Selection of invalid instruments can improve estimation in Mendelian randomization

Ashish Patel a, Francis J. DiTraglia b, Verena Zuber c, & Stephen Burgess a,d

a MRC Biostatistics Unit, University of Cambridge
b Department of Economics, University of Oxford
c Department of Epidemiology and Biostatistics, Imperial College London
d Cardiovascular Epidemiology Unit, University of Cambridge

Abstract

Mendelian randomization (MR) is a widely-used method to identify causal links between a risk factor and disease. A fundamental part of any MR analysis is to choose appropriate genetic variants as instrumental variables. Current practice usually involves selecting only those genetic variants that are deemed to satisfy certain exclusion restrictions, in a bid to remove bias from unobserved confounding. Many more genetic variants may violate these exclusion restrictions due to unknown pleiotropic effects (i.e. direct effects on the outcome not via the exposure), but their inclusion could increase the precision of causal effect estimates at the cost of allowing some bias. We explore how to optimally tackle this bias-variance trade-off by carefully choosing from many weak (Zhao et al., 2020) and locally invalid (DiTraglia, 2016) instruments. Specifically, we study a focused instrument selection approach for publicly available two-sample summary data on genetic associations, whereby genetic variants are selected on the basis of how they impact the asymptotic mean square error of causal effect estimates. We show how different restrictions on the nature of pleiotropic effects have important implications for the quality of post-selection inferences. In particular, a focused selection approach under systematic pleiotropy allows for consistent model selection, but in practice can be susceptible to winner’s curse biases. Whereas a more general form of idiosyncratic pleiotropy allows only conservative model selection, but offers uniformly valid confidence intervals. We propose a novel method to tighten honest confidence intervals through support restrictions on pleiotropy. We apply our results to several real data examples which suggest that the optimal selection of instruments does not only involve biologically-justified valid instruments, but additionally hundreds of potentially pleiotropic variants.

*We thank Dipender Gill for helpful discussions.
Ashish Patel (ashish.patel@mrc-bsu.cam.ac.uk); Francis J. DiTraglia (francis.ditraglia@economics.ox.ac.uk);
Verena Zuber (v.zuber@imperial.ac.uk); Stephen Burgess (sb452@medschl.cam.ac.uk)
1 Introduction

1.1 Background

MR uses genetic variants as instrumental variables (IVs) to estimate the causal effect of an exposure on an outcome in the presence of unobserved confounding. By Mendel’s second law, genetic variants sort independently of other traits. Thus, genetic variants, which are fixed at conception, can provide a source of exogenous variation in a risk factor of interest, allowing analyses which are less vulnerable to reverse causality and confounding (Davey Smith and Ebrahim, 2003).

Large-scale consortia genome-wide association studies (meta-GWASs) have identified large numbers of genetic variants that are robustly associated with a wide range of traits. Due in part to privacy issues, often only summary statistics of these genetic associations are made publicly available. Since these results are easily accessible, MR investigations increasingly rely on inferential methods that require only two-sample summary data (Burgess et al., 2015). In such applications, variant–exposure associations are obtained from a non-overlapping sample drawn from the same population used to compute variant–outcome associations.

As in usual IV analyses, identifying causal effects through MR requires some key assumptions. For a genetic variant to be a valid instrument, it must be associated with the exposure; \( \beta_{X_j} \neq 0 \) in Fig. 1 (relevance). Second, the variant must be uncorrelated with unobserved confounders. Third, any effect that the variant has on the outcome must be mediated by its effect on the exposure; \( \tau_j = 0 \) in Fig. 1 (exclusion).

![Diagram](https://via.placeholder.com/150)

Figure 1. The effect of genetic variant \( Z_j \) on the exposure \( X \) and outcome \( Y \), where \( U \) is an unobserved confounder.

