}}$ as in (26)                ▷ Construction of Sampling CI
```

B Proofs

B.1 Proof Preparation

Throughout the proof, we focus on the low-dimension setting with homoskedastic errors. To estimate γ and Γ, we focus on the OLS estimator $\widehat{\gamma}$ and $\widehat{\Gamma}$ defined in (9), which satisfy

$$\widehat{\gamma}_j - \gamma_j^* = \widehat{\Omega}_{j\cdot}^{\intercal} \frac{1}{n} W^{\intercal}\delta \quad \text{and} \quad \widehat{\Gamma}_j - \Gamma_j^* = \widehat{\Omega}_{j\cdot}^{\intercal} \frac{1}{n} W^{\intercal}\epsilon \quad \text{for} \quad 1 \leq j \leq p. \tag{46}$$

For the OLS estimators, they satisfy the limiting distribution in (8) with $\mathbf{V}_{jj}^{\gamma} = \sigma_{\delta}^2 \Omega_{jj}$, $\mathbf{V}_{jj}^{\Gamma} = \sigma_{\epsilon}^2 \Omega_{jj}$, and $\mathbf{C}_{jj} = \sigma_{\epsilon,\delta}\Omega_{jj}$ for $1 \leq j \leq p_{\mathrm{z}}$ where $\sigma_{\epsilon,\delta} = \mathrm{Cov}(\epsilon_1, \delta_1)$ and $\Omega = \Sigma^{-1}$. The noise levels $\sigma_{\delta}^2, \sigma_{\epsilon}^2$ and $\sigma_{\epsilon,\delta}$ are estimated in (10). Since Cov in (8) can be expressed as

$$\begin{pmatrix} \mathbf{V}^{\Gamma} & \mathbf{C} \\ \mathbf{C}^{\intercal} & \mathbf{V}^{\gamma} \end{pmatrix} = \begin{pmatrix} \mathrm{I} & \frac{\sigma_{\epsilon,\delta}}{\sigma_{\delta}^2}\mathrm{I} \\ 0 & \mathrm{I} \end{pmatrix} \begin{pmatrix} \left(1 - \frac{\sigma_{\epsilon,\delta}^2}{\sigma_{\epsilon}^2\sigma_{\delta}^2}\right)\mathbf{V}^{\Gamma} & 0 \\ 0^{\intercal} & \mathbf{V}^{\gamma} \end{pmatrix} \begin{pmatrix} \mathrm{I} & \frac{\sigma_{\epsilon,\delta}}{\sigma_{\delta}^2}\mathrm{I} \\ 0 & \mathrm{I} \end{pmatrix}^{\intercal}$$

we have

$$\lambda_{\min}(\mathrm{Cov}) \geq \min\left\{\sigma_{\epsilon}^2 - \sigma_{\epsilon,\delta}^2/\sigma_{\delta}^2, \sigma_{\delta}^2\right\} \cdot \lambda_{\min}(\Sigma^{-1}). \tag{47}$$

Define

$$\mathcal{S}^0 = \left\{1 \leq j \leq p_{\mathrm{z}} : |\gamma_j^*| \geq (\sqrt{\log n} - C(\log n)^{1/4}) \cdot \sqrt{\widehat{\mathbf{V}}_{jj}^{\gamma}/n}\right\}. \tag{48}$$

Define the following events

$$\begin{aligned}
\mathcal{G}_0 &= \left\{\max\left\{\left\|\frac{1}{n}W^{\intercal}\epsilon\right\|_{\infty}, \left\|\frac{1}{n}W^{\intercal}\delta\right\|_{\infty}\right\} \leq C\frac{(\log n)^{1/4}}{\sqrt{n}}\right\} \\
\mathcal{G}_1 &= \left\{\max_{1\leq j\leq p} \max\left\{|\widehat{\gamma}_j - \gamma_j^*|/\sqrt{\mathbf{V}_{jj}^{\gamma}/n}, |\widehat{\Gamma}_j - \Gamma_j^*|/\sqrt{\mathbf{V}_{jj}^{\Gamma}/n}\right\} \leq C(\log n)^{1/4}\right\} \\
\mathcal{G}_2 &= \left\{\max\{\left|\widehat{\sigma}_{\epsilon}^2 - \sigma_{\epsilon}^2\right|, \left|\widehat{\sigma}_{\delta}^2 - \sigma_{\delta}^2\right|, |\widehat{\sigma}_{\epsilon,\delta} - \sigma_{\epsilon,\delta}|\} \leq C\sqrt{\frac{\log n}{n}}\right\} \\
\mathcal{G}_3 &= \left\{\|\widehat{\Omega} - \Sigma^{-1}\|_2 \leq C\sqrt{\frac{\log n}{n}}\right\} \\
\mathcal{G}_4 &= \left\{\max\left\{\|\widehat{\mathbf{V}}^{\Gamma} - \mathbf{V}^{\Gamma}\|_2, \|\widehat{\mathbf{V}}^{\gamma} - \mathbf{V}^{\gamma}\|_2, \|\widehat{\mathbf{C}} - \mathbf{C}\|_2\right\} \leq C\sqrt{\frac{\log n}{n}}\right\} \\
\mathcal{G}_5 &= \left\{\mathcal{S}_{\mathrm{str}} \subset \widehat{\mathcal{S}} \subset \mathcal{S}^0 \subset \mathcal{S}\right\} \\
\mathcal{G}_6 &= \left\{\max_{j,k\in\widehat{\mathcal{S}}} \left|\frac{\widehat{\gamma}_k/\widehat{\gamma}_j}{\gamma_k^*/\gamma_j^*} - 1\right| \leq C\frac{1}{(\log n)^{1/4}}\right\} \\
\mathcal{G}_7 &= \left\{\max_{j\in\widehat{\mathcal{S}}} \left|\frac{\widehat{\Gamma}_j}{\widehat{\gamma}_j} - \frac{\Gamma_j^*}{\gamma_j^*}\right| \leq C\left(1 + \left|\frac{\Gamma_j^*}{\gamma_j^*}\right|\right)\frac{1}{(\log n)^{1/4}}\right\}.
\end{aligned} \tag{49}$$

Define

$$\mathcal{G} = \cap_{j=0}^{7} \mathcal{G}_j.$$

The following lemma controls the probability of $\mathcal{G}$, whose proof can be found in Section C.1.

Lemma 1 *Suppose that conditions* (C1) *and* (C2) *hold, then for a sufficiently large n,*

$$\mathbb{P}(\mathcal{G}) \geq 1 - \exp(-c\sqrt{\log n})$$

for some positive constant $c > 0$.

We define the event

$$\mathcal{E}_0(\alpha) = \left\{ \max_{\beta \in \mathcal{B}} \max_{j \in \widehat{\mathcal{S}}} \frac{|\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2 \widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta \widehat{\mathbf{C}}_{jj})/n}} \leq \widehat{\rho}(\alpha) \right\}. \tag{50}$$

The following lemma justifies our theoretical choices of $\widehat{\rho}(\alpha)$, whose proof can be found in Section C.2.

Lemma 2 *Suppose that conditions* (C1) *and* (C2) *hold. There exists a positive constant $C > 0$ and a positive integer $N_0 > 0$ such that for $n \geq N_0$ and $\widehat{\rho}(\alpha) = C\sqrt{\log |\mathcal{B}|}$, the event $\mathcal{E}_0(\alpha)$ defined in* (50) *satisfies $\mathbb{P}(\mathcal{E}_0(\alpha)) \geq 1 - \alpha$. Furthermore, if $(\epsilon_i, \delta_i)^\intercal$ is bivariate normal and independent of $W_{i\cdot}$, then the event $\mathcal{E}_0(\alpha)$ defined in* (50) *satisfies $\mathbb{P}(\mathcal{E}_0(\alpha)) \geq 1 - \alpha$, with the threshold $\widehat{\rho}(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2|\mathcal{B}| \cdot p_z}\right)$ or $\widehat{\rho}(\alpha) = \sqrt{2.005 \log |\mathcal{B}|}$.*

B.2 Proof of Theorem 1

Coverage property of $\mathrm{CI}^{\rm search}$ in (20). By the decomposition (16), if β is taken as β^*, then

$$\widehat{\Gamma}_j - \beta^* \widehat{\gamma}_j = \widehat{\Gamma}_j - \Gamma_j^* - \beta^*(\widehat{\gamma}_j - \gamma_j^*) + \pi_j^*.$$

Hence, on the event $\mathcal{E}_0(\alpha)$ defined in (50), we have

$$\left| \widehat{\Gamma}_j - \beta^* \widehat{\gamma}_j \right| \leq \widehat{\rho}_j(\beta, \alpha) \quad \text{for all} \quad j \in \mathcal{V} \cap \widehat{\mathcal{S}}$$

where $\widehat{\rho}_j(\beta, \alpha)$ is defined in (18). This leads to

$$\left| \left\{ j \in \widehat{\mathcal{S}} : \left| \widehat{\Gamma}_j - \beta^* \widehat{\gamma}_j \right| \leq \widehat{\rho}_j(\beta, \alpha) \right\} \right| \geq \left| \mathcal{V} \cap \widehat{\mathcal{S}} \right|. \tag{51}$$

On the event $\mathcal{G}_5$ defined in (50), we have

$$|\mathcal{V} \cap \widehat{\mathcal{S}}| \geq |\mathcal{V} \cap \mathcal{S}_{\rm str}| > \frac{|\mathcal{S}|}{2} \geq \frac{|\widehat{\mathcal{S}}|}{2}. \tag{52}$$

where the second inequality follows from the finite-sample majority rule and the last inequality follows from the definition of $\mathcal{G}_5$ in (50).

By combining (51) and (52), we show that, on the event $\mathcal{E}_0(\alpha) \cap \mathcal{G}_5$,

$$\|\widehat{\pi}(\beta^*)\|_0 \leq \left|\widehat{\mathcal{S}}\right| - \left|\mathcal{V} \cap \widehat{\mathcal{S}}\right| < \frac{|\widehat{\mathcal{S}}|}{2}.$$

It follows from the definition of $\mathrm{CI}^{\mathrm{search}}$ in (20) that on the event $\mathcal{E}_0(\alpha)\cap\mathcal{G}_5$, $\beta^* \in (\beta_{\min}, \beta_{\max})$. Hence,

$$\mathbf{P}\left(\beta^* \in (\beta_{\min}, \beta_{\max})\right) \geq \mathbf{P}\left(\mathcal{E}_0(\alpha) \cap \mathcal{G}_5\right).$$

We establish the coverage property by applying Lemma 1 and $\widehat{\rho}(\alpha)$ satisfying (17).

Length of $\mathrm{CI}^{\mathrm{search}}$ in (20). We consider the j-th IV such that $\pi_j^* = 0$ and simplify the decomposition in (16) as

$$\widehat{\Gamma}_j - \beta\widehat{\gamma}_j = \widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*) + (\beta^* - \beta)\gamma_j^* \tag{53}$$

For β satisfying $|\gamma_j^*| \cdot |\beta - \beta^*| \geq 2\widehat{\rho}_j(\beta, \alpha)$, we have $\widehat{\pi}_j(\beta) \neq 0$ if the event $\mathcal{E}_0(\alpha)$ holds. If β satisfies

$$|\beta - \beta^*| \geq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{2\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|} \tag{54}$$

then on the event $\mathcal{E}_0(\alpha) \cap \mathcal{G}_5$,

$$\|\widehat{\pi}(\beta)\|_0 \geq |\widehat{\mathcal{S}} \cap \mathcal{V}| > \frac{|\widehat{\mathcal{S}}|}{2},$$

where the second inequality follows from (52). Hence, on the event $\mathcal{E}_0(\alpha) \cap \mathcal{G}_5$, if β satisfies (54), then $\beta \notin \mathrm{CI}^{\mathrm{search}}$ and hence

$$|\beta_{\max} - \beta_{\min}| \leq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{4\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|}.$$

Together with the fact that $\widehat{\rho}_j(\beta, \alpha) \leq \sqrt{\log n / n}$, we establish the high probability upper bound for the length of $\mathrm{CI}^{\mathrm{search}}$.

B.3 Proof of Proposition 1

Denote all data by $\mathcal{O}$, that is, $\mathcal{O} = \{Y_{i\cdot}, D_{i\cdot}, Z_{i\cdot}, X_{i\cdot}\}_{1 \leq i \leq n}$. Define

$$\widehat{U} = \sqrt{n}\left[\begin{pmatrix}\widehat{\Gamma} \\ \widehat{\gamma}\end{pmatrix} - \begin{pmatrix}\Gamma^* \\ \gamma^*\end{pmatrix}\right] \quad \text{and} \quad U^{[m]} = \sqrt{n}\left[\begin{pmatrix}\widehat{\Gamma} \\ \widehat{\gamma}\end{pmatrix} - \begin{pmatrix}\widehat{\Gamma}^{[m]} \\ \widehat{\gamma}^{[m]}\end{pmatrix}\right] \quad \text{for} \quad 1 \leq m \leq M.$$

Define

$$\mathrm{Cov} = \begin{pmatrix}\mathbf{V}^{\Gamma} & \mathbf{C} \\ \mathbf{C}^{\intercal} & \mathbf{V}^{\gamma}\end{pmatrix} \quad \text{and} \quad \widehat{\mathrm{Cov}} = \begin{pmatrix}\widehat{\mathbf{V}}^{\Gamma} & \widehat{\mathbf{C}} \\ \widehat{\mathbf{C}}^{\intercal} & \widehat{\mathbf{V}}^{\gamma}\end{pmatrix}$$

Recall that $\widehat{U}$ is a function of the observed data $\mathcal{O}$

$$\widehat{U} \xrightarrow{d} N\left(\mathbf{0}, \mathrm{Cov}\right) \quad \text{and} \quad U^{[m]}|\mathcal{O} \overset{i.i.d.}{\sim} N\left(\mathbf{0}, \widehat{\mathrm{Cov}}\right) \quad \text{for} \quad 1 \leq m \leq M.$$

Let $f(U \mid \mathcal{O})$ denote the conditional density function of $U^{[m]}$ given the data $\mathcal{O}$, that is,

$$f(U \mid \mathcal{O}) = \frac{1}{\sqrt{(2\pi)^{2p_z}\det(\widehat{\mathrm{Cov}})}} \exp\left(-\frac{1}{2}U^{\intercal}\widehat{\mathrm{Cov}}^{-1}U\right).$$

We define the following event for the data $\mathcal{O}$,

$$\mathcal{E}_1 = \left\{\|\widehat{\mathrm{Cov}} - \mathrm{Cov}\|_2 < c_2\right\} \tag{55}$$

where $\|\widehat{\mathrm{Cov}} - \mathrm{Cov}\|_2$ denotes the spectral norm of the matrix $\widehat{\mathrm{Cov}} - \mathrm{Cov}$ and $0 < c_2 < \lambda_{\min}(\mathrm{Cov})/2$ is a small positive constant.

We define the following function to facilitate the proof,

$$g(U) = \frac{1}{\sqrt{(2\pi)^{2p_z}\det(\mathrm{Cov} + c_2\mathbf{I})}} \exp\left(-\frac{1}{2}U^{\intercal}(\mathrm{Cov} - c_2\mathbf{I})^{-1}U\right). \tag{56}$$

We define the following event for the data $\mathcal{O}$,

$$\mathcal{E}_2 = \left\{g(\widehat{U}) \cdot \mathbf{1}_{\mathcal{E}_1} \geq c^*(\alpha_0)\right\} \tag{57}$$

with $c^*(\alpha_0)$ defined in (42). The following lemma shows that the event $\mathcal{E}_1 \cap \mathcal{E}_2$ holds with a high probability and the proof is postponed to Section C.3.

Lemma 3 *Suppose that conditions* (C1) *and* (C2) *hold, then we have*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \alpha_0.$$

Since

$$\|U^{[m]} - \widehat{U}\|_\infty = \sqrt{n} \cdot \max\left\{\left\|\widehat{\gamma}^{[m]} - \gamma^*\right\|_\infty, \left\|\widehat{\Gamma}^{[m]} - \Gamma^*\right\|_\infty\right\},$$

it is sufficient to control

$$\mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M, \alpha_0)\right).$$

We use $\mathbb{P}(\cdot \mid \mathcal{O})$ to denote the conditional probability with respect to the observed data $\mathcal{O}$. Note that

$$\begin{aligned}
&\mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right) \\
&= 1 - \mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \geq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right) \\
&= 1 - \prod_{m=1}^{M}\left[1 - \mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\right] \\
&\geq 1 - \exp\left[-\sum_{m=1}^{M}\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\right],
\end{aligned} \tag{58}$$

where the second equality follows from the conditional independence of $\{U^{[m]}\}_{1\leq m\leq M}$ given the data $\mathcal{O}$ and the last inequality follows from $1 - x \leq e^{-x}$. Hence, we have

$$\begin{aligned}
&\mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq \left(1 - \exp\left[-\sum_{m=1}^{M}\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\right]\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&= 1 - \exp\left[-\sum_{m=1}^{M}\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}\right].
\end{aligned} \tag{59}$$

In the following, we first establish an lower bound for

$$\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}. \tag{60}$$

and then apply (58) and (59) to establish a lower bound for

$$\mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right).$$

On the event $\mathcal{O}\in\mathcal{E}_1$, we have $\mathrm{Cov} + c_2\mathbf{I} \succ \widehat{\mathrm{Cov}} \succ \mathrm{Cov} - c_2\mathbf{I} \succ \frac{1}{2}\lambda_{\min}(\mathrm{Cov})\cdot\mathbf{I}$ and hence

$$f(U^{[m]} \mid \mathcal{O})\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1} \geq g(U^{[m]})\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1}.$$

We apply the above inequality and further lower bound the targeted probability in (60) as

$$\begin{aligned}
&\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&= \int f(U^{[m]} \mid \mathcal{O})\mathbf{1}_{\left\{\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]}\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq \int g(U^{[m]})\mathbf{1}_{\left\{\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]}\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&= \int g(\widehat{U})\mathbf{1}_{\left\{\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]}\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\quad + \int [g(U^{[m]}) - g(\widehat{U})]\mathbf{1}_{\left\{\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]}\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}.
\end{aligned} \tag{61}$$

By the definition of $\mathcal{E}_2$ in (57), we establish

$$
\begin{aligned}
&\int g(\widehat{U})\mathbf{1}_{\left\{\|U^{[m]}-\widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq c^*(\alpha_0) \cdot \int \mathbf{1}_{\left\{\|U^{[m]}-\widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq c^*(\alpha_0) \cdot [2\mathrm{err}_n(M,\alpha_0)]^{2p_z} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}.
\end{aligned}
\tag{62}
$$

Since $\lambda_{\min}(\mathrm{Cov} - c_2\mathbf{I}) \geq \lambda_{\min}(\mathrm{Cov})/2$ and there exists $t \in (0,1)$ such that

$$
g(U^{[m]}) - g(\widehat{U}) = [\nabla g(\widehat{U} + t(U^{[m]} - \widehat{U}))]^\intercal (U^{[m]} - \widehat{U})
$$

with

$$
\nabla g(u) = \frac{1}{\sqrt{(2\pi)^{2p_z}\det(\mathrm{Cov} + c_2\mathbf{I})}} \exp\left(-\frac{1}{2}u^\intercal(\mathrm{Cov} - c_2\mathbf{I})^{-1}u\right)^{-1} (\mathrm{Cov} - c_2\mathbf{I})^{-1}u,
$$

then ∇g is bounded and there exists a positive constant $C > 0$ such that

$$
\left|g(U^{[m]}) - g(\widehat{U})\right| \leq C\sqrt{2p_z}\|U^{[m]} - \widehat{U}\|_\infty.
$$

Then we establish

$$
\begin{aligned}
&\left|\int [g(U^{[m]}) - g(\widehat{U})]\mathbf{1}_{\left\{\|U^{[m]}-\widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}\right| \\
&\leq C\sqrt{2p_z} \cdot \mathrm{err}_n(M,\alpha_0) \cdot \int \mathbf{1}_{\left\{\|U^{[m]}-\widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0)\right\}} dU^{[m]} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&= C\sqrt{2p_z} \cdot \mathrm{err}_n(M,\alpha_0) \cdot [2\mathrm{err}_n(M,\alpha_0)]^{2p_z} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}.
\end{aligned}
\tag{63}
$$

By assuming $C\sqrt{2p_z} \cdot \mathrm{err}_n(M,\alpha_0) \leq \frac{1}{2}c^*(\alpha_0)$, we combine (61), (62) and (63) and obtain

$$
\begin{aligned}
&\mathbb{P}\left(\|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right) \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq \frac{1}{2}c^*(\alpha_0) \cdot [2\mathrm{err}_n(M,\alpha_0)]^{2p_z} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}
\end{aligned}
$$

Together with (59), we establish

$$
\begin{aligned}
&\mathbb{P}\left(\min_{1\leq m\leq M} \|U^{[m]} - \widehat{U}\|_\infty \leq \mathrm{err}_n(M,\alpha_0) \mid \mathcal{O}\right) \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2} \\
&\geq 1 - \exp\left[-M \cdot \frac{1}{2}c^*(\alpha_0) \cdot [2\mathrm{err}_n(M,\alpha_0)]^{2p_z} \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}\right] \\
&= \left(1 - \exp\left[-M \cdot \frac{1}{2}c^*(\alpha_0) \cdot [2\mathrm{err}_n(M,\alpha_0)]^{2p_z}\right]\right) \cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}
\end{aligned}
\tag{64}
$$

With $\mathbf{E}_{\mathcal{O}}$ denoting the expectation taken with respect to the observed data $\mathcal{O}$, we further integrate with respect to $\mathcal{O}$ and establish

$$\begin{aligned}
&\mathbb{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_{\infty}\leq \mathrm{err}_n(M,\alpha_0)\right)\\
&=\mathbf{E}_{\mathcal{O}}\left[\mathbb{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_{\infty}\leq \mathrm{err}_n(M,\alpha_0)\mid \mathcal{O}\right)\right]\\
&\geq \mathbf{E}_{\mathcal{O}}\left[\mathbb{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_{\infty}\leq \mathrm{err}_n(M,\alpha_0)\mid \mathcal{O}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1\cap\mathcal{E}_2}\right]\\
&\geq \left(1-\exp\left[-M\cdot\frac{1}{2}c^*(\alpha_0)\cdot[2\mathrm{err}_n(M,\alpha_0)]^{2p_{\mathrm{z}}}\right]\right)\cdot\mathbb{P}\left(\mathcal{E}_1\cap\mathcal{E}_2\right)
\end{aligned}$$

We choose

$$\mathrm{err}_n(M,\alpha_0)=\frac{1}{2}\left[\frac{2\log n}{c^*(\alpha_0)M}\right]^{\frac{1}{2p_{\mathrm{z}}}}$$

and establish

$$\mathbb{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_{\infty}\leq \mathrm{err}_n(M,\alpha_0)\right)\geq (1-n^{-1})\cdot\mathbb{P}\left(\mathcal{E}_1\cap\mathcal{E}_2\right).$$

We further apply Lemma 3 and establish

$$\liminf_{n\rightarrow\infty}\mathbb{P}\left(\min_{1\leq m\leq M}\|U^{[m]}-\widehat{U}\|_{\infty}\leq \mathrm{err}_n(M,\alpha_0)\right)\geq 1-\alpha_0.$$

B.4 Proof of Theorem 2

Recall that the set $\mathcal{M}$ is defined in (25). We define the events

$$\begin{aligned}
\mathcal{E}_3&=\left\{\min_{1\leq m\leq M}\max\left\{\left\|\widehat{\gamma}^{[m]}-\gamma^*\right\|_{\infty},\left\|\widehat{\Gamma}^{[m]}-\Gamma^*\right\|_{\infty}\right\}\leq\frac{\mathrm{err}_n(M,\alpha_0)}{\sqrt{n}}\right\}\\
\mathcal{E}_4&=\left\{\max_{m\in\mathcal{M}}\max_{\beta\in\mathcal{B}}\max_{j\in\widehat{\mathcal{S}}}\left|\widehat{\Gamma}_j^{[m]}-\Gamma_j^*-\beta(\widehat{\gamma}_j^{[m]}-\gamma_j^*)\right|\leq C\sqrt{\frac{\log|\mathcal{B}|+\log|\mathcal{M}|}{n}}\right\}
\end{aligned}\tag{65}$$

for some positive constant $C>0$ independent of n. The control of the event $\mathcal{E}_3$ is established in Proposition 1. The following lemma controls the probability of $\mathcal{E}_4$, whose proof can be found in Section C.4.

Lemma 4 *Suppose that conditions* (C1) *and* (C2) *hold, then* $\mathbb{P}\left(\mathcal{E}_4\right)\geq\mathbb{P}(\mathcal{G})-|\mathcal{B}|^{-c}$.

Recall that, on the event $\mathcal{G}_5$ defined in (50), Condition 1 implies (52), that is,

$$|\mathcal{V}\cap\widehat{\mathcal{S}}|\geq|\mathcal{V}\cap\mathcal{S}_{\mathrm{str}}|>\frac{|\mathcal{S}|}{2}\geq\frac{|\widehat{\mathcal{S}}|}{2}.$$

Coverage property of $\mathrm{CI}^{\mathrm{sample}}$ in (26). Using a similar decomposition as (16), we have

$$\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]} = \widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*) + (\beta^* - \beta)\gamma_j^* + \pi_j^* \quad \text{for} \quad 1 \leq m \leq M. \tag{66}$$

If β is taken as β^*, then

$$\widehat{\Gamma}_j^{[m]} - \beta^*\widehat{\gamma}_j^{[m]} = \widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta^*(\widehat{\gamma}_j^{[m]} - \gamma_j^*) + \pi_j^*. \tag{67}$$

For all $j \in \mathcal{V} \cap \widehat{\mathcal{S}}$,

$$\begin{aligned} \left|\widehat{\Gamma}_j^{[m]} - \beta^*\widehat{\gamma}_j^{[m]}\right| &\leq \|\widehat{\Gamma}^{[m]} - \Gamma\|_\infty + |\beta^*| \cdot \|\widehat{\gamma}^{[m]} - \gamma\|_\infty \\ &\leq (1 + |\beta^*|) \max\{\|\widehat{\Gamma}^{[m]} - \Gamma\|_\infty, \|\widehat{\gamma}^{[m]} - \gamma\|_\infty\} \end{aligned} \tag{68}$$

On the event $\mathcal{E}_3$, we have

$$\begin{aligned} \min_{1\leq m\leq M} \max_{j\in\mathcal{V}\cap\widehat{\mathcal{S}}} \left|\widehat{\Gamma}_j^{[m]} - \beta^*\widehat{\gamma}_j^{[m]}\right| &\leq (1 + |\beta^*|) \min_{1\leq m\leq M} \max\{\|\widehat{\Gamma}^{[m]} - \Gamma\|_\infty, \|\widehat{\gamma}^{[m]} - \gamma\|_\infty\} \\ &\leq (1 + |\beta^*|)\frac{\mathrm{err}_n(M, \alpha_0)}{\sqrt{n}}. \end{aligned} \tag{69}$$

For λ satisfying the property (43), we have

$$\lambda\widehat{\rho}_j(\beta^*, \alpha) \geq (1 + |\beta^*|)\frac{\mathrm{err}_n(M, \alpha_0)}{\sqrt{n}} \quad \text{for all} \quad j \in \widehat{\mathcal{S}}$$

By combining the above equation with (68) and (69), on the event $\mathcal{E}_3$, we show that there exists $1 \leq m^* \leq M$ such that

$$\left|\widehat{\Gamma}_j^{[m^*]} - \beta^*\widehat{\gamma}_j^{[m^*]}\right| \leq (1 + |\beta^*|)\frac{\mathrm{err}_n(M, \alpha_0)}{\sqrt{n}} \leq \lambda\widehat{\rho}_j(\beta^*, \alpha) \quad \text{for any } j \in \mathcal{V} \cap \widehat{\mathcal{S}} \tag{70}$$

Combined with the definition in (23), we establish

$$\left|\left\{j \in \widehat{\mathcal{S}} : \left|\widehat{\Gamma}_j^{[m^*]} - \beta^*\widehat{\gamma}_j^{[m^*]}\right| \leq \lambda\widehat{\rho}_j(\beta, \alpha)\right\}\right| \geq \left|\mathcal{V} \cap \widehat{\mathcal{S}}\right| \tag{71}$$

By combining (71) and (52), we show that, on the event $\mathcal{G}_5 \cap \mathcal{E}_3$,

$$\|\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m^*]}(\beta^*, \lambda)\|_0 \leq |\widehat{\mathcal{S}}| - |\mathcal{V} \cap \widehat{\mathcal{S}}| < \frac{|\widehat{\mathcal{S}}|}{2}.$$

It follows from the definition of $\mathrm{CI}^{\mathrm{sample}}$ in (26) that on the event $\mathcal{G}_5 \cap \mathcal{E}_3$,

$$\beta^* \in \left(\beta_{\min}^{[m^*]}, \beta_{\max}^{[m^*]}\right) \subset \mathrm{CI}^{\mathrm{sample}}.$$

Hence $\mathbf{P}\left(\beta^* \in (\beta_{\min}, \beta_{\max})\right) \geq \mathbf{P}\left(\mathcal{G}_5 \cap \mathcal{E}_3\right)$. Together with Lemma 1 and Proposition 1, we establish the coverage property.

Length of $\mathrm{CI}^{\mathrm{sample}}$ in (26). Suppose that $\left(\beta_{\min}^{[m]}, \beta_{\max}^{[m]}\right)$ for some $1 \leq m \leq M$ is not an empty set, then it follows from (52) that there exists $j \in \mathcal{V} \cap \widehat{\mathcal{S}}$ such that

$$\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]} = \widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*) + (\beta^* - \beta)\gamma_j^*.$$

For β satisfying $|\gamma_j^*| \cdot |\beta - \beta^*| \geq \left|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)\right| + \lambda\widehat{\rho}_j(\beta, \alpha)$, we have $\widehat{\pi}_j(\beta) \neq 0$. That is, for β satisfying

$$|\beta - \beta^*| \geq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{\left|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)\right| + \lambda\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|},$$

we have

$$\|\widehat{\pi}(\beta)\|_0 \geq |\widehat{\mathcal{S}} \cap \mathcal{V}| > \frac{|\widehat{\mathcal{S}}|}{2},$$

where the second inequality follows from (52). That is,

$$\beta \notin \left(\beta_{\min}^{[m]}, \beta_{\max}^{[m]}\right).$$

Hence, if β satisfies

$$|\beta - \beta^*| \geq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{\max_{m \in \mathcal{M}} \max_{\beta \in \mathcal{B}} \left|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)\right| + \lambda\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|}$$

then $\beta \notin \mathrm{CI}^{\mathrm{sample}}$.

On the event $\mathcal{G}_4$, we establish

$$\begin{aligned}
\max\left\{\left|\beta_{\max}^{[m]} - \beta^*\right|, \left|\beta_{\min}^{[m]} - \beta^*\right|\right\} &\leq 2 \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{\max_{\beta \in \mathcal{B}, m \in \mathcal{M}} \left|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)\right| + \lambda\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|} \\
&\leq \max_{j \in \widehat{\mathcal{S}} \cap \mathcal{V}} \frac{C\sqrt{\frac{\log|\mathcal{B}| + \log|\mathcal{M}|}{n}} + \lambda\widehat{\rho}_j(\beta, \alpha)}{|\gamma_j^*|}.
\end{aligned}$$

Together with the fact that $\widehat{\rho}_j(\beta, \alpha) \leq \sqrt{\log n/n}$, we establish the high probability upper bound for the length.

B.5 Proof of Proposition 2

For $j, k \in \widehat{\mathcal{S}}$, we conduct the following error decomposition of $\widehat{\beta}^{[j]} = \widehat{\Gamma}_j/\widehat{\gamma}_j$ and $\widehat{\pi}_k^{[j]}$,

$$\widehat{\beta}^{[j]} - \frac{\Gamma_j^*}{\gamma_j^*} = \frac{1}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\frac{1}{n}W^\intercal\left(\epsilon - \frac{\Gamma_j^*}{\gamma_j^*} \cdot \delta\right) + \frac{1}{\gamma_j^*}\left(\frac{\widehat{\Gamma}_j}{\widehat{\gamma}_j} - \frac{\Gamma_j^*}{\gamma_j^*}\right)(\gamma_j^* - \widehat{\gamma}_j), \tag{72}$$

and

$$
\begin{aligned}
&\widehat{\pi}_k^{[j]}-\left(\Gamma_k^*-\frac{\Gamma_j^*}{\gamma_j^*}\gamma_k^*\right)=\left(\widehat{\Gamma}_k-\widehat{\beta}^{[j]}\widehat{\gamma}_k\right)-\left(\Gamma_k^*-\frac{\Gamma_j^*}{\gamma_j^*}\gamma_k^*\right)\\
&=\left(\widehat{\Gamma}_k-\Gamma_k^*\right)-\frac{\Gamma_j^*}{\gamma_j^*}\left(\widehat{\gamma}_k-\gamma_k^*\right)-\gamma_k^*\left(\widehat{\beta}^{[j]}-\frac{\Gamma_j^*}{\gamma_j^*}\right)-\left(\widehat{\beta}^{[j]}-\frac{\Gamma_j^*}{\gamma_j^*}\right)\left(\widehat{\gamma}_k-\gamma_k^*\right).