In practice, violations of both the relevance and exclusion conditions are common (Lawlor et al., 2008). Genetics generally explains a low proportion of variation in complex traits, and thus MR routinely makes use of genetic variants that are only weakly associated with the exposure (Burgess and Thompson, 2011). In popular MR methods for two-sample summary data, the use of weak instruments leads to low-powered testing, i.e. a tendency to maintain a null hypothesis of no causal effect (\( \theta_0 = 0 \)) even when one exists (Lawlor, 2016).
While the strength of instruments can be empirically tested, the exclusion restriction cannot be verified in a homogeneous effects setting. GWASs have highlighted the widespread phenomenon of pleiotropy, where a single genetic variant may influence several traits (Solovieff et al., 2013; Hemani et al., 2018). In this paper, we will refer to pleiotropy as the problematic case in which a genetic variant has independent effects on the exposure and outcome. The biological mechanism of certain genetic effects may be well understood (for example, when using protein risk factors for drug target validation; see Schmidt et al., 2020), but it is generally difficult to rule out the possibility that many candidate instruments violate the exclusion restriction (Verbanck et al., 2018).

1.2 Recent developments

Reflecting demand from applied researchers, recent methodological developments in MR have proposed estimators that allow at least some of the following features: (i) a large number of genetic variants; (ii) instruments that have weak effects on the exposure; (iii) instruments that may be pleiotropic; and (iv) two-sample summary data as inputs. We briefly note some related work in this regard.

Popular methods that have tackled these problems can be categorised into two general approaches. The first group of methods allows inclusion of weak and/or pleiotropic instruments while remaining robust to their harmful effects. In the many weak instruments and summary-data setting, Zhao et al. (2020) and Ye et al. (2020) study valid inference under systematic pleiotropy (also called balanced pleiotropy), in which a genetic variant that breaks the exclusion restriction is equally like to have a positive effect on the outcome as they are to have a negative one, on average across all instruments. Sanderson et al. (2020) extend the systematic pleiotropy model to allow several risk factors to be considered, and derive a heterogeneity statistic to assess the validity of weak instruments. Burgess et al. (2020) study robust estimation under systematic pleiotropy by grouping together instruments offering similar causal effect estimates.

For robustness to a more severe pleiotropy problem, the weighted-median estimator studied by Bowden et al. (2016) offers consistent estimation by down-weighting misleading effects from outliers under the assumption that 50% of instruments are valid. Morrison et al. (2020) study a Bayesian model which permits some variants to have correlated pleiotropic effects. An even more general model studies the case where the exclusion restriction is not satisfied, even on average, by any instrument. With individual-level data, Tchetgen Tchetgen et al. (2019) and Liu et al. (2020) exploit heteroskedasticity in genetic associations for identification and estimation of causal effects, in the spirit of Lewbel (2012).

The second group of methods aim to directly screen out problematic candidate instruments.
Kang et al. (2016) derive performance bounds of a penalised regression estimator which selects valid instruments under the assumption that 50% of them are valid. In a similar setting, Windmeijer et al. (2019) study a median-based adaptive Lasso method and provide conditions for asymptotic normality. Guo et al. (2018) propose valid inference under a plurality rule, which requires that, of subsets of candidate instruments which report the same asymptotic causal effect estimate, the largest set contains only valid instruments. In some cases this plurality rule can be weaker than the majority rules required by median-based approaches. Lastly, Bi et al. (2019) propose a selective inference approach with data-dependent critical values which account for first-stage selection of instruments based on their validity.

1.3 Scope of this paper

MR analysis under pleiotropy is statistically challenging, with several works noted above imposing strong restrictions on pleiotropic effects in order to obtain asymptotically unbiased estimates of causal effects.

Our focus in this paper is different from existing MR methods. We assume a known set of genetic variants is able to identify causal effects, but we abandon the goal of asymptotically unbiased estimation in favour of minimising estimation risk. In typical applications, relatively few genetic variants may be plausibly valid instruments. By only choosing those valid instruments, we will have asymptotically unbiased, but potentially imprecise causal effect estimates. The additional inclusion of slightly pleiotropic variants that are robustly associated with the risk factor could significantly improve precision at the cost of allowing a small amount of bias. Our work explores two questions: (i) how can we optimally tackle this asymptotic bias-variance trade-off to improve estimation with many potentially pleiotropic genetic variants?; (ii) how do assumptions on pleiotropy influence the quality of post-selection inferences?