\end{aligned}
\tag{73}
$$

Note that

$$
\Gamma_k^*-\frac{\Gamma_j^*}{\gamma_j^*}\gamma_k^*=\pi_k^*-\frac{\pi_j^*}{\gamma_j^*}\gamma_k^*. \tag{74}
$$

By plugging (74) and (72) into (73), we have the following decomposition of $\widehat{\pi}_k^{[j]}-\pi_k^*$

$$
\widehat{\pi}_k^{[j]}-\left(\pi_k^*-\frac{\pi_j^*}{\gamma_j^*}\gamma_k^*\right)=\mathcal{M}_k^{[j]}+\mathcal{R}_k^{[j]}, \tag{75}
$$

where

$$
\mathcal{M}_k^{[j]}=\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right),
$$

and

$$
\mathcal{R}_k^{[j]}=-\frac{\gamma_k^*}{\gamma_j^*}\left(\widehat{\beta}^{[j]}-\frac{\Gamma_j^*}{\gamma_j^*}\right)\left(\gamma_j^*-\widehat{\gamma}_j\right)-\left(\widehat{\beta}^{[j]}-\frac{\Gamma_j^*}{\gamma_j^*}\right)\left(\widehat{\gamma}_k-\gamma_k^*\right). \tag{76}
$$

Define the event

$$
\mathcal{F}_k^{[j]}=\left\{\left|\mathcal{M}_k^{[j]}\right|\leq 0.9\sqrt{\log n}\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{n}W\right\|_2\right\}
$$

and

$$
\mathcal{F}=\cap_{j,k\in\widehat{\mathcal{S}}}\mathcal{F}_k^{[j]} \tag{77}
$$

The following lemma controls the probability of the event $\mathcal{F}$, whose proof can be found in Section C.6.

Lemma 5 *Suppose that conditions* (C1) *and* (C2) *hold, then* $\mathbb{P}\left(\mathcal{F}\right)\geq\mathbb{P}(\mathcal{G})-p_z^2n^{-c}$.

We also need the following lemma, whose proof can be found in Section C.5.

Lemma 6 *On the event $\mathcal{G}$, for a sufficiently large n, we have*

$$
0.99\sqrt{\frac{c_1}{C_0 n}}\leq\frac{\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{n}W\right\|_2}{\sqrt{1+\left(\Gamma_j^*/\gamma_j^*\right)^2}\sqrt{1+(\gamma_k^*/\gamma_j^*)^2}}\leq 1.01\sqrt{\frac{C_1}{c_0 n}}; \tag{78}
$$

$$0.99 \le \frac{\sqrt{\widehat{\sigma}_{\epsilon}^2 + (\widehat{\beta}^{[j]})^2 \widehat{\sigma}_{\delta}^2 - 2\widehat{\beta}^{[j]}\widehat{\sigma}_{\epsilon,\delta}} \left\| \left(\widehat{\Omega}_{k\cdot} - \frac{\widehat{\gamma}_k}{\widehat{\gamma}_j}\widehat{\Omega}_{j\cdot}\right) \frac{1}{n} W \right\|_2}{\sqrt{\mathrm{Var}\left(\epsilon_1 - \Gamma_j^*/\gamma_j^* \delta_1\right)} \left\| \left(\widehat{\Omega}_{k\cdot} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right) \frac{1}{n} W \right\|_2} \le 1.01; \tag{79}$$

and

$$\max_{j,k \in \widehat{\mathcal{S}}} \left| \mathcal{R}_k^{[j]} \right| \le 0.05\sqrt{\log n} \sqrt{\mathrm{Var}\left(\epsilon_1 - \Gamma_j^*/\gamma_j^* \delta_1\right)} \left\| \left(\widehat{\Omega}_{k\cdot} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right) \frac{1}{n} W \right\|_2. \tag{80}$$

We shall consider two cases: (1) $\left|\pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^*\right| = 0$; (2) $\left|\pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^*\right| \ge \mathrm{sep}(n)$.

For the setting $\left|\pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^*\right| = 0$, we simplify (75) as

$$\widehat{\pi}_k^{[j]} = \mathcal{M}_k^{[j]} + \mathcal{R}_k^{[j]}. \tag{81}$$

By (79) and (80), on the event $\mathcal{G} \cap \mathcal{F}$,

$$\max_{j,k \in \widehat{\mathcal{S}}} \left| \mathcal{M}_k^{[j]} + \mathcal{R}_k^{[j]} \right| \le \sqrt{\log n} \cdot \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]}) \tag{82}$$

where $\widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]})$ is defined in (34). Hence, by the definition in (35), we have $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 1$.

We now consider the setting $\left|\pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^*\right| \ge \mathrm{sep}(n)$. By (75), the event $\{\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 0\}$ is equivalent to that at least one of the following two events happens

$$\left| \pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^* + \mathcal{M}_k^{[j]} + \mathcal{R}_k^{[j]} \right| > \sqrt{\log n} \cdot \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]}) \tag{83}$$

$$\left| \pi_j^* - \frac{\pi_k^*}{\gamma_k^*}\gamma_j^* + \mathcal{M}_j^{[k]} + \mathcal{R}_j^{[k]} \right| > \sqrt{\log n} \cdot \widehat{\mathrm{SE}}(\widehat{\pi}_j^{[k]}) \tag{84}$$

By the upper bound in (82), on the event $\mathcal{G} \cap \mathcal{F}$, the event in (83) happens if

$$\left| \pi_k^* - \frac{\pi_j^*}{\gamma_j^*}\gamma_k^* \right| > 2\sqrt{\log n} \cdot \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]}); \tag{85}$$

the event (84) happens if

$$\left| \pi_j^* - \frac{\pi_k^*}{\gamma_k^*}\gamma_j^* \right| > 2\sqrt{\log n} \cdot \widehat{\mathrm{SE}}(\widehat{\pi}_j^{[k]}). \tag{86}$$

It follows from (78) and (79) that

$$\begin{aligned} \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]}) &\le 1.01 \sqrt{\mathrm{Var}\left(\epsilon_1 - \Gamma_j^*/\gamma_j^* \delta_1\right)} \left\| \left(\widehat{\Omega}_{k\cdot} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right) \frac{1}{n} W \right\|_2 \\ &\le 1.01^2 \sqrt{\frac{C_1}{c_0 n}} \sqrt{1 + \left(\Gamma_j^*/\gamma_j^*\right)^2} \sqrt{1 + (\gamma_k^*/\gamma_j^*)^2} \end{aligned} \tag{87}$$

Note that as long as

$$\left|\frac{\pi_k^*}{\gamma_k^*} - \frac{\pi_j^*}{\gamma_j^*}\right| > \frac{2}{|\gamma_k^*|}\sqrt{\log n} \cdot 1.01^2 \sqrt{\frac{C_1}{c_0 n}} \sqrt{1 + \left(\Gamma_j^*/\gamma_j^*\right)^2} \sqrt{1 + (\gamma_k^*/\gamma_j^*)^2} \tag{88}$$

we apply (87) and establish (85). By the definition of $\text{sep}(n)$ in (30), we establish (88) and hence (85) and (86), we establish $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 0$.

B.6 Proof of Proposition 3

We apply the definitions of $\widetilde{\mathcal{V}}$ in (44) and $\widehat{\mathcal{V}} = \widehat{\mathcal{V}}^{\texttt{TSHT}}$ in (45), which is equivalent to the definition in (36).

We apply Proposition 2 and establish the following results holds with probability larger than $1 - \exp(-c\sqrt{\log n})$. For any $l \in \widehat{\mathcal{V}}$, there exists $k \in \widetilde{\mathcal{V}}$ such that

$$\left|\frac{\pi_l^*}{\gamma_l^*} - \frac{\pi_k^*}{\gamma_k^*}\right| \leq \text{sep}(n). \tag{89}$$

For any $k \in \widetilde{\mathcal{V}}$, there exists $j \in \widehat{\mathcal{W}}$ such that

$$\left|\frac{\pi_k^*}{\gamma_k^*} - \frac{\pi_j^*}{\gamma_j^*}\right| \leq \text{sep}(n). \tag{90}$$

A combination of (89) and (90) leads to the fact that: for any $l \in \widehat{\mathcal{V}}$, there exists $j \in \widehat{\mathcal{W}}$ such that

$$\left|\frac{\pi_l^*}{\gamma_l^*} - \frac{\pi_j^*}{\gamma_j^*}\right| \leq 2\text{sep}(n). \tag{91}$$

With the similar argument, we also show that: for any $j \in \widehat{\mathcal{W}}$, there exists $l \in \widehat{\mathcal{V}}$ such that (91) holds.

We apply the above facts and consider two settings in the following.

1. We first consider the case that all elements in the winner set belong to the set of valid IV, that is, $\widehat{\mathcal{W}} \subset \mathcal{V}$. Hence, for $k \in \widehat{\mathcal{S}} \cap \mathcal{V}$, we have $\frac{\pi_k^*}{\gamma_k^*} = \frac{\pi_j^*}{\gamma_j^*} = 0$ for any $j \in \widehat{\mathcal{W}}$. Hence, with a high probability, $\widehat{\Pi}_{k,j} = \widehat{\Pi}_{j,k} = 1$. That is, $k \in \widetilde{\mathcal{V}}$ and hence

$$\mathcal{V} \cap \widehat{\mathcal{S}} \subset \widetilde{\mathcal{V}} \subset \widehat{\mathcal{V}}. \tag{92}$$

It follows from (91) that

$$\widehat{\mathcal{V}} \subset \mathcal{I}(0, 2\text{sep}(n)). \tag{93}$$

Since $\mathcal{V} \cap \mathcal{S}_{\text{Str}} \subset \mathcal{V} \cap \widehat{\mathcal{S}}$, we combine (92) and (93) and establish

$$\mathcal{V} \cap \mathcal{S}_{\text{Str}} \subset \widehat{\mathcal{V}} \subset \mathcal{I}(0, 2\text{sep}(n)). \tag{94}$$

2. We then consider the setting that $\widehat{\mathcal{W}}$ contains some invalid IV, that is, $\widehat{\mathcal{W}} \not\subset \mathcal{V}$.

   In this case, there exists $j \notin \mathcal{V}$ but $j \in \widehat{\mathcal{W}}$. Note that the support of $\widehat{\Pi}_{j\cdot}$, denoted as $\text{supp}(\widehat{\Pi}_{j\cdot})$, satisfies

$$\text{supp}(\widehat{\Pi}_{j\cdot}) \subset \mathcal{I}\left(\frac{\pi_j^*}{\gamma_j^*}, \text{sep}(n)\right) \cap \widehat{\mathcal{S}}. \tag{95}$$

   In the following, we show by contradiction that

$$\text{supp}(\widehat{\Pi}_{j\cdot}) \cap \mathcal{V} \neq \varnothing.$$

   Assume that

$$\text{supp}(\widehat{\Pi}_{j\cdot}) \cap \mathcal{V} = \varnothing. \tag{96}$$

   For any $l \in \mathcal{V} \cap \widehat{\mathcal{S}}$ and $l \neq j$, with a high probability,

$$\mathcal{V} \cap \widehat{\mathcal{S}} \subset \text{supp}(\widehat{\Pi}_{l\cdot}). \tag{97}$$

   Since the plurality rule implies $\|\widehat{\Pi}_{l\cdot}\|_0 \leq \|\widehat{\Pi}_{j\cdot}\|_0$, we apply (97) and establish

$$|\mathcal{V} \cap \widehat{\mathcal{S}}| \leq \left|\text{supp}(\widehat{\Pi}_{j\cdot})\right| = \left|\text{supp}(\widehat{\Pi}_{j\cdot}) \backslash \mathcal{V}\right| \leq \left|\mathcal{I}\left(\frac{\pi_j^*}{\gamma_j^*}, \text{sep}(n)\right) \backslash \mathcal{V}\right|,$$

   where the first equality follows from the assumption (96) and the last inequality follows from (95). Note that

$$|\mathcal{V} \cap \mathcal{S}_{\text{str}}| \leq |\mathcal{V} \cap \widehat{\mathcal{S}}| \leq \left|\mathcal{I}\left(\frac{\pi_j^*}{\gamma_j^*}, \text{sep}(n)\right) \backslash \mathcal{V}\right|,$$

   which contradicts the finite-sample plurality rule and hence the conjecture (96) does not holds. That is,

$$\text{there exists } k \in \text{supp}(\widehat{\Pi}_{j\cdot}) \cap \mathcal{V}. \tag{98}$$

   Then (98) implies that $k \in \widetilde{\mathcal{V}}$ and $\mathcal{V} \cap \widehat{\mathcal{S}} \subset \text{supp}(\widehat{\Pi}_{k\cdot})$, which implies

$$\mathcal{V} \cap \widehat{\mathcal{S}} \subset \widehat{\mathcal{V}}. \tag{99}$$

   Furthermore, (98) implies

$$\left|\frac{\pi_j^*}{\gamma_j^*} - \frac{\pi_k^*}{\gamma_k^*}\right| \leq \text{sep}(n). \tag{100}$$

   We combine (100) and (91) and show that

$$\widehat{\mathcal{V}} \subset \mathcal{I}(0, 3\text{sep}(n)) \tag{101}$$

   Since $\mathcal{V} \cap \mathcal{S}_{\text{Str}} \subset \mathcal{V} \cap \widehat{\mathcal{S}}$, we combine (99) and (101) and establish

$$\mathcal{V} \cap \mathcal{S}_{\text{Str}} \subset \widehat{\mathcal{V}} \subset \mathcal{I}(0, 3\text{sep}(n)). \tag{102}$$

We combine (94) and (102) and establish that $\widehat{\mathcal{V}}$ satisfies (32).

B.7 Proofs of Theorems 3 and 4

The proof of Theorem 3 follows from that of Theorem 1 and Proposition 3. Note that, on the event $\mathcal{G}$, we have $\mathcal{V} \cap \mathcal{S}_{\rm Str} \subset \widehat{\mathcal{V}} \subset \mathcal{I}(0, 3{\rm sep}(n))$. By the finite sample plurality rule (Condition 2), $\mathcal{V} \cap \mathcal{S}_{\rm Str}$ is the majority of the set $\mathcal{I}(0, 3{\rm sep}(n))$ and also the majority of $\widehat{\mathcal{V}}$. We then apply the same argument for Theorem 1 by replacing $\widehat{\mathcal{S}}$ with $\widehat{\mathcal{V}}$ and $\mathcal{S}$ with $\mathcal{I}(0, 3{\rm sep}(n))$.

The proof of Theorem 4 follows from that of Theorem 2 and Proposition 3. Note that, on the event $\mathcal{G}$, we have $\mathcal{V} \subset \widehat{\mathcal{V}} \subset \mathcal{I}(0, 3{\rm sep}(n))$. By the finite sample plurality rule (Condition 2), $\mathcal{V} \cap \mathcal{S}_{\rm Str}$ is the majority of the set $\mathcal{I}(0, 3{\rm sep}(n))$ and also the majority of $\widehat{\mathcal{V}}$. We then apply the same argument for Theorem 2 by replacing $\widehat{\mathcal{S}}$ with $\widehat{\mathcal{V}}$ and $\mathcal{S}$ with $\mathcal{I}(0, 3{\rm sep}(n))$.

C Proofs of Lemmas

C.1 Proof of Lemma 1

Control of $\mathcal{G}_0$. Note that $\mathbf{E} W_{ij}\delta_i = 0$ for any $1 \leq i \leq n$ and $1 \leq j \leq p$ and $W_{ij}\delta_i$ is Sub-exponential random variable. Since $(W^{\intercal}\delta)_j = \sum_{i=1}^{n} W_{ij}\delta_i$, we apply Proposition 5.16 of [32] with the corresponding $t = \sqrt{n\sqrt{\log n}}$ and establish

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} W_{ij}\delta_i\right| \leq C\sqrt{n\sqrt{\log n}}\right) \geq 1 - \exp(-c'\sqrt{\log n}).$$

where C and c' are positive constants independent of n. For the fixed $p_{\rm z}$ setting, we apply the union bound and establish that

$$\begin{aligned}\mathbb{P}\left(\|W^{\intercal}\delta\|_{\infty} = \max_{1\leq j\leq p_{\rm z}}\left|\sum_{i=1}^{n} W_{ij}\delta_i\right| \leq C\sqrt{n\sqrt{\log n}}\right) \geq & 1 - p_{\rm z}\exp(-c'\sqrt{\log n}) \\ \geq & 1 - \exp(-c\sqrt{\log n}).\end{aligned} \tag{103}$$

for some positive constant $c > 0$.