In order to identify and estimate the asymptotic bias associated with each chosen set of instruments, we assume some proportion of genetic variants are known a priori to be valid instruments. This describes a common setting in which researchers may be confident in the validity of some instruments, but they are unsure whether any additional instruments should be included for extra precision. We discuss several examples where a set of genetic variants can be biologically-justified as valid instruments.

To draw a connection with current practices, we start by considering the problem of choosing from sets of additional variants that satisfy systematic pleiotropy. Since systematic pleiotropy allows asymptotically unbiased estimation, the selection problem reduces to finding the asymptotic variance-minimising set of instruments. Here an asymptotic framework with many additional instruments yields consistent model selection, in which the uncertainty
from model selection should be negligible for asymptotic analysis. However, in finite-samples, ignoring this uncertainty will lead to inflated type 1 error rates, which we illustrate with a simple simulation exercise.

While assumptions such as systematic pleiotropy are convenient for statistical analysis, they generally lack plausible justification across all variants genome-wide. In some applications, we may not be comfortable imposing any group structure on pleiotropy. For this more general setting, we study estimation under *idiosyncratic pleiotropy*, where each additional variant potentially has its own independent effect on the outcome, with no restrictions on its direction or relative magnitude. This induces a non-ignorable asymptotic bias for causal effect estimation when using additional instruments.

Building on DiTraglia (2016)’s framework with a finite number of strong instruments, we develop a *focused* instrument selection approach for two-sample summary data on genetic associations, where the number of weak instruments can grow at the same rate as the sample size. In particular, we choose from sets of additional genetic variants which compete on the basis of how they impact the asymptotic mean square error (AMSE) of causal effect estimates. These sets may be data-driven, and have overlapping variants. We then derive the asymptotic properties of post-selection estimators and study how to make honest inferences which account for the extra uncertainty from first-stage selection. In a locally invalid instruments framework, the asymptotic bias can only be estimated up to its asymptotic distribution. We thus have *conservative* model selection for the idiosyncratic pleiotropy model, rather than consistent model selection as obtained under systematic pleiotropy.

Consistent model selection is a feature of several existing MR methods which try to screen and remove ‘very invalid’ candidate instruments. The approaches of Kang et al. (2016), Guo et al. (2018), Windmeijer et al. (2019) and others are able to retain only valid instruments in large samples, whereas our focus here is on allowing the inclusion of invalid instruments to the extent that they lower the AMSE. For estimation, the focused selection approach offers a clear improvement in terms of AMSE under the idiosyncratic pleiotropy model.

This distinction between conservative and consistent model selection also has important consequences for the accuracy of post-selection inferences. Under conservative model selection, we can construct confidence intervals with uniformly correct coverage across the parameter space of causal effects and pleiotropy parameters. However, actual coverage probabilities will be greater than nominal coverage, suggesting the intervals are relatively wide. On the other hand, confidence intervals based on consistent selection procedures are not uniformly consistent over parameter space, potentially leading to confidence intervals with under-coverage and size-distorted tests (Leeb and Pötscher, 2005; Yang, 2005).

While exact inference is not possible, we consider two approaches to improve post-selection inference based on DiTraglia (2016). First, confidence intervals which account for uncertainty
in model selection, but not uncertainty in bias estimation, may offer narrower intervals with competitive coverage when the set of valid instruments is weak.

Second, we propose a novel way to tighten DiTraglia (2016)’s uniformly valid confidence intervals by using support restrictions on pleiotropy. The support restrictions may be motivated in practice in several ways; for example, through assumptions on the direction of pleiotropic effects (Nevo and Rosen, 2012), or by knowledge of low heritability of a disease. By exploiting support restrictions, our method is shown to substantially reduce the width of honest confidence intervals, thus highlighting the potential of the focused instrument selection approach to improve inference in MR, as well as estimation.

Our theoretical results are applied to four real data examples, which further suggest that the optimal choice of instruments does not only involve a small set of biologically-justified valid instruments, but additionally hundreds of potentially pleiotropic variants.

Finally, the requirement that some genetic variants are known to be valid could be undesirable for more exploratory investigations in which little may be known about the functions of relevant genes. In such cases, the approach considered here could be useful for sensitivity analyses post-selection, by assessing the bias-variance trade-off relative to the set of instruments chosen by the researcher.

We use the following notation and abbreviations: $\xrightarrow{P}$ ‘converges in probability to’; $\xrightarrow{D}$ ‘converges in distribution to’; $\sim$ ‘is asymptotically distributed as’. For any sequences $a_n$ and $b_n$, if $a_n = O(b_n)$, then there exists a positive constant $C$ and a positive integer $N$ such that for all $n \geq N$, $b_n > 0$ and $|a_n| \leq C b_n$. If $a_n = o(b_n)$, then $|a_n|/b_n \to 0$ as $n \to \infty$. Also, if $a_n = \Theta(b_n)$, then there exist positive constants $C_1$ and $C_2$, $C_1 \leq C_2 < \infty$, and a positive integer $N$ such that $C_1 b_n \leq a_n \leq C_2 b_n$ for all $n \geq N$. The proofs of theoretical results are given in Supplementary Material.

## 2 Many weak instruments and summary data

### 2.1 Two-sample summary data

We first outline our assumptions on genetic association summary data. Let $Z = \{Z_1, \ldots, Z_p\}$ be a set of mutually uncorrelated genetic variants. The parameter of interest is the causal effect $\theta_0$ of the exposure $X$ on the outcome $Y$. For any valid instrument $Z_j$, under a linear model we have $\beta_{Yj} = \theta_0 \beta_{Xj}$, where $\beta_{Xj}$ denotes the association of the $j$-th instrument with the exposure, and $\beta_{Yj}$ its association with the outcome.

For each $j \in [p]$, we observe the estimated effect size $\hat{\beta}_{Xj}$ and standard error $\sigma_{Xj}$ from the $X$ on $Z_j$ linear regression with a constant term included. The quantities $\hat{\beta}_{Yj}$ and $\sigma_{Yj}$ for the
variant–outcome associations are defined analogously.

We assume that variant–exposure associations \((\hat{\beta}_X, \sigma_X)\) are taken from a mutually independent \(n_X\)-sized sample from an \(n_Y\)-sized one used to compute the variant–outcome associations \((\hat{\beta}_Y, \sigma_Y)\). For ease of exposition, we also assume \(n := n_Y = cn_X\), for some unknown constant \(0 < c < \infty\), so that our asymptotic analysis does not consider the information from one study to be negligible relative to the other. Both samples are drawn from the same joint distribution of \((X, Y, Z)\).

**Assumption 1** (precision of genetic association studies). For each \(j \in [p]\), let \(\hat{\beta}_X \sim N(\beta_X, \sigma^2_X)\) and \(\hat{\beta}_Y \sim N(\beta_Y, \sigma^2_Y)\), where all estimated associations \(\{\hat{\beta}_X\}_{j=1}^p\) and \(\{\hat{\beta}_Y\}_{j=1}^p\) are mutually independent. The variances \(\sigma^2_X\) and \(\sigma^2_Y\) are assumed known, and satisfy \(\sigma^2_X = \Theta(1/n)\) and \(\sigma^2_Y = \Theta(1/n)\).

Assumption 1 is taken from Zhao et al. (2020) (henceforth ZWHBS) and states a normal approximation of estimated genetic associations which is typically justified by asymptotic arguments under large random sampling expected in genetic association studies. The assumption that the true variances of the estimated marginal associations exactly match the reported standard errors is commonly maintained in MR given the large number of participants in GWASs. A theoretical justification for this is given by Ye et al. (2020). To simplify notation, the dependence of the standard errors on sample sizes is not made explicit, but they are assumed to decrease at the usual parametric rate.

The widely-used two-sample design obtains variant–exposure associations from a representative and non-overlapping sample from variant–outcome associations, and therefore independence of estimated associations between the two studies is reasonable. It is also common practice for researchers to consider only uncorrelated genetic variants, where the uncorrelatedness of genetic variants can be easily verified using commands in popular open-source software (Hemani et al., 2018; Yavorska et al., 2020). Then, joint normality of genetic associations within each study implies the independence restrictions in Assumption 1.