Control of $\mathcal{G}_3$. Since $\{W_{i\cdot}\}_{1\leq i\leq n}$ are i.i.d Sub-gaussian random vectors and the dimension p is fixed, then we apply equation (5.25) in [32] and establish the following concentration results for $\widehat{\Sigma} - \Sigma$: with probability larger than $1 - n^{-c}$,

$$\|\widehat{\Sigma} - \Sigma\|_2 \leq C\sqrt{\frac{\log n}{n}} \tag{104}$$

where c and C are positive constants independent of n. As a consequence, we have

$$|\lambda_{\min}(\widehat{\Sigma}) - \lambda_{\min}(\Sigma)| \leq \|\widehat{\Sigma} - \Sigma\|_2 \leq C\sqrt{\frac{\log n}{n}} \tag{105}$$

Since $\widehat{\Sigma}^{-1} - \Sigma^{-1} = \widehat{\Sigma}^{-1}(\Sigma - \widehat{\Sigma})\Sigma^{-1}$, we have

$$\|\widehat{\Sigma}^{-1} - \Sigma^{-1}\|_2 \leq \frac{1}{\lambda_{\min}(\widehat{\Sigma}) \cdot \lambda_{\min}(\Sigma)} \|\Sigma - \widehat{\Sigma}\|_2 \leq \frac{C\sqrt{\frac{\log n}{n}}}{\left(\lambda_{\min}(\Sigma) - C\sqrt{\frac{\log n}{n}}\right) \cdot \lambda_{\min}(\Sigma)}.$$

Then we have

$$\mathbb{P}(\mathcal{G}_3) \geq 1 - n^{-c}. \tag{106}$$

Control of $\mathcal{G}_1$. We shall focus on the analysis of $\widehat{\gamma}_j - \gamma_j^*$ and the analysis of $\widehat{\Gamma}_j - \Gamma_j^*$ is similar. We decompose the expression (46) as

$$\frac{\widehat{\Omega}_{j\cdot}^{\intercal} \frac{1}{n} W^{\intercal}\delta}{\sqrt{\mathbf{V}_{jj}^{\gamma}/n}} = \frac{\Omega_{j\cdot}^{\intercal} \frac{1}{n} W^{\intercal}\delta}{\sqrt{\mathbf{V}_{jj}^{\gamma}/n}} + \frac{(\widehat{\Omega}_{j\cdot} - \Omega_{j\cdot})^{\intercal} \frac{1}{n} W^{\intercal}\delta}{\sqrt{\mathbf{V}_{jj}^{\gamma}/n}}. \tag{107}$$

By the central limit theorem (c.f. Theorem 3.2 in [35]) and Condition (C2), we have

$$\frac{\Omega_{j\cdot}^{\intercal} \frac{1}{n} W^{\intercal}\delta}{\sqrt{\mathbf{V}_{jj}^{\gamma}/n}} \xrightarrow{d} N(0,1).$$

On the event $\mathcal{G}_0 \cap \mathcal{G}_3$, we apply (105), (107) and (103) and establish

$$\mathbb{P}\left(\max_{1 \leq j \leq p_z} |\widehat{\gamma}_j - \gamma_j^*| / \sqrt{\mathbf{V}_{jj}^{\gamma}/n} \leq C(\log n)^{1/4}\right) \geq 1 - \exp(-c\sqrt{\log n}).$$

We can apply a similar argument to control $\widehat{\Gamma}_j - \Gamma_j^*$ and then establish

$$\mathbb{P}(\mathcal{G}_1) \geq 1 - \exp(-c\sqrt{\log n}). \tag{108}$$

By a similar argument, for the fixed p setting, we can establish

$$\mathbb{P}\left(\|\Psi^* - \widehat{\Psi}\|_1 + \|\psi^* - \widehat{\psi}\|_1 \leq C\frac{(\log n)^{1/4}}{\sqrt{n}}\right) \geq 1 - \exp(-c\sqrt{\log n}). \tag{109}$$

Control of $\mathcal{G}_2$. Recall that the variance and covariance estimators are defined in (10). We shall detail the proof for $\widehat{\sigma}_{\epsilon,\delta} - \sigma_{\epsilon,\delta}$ and the other two terms can be controlled using a similar argument. We start with the decomposition of $\widehat{\sigma}_{\epsilon,\delta} - \sigma_{\epsilon,\delta}$

$$\begin{aligned}
&\frac{1}{n-1}(Y - Z\widehat{\Gamma} - X\widehat{\Psi})^{\intercal}(D - Z\widehat{\gamma} - X\widehat{\psi}) - \sigma_{\epsilon,\delta} \\
&= \frac{\epsilon^{\intercal}\delta - n\sigma_{\epsilon,\delta}}{n-1} + \frac{1}{n-1}\epsilon^{\intercal}\left[Z(\gamma^* - \widehat{\gamma}) + X(\psi^* - \widehat{\psi})\right] + \frac{1}{n-1}\delta^{\intercal}\left[Z(\Gamma^* - \widehat{\Gamma}) + X(\Psi^* - \widehat{\Psi})\right] \\
&+ \frac{1}{n-1}\left[Z(\gamma^* - \widehat{\gamma}) + X(\psi^* - \widehat{\psi})\right]^{\intercal}\left[Z(\Gamma^* - \widehat{\Gamma}) + X(\Psi^* - \widehat{\Psi})\right] + \frac{1}{n-1}\sigma_{\epsilon,\delta}.
\end{aligned} \tag{110}$$

Since ϵ_i and δ_i are Sub-gaussian random variables and $\mathbf{E}\epsilon_i\delta_i - \sigma_{\epsilon,\delta} = 0$, we apply Proposition 5.16 of [32] with the corresponding $t = \sqrt{n\log n}$ and establish

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \epsilon_i\delta_i - n\sigma_{\epsilon,\delta}\right| \leq C\sqrt{n\log n}\right) \geq 1 - n^{-c}. \tag{111}$$

By the decomposition (110), we apply (104), (111) and (109) and establish

$$\mathbb{P}\left(|\widehat{\sigma}_{\epsilon,\delta} - \sigma_{\epsilon,\delta}| \leq C\sqrt{\frac{\log n}{n}}\right) \geq 1 - \exp(-c\sqrt{\log n}),$$

for some positive constants $C > 0$ and $c > 0$. Then we apply a similar argument to control $\left|\widehat{\sigma}_{\epsilon}^2 - \sigma_{\epsilon}^2\right|$ and $\left|\widehat{\sigma}_{\delta}^2 - \sigma_{\delta}^2\right|$ and establish

$$\mathbb{P}\left(\mathcal{G}_3\right) \geq 1 - \exp(-c\sqrt{\log n}). \tag{112}$$

Control of events $\mathcal{G}_4$ and $\mathcal{G}_5$. Note that $\mathbf{V}^{\gamma}_{jj} = \sigma_{\delta}^2\Omega_{jj}$, $\mathbf{V}^{\Gamma}_{jj} = \sigma_{\epsilon}^2\Omega_{jj}$ and $\mathbf{C}_{jj} = \sigma_{\epsilon,\delta}\Omega_{jj}$. On the event $\mathcal{G}_2$ and $\mathcal{G}_3$, then the event $\mathcal{G}_4$ holds when the dimension p is fixed. Recall that $\widehat{\mathcal{S}}$ is defined in (12) and $\mathcal{S}_{\rm str}$ is defined in (14). Then for $j \in \mathcal{S}_{\rm str}$, on the event $\mathcal{G}_1 \cap \mathcal{G}_4$, if $\sqrt{\log n} > C(\log n)^{1/4}$, we have

$$|\widehat{\gamma}_j| \geq \sqrt{\log n} \cdot \sqrt{\widehat{\mathbf{V}}^{\gamma}_{jj}/n},$$

that is $\mathcal{S}_{\rm str} \subset \widehat{\mathcal{S}}$. Furthermore, for $j \in \widehat{\mathcal{S}}$, on the event $\mathcal{G}_1 \cap \mathcal{G}_4$, we have

$$|\gamma^*_j| \geq (\sqrt{\log n} - C(\log n)^{1/4}) \cdot \sqrt{\widehat{\mathbf{V}}^{\gamma}_{jj}/n}$$

that is $\widehat{\mathcal{S}} \subset \mathcal{S}^0$. Hence, we have

$$\mathbb{P}(\mathcal{G}_4 \cap \mathcal{G}_5) \geq \mathbb{P}\left(\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3\right). \tag{113}$$

Control of events $\mathcal{G}_6$ and $\mathcal{G}_8$. On the event $\mathcal{G}_1 \cap \mathcal{G}_4 \cap \mathcal{G}_5$, for $j \in \widehat{\mathcal{S}}$

$$\left|\frac{\widehat{\gamma}_j}{\gamma^*_j} - 1\right| \leq \frac{C(\log n)^{1/4}\sqrt{\mathbf{V}^{\gamma}_{jj}/n}}{(\sqrt{\log n} - C(\log n)^{1/4})\sqrt{\widehat{\mathbf{V}}^{\gamma}_{jj}/n}} \lesssim \frac{1}{(\log n)^{1/4}}$$

and hence

$$\max_{j \in \widehat{\mathcal{S}}} \left|\frac{\gamma^*_j}{\widehat{\gamma}_j} - 1\right| \lesssim \frac{1}{(\log n)^{1/4}} \tag{114}$$

By the decomposition

$$\begin{aligned}(\widehat{\gamma}_k/\widehat{\gamma}_j - \gamma^*_k/\gamma^*_j)\gamma^*_j =& (\widehat{\gamma}_k - \gamma^*_k)\frac{\gamma^*_j}{\widehat{\gamma}_j} + \gamma^*_k\left(\frac{\gamma^*_j}{\widehat{\gamma}_j} - 1\right) \\ =& \left(\frac{\widehat{\gamma}_k}{\gamma^*_k} - 1\right) \cdot \gamma^*_k \cdot \frac{\gamma^*_j}{\widehat{\gamma}_j} + \gamma^*_k\left(\frac{\gamma^*_j}{\widehat{\gamma}_j} - 1\right)\end{aligned}$$

we apply (114) and establish

$$\left|(\widehat{\gamma}_k/\widehat{\gamma}_j-\gamma_k^*/\gamma_j^*)\gamma_j^*\right|\lesssim\frac{|\gamma_k^*|}{(\log n)^{1/4}}\quad\text{and}\quad\left|\frac{\widehat{\gamma}_k}{\widehat{\gamma}_j}-\frac{\gamma_k^*}{\gamma_j^*}\right|\lesssim\left|\frac{\gamma_k^*}{\gamma_j^*}\right|\cdot\frac{1}{(\log n)^{1/4}}.$$

That is, the event $\mathcal{G}_6$ holds and

$$\mathbb{P}(\mathcal{G}_6)\geq\mathbb{P}(\mathcal{G}_1\cap\mathcal{G}_4\cap\mathcal{G}_5).\tag{115}$$

By the decomposition

$$(\widehat{\Gamma}_j/\widehat{\gamma}_j-\Gamma_j^*/\gamma_j^*)\gamma_j^*=(\widehat{\Gamma}_j-\Gamma_j^*)\frac{\gamma_j^*}{\widehat{\gamma}_j}+\Gamma_j^*\left(\frac{\gamma_j^*}{\widehat{\gamma}_j}-1\right)$$

we apply (114) and establish on the event $\mathcal{G}_1$ that

$$\left|(\widehat{\Gamma}_j/\widehat{\gamma}_j-\Gamma_j^*/\gamma_j^*)\gamma_j^*\right|\leq C\sqrt{\mathbf{V}^{\Gamma}_{jj}/n}(\log n)^{1/4}+\left|\Gamma_j^*\right|\frac{1}{(\log n)^{1/4}}.$$

For $j\in\widehat{\mathcal{S}}$, on the event $\mathcal{G}_5$, $j\in\mathcal{S}^0$ and hence

$$\left|\widehat{\Gamma}_j/\widehat{\gamma}_j-\Gamma_j^*/\gamma_j^*\right|\leq C\left(1+\left|\frac{\Gamma_j^*}{\gamma_j^*}\right|\right)\frac{1}{(\log n)^{1/4}}.$$

That is, the event $\mathcal{G}_8$ holds and

$$\mathbb{P}(\mathcal{G}_8)\geq\mathbb{P}(\mathcal{G}_1\cap\mathcal{G}_4\cap\mathcal{G}_5).\tag{116}$$

We establish the lemma by combining (106), (108), (112), (115) and (116).

C.2 Proof of Lemma 2

Note that

$$\widehat{\Gamma}_j-\Gamma_j^*-\beta(\widehat{\gamma}_j-\gamma_j^*)=\widehat{\Omega}_{j\cdot}^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta).\tag{117}$$

We first assume that $(\epsilon_i,\delta_i)^{\intercal}$ is bivariate normal and independent of $W_{i\cdot}$. In this case, by conditioning on W, we have

$$\frac{|\widehat{\Gamma}_j-\Gamma_j^*-\beta(\widehat{\gamma}_j-\gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj}+\beta^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta\widehat{\mathbf{C}}_{jj})/n}}\mid W\sim N(0,1).$$

By the Bonferroni correction, we have $\mathbb{P}(\mathcal{E}_0(\alpha))\geq 1-\alpha$ with $\widehat{\rho}(\alpha)=\Phi^{-1}\left(1-\frac{\alpha}{2|\mathcal{B}|\cdot p_{\mathrm{z}}}\right)$. Also, we have

$$\mathbb{P}\left(\frac{|\widehat{\Gamma}_j-\Gamma_j^*-\beta(\widehat{\gamma}_j-\gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj}+\beta^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta\widehat{\mathbf{C}}_{jj})/n}}\geq\sqrt{2.005\log|\mathcal{B}|}\mid W\right)\leq\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{2.005\log|\mathcal{B}|}{2}\right)$$

Hence we have

$$\mathbb{P}\left(\max_{\beta\in\mathcal{B}}\max_{j\in\widehat{\mathcal{S}}}\frac{|\widehat{\Gamma}_j-\Gamma^*_j-\beta(\widehat{\gamma}_j-\gamma^*_j)|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj}+\beta^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta\widehat{\mathbf{C}}_{jj})/n}}\geq\sqrt{2.005\log|\mathcal{B}|}\mid W\right)\leq|\mathcal{B}|^{-0.0025}\cdot p_{\mathrm{z}}$$

This leads to $\mathbb{P}(\mathcal{E}_0(\alpha))\geq 1-\alpha$ with $\widehat{\rho}(\alpha)=\sqrt{2.005\log|\mathcal{B}|}$.