### 2.2 Many weak instruments

In typical applications, we may expect many genetic variants to have weak effects on the exposure where, loosely speaking, the \(j\)-th instrument is weak if \(\beta_X \approx 0\). This can cause difficulties for identifying and estimating the causal effect.

While we assume the causal effect is point identified, we work with the ‘many weak instruments’ set-up (see, for example, ZWHBS and Newey and Windmeijer, 2009) under which standard errors based on strong instruments and ‘fixed \(p\)’ asymptotics are too small. This is
of particular concern given our focus, since under-estimated variances could cause an MSE selection criteria to falsely recommend the inclusion of irrelevant instruments.

**Assumption 2** (many weak instruments). For $\beta_X = (\beta_{X_1}, \ldots, \beta_{X_p})'$, (i) $\|\beta_X\|_2 = \Theta(1)$ as $p \to \infty$; (ii) $\|\beta_X\|_3 \to 0$ as $p \to \infty$; (iii) $p/n = O(1)$ as $n, p \to \infty$.

Assumption 2 describes the many weak instruments set-up of ZWHBS. Assumption 2(i) limits the magnitude of all genetic associations with the exposure, and implies that the explanatory power of any individual variant is decreasing as $p \to \infty$. Assumption 2(ii) limits the skewness of the distribution of variant–exposure associations across all variants. Assumption 2(iii) is a rate restriction which ensures asymptotic normality for MR estimates using any set of instruments; it is usually trivially satisfied if at least some genetic variants are robustly associated with the exposure, since GWAS sample sizes are expected to be larger than the number of candidate instruments in an MR analysis.

### 2.3 Valid and additional instrument sets

In order to identify the bias associated with choosing potentially pleiotropic variants as instruments, we assume a subset of the $p$ genetic variants $V$ are known to be valid instruments a priori. For example, a few genes may be well-known to have functional effects on an exposure; genetic variants from those gene regions may then be considered more likely to satisfy the exclusion restriction (Schmidt et al., 2020).

**Assumption 3** (valid instruments). $V$ is a subset of valid instruments, $1 \leq |V| < p$.

The set of valid instruments $V$ is always included for estimation of $\theta_0$. Furthermore, an estimate of $\theta_0$ that only uses those $|V|$ variants is asymptotically unbiased and therefore forms a basis from which to measure the bias from including additional variants. Our interest is in whether any of the remaining $p - |V|$ genetic variants should be included to improve estimation. Searching over all possible combinations of additional variants results in $2^{(p - |V|)} - 1$ additional instrument sets, which, for our $p \to \infty$ setting provides a huge computational burden.

Instead, a practical way to handle many additional variants is to group them into a smaller number of additional instrument sets. These sets may be chosen based on biological considerations, by grouping together variants from specific genes known to have similar functional effects on the exposure. They may be chosen in a data-driven way, by grouping together variants with similar measured associations with the exposure and outcome. The sets may be partitions of the $p - |V|$ additional variants, or they could be overlapping so that some variants with stronger effects are prioritised with inclusion into several sets.
Thus we do not restrict how the additional instrument sets are formed. Furthermore, the total number of sets may grow with the total number of variants $p$.

**Assumption 4** (additional instrument sets). $S_1, \ldots, S_K$ are additional instrument sets of the remaining $p - |V|$ additional variants.

For our simulation exercises and real-data examples, we construct $S_1, \ldots, S_K$ by forming sets based on k-means clustering on the variant-specific ratio estimates $\hat{\beta}_{Y_j}/\hat{\beta}_{X_j}$, $j \notin V$. The motivation for this is partly dimension reduction: genetic variants which belong to the same cluster suggest a similar value for the causal effect $\theta_0$, and would therefore impact the asymptotic bias in a similar way. However, the idea of forming candidate instrument sets by clustering variants offering similar ratio estimates has also been used to shed light on biological mechanisms that may drive heterogeneous causal effects; see, for example, Jong et al. (2020) and Foley et al. (2021).