Now we turn to the model general setting without assuming normal errors. We further decompose (117) as

$$\widehat{\Gamma}_j-\Gamma^*_j-\beta(\widehat{\gamma}_j-\gamma^*_j)=\Omega_{j\cdot}^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta)+(\widehat{\Omega}_{j\cdot}-\Omega_{j\cdot})^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta) \tag{118}$$

Since $\Omega_{j\cdot}^{\intercal}W_{i\cdot}\cdot(\epsilon_i-\beta\delta_i)$ is Sub-exponential with zero mean and

$$\Omega_{j\cdot}^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta)=\frac{1}{n}\sum_{i=1}^{n}\Omega_{j\cdot}^{\intercal}W_{i\cdot}\cdot(\epsilon_i-\beta\delta_i),$$

we apply Proposition 5.16 of [32] with the corresponding $t=C\sqrt{\frac{\log|\mathcal{B}|}{n}}$ and establish

$$\mathbb{P}\left(\left|\Omega_{j\cdot}^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta)\right|\geq C\sqrt{\frac{\log|\mathcal{B}|}{n}}\right)\leq|\mathcal{B}|^{-c}, \tag{119}$$

where C and $c>1$ are positive constants independent of n. Furthermore, on the event $\mathcal{G}_0\cap\mathcal{G}_3$, we have

$$\left|(\widehat{\Omega}_{j\cdot}-\Omega_{j\cdot})^{\intercal}\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta)\right|\leq\|\widehat{\Omega}_{j\cdot}-\Omega_{j\cdot}\|_1\left\|\frac{1}{n}W^{\intercal}(\epsilon-\beta\delta)\right\|_{\infty}\leq\frac{\log n}{n}. \tag{120}$$

Note that

$$\widehat{\mathbf{V}}^{\Gamma}_{jj}+\beta^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta\widehat{\mathbf{C}}_{jj}=\widehat{\Omega}_{jj}\cdot\begin{pmatrix}1 & -\beta\end{pmatrix}\begin{pmatrix}\widehat{\sigma}^2_{\epsilon} & \widehat{\sigma}_{\epsilon,\delta}\\ \widehat{\sigma}_{\epsilon,\delta} & \widehat{\sigma}^2_{\delta}\end{pmatrix}\begin{pmatrix}1\\ -\beta\end{pmatrix}$$

On the event $\mathcal{G}_2$, Condition (C1) implies that

$$\begin{aligned}\widehat{\mathbf{V}}^{\Gamma}_{jj}+\beta^2\widehat{\mathbf{V}}^{\gamma}_{jj}-2\beta\widehat{\mathbf{C}}_{jj}&\geq\widehat{\Omega}_{jj}(1+\beta^2)(c_1-C\sqrt{\log n/n})\\&\geq(c_0-C\sqrt{\log n/n})(1+\beta^2)(c_1-C\sqrt{\log n/n}).\end{aligned} \tag{121}$$

Combined with (119) and (120), we apply the union bound and establish that

$$\mathbb{P}(\mathcal{E}_0(\alpha))\geq 1-\mathbb{P}(\mathcal{G}^c)-p_{\mathrm{z}}\cdot|\mathcal{B}|\cdot|\mathcal{B}|^{-c}.$$

where the constant $c>1$ is used in (119). Since $|\mathcal{B}|\asymp n^a$, for a sufficiently large n, we have $\mathbb{P}(\mathcal{E}_0(\alpha))\geq 1-\alpha$.

C.3 Proof of Lemma 3

Recall that $\widehat{U} \xrightarrow{d} U^*$ where $U^* \sim N(0, \mathrm{Cov})$. For a small constant $0 < \alpha_0 < 1/2$, we define the positive constant

$$c_3 = \exp\left(-F^{-1}_{\chi^2_{2p_{\mathrm{z}}}}(1-\alpha_0)\right). \tag{122}$$

where $F^{-1}_{\chi^2_{2p_{\mathrm{z}}}}(1-\alpha_0)$ denotes the $1-\alpha_0$ quantile of the χ^2 distribution with degree of freedom $2p_{\mathrm{z}}$. On the event $\mathcal{E}_1$, we have

$$\exp\left(-\frac{1}{2}[U^*]^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}U^*\right) \geq \exp\left(-\frac{1}{2}[U^*]^\intercal(\mathrm{Cov}/2)^{-1}U^*\right),$$

and

$$\mathbb{P}\left(\exp\left(-\frac{1}{2}[U^*]^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}U^*\right) \geq c_3\right) \geq \mathbb{P}\left(\exp\left(-[U^*]^\intercal\mathrm{Cov}^{-1}U^*\right) \geq c_3\right) = 1-\alpha_0, \tag{123}$$

where c_3 is defined in (122).

For any constant $0 < c < 1$, we apply the union bound and establish

$$\begin{aligned}
&\mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1} \geq (1-c)\cdot c_3\right)\\
&\geq \mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right) \geq c_3\right)\\
&-\mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right)\cdot \mathbf{1}_{\mathcal{O}\notin\mathcal{E}_1} \geq c\cdot c_3\right)
\end{aligned}$$

Together with

$$\mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right)\cdot \mathbf{1}_{\mathcal{O}\notin\mathcal{E}_1} \geq c\cdot c_3\right) \leq \mathbb{P}(\mathcal{E}_1^c),$$

we establish

$$\begin{aligned}
&\mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right)\cdot \mathbf{1}_{\mathcal{O}\in\mathcal{E}_1} \geq (1-c)\cdot c_3\right)\\
&\geq \mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right) \geq c_3\right) - \mathbb{P}(\mathcal{E}_1^c).
\end{aligned} \tag{124}$$

Since $\widehat{U} \xrightarrow{d} U^*$, we establish

$$\mathbb{P}\left(\exp\left(-\frac{1}{2}\widehat{U}^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}\widehat{U}\right) \geq c_3\right) \to \mathbb{P}\left(\exp\left(-\frac{1}{2}[U^*]^\intercal(\mathrm{Cov}-c_2\mathbf{I})^{-1}U^*\right) \geq c_3\right)$$

Together with (123) and (124), we establish

$$\lim_{n\to\infty}\mathbb{P}(\mathcal{E}_2) \geq 1-\alpha_0 - \lim_{n\to\infty}\mathbb{P}(\mathcal{E}_1^c)$$

Since $\|\widehat{\mathrm{Cov}} - \mathrm{Cov}\|_2 \lesssim \max\left\{\|\widehat{\mathbf{V}}^{\Gamma} - \mathbf{V}^{\Gamma}\|_2, \|\widehat{\mathbf{V}}^{\gamma} - \mathbf{V}^{\gamma}\|_2, \|\widehat{\mathbf{C}} - \mathbf{C}\|_2\right\}$, we establish

$$\mathbb{P}\left(\mathcal{E}_1\right) \geq \mathbb{P}\left(\mathcal{G}_4\right) \geq 1 - \exp(-c\sqrt{\log n}).$$

Hence, we have

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \alpha_0.$$

C.4 Proof of Lemma 4

Note that

$$\left|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)\right| \leq \left|\widehat{\Gamma}_j^{[m]} - \widehat{\Gamma}_j - \beta(\widehat{\gamma}_j^{[m]} - \widehat{\gamma}_j)\right| + \left|\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)\right|$$

Following from the argument of (119) and (120), we have

$$\mathbb{P}\left(\max_{m\in\mathcal{M}} \max_{\beta\in\mathcal{B}} \max_{j\in\widehat{\mathcal{S}}} \left|\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)\right| \lesssim \sqrt{\frac{\log|\mathcal{B}|}{n}}\right) \geq \mathbb{P}\left(\mathcal{G}_0 \cap \mathcal{G}_3\right) - |\mathcal{B}|^{-c}. \tag{125}$$

Since

$$\frac{|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2\widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta\widehat{\mathbf{C}}_{jj})/n}} \mid \mathcal{O} \sim N(0,1),$$

we apply the union bound and establish

$$\mathbb{P}\left(\max_{m\in\mathcal{M}} \max_{\beta\in\mathcal{B}} \max_{j\in\widehat{\mathcal{S}}} \left|\frac{|\widehat{\Gamma}_j^{[m]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m]} - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2\widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta\widehat{\mathbf{C}}_{jj})/n}}\right| \lesssim \sqrt{\frac{\log(|\mathcal{B}|\cdot|\mathcal{M}|)}{n}} \mid \mathcal{O}\right) \geq 1 - (|\mathcal{B}|\cdot|\mathcal{M}|)^{-c}. \tag{126}$$

On the event $\mathcal{G}_2$, we have (121). Combined with (125) and (126), we establish Lemma 4.

C.5 Proof of Lemma 6

Since the spectrum of the covariance matrix of (ϵ_1, δ_1) is bounded within the range $[c_1, C_1]$, we have

$$\sqrt{c_1}\sqrt{1 + \left(\Gamma_j^*/\gamma_j^*\right)^2} \leq \sqrt{\mathrm{Var}\left(\epsilon_1 - \Gamma_j^*/\gamma_j^*\delta_1\right)} \leq \sqrt{C_1}\sqrt{1 + \left(\Gamma_j^*/\gamma_j^*\right)^2}.$$

Note that

$$\left\|\left(\widehat{\Omega}_{k\cdot} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{n}W\right\|_2 = \frac{1}{\sqrt{n}}\sqrt{\widehat{\Omega}_{kk} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{jk} + (\frac{\gamma_k^*}{\gamma_j^*})^2\widehat{\Omega}_{jj}}$$

Hence, we have

$$\sqrt{\lambda_{\min}(\widehat{\Omega})\left(1 + (\gamma_k^*/\gamma_j^*)^2\right)} \leq \left\|\left(\widehat{\Omega}_{k\cdot} - \frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{\sqrt{n}}W\right\|_2 \leq \sqrt{\lambda_{\max}(\widehat{\Omega})\left(1 + (\gamma_k^*/\gamma_j^*)^2\right)} \tag{127}$$

On the event $\mathcal{G}$, we establish (78).

On the event $\mathcal{G}$, the difference between $\sqrt{\widehat{\sigma}_{\epsilon}^2+(\widehat{\beta}^{[j]})^2\widehat{\sigma}_{\delta}^2-2\widehat{\beta}^{[j]}\widehat{\sigma}_{\epsilon,\delta}}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\widehat{\gamma}_k}{\widehat{\gamma}_j}\widehat{\Omega}_{j\cdot}\right)\frac{1}{\sqrt{n}}W\right\|_2$ and $\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{\sqrt{n}}W\right\|_2$ converges to zero. By the lower bound in (78), we establish (79) holds for a sufficiently large n.

On the event $\mathcal{G}$, we apply the expression (76) and establish

$$\left|\mathcal{R}_k^{[j]}\right|\lesssim\left(1+\left|\frac{\gamma_k^*}{\gamma_j^*}\right|\right)\left(1+\left|\frac{\Gamma_j^*}{\gamma_j^*}\right|\right)\frac{1}{(\log n)^{1/4}}\cdot\frac{(\log n)^{1/4}}{\sqrt{n}}.$$

For a sufficiently large n, we further apply (78) and bound the above expression by

$$\left|\mathcal{R}_k^{[j]}\right|\leq 0.05\sqrt{\log n}\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{n}W\right\|_2.$$

C.6 Proof of Lemma 5

Note that

$$\begin{aligned}&\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)=\left(\Omega_{k,\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\Omega_{j,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)\\&+\left(\widehat{\Omega}_{k\cdot}-\Omega_{k,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)+\left(\widehat{\Omega}_{j\cdot}-\Omega_{j,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right).\end{aligned}$$

Following the same argument as (120), we establish that, on the event $\mathcal{G}$,

$$\begin{aligned}&\left|\left(\widehat{\Omega}_{k\cdot}-\Omega_{k,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)+\frac{\gamma_k^*}{\gamma_j^*}\left(\widehat{\Omega}_{j\cdot}-\Omega_{j,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)\right|\\&\lesssim\frac{\log n}{n}\left(1+\left|\frac{\gamma_k^*}{\gamma_j^*}\right|\right)\left(1+\left|\frac{\Gamma_j^*}{\gamma_j^*}\right|\right).\end{aligned}\tag{128}$$

Following the same argument as (119), we have

$$\mathbb{P}\left(\left|\frac{\left(\Omega_{k,\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\Omega_{j,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)}{\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}}\right|\geq C\left(1+\left|\frac{\gamma_k^*}{\gamma_j^*}\right|\right)\sqrt{\frac{\log n}{n}}\right)\leq n^{-c}.$$

Combined with (127), we have

$$\mathbb{P}\left(\left|\frac{\left(\Omega_{k,\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\Omega_{j,\cdot}\right)^{\intercal}\frac{1}{n}W^{\intercal}\left(\epsilon-\frac{\Gamma_j^*}{\gamma_j^*}\delta\right)}{\sqrt{\mathrm{Var}\left(\epsilon_1-\Gamma_j^*/\gamma_j^*\delta_1\right)}\left\|\left(\widehat{\Omega}_{k\cdot}-\frac{\gamma_k^*}{\gamma_j^*}\widehat{\Omega}_{j\cdot}\right)\frac{1}{n}W\right\|_2}\right|\geq 0.8\sqrt{\log n}\right)\geq\mathbb{P}(\mathcal{G})-n^{-c}.$$

Together with (128) and (127), we establish the lemma by applying the union bound.

D Additional Simulation Analysis

D.1 Additional Simulation Results for Settings S1 to S5

We present the complete simulation results for settings **S1** to **S5** detailed in Section 7.2. We vary γ_0 across $\{0.25, 0.5\}$, τ across $\{0.1, 0.2, 0.3, 0.4\}$ and n across $\{500, 1000, 2000, 5000\}$. The results are reported from Table D.1 to Table D.10. The main observation is similar to those in Section 7.2 in the main paper, which is summarized in the following.

1. The CIs by `TSHT` [14] and `CIIV` [34] achieve the 95% coverage level for a large sample size and a relatively large violation level, such as $n = 5000$ and $\tau = 0.3, 0.4$. For many settings with $\tau = 0.1, 0.2$, the CIs by `TSHT` and `CIIV` do not even have coverage when $n = 5000$. The CI by `CIIV` is more robust in the sense that its validity may require a smaller sample size than `TSHT`.

2. The CIs by the `Union` method [18] with $\bar{s} = p_{\rm z} - 1$ (assuming there are two valid IVs) achieve the desired coverage levels for all settings. The CIs by the `Union` method with $\bar{s} = \lceil p_{\rm z}/2 \rceil$ (assuming the majority rule) do not achieve the desired coverage level, except for the setting **S1** where the majority rule holds.

3. Our proposed searching and sampling CIs achieve the desired coverage levels in most settings. Settings **S1**, **S2** and **S4** are relatively easier as the corresponding finite-sample majority and plurality rules hold more plausibly. For the more challenging settings **S3** and **S5**, the combined intervals in general achieve the desired coverage level in most settings.

4. When the CIs by `TSHT` [14] and `CIIV` [34] are valid, their lengths are similar to the length of the CI by `oracle` TSLS, which has been justified in [14, 34]. The sampling CI, searching CI and CI by the `Union` are in general longer than the CI by the `oracle` TSLS, which is a price to pay for constructing uniformly valid CIs.

5. Among all CIs achieving the desired coverage level, the sampling CIs are typically the shortest CIs achieving the desired coverage levels. Both searching and sampling CIs are in general shorter than the CIs by the `Union` method.

We shall remark that even when the majority rule holds for the setting **S1**, we implement `TSHT`, `CIIV` and our proposed sampling and searching CIs by only assuming the plurality rule to hold. Even when the majority rule holds, CIs by `TSHT` and `CIIV` are under-coverage in the presence of weakly invalid instruments (e.g. $\tau = 0.1, 0.2$).