### 2.4 Example: Vitamin D and Coronary heart disease

One of the examples in Section 6 estimates the effect of vitamin D levels on coronary heart disease (CHD). GWASs have identified strong genetic associations with vitamin D in biologically plausible genes; $GC$, $DHCR7$, $CYP2R1$, and $CYP24A1$. Each of these genes is known influence vitamin D level through different mechanisms. In order to investigate the effect of vitamin D supplementation on a range of traits and diseases, MR studies have taken genetic variants near those genes to instrument vitamin D as they are considered more likely to satisfy the exclusion restriction than other genome-wide significant variants (Mokry et al., 2015; Revez et al., 2020).

The left panel of Figure 2 shows genetic associations with vitamin D levels and CHD of 7 uncorrelated genetic variants taken from $GC$, $DHCR7$, $CYP2R1$, and $CYP24A1$; these 7 variants form the set of valid instruments $V$. The right panel shows an additional 102 uncorrelated variants associated with vitamin D levels and CHD, which are partitioned into 4 groups by k-means clustering on the ratio of effects $(\hat{\beta}_{Y_j}/\hat{\beta}_{X_j})$, $j \notin V$. By taking all possible combinations of those 4 groups, we can form $2^4 - 1 = 15$ additional instrument sets $S_1, \ldots, S_{15}$ of the additional variants by taking all possible combinations of these 4 groups.

An MR analysis only using the 7 valid instruments $V$ suggests an insignificant effect of vitamin D on CHD, $\hat{\theta}_V = -0.01$ log odds per nmol/L. Our aim is to discover whether or not one of the additional instrument sets $S_1, \ldots, S_{15}$ should be used along with $V$ to provide a better estimate in terms of AMSE, and to suggest an honest confidence interval which takes into account the model uncertainty from selecting the AMSE-minimising set of instruments.
Figure 2. Genetic associations with vitamin D (nmol/L) and CHD (log odds ratio). On the right panel, associations of additional variants are colour and shape-coded by their membership into groups formed by \(k\)-means clustering on the ratio of effects \(\hat{\beta}_Y \big/ \hat{\beta}_X\), \(j \notin V\).

3 Consistent selection under systematic pleiotropy

A key part of the decision problem of choosing from additional instrument sets will be the \textit{Valid estimator} which uses only the known valid instruments \(V\); this estimator will form a reference point from which to assess the bias-variance trade-off from including additional instruments.

In the linear model, the valid set of instruments imply the \(|V|\) restrictions \(\beta_Y = \theta_0\beta_X\), \(j \in V\). The Valid estimator of \(\theta_0\) is the limited information maximum likelihood estimator (Anderson and Rubin, 1949) given by

\[
\hat{\theta}_V = \arg \min_{\theta} \sum_{j \in V} \frac{(\hat{\beta}_Y - \theta \hat{\beta}_X)_j^2}{\sigma^2_{\hat{\beta}_Y} + \theta^2 \sigma^2_{\hat{\beta}_X}},
\]

which in the context of MR has been motivated by minimising a heterogeneity statistic used to detect the presence of pleiotropy (Bowden et al., 2019), and as a profile score estimator based on normality assumptions (ZWHBS).

Before we tackle the case of idiosyncratic pleiotropy, we study a model selection problem under the widely-deployed systematic pleiotropy assumption. Here, we assume all additional instrument sets \(S_1, \ldots, S_K\) satisfy systematic pleiotropy, meaning that there is no asymptotic bias from using any additional instrument set. The goal of minimising AMSE then reduces
to one of efficiency; we aim to choose the set of instruments that corresponds to the lowest asymptotic variance.

**Assumption 5 (systematic pleiotropy).** For any instrument \( j \) in set \( S_k \), \( \beta_{Yj} = \theta_0 \beta_{Xj} + \alpha_j \), where \( \alpha_j \sim N(0, \kappa^2_s) \), and \( \kappa^2_s = \Theta(1/n) \), \( k \in [K] \).