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.93	0.78	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.95	0.87	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.96	0.81	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5000	0.95	0.64	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.2	500	0.93	0.68	0.72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.93	0.62	0.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.96	0.45	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	5000	0.94	0.20	0.78	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99
0.3	500	0.95	0.55	0.65	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	1000	0.94	0.42	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.93	0.19	0.72	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98
	5000	0.93	0.57	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.96
0.4	500	0.92	0.38	0.64	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	1000	0.93	0.18	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
	2000	0.95	0.28	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
	5000	0.95	0.84	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.94

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.19	0.16	0.16	2.27	1.90	2.30	0.63	0.65	0.72	5.22	4.71	2.15	0.48
	1000	0.14	0.11	0.11	1.05	1.05	1.07	0.41	0.42	0.46	4.25	2.27	1.89	0.33
	2000	0.10	0.08	0.08	0.66	0.66	0.67	0.29	0.30	0.32	3.64	1.10	1.74	0.24
	5000	0.06	0.05	0.05	0.38	0.39	0.39	0.19	0.19	0.21	2.79	0.44	1.59	0.16
0.2	500	0.19	0.16	0.16	2.15	1.85	2.18	0.64	0.65	0.74	5.17	4.65	2.16	0.48
	1000	0.14	0.11	0.11	1.02	1.03	1.04	0.42	0.43	0.47	4.25	2.17	1.90	0.34
	2000	0.10	0.08	0.08	0.63	0.66	0.66	0.31	0.32	0.35	3.69	1.09	1.76	0.24
	5000	0.06	0.05	0.06	0.36	0.40	0.40	0.19	0.22	0.23	2.87	0.50	1.61	0.14
0.3	500	0.20	0.16	0.16	2.18	1.86	2.23	0.66	0.66	0.76	5.33	4.81	2.18	0.49
	1000	0.14	0.11	0.12	0.99	1.04	1.05	0.44	0.47	0.51	4.31	2.26	1.93	0.34
	2000	0.10	0.08	0.09	0.61	0.66	0.67	0.30	0.34	0.37	3.75	1.11	1.77	0.22
	5000	0.06	0.06	0.06	0.36	0.38	0.39	0.18	0.20	0.21	3.00	0.57	1.63	0.09
0.4	500	0.20	0.15	0.17	2.07	1.85	2.15	0.65	0.69	0.76	5.32	4.66	2.19	0.48
	1000	0.14	0.11	0.13	0.95	1.05	1.06	0.44	0.51	0.54	4.38	2.32	1.95	0.31
	2000	0.10	0.08	0.09	0.59	0.66	0.67	0.29	0.33	0.35	3.85	1.21	1.79	0.17
	5000	0.06	0.07	0.06	0.35	0.37	0.37	0.18	0.19	0.21	3.12	0.53	1.66	0.07

Table D.1: Empirical coverage and average lengths of CIs for setting **S1** with $\gamma_0 = 0.25$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.93	0.79	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	1000	0.96	0.70	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.95	0.50	0.62	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	5000	0.95	0.25	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
0.2	500	0.94	0.45	0.60	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99
	1000	0.95	0.26	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
	2000	0.96	0.38	0.82	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98
	5000	0.94	0.87	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95
0.3	500	0.95	0.19	0.77	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99
	1000	0.97	0.42	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
	2000	0.96	0.84	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
	5000	0.95	0.95	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
0.4	500	0.95	0.30	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
	1000	0.95	0.77	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
	2000	0.97	0.95	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
	5000	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.10	0.08	0.08	0.68	0.67	0.70	0.29	0.29	0.32	2.13	1.15	1.00	0.24
	1000	0.07	0.06	0.06	0.44	0.44	0.45	0.20	0.21	0.22	1.89	0.60	0.90	0.17
	2000	0.05	0.04	0.04	0.29	0.30	0.31	0.15	0.16	0.17	1.57	0.35	0.84	0.12
	5000	0.03	0.03	0.03	0.18	0.19	0.20	0.10	0.10	0.11	1.07	0.24	0.78	0.07
0.2	500	0.10	0.08	0.08	0.65	0.66	0.67	0.30	0.31	0.34	2.17	1.17	1.01	0.25
	1000	0.07	0.06	0.06	0.42	0.45	0.46	0.22	0.23	0.25	1.96	0.62	0.92	0.17
	2000	0.05	0.04	0.05	0.28	0.30	0.31	0.15	0.16	0.17	1.68	0.42	0.86	0.09
	5000	0.03	0.03	0.03	0.16	0.18	0.18	0.09	0.09	0.10	1.23	0.27	0.80	0.04
0.3	500	0.10	0.08	0.09	0.63	0.67	0.69	0.30	0.34	0.36	2.23	1.20	1.03	0.22
	1000	0.07	0.06	0.07	0.41	0.44	0.45	0.20	0.22	0.24	2.06	0.70	0.94	0.12
	2000	0.05	0.05	0.05	0.27	0.29	0.30	0.13	0.15	0.16	1.82	0.42	0.88	0.06
	5000	0.03	0.03	0.03	0.18	0.18	0.18	0.09	0.09	0.10	1.36	0.25	0.83	0.03
0.4	500	0.10	0.08	0.09	0.61	0.66	0.68	0.28	0.32	0.35	2.33	1.31	1.05	0.18
	1000	0.07	0.07	0.07	0.40	0.42	0.43	0.19	0.21	0.22	2.16	0.70	0.96	0.09
	2000	0.05	0.05	0.05	0.28	0.29	0.29	0.14	0.15	0.16	1.92	0.40	0.90	0.05
	5000	0.03	0.03	0.03	0.18	0.18	0.18	0.09	0.09	0.10	1.45	0.25	0.85	0.03

Table D.2: Empirical coverage and average lengths of CIs for setting **S1** with $\gamma_0 = 0.5$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2\rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.92	0.64	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86
	1000	0.93	0.80	0.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81
	2000	0.94	0.74	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.71
	5000	0.95	0.51	0.58	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.62	0.49
0.2	500	0.95	0.67	0.70	0.99	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.73
	1000	0.94	0.58	0.59	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52
	2000	0.93	0.33	0.52	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.95	0.23
	5000	0.95	0.14	0.70	0.99	1.00	1.00	0.99	0.98	1.00	1.00	1.00	0.33	0.00
0.3	500	0.94	0.55	0.52	0.99	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.48
	1000	0.93	0.31	0.50	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.17
	2000	0.96	0.14	0.64	1.00	0.99	1.00	0.99	0.98	1.00	1.00	1.00	0.95	0.00
	5000	0.96	0.23	0.82	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	0.65	0.00
0.4	500	0.93	0.48	0.49	0.96	0.98	0.99	0.99	0.97	0.99	1.00	1.00	1.00	0.23
	1000	0.95	0.19	0.66	0.99	0.99	1.00	0.99	0.97	1.00	1.00	1.00	1.00	0.02
	2000	0.96	0.11	0.74	0.99	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.00
	5000	0.94	0.69	0.92	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.26	0.19	0.19	3.11	1.86	3.20	1.26	0.71	1.36	7.46	7.46	3.73	0.26
	1000	0.18	0.14	0.13	0.98	0.95	1.01	0.45	0.44	0.49	6.29	5.07	3.45	0.15
	2000	0.13	0.11	0.10	0.60	0.60	0.61	0.30	0.30	0.33	5.70	3.12	3.26	0.10
	5000	0.08	0.07	0.06	0.35	0.36	0.36	0.20	0.21	0.22	4.97	1.31	3.07	0.05
0.2	500	0.26	0.19	0.21	2.84	1.82	2.96	1.31	0.72	1.42	7.45	7.46	3.77	0.23
	1000	0.18	0.14	0.14	0.95	0.94	1.00	0.47	0.47	0.53	6.32	5.09	3.46	0.12
	2000	0.13	0.12	0.10	0.58	0.61	0.61	0.32	0.35	0.37	5.71	3.04	3.26	0.06
	5000	0.08	0.12	0.07	0.32	0.36	0.37	0.22	0.25	0.27	5.09	1.33	3.08	0.00
0.3	500	0.26	0.18	0.21	2.66	1.81	2.78	1.30	0.74	1.42	7.52	7.44	3.77	0.19
	1000	0.18	0.14	0.14	0.90	0.93	0.98	0.47	0.51	0.56	6.39	5.17	3.49	0.08
	2000	0.13	0.17	0.11	0.54	0.61	0.62	0.33	0.39	0.42	5.80	3.19	3.29	0.01
	5000	0.08	0.18	0.08	0.30	0.34	0.36	0.20	0.21	0.24	5.22	1.41	3.08	0.00
0.4	500	0.26	0.18	0.23	2.57	1.81	2.72	1.43	0.80	1.55	7.61	7.55	3.81	0.13
	1000	0.18	0.15	0.16	0.85	0.94	0.98	0.48	0.56	0.62	6.46	5.16	3.52	0.02
	2000	0.13	0.26	0.12	0.50	0.59	0.61	0.35	0.37	0.43	5.95	3.25	3.33	0.00
	5000	0.08	0.12	0.08	0.30	0.33	0.33	0.20	0.20	0.22	5.34	1.38	3.11	0.00

Table D.3: Empirical coverage and average lengths of CIs for setting **S2** with $\gamma_0 = 0.25$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.94	0.74	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.76
	1000	0.93	0.68	0.65	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60
	2000	0.94	0.39	0.49	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.94	0.28
	5000	0.95	0.26	0.61	0.99	0.99	1.00	0.98	0.96	0.99	1.00	1.00	0.33	0.01
0.2	500	0.95	0.41	0.45	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.26
	1000	0.95	0.24	0.63	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.04
	2000	0.94	0.18	0.72	0.98	0.99	0.99	1.00	0.97	1.00	1.00	1.00	0.99	0.00
	5000	0.95	0.70	0.92	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.00
0.3	500	0.94	0.24	0.63	0.99	0.99	1.00	0.99	0.97	1.00	1.00	1.00	1.00	0.03
	1000	0.96	0.25	0.77	0.98	0.99	1.00	0.98	0.98	1.00	1.00	1.00	1.00	0.00
	2000	0.95	0.56	0.91	0.97	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.00
	5000	0.95	0.94	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
0.4	500	0.95	0.36	0.74	0.90	0.99	0.99	0.97	0.97	1.00	1.00	1.00	1.00	0.01
	1000	0.95	0.51	0.90	0.97	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.00
	2000	0.96	0.93	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
	5000	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.13	0.10	0.10	0.62	0.62	0.66	0.31	0.31	0.35	3.12	2.45	1.74	0.10
	1000	0.09	0.08	0.07	0.41	0.41	0.42	0.22	0.22	0.24	2.83	1.42	1.63	0.06
	2000	0.06	0.06	0.05	0.27	0.28	0.28	0.16	0.17	0.18	2.63	0.72	1.56	0.03
	5000	0.04	0.06	0.04	0.16	0.18	0.18	0.11	0.12	0.13	2.24	0.27	1.50	0.00
0.2	500	0.13	0.10	0.10	0.58	0.61	0.65	0.32	0.34	0.38	3.13	2.45	1.74	0.07
	1000	0.09	0.11	0.08	0.38	0.42	0.43	0.24	0.26	0.29	2.93	1.45	1.66	0.01
	2000	0.06	0.13	0.06	0.25	0.28	0.29	0.16	0.18	0.20	2.72	0.75	1.58	0.00
	5000	0.04	0.08	0.04	0.13	0.16	0.16	0.09	0.10	0.11	2.42	0.28	1.51	0.00
0.3	500	0.13	0.11	0.11	0.55	0.61	0.64	0.35	0.37	0.42	3.23	2.50	1.77	0.02
	1000	0.09	0.20	0.09	0.34	0.40	0.41	0.22	0.25	0.28	3.01	1.51	1.68	0.00
	2000	0.06	0.12	0.06	0.22	0.25	0.26	0.15	0.15	0.17	2.84	0.77	1.61	0.00
	5000	0.04	0.04	0.04	0.16	0.16	0.16	0.09	0.09	0.10	2.53	0.26	1.55	0.00
0.4	500	0.13	0.18	0.12	0.51	0.60	0.64	0.37	0.36	0.44	3.31	2.57	1.80	0.00
	1000	0.09	0.21	0.09	0.34	0.38	0.39	0.23	0.22	0.27	3.10	1.49	1.70	0.00
	2000	0.06	0.07	0.06	0.25	0.25	0.26	0.15	0.15	0.17	2.89	0.73	1.63	0.00
	5000	0.04	0.04	0.04	0.16	0.16	0.16	0.09	0.09	0.10	2.55	0.27	1.59	0.00

Table D.4: Empirical coverage and average lengths of CIs for setting **S2** with $\gamma_0 = 0.5$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.94	0.50	0.82	0.99	0.97	1.00	0.99	0.96	1.00	1.00	1.00	1.00	0.94
	1000	0.95	0.76	0.73	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83
	2000	0.94	0.76	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.77
	5000	0.95	0.52	0.54	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.47
0.2	500	0.95	0.59	0.71	0.98	0.97	0.99	1.00	0.95	1.00	1.00	1.00	1.00	0.84
	1000	0.96	0.76	0.60	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.61
	2000	0.96	0.42	0.47	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.26
	5000	0.94	0.15	0.62	1.00	0.99	1.00	0.99	0.97	1.00	1.00	1.00	1.00	0.01
0.3	500	0.94	0.59	0.63	0.96	0.95	0.99	0.99	0.94	0.99	1.00	1.00	1.00	0.74
	1000	0.95	0.59	0.53	0.93	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.32
	2000	0.95	0.27	0.67	0.97	0.99	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.03
	5000	0.94	0.24	0.81	0.97	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.00
0.4	500	0.95	0.52	0.56	0.91	0.94	0.99	0.99	0.94	1.00	1.00	1.00	1.00	0.62
	1000	0.95	0.40	0.65	0.90	0.99	0.99	0.99	0.98	1.00	1.00	1.00	1.00	0.19
	2000	0.93	0.34	0.72	0.90	0.98	0.99	0.98	0.96	1.00	1.00	1.00	1.00	0.01
	5000	0.94	0.70	0.92	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.26	0.18	0.20	3.44	2.31	3.59	1.56	0.87	1.71	4.29	4.44	2.02	0.53
	1000	0.18	0.13	0.13	1.17	1.04	1.29	0.66	0.49	0.75	3.48	3.39	1.84	0.19
	2000	0.13	0.10	0.10	0.62	0.62	0.65	0.31	0.31	0.35	3.08	2.42	1.70	0.10
	5000	0.08	0.07	0.06	0.35	0.36	0.36	0.20	0.20	0.22	2.77	1.12	1.59	0.05
0.2	500	0.26	0.18	0.21	3.43	2.23	3.59	1.69	0.87	1.83	4.41	4.56	2.05	0.48
	1000	0.18	0.13	0.14	1.09	1.01	1.21	0.66	0.50	0.73	3.51	3.40	1.85	0.17
	2000	0.13	0.10	0.10	0.58	0.61	0.64	0.33	0.35	0.39	3.13	2.42	1.72	0.06
	5000	0.08	0.12	0.07	0.32	0.37	0.37	0.21	0.24	0.27	2.84	1.20	1.61	0.01
0.3	500	0.26	0.18	0.22	3.32	2.23	3.52	1.80	0.87	1.91	4.34	4.46	2.08	0.44
	1000	0.18	0.12	0.15	1.01	1.02	1.18	0.77	0.54	0.85	3.58	3.44	1.88	0.11
	2000	0.13	0.12	0.11	0.54	0.62	0.64	0.36	0.38	0.43	3.21	2.49	1.75	0.01
	5000	0.08	0.19	0.08	0.29	0.34	0.35	0.20	0.22	0.24	2.94	1.24	1.64	0.00
0.4	500	0.26	0.17	0.24	3.22	2.23	3.46	2.04	0.92	2.16	4.42	4.52	2.11	0.35
	1000	0.18	0.13	0.16	0.97	1.02	1.17	0.85	0.58	0.93	3.62	3.52	1.90	0.06
	2000	0.13	0.18	0.12	0.51	0.59	0.63	0.37	0.38	0.45	3.28	2.55	1.77	0.00
	5000	0.08	0.15	0.08	0.31	0.33	0.33	0.19	0.20	0.22	3.02	1.20	1.66	0.00

Table D.5: Empirical coverage and average lengths of CIs for setting **S3** with $\gamma_0 = 0.25$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.94	0.64	0.75	0.96	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.85
	1000	0.94	0.72	0.65	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.62
	2000	0.95	0.42	0.50	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.28
	5000	0.94	0.23	0.64	1.00	0.99	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.01
0.2	500	0.97	0.63	0.64	0.90	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.63
	1000	0.95	0.40	0.63	0.92	0.99	0.99	0.99	0.96	1.00	1.00	1.00	1.00	0.17
	2000	0.95	0.38	0.73	0.96	0.98	0.99	0.98	0.98	1.00	1.00	1.00	1.00	0.00
	5000	0.96	0.72	0.93	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.00
0.3	500	0.95	0.42	0.68	0.72	0.98	0.98	0.99	0.94	1.00	1.00	1.00	1.00	0.39
	1000	0.96	0.52	0.71	0.73	0.99	1.00	0.98	0.97	0.99	1.00	1.00	1.00	0.08
	2000	0.94	0.73	0.91	0.93	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	0.00
	5000	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
0.4	500	0.93	0.45	0.73	0.60	0.96	0.97	0.93	0.94	0.98	1.00	1.00	1.00	0.22
	1000	0.95	0.66	0.87	0.71	0.99	1.00	0.92	0.98	1.00	1.00	1.00	1.00	0.01
	2000	0.94	0.86	0.93	0.98	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.00
	5000	0.94	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.13	0.09	0.09	0.63	0.69	0.76	0.39	0.35	0.46	1.79	1.76	0.95	0.16
	1000	0.09	0.07	0.07	0.39	0.42	0.44	0.22	0.23	0.25	1.59	1.30	0.87	0.08
	2000	0.06	0.05	0.05	0.27	0.28	0.30	0.16	0.17	0.18	1.47	0.80	0.83	0.03
	5000	0.04	0.06	0.04	0.16	0.18	0.18	0.11	0.12	0.13	1.35	0.33	0.79	0.00
0.2	500	0.13	0.09	0.10	0.57	0.66	0.72	0.45	0.36	0.51	1.83	1.77	0.97	0.13
	1000	0.09	0.08	0.08	0.35	0.42	0.43	0.26	0.26	0.30	1.65	1.36	0.90	0.03
	2000	0.06	0.11	0.06	0.25	0.28	0.29	0.17	0.18	0.21	1.55	0.87	0.85	0.00
	5000	0.04	0.08	0.04	0.14	0.16	0.16	0.09	0.10	0.11	1.45	0.33	0.81	0.00
0.3	500	0.13	0.09	0.12	0.46	0.65	0.70	0.53	0.38	0.58	1.90	1.85	1.01	0.06
	1000	0.09	0.13	0.09	0.32	0.41	0.44	0.27	0.25	0.31	1.74	1.41	0.93	0.01
	2000	0.06	0.15	0.06	0.23	0.26	0.27	0.16	0.15	0.18	1.64	0.85	0.88	0.00
	5000	0.04	0.05	0.04	0.16	0.16	0.16	0.09	0.09	0.10	1.49	0.33	0.84	0.00
0.4	500	0.13	0.10	0.13	0.43	0.66	0.72	0.57	0.40	0.64	2.00	1.90	1.04	0.03
	1000	0.09	0.23	0.09	0.30	0.39	0.42	0.26	0.24	0.31	1.83	1.42	0.96	0.00
	2000	0.06	0.15	0.06	0.25	0.26	0.28	0.16	0.15	0.18	1.69	0.82	0.92	0.00
	5000	0.04	0.05	0.04	0.16	0.16	0.16	0.09	0.09	0.10	1.51	0.32	0.88	0.00

Table D.6: Empirical coverage and average lengths of CIs for setting **S3** with $\gamma_0 = 0.5$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.94	0.90	0.79	0.97	0.88	0.98	0.97	0.88	0.98	1.00	0.99	0.99	0.00
	1000	0.94	0.94	0.86	0.99	0.97	0.99	0.99	0.97	0.99	1.00	0.98	0.89	0.00
	2000	0.96	0.87	0.87	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.57	0.00
	5000	0.94	0.76	0.78	0.99	0.98	0.99	0.99	0.98	0.99	1.00	0.98	0.16	0.00
0.2	500	0.94	0.82	0.70	0.94	0.87	0.95	0.95	0.86	0.96	1.00	0.97	0.99	0.00
	1000	0.94	0.87	0.75	0.97	0.97	0.98	0.97	0.96	0.98	1.00	0.97	0.91	0.00
	2000	0.94	0.57	0.64	0.97	0.97	0.98	0.97	0.96	0.98	0.99	0.95	0.58	0.00
	5000	0.95	0.43	0.55	0.98	0.96	0.99	0.98	0.95	0.99	0.98	0.98	0.22	0.00
0.3	500	0.96	0.76	0.62	0.93	0.86	0.96	0.94	0.85	0.96	1.00	0.98	0.99	0.00
	1000	0.96	0.79	0.61	0.94	0.94	0.98	0.95	0.93	0.98	1.00	0.98	0.93	0.00
	2000	0.96	0.32	0.55	0.95	0.95	0.98	0.95	0.94	0.99	0.99	0.97	0.66	0.00
	5000	0.96	0.43	0.75	0.97	0.93	0.97	0.98	0.94	0.98	0.96	0.96	0.22	0.00
0.4	500	0.95	0.64	0.49	0.88	0.81	0.92	0.91	0.81	0.92	1.00	0.98	1.00	0.00
	1000	0.94	0.65	0.48	0.86	0.91	0.95	0.87	0.89	0.95	0.98	0.95	0.95	0.00
	2000	0.95	0.22	0.62	0.92	0.93	0.97	0.95	0.91	0.98	0.97	0.96	0.74	0.00
	5000	0.96	0.64	0.93	0.98	0.97	0.98	0.98	0.97	0.99	0.97	0.96	0.31	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.47	0.38	0.44	2.04	1.58	2.15	1.64	1.34	1.85	5.04	3.28	2.33	0.02
	1000	0.32	0.64	0.25	0.91	0.89	0.95	0.81	0.76	0.88	4.20	1.64	2.08	0.00
	2000	0.23	0.19	0.16	0.58	0.58	0.59	0.50	0.49	0.53	3.55	0.80	1.84	0.00
	5000	0.14	0.12	0.10	0.34	0.34	0.35	0.30	0.30	0.33	2.69	0.28	1.48	0.00
0.2	500	0.47	0.38	0.43	1.91	1.61	2.02	1.57	1.30	1.75	5.14	3.25	2.32	0.04
	1000	0.32	0.59	0.24	0.87	0.87	0.92	0.80	0.74	0.88	4.27	1.60	2.10	0.00
	2000	0.23	0.20	0.17	0.56	0.55	0.58	0.51	0.47	0.55	3.59	0.78	1.84	0.00
	5000	0.14	0.14	0.11	0.32	0.30	0.32	0.28	0.26	0.30	2.75	0.27	1.49	0.00
0.3	500	0.46	0.36	0.44	1.81	1.49	1.93	1.62	1.25	1.82	5.13	3.19	2.30	0.02
	1000	0.32	0.58	0.27	0.84	0.85	0.94	0.81	0.73	0.92	4.24	1.63	2.08	0.00
	2000	0.23	0.22	0.19	0.58	0.51	0.61	0.61	0.44	0.66	3.66	0.76	1.85	0.00
	5000	0.14	0.16	0.14	0.28	0.25	0.28	0.29	0.21	0.30	2.80	0.23	1.49	0.00
0.4	500	0.48	0.36	0.52	1.80	1.57	1.97	1.68	1.34	1.88	5.14	3.14	2.30	0.02
	1000	0.32	0.52	0.30	0.86	0.84	1.00	0.90	0.70	1.03	4.32	1.55	2.10	0.00
	2000	0.23	0.25	0.22	0.58	0.47	0.63	0.77	0.38	0.83	3.71	0.67	1.86	0.00
	5000	0.14	0.16	0.14	0.26	0.23	0.27	0.34	0.18	0.36	2.84	0.19	1.47	0.00

Table D.7: Empirical coverage and average lengths of CIs for setting **S4** with $\gamma_0 = 0.25$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2\rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.96	0.89	0.83	0.98	0.98	0.98	0.99	0.98	0.99	1.00	0.99	1.00	0.00
	1000	0.95	0.78	0.78	0.99	0.99	1.00	0.99	0.98	1.00	1.00	0.99	0.89	0.00
	2000	0.93	0.67	0.68	0.99	0.97	0.99	0.99	0.96	0.99	0.99	0.98	0.55	0.00
	5000	0.95	0.49	0.49	0.97	0.94	0.97	0.98	0.94	0.98	0.98	0.95	0.10	0.00
0.2	500	0.93	0.75	0.64	0.94	0.96	0.98	0.96	0.94	0.98	1.00	0.97	1.00	0.00
	1000	0.93	0.49	0.56	0.96	0.95	0.98	0.97	0.93	0.98	0.99	0.97	0.96	0.00
	2000	0.95	0.41	0.60	0.97	0.92	0.97	0.96	0.92	0.98	0.97	0.94	0.67	0.00
	5000	0.93	0.77	0.88	0.95	0.93	0.95	0.96	0.94	0.98	0.94	0.94	0.20	0.00
0.3	500	0.96	0.57	0.55	0.82	0.92	0.96	0.85	0.91	0.96	1.00	0.97	1.00	0.00
	1000	0.95	0.34	0.64	0.94	0.92	0.96	0.95	0.91	0.98	0.97	0.95	0.99	0.00
	2000	0.94	0.64	0.87	0.97	0.94	0.97	0.97	0.94	0.98	0.95	0.94	0.91	0.00
	5000	0.95	0.95	0.95	0.96	0.95	0.96	0.98	0.98	0.99	0.95	0.96	0.78	0.00
0.4	500	0.94	0.48	0.54	0.69	0.84	0.91	0.73	0.84	0.92	1.00	0.94	1.00	0.00
	1000	0.94	0.33	0.81	0.92	0.90	0.96	0.92	0.88	0.96	0.98	0.93	0.99	0.00
	2000	0.93	0.73	0.91	0.95	0.95	0.95	0.96	0.94	0.97	0.94	0.93	0.98	0.00
	5000	0.98	0.97	0.97	0.98	0.98	0.98	0.98	0.97	0.99	0.98	0.97	0.92	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.23	0.35	0.17	0.57	0.58	0.60	0.48	0.47	0.53	2.15	0.84	1.10	0.00
	1000	0.16	0.14	0.12	0.39	0.38	0.39	0.33	0.32	0.35	1.89	0.42	1.01	0.00
	2000	0.11	0.10	0.08	0.26	0.26	0.27	0.24	0.22	0.25	1.60	0.23	0.88	0.00
	5000	0.07	0.07	0.06	0.16	0.15	0.16	0.14	0.13	0.15	1.15	0.13	0.68	0.00
0.2	500	0.23	0.34	0.17	0.53	0.55	0.59	0.51	0.45	0.57	2.19	0.88	1.09	0.00
	1000	0.16	0.15	0.13	0.38	0.35	0.39	0.37	0.29	0.39	1.94	0.42	1.01	0.00
	2000	0.11	0.12	0.10	0.23	0.21	0.24	0.23	0.18	0.25	1.64	0.20	0.89	0.00
	5000	0.07	0.08	0.07	0.12	0.11	0.12	0.10	0.09	0.11	1.21	0.09	0.67	0.00
0.3	500	0.23	0.31	0.20	0.49	0.52	0.59	0.50	0.42	0.58	2.27	0.83	1.09	0.00
	1000	0.16	0.18	0.15	0.38	0.30	0.40	0.50	0.24	0.53	1.98	0.38	0.99	0.00
	2000	0.11	0.13	0.11	0.21	0.19	0.22	0.23	0.14	0.25	1.67	0.15	0.88	0.00
	5000	0.07	0.08	0.07	0.11	0.11	0.11	0.09	0.09	0.10	1.23	0.08	0.66	0.00
0.4	500	0.23	0.30	0.23	0.45	0.49	0.60	0.51	0.39	0.63	2.29	0.80	1.09	0.00
	1000	0.16	0.19	0.16	0.39	0.28	0.41	0.61	0.22	0.64	2.01	0.33	1.00	0.00
	2000	0.11	0.12	0.11	0.20	0.18	0.20	0.26	0.14	0.27	1.69	0.14	0.87	0.00
	5000	0.07	0.08	0.07	0.11	0.11	0.11	0.09	0.09	0.10	1.24	0.08	0.67	0.00

Table D.8: Empirical coverage and average lengths of CIs for setting **S4** with $\gamma_0 = 0.5$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.95	0.69	0.69	0.96	0.94	0.98	0.98	0.93	0.99	1.00	0.99	1.00	0.57
	1000	0.94	0.63	0.74	0.97	0.98	0.99	0.99	0.97	1.00	1.00	0.99	0.91	0.30
	2000	0.94	0.63	0.81	0.96	0.99	0.99	0.98	0.97	0.99	1.00	0.98	0.46	0.06
	5000	0.95	0.77	0.80	0.99	0.98	0.99	0.99	0.97	0.99	1.00	0.98	0.07	0.00
0.2	500	0.93	0.56	0.60	0.92	0.92	0.96	0.96	0.88	0.97	1.00	0.99	1.00	0.42
	1000	0.95	0.60	0.67	0.94	0.96	0.97	0.96	0.95	0.99	1.00	0.99	0.88	0.13
	2000	0.95	0.55	0.68	0.91	0.97	0.98	0.96	0.95	0.98	0.99	0.98	0.28	0.00
	5000	0.95	0.42	0.54	0.97	0.94	0.98	0.98	0.93	0.99	0.99	0.97	0.01	0.00
0.3	500	0.96	0.41	0.54	0.89	0.90	0.95	0.95	0.86	0.98	1.00	0.99	1.00	0.20
	1000	0.95	0.55	0.57	0.83	0.94	0.96	0.93	0.93	0.98	0.99	0.98	0.90	0.03
	2000	0.97	0.33	0.54	0.87	0.96	0.97	0.92	0.95	0.97	0.99	0.97	0.40	0.00
	5000	0.94	0.45	0.68	0.97	0.94	0.98	0.97	0.94	0.98	0.95	0.95	0.04	0.00
0.4	500	0.95	0.32	0.42	0.83	0.82	0.90	0.88	0.79	0.93	1.00	0.98	1.00	0.07
	1000	0.96	0.42	0.48	0.69	0.92	0.94	0.85	0.89	0.95	0.98	0.97	0.94	0.00
	2000	0.95	0.19	0.58	0.85	0.92	0.96	0.87	0.91	0.97	0.96	0.95	0.74	0.00
	5000	0.95	0.63	0.90	0.95	0.94	0.95	0.97	0.95	0.98	0.95	0.93	0.25	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$		$\bar{s} = \lceil p_z/2 \rceil$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`	`TSLS`	`S-TSLS`
0.1	500	0.47	0.30	0.35	1.71	1.43	1.87	1.31	0.96	1.48	4.15	3.52	2.13	0.32
	1000	0.32	0.36	0.22	0.76	0.81	0.84	0.60	0.62	0.70	3.51	1.82	1.92	0.10
	2000	0.23	0.22	0.16	0.52	0.56	0.57	0.45	0.46	0.51	3.03	0.81	1.70	0.02
	5000	0.14	0.12	0.10	0.34	0.34	0.35	0.30	0.30	0.32	2.37	0.29	1.46	0.00
0.2	500	0.47	0.29	0.39	1.59	1.40	1.77	1.29	1.00	1.48	4.16	3.37	2.13	0.24
	1000	0.32	0.45	0.23	0.73	0.84	0.86	0.63	0.67	0.76	3.59	1.78	1.91	0.05
	2000	0.23	0.22	0.17	0.51	0.54	0.57	0.51	0.47	0.57	3.10	0.82	1.68	0.00
	5000	0.14	0.14	0.11	0.32	0.30	0.32	0.30	0.26	0.32	2.49	0.26	1.42	0.00
0.3	500	0.47	0.29	0.41	1.44	1.40	1.71	1.34	1.07	1.60	4.29	3.44	2.14	0.15
	1000	0.32	0.48	0.25	0.65	0.80	0.84	0.66	0.67	0.80	3.62	1.68	1.90	0.02
	2000	0.23	0.23	0.19	0.53	0.51	0.60	0.57	0.43	0.64	3.22	0.77	1.66	0.00
	5000	0.14	0.16	0.13	0.28	0.26	0.29	0.28	0.22	0.30	2.61	0.22	1.36	0.00
0.4	500	0.46	0.30	0.49	1.40	1.45	1.72	1.25	1.20	1.59	4.34	3.30	2.13	0.09
	1000	0.32	0.45	0.30	0.64	0.82	0.92	0.82	0.67	0.99	3.73	1.67	1.89	0.00
	2000	0.23	0.25	0.22	0.56	0.48	0.64	0.75	0.40	0.83	3.32	0.73	1.65	0.00
	5000	0.14	0.16	0.14	0.26	0.23	0.26	0.34	0.18	0.35	2.68	0.19	1.30	0.00

Table D.9: Empirical coverage and average lengths of CIs for setting **S5** with $\gamma_0 = 0.25$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.95	0.64	0.68	0.95	0.98	0.98	0.96	0.95	0.98	1.00	1.00	1.00	0.42
	1000	0.94	0.46	0.64	0.89	0.96	0.96	0.94	0.92	0.97	1.00	0.98	0.85	0.13
	2000	0.96	0.54	0.68	0.84	0.96	0.97	0.95	0.95	0.99	1.00	0.98	0.24	0.00
	5000	0.94	0.47	0.51	0.95	0.95	0.97	0.95	0.94	0.97	0.97	0.96	0.01	0.00
0.2	500	0.94	0.45	0.50	0.83	0.91	0.93	0.87	0.84	0.93	1.00	0.98	1.00	0.12
	1000	0.94	0.29	0.54	0.68	0.90	0.93	0.84	0.88	0.95	0.98	0.97	0.92	0.00
	2000	0.96	0.28	0.57	0.75	0.91	0.95	0.81	0.89	0.96	0.97	0.96	0.63	0.00
	5000	0.95	0.79	0.89	0.96	0.93	0.96	0.96	0.94	0.97	0.95	0.95	0.22	0.00
0.3	500	0.93	0.31	0.38	0.59	0.84	0.87	0.79	0.81	0.92	0.99	0.96	1.00	0.02
	1000	0.94	0.20	0.55	0.52	0.86	0.90	0.63	0.84	0.92	0.96	0.95	1.00	0.00
	2000	0.94	0.43	0.85	0.87	0.92	0.96	0.87	0.91	0.97	0.94	0.95	0.99	0.00
	5000	0.95	0.93	0.94	0.94	0.94	0.94	0.98	0.96	0.99	0.95	0.95	0.98	0.00
0.4	500	0.95	0.26	0.43	0.38	0.73	0.81	0.63	0.70	0.86	1.00	0.93	1.00	0.00
	1000	0.95	0.25	0.77	0.67	0.86	0.93	0.66	0.86	0.93	0.98	0.93	1.00	0.00
	2000	0.95	0.58	0.92	0.97	0.94	0.97	0.99	0.95	0.99	0.96	0.96	1.00	0.00
	5000	0.94	0.92	0.93	0.97	0.96	0.97	0.97	0.96	0.99	0.94	0.96	0.99	0.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s}=p_z-1$		$\bar{s}=\lceil p_z/2\rceil$	
τ	n	oracle	TSHT	CIIV	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	TSLS	S-TSLS	TSLS	S-TSLS
0.1	500	0.23	0.21	0.15	0.49	0.52	0.55	0.36	0.37	0.42	1.82	0.99	0.99	0.10
	1000	0.16	0.18	0.11	0.31	0.36	0.37	0.27	0.29	0.32	1.62	0.51	0.89	0.03
	2000	0.11	0.12	0.08	0.22	0.26	0.27	0.21	0.22	0.25	1.37	0.26	0.80	0.00
	5000	0.07	0.07	0.06	0.16	0.15	0.16	0.14	0.13	0.15	1.04	0.13	0.66	0.00
0.2	500	0.23	0.25	0.17	0.43	0.52	0.55	0.39	0.41	0.49	1.90	1.00	1.00	0.05
	1000	0.16	0.19	0.12	0.26	0.35	0.37	0.29	0.29	0.36	1.72	0.50	0.89	0.00
	2000	0.11	0.13	0.10	0.20	0.22	0.25	0.21	0.18	0.24	1.48	0.23	0.77	0.00
	5000	0.07	0.08	0.07	0.12	0.11	0.12	0.10	0.09	0.11	1.18	0.09	0.60	0.00
0.3	500	0.23	0.29	0.18	0.37	0.50	0.56	0.43	0.41	0.55	1.99	0.96	1.00	0.01
	1000	0.16	0.18	0.15	0.25	0.32	0.38	0.34	0.25	0.43	1.81	0.44	0.89	0.00
	2000	0.11	0.11	0.11	0.20	0.18	0.22	0.25	0.14	0.28	1.58	0.17	0.74	0.00
	5000	0.07	0.08	0.07	0.11	0.11	0.11	0.10	0.09	0.10	1.24	0.08	0.57	0.00
0.4	500	0.23	0.31	0.22	0.33	0.48	0.60	0.51	0.38	0.67	2.08	0.97	1.01	0.00
	1000	0.16	0.15	0.17	0.35	0.29	0.44	0.60	0.22	0.67	1.88	0.40	0.88	0.00
	2000	0.11	0.11	0.11	0.22	0.18	0.22	0.34	0.14	0.35	1.61	0.15	0.73	0.00
	5000	0.07	0.08	0.07	0.11	0.11	0.11	0.09	0.09	0.10	1.27	0.08	0.57	0.00

Table D.10: Empirical coverage and average lengths of CIs for setting **S5** with $\gamma_0 = 0.5$. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ and $\lceil p_z/2\rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

D.2 Additional Simulation Results for Settings CIIV-1 to CIIV-2

We now present the complete simulation results for settings **CIIV-1** to **CIIV-2**. The results are similar to those for settings **S1** to **S5** and the setting **CIIV-1** in Section 7.2 in the main paper. The empirical coverage of our proposed searching and sampling CIs in Table D.12 (corresponding to setting **CIIV-2**) is better than that in Table D.11 (corresponding to setting **CIIV-1**). This happens since the finite-sample plurality rule holds more plausibly in the setting **CIIV-2**, in comparison to the setting **CIIV-1**.

D.3 Computation Time Comparison

We report the computational time for setting **S1** in Table D.13 and observe that our proposed methods are computationally feasible. The `Union` method takes more time other algorithms as they search over a large number of sub-models. The most time-consuming algorithm is `Union` method with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ as it is involved with searching over all CIs constructed by $\lfloor p_{\mathrm{z}}/2 \rfloor$ candidate IVs. The computational time for settings **S1** to **S5** is similar and hence the computational time for other settings is omitted here for the sake of space.

We further report the computational time for the setting **CIIV-1** in Table D.14 and observe that our proposed methods are much faster than the `Union` method with $\bar{s} = p_{\mathrm{z}} - 1$. We do not implement the `Union` method with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ since the majority rule is not satisfied for the setting **CIIV-1**. From Table D.13, it is known that the `Union` method with $\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$ takes even longer time than that with $\bar{s} = p_{\mathrm{z}} - 1$.

Empirical Coverage

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`
0.1	500	0.95	0.00	0.08	1.00	1.00	1.00	0.98	0.92	0.99	1.00	1.00
	1000	0.96	0.00	0.06	1.00	1.00	1.00	0.91	0.75	0.94	1.00	1.00
	2000	0.94	0.00	0.12	1.00	0.99	1.00	0.88	0.52	0.89	1.00	1.00
	5000	0.93	0.00	0.51	0.99	0.92	1.00	0.95	0.79	0.96	1.00	1.00
0.2	500	0.94	0.00	0.13	1.00	1.00	1.00	0.84	0.55	0.88	1.00	1.00
	1000	0.95	0.00	0.44	1.00	0.94	1.00	0.92	0.73	0.94	1.00	1.00
	2000	0.96	0.00	0.76	0.73	0.95	0.98	0.92	0.92	0.97	1.00	1.00
	5000	0.96	0.01	0.93	0.06	1.00	1.00	0.11	1.00	1.00	1.00	1.00
0.3	500	0.95	0.00	0.47	1.00	0.93	1.00	0.92	0.70	0.93	1.00	1.00
	1000	0.95	0.00	0.79	0.59	0.96	0.97	0.89	0.95	0.97	1.00	1.00
	2000	0.95	0.00	0.92	0.01	1.00	1.00	0.05	1.00	1.00	1.00	1.00
	5000	0.94	0.77	0.93	0.98	0.99	0.99	0.98	0.99	0.99	1.00	1.00
0.4	500	0.94	0.00	0.65	0.85	0.89	0.96	0.94	0.85	0.96	1.00	1.00
	1000	0.94	0.00	0.89	0.02	0.99	0.99	0.12	0.99	0.99	1.00	1.00
	2000	0.94	0.13	0.94	0.58	0.92	0.92	0.59	0.92	0.92	1.00	1.00
	5000	0.95	0.91	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Average Length of Confidence Intervals

					Proposed Searching			Proposed Sampling			$\bar{s} = p_z - 1$	
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	`TSLS`	`S-TSLS`
0.1	500	0.09	0.06	0.07	1.08	1.05	1.10	0.35	0.33	0.38	1.08	1.15
	1000	0.07	0.04	0.05	0.68	0.66	0.70	0.28	0.24	0.29	0.79	0.84
	2000	0.05	0.03	0.04	0.46	0.43	0.48	0.24	0.18	0.25	0.62	0.65
	5000	0.03	0.02	0.03	0.28	0.26	0.33	0.20	0.12	0.21	0.48	0.50
0.2	500	0.09	0.06	0.09	1.07	1.01	1.12	0.48	0.36	0.51	1.33	1.40
	1000	0.07	0.04	0.07	0.68	0.62	0.77	0.42	0.26	0.45	1.04	1.09
	2000	0.05	0.03	0.05	0.40	0.42	0.57	0.34	0.19	0.38	0.88	0.91
	5000	0.03	0.05	0.03	0.05	0.26	0.27	0.26	0.12	0.35	0.73	0.72
0.3	500	0.09	0.06	0.10	1.07	0.97	1.24	0.67	0.39	0.71	1.63	1.71
	1000	0.07	0.05	0.07	0.56	0.62	0.84	0.48	0.27	0.55	1.33	1.38
	2000	0.05	0.05	0.05	0.08	0.42	0.44	0.42	0.19	0.56	1.15	1.16
	5000	0.03	0.06	0.03	0.24	0.26	0.26	0.12	0.12	0.13	0.99	0.74
0.4	500	0.09	0.06	0.10	1.04	0.96	1.39	0.89	0.39	0.95	1.96	2.04
	1000	0.07	0.06	0.07	0.22	0.63	0.70	0.48	0.27	0.67	1.62	1.66
	2000	0.05	0.22	0.05	0.20	0.39	0.41	0.19	0.18	0.29	1.43	1.27
	5000	0.03	0.04	0.03	0.26	0.26	0.26	0.12	0.12	0.13	1.24	0.77

Table D.11: Empirical coverage and average lengths of CIs for setting **CIIV-1**. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ correspond to the union methods assuming only two valid IVs.

Empirical Coverage

τ	n	oracle	TSHT	CIIV	Proposed Searching $\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	Proposed Sampling $\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\bar{s}=p_z-1$ TSLS	S-TSLS
0.1	500	0.95	0.91	0.68	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.94	0.85	0.62	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	2000	0.95	0.65	0.60	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00
	5000	0.94	0.66	0.80	0.99	1.00	1.00	1.00	0.98	1.00	1.00	1.00
0.2	500	0.95	0.61	0.59	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00
	1000	0.96	0.56	0.81	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00
	2000	0.93	0.62	0.85	0.79	1.00	1.00	0.98	1.00	1.00	1.00	1.00
	5000	0.96	0.86	0.94	0.54	1.00	1.00	0.87	1.00	1.00	1.00	1.00
0.3	500	0.96	0.58	0.79	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00
	1000	0.97	0.59	0.88	0.63	1.00	1.00	0.96	1.00	1.00	1.00	1.00
	2000	0.94	0.73	0.91	0.22	1.00	1.00	0.81	1.00	1.00	1.00	1.00
	5000	0.96	0.72	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.4	500	0.94	0.61	0.82	0.86	1.00	1.00	0.98	0.99	1.00	1.00	1.00
	1000	0.95	0.60	0.88	0.11	1.00	1.00	0.76	1.00	1.00	1.00	1.00
	2000	0.97	0.75	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5000	0.95	0.88	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Average Length of Confidence Intervals

τ	n	oracle	TSHT	CIIV	Proposed Searching $\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	Proposed Sampling $\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	Comb	$\bar{s}=p_z-1$ TSLS	S-TSLS
0.1	500	0.09	0.06	0.07	1.10	1.07	1.12	0.38	0.35	0.40	1.20	1.27
	1000	0.07	0.04	0.05	0.70	0.67	0.71	0.30	0.26	0.31	0.93	0.97
	2000	0.05	0.03	0.04	0.47	0.44	0.48	0.23	0.19	0.24	0.77	0.80
	5000	0.03	0.02	0.03	0.25	0.26	0.28	0.15	0.12	0.16	0.66	0.68
0.2	500	0.09	0.06	0.09	1.11	1.04	1.14	0.50	0.38	0.52	1.68	1.75
	1000	0.07	0.05	0.06	0.68	0.64	0.70	0.37	0.27	0.39	1.42	1.46
	2000	0.05	0.04	0.05	0.30	0.42	0.43	0.19	0.18	0.22	1.28	1.31
	5000	0.03	0.05	0.03	0.13	0.26	0.26	0.12	0.12	0.16	1.18	1.03
0.3	500	0.09	0.07	0.09	1.08	1.01	1.11	0.55	0.39	0.57	2.23	2.30
	1000	0.07	0.06	0.07	0.42	0.63	0.64	0.28	0.26	0.32	1.96	2.00
	2000	0.05	0.05	0.05	0.13	0.42	0.42	0.25	0.19	0.30	1.81	1.71
	5000	0.03	0.10	0.03	0.24	0.26	0.26	0.11	0.12	0.13	1.68	1.03
0.4	500	0.09	0.08	0.09	0.88	0.99	1.06	0.50	0.39	0.55	2.81	2.88
	1000	0.07	0.06	0.07	0.16	0.63	0.63	0.41	0.26	0.49	2.53	2.49
	2000	0.05	0.19	0.05	0.34	0.43	0.43	0.15	0.20	0.20	2.35	1.82
	5000	0.03	0.04	0.03	0.26	0.26	0.26	0.12	0.12	0.13	2.15	1.10

Table D.12: Empirical coverage and average lengths of CIs for setting **CIIV-2**. The columns indexed with `oracle`, `TSHT` and `CIIV` represent the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ represent our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively; the column indexed with "Comb" is a union of the corresponding two intervals. `TSLS` and `S-TSLS` denote the union method with TSLS estimators and TSLS estimators (passing a Sargan test), respectively. The columns indexed with $p_z - 1$ correspond to the union methods assuming only two valid IVs.

					Proposed Searching		Proposed Sampling			
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	$\bar{s} = p_{\mathrm{z}} - 1$	$\bar{s} = \lceil p_{\mathrm{z}}/2 \rceil$
0.1	500	0.01	0.00	0.04	0.32	0.33	1.08	0.96	5.22	22.61
	1000	0.01	0.01	0.05	0.31	0.35	1.08	0.99	7.41	30.67
	2000	0.01	0.01	0.07	0.32	0.37	1.21	0.99	11.60	45.94
	5000	0.03	0.03	0.15	0.36	0.50	1.59	0.99	27.55	105.68
0.2	500	0.01	0.00	0.05	0.31	0.34	1.22	0.94	5.17	22.31
	1000	0.01	0.01	0.06	0.31	0.35	1.37	0.85	7.34	30.35
	2000	0.01	0.01	0.08	0.32	0.39	1.64	0.88	11.98	47.64
	5000	0.03	0.03	0.18	0.36	0.52	1.53	1.12	27.90	106.89
0.3	500	0.01	0.00	0.06	0.32	0.36	1.47	0.85	5.48	23.63
	1000	0.01	0.01	0.07	0.32	0.38	1.66	0.90	7.75	32.16
	2000	0.01	0.01	0.10	0.33	0.42	1.39	0.99	12.43	49.52
	5000	0.03	0.03	0.18	0.36	0.52	0.98	1.10	27.23	104.13
0.4	500	0.01	0.00	0.06	0.32	0.36	1.63	0.87	5.54	23.99
	1000	0.01	0.01	0.08	0.32	0.39	1.40	0.98	7.91	32.80
	2000	0.02	0.01	0.11	0.33	0.42	0.98	1.01	12.61	50.15
	5000	0.03	0.03	0.21	0.38	0.58	1.06	1.21	30.15	116.13

Table D.13: Computation time comparison for setting **S1** with $\gamma_0 = 0.5$. All computation time are reported in the unit of second. The columns indexed with `oracle`, `TSHT` and `CIIV` correspond the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ correspond to our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively. The columns indexed with $p_{\mathrm{z}} - 1$ and $\lceil p_{\mathrm{z}}/2 \rceil$ correspond to the union methods assuming only two valid IVs and the majority rule, respectively.

					Proposed Searching		Proposed Sampling		
τ	n	`oracle`	`TSHT`	`CIIV`	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	$\widehat{\mathcal{V}}^{\texttt{TSHT}}$	$\widehat{\mathcal{V}}^{\texttt{CIIV}}$	$\bar{s} = p_{\mathrm{z}} - 1$
0.1	500	0.01	0.01	0.05	0.50	0.49	2.45	1.98	24.33
	1000	0.01	0.01	0.11	0.55	0.59	3.32	2.24	40.24
	2000	0.01	0.02	0.24	0.56	0.72	4.36	2.38	64.86
	5000	0.03	0.06	0.79	0.76	1.44	8.13	3.54	171.54
0.2	500	0.01	0.01	0.16	0.68	0.75	5.22	2.72	32.22
	1000	0.01	0.02	0.28	0.63	0.83	6.63	2.70	45.53
	2000	0.02	0.02	0.42	0.60	0.93	7.67	2.49	67.66
	5000	0.03	0.05	0.84	0.54	1.41	6.87	2.63	148.46
0.3	500	0.01	0.01	0.21	0.58	0.71	6.17	2.35	27.49
	1000	0.01	0.02	0.33	0.62	0.87	8.18	2.56	43.82
	2000	0.02	0.03	0.50	0.55	1.06	7.49	2.45	74.75
	5000	0.03	0.06	0.99	0.66	1.60	2.97	2.83	166.95
0.4	500	0.01	0.01	0.32	0.78	0.98	10.14	3.07	36.55
	1000	0.01	0.02	0.44	0.69	1.11	9.51	2.90	54.29
	2000	0.02	0.02	0.50	0.55	1.04	6.14	2.57	72.03
	5000	0.03	0.06	0.95	0.64	1.54	1.77	2.66	158.15

Table D.14: Computation time comparison for setting **CIIV-1**. All computation time are reported in the unit of second. The columns indexed with `oracle`, `TSHT` and `CIIV` correspond the oracle TSLS estimator with the knowledge of valid IVs, the `TSHT` estimator and the `CIIV` estimator, respectively. Under the columns indexed with "Proposed Searching" (or "Proposed Sampling"), the columns indexed with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$ correspond to our proposed searching CI (or sampling CI) with $\widehat{\mathcal{V}}^{\texttt{TSHT}}$ and $\widehat{\mathcal{V}}^{\texttt{CIIV}}$, respectively. The column indexed with $p_{\mathrm{z}}-1$ corresponds to the union methods assuming only two valid IVs.