Systematic pleiotropy assumes each additional instrument set \( S_k \), \( k \in [K] \) has a common group-specific effect up to an unknown overdispersion parameter \( \kappa^2_s \), \( k \in [K] \). The rate restriction on \( \kappa^2_s \) ensures the uncertainty from the random pleiotropic effects contributes to the asymptotic variance. The assumption that pleiotropic effects are normally distributed with zero mean is commonly maintained in MR studies; see, for example, Yuan et al. (2020).

### 3.1 Estimation under systematic pleiotropy

By Assumption 5, the overdispersion parameter \( \kappa^2_s \) is the variance of the random effects \( \alpha_j = \beta_{Yj} - \theta_0 \beta_{Xj} \) across \( j \in S \). Given our genetic associations and the Valid estimator \( \hat{\theta}_V \), we estimate the overdispersion parameters by calculating the variance of \( \hat{\alpha}_j = \beta_{Yj} - \hat{\theta}_V \hat{\beta}_{Xj} \), \( j \in S \), and adjusting for bias-correction terms. For any additional instrument set \( S \), a consistent estimator for \( \kappa^2_s \) is given by

\[
\hat{\kappa}^2_s = \frac{1}{|S|} - \frac{1}{|S|} \sum_{j \in S} (\hat{\alpha}_j - \bar{\alpha}_S)^2 - \frac{1}{|S|} \sum_{j \in S} \hat{\Omega}_j - \frac{1}{|S|} \sum_{j \in S} \left( \hat{\beta}_{Xj}^2 - \frac{1}{(|S| - 1)} \sum_{k \neq j} \hat{\beta}_{Xj} \beta_{Xk} \right),
\]

where \( \bar{\alpha}_S = \sum_{j \in S} (\hat{\beta}_{Yj} - \hat{\theta}_V \hat{\beta}_{Xj}) / |S| \), \( \hat{\Omega}_j = \sigma^2_{Yj} + \hat{\theta}_V^2 \sigma^2_{Xj} \), and \( \hat{\eta}_V = \sum_{j \in V} \hat{\Omega}_j^{-1} (\hat{\beta}_{Xj}^2 - \sigma^2_{Xj}) \).

Using estimates \( \hat{\kappa}^2_s \), we can construct causal effect estimators which appropriately adjust for the extra variance resulting from including potentially pleiotropic variants. In particular, an estimator of \( \theta_0 \) which uses the additional instrument set \( S \) is given by

\[
\hat{\theta}_S = \arg \min_{\theta} \sum_{j \in V \cup S} \frac{(\hat{\beta}_{Yj} - \theta \hat{\beta}_{Xj})^2}{\sigma^2_{Yj} + \theta^2 \sigma^2_{Xj} + \mathbb{I}(j \in S) \hat{\kappa}^2_s}.
\]

**Theorem 1 (Estimation under systematic pleiotropy).** Under Assumptions 1-5, if \( |V| = \Theta(1) \), then \( n(\hat{\kappa}^2_s - \kappa^2_s) \xrightarrow{p} 0 \), and

\[
\frac{\eta_V + \Lambda_S}{\sqrt{\eta_V + \Lambda_S + \delta_S}} (\hat{\theta}_S - \theta_0) \xrightarrow{D} N(0, 1)
\]

as \( n, p \to \infty \), where \( \eta_V = \sum_{j \in V} \Omega_j^{-1} \beta_{Xj}^2 \), \( \Lambda_S = \sum_{j \in S} \Gamma_{js}^{-1} \beta_{Xj}^2 \), \( \delta_S = \sum_{j \in S} \Gamma_{js}^{-2} \sigma^2_{Xj} (\sigma^2_{Yj} + \kappa^2_s) \), \( \Gamma_{js} = \Omega_j + \kappa^2_s \), and \( \Omega_j = \sigma^2_{Yj} + \theta_0 \sigma^2_{Xj} \).
Theorem 1 considers the case where $V$ contains a finite number of strong instruments, although the results can be easily adjusted to handle $V$ containing many weak instruments. Under our assumptions, $\eta_V = \Theta(n)$, $\Lambda_S = \Theta(n)$, and $\delta_S = \Theta(p)$. Hence, the convergence rate of $\hat{\theta}_S$ is $\sqrt{n}$ if $p = O(n)$. Intuitively, $\eta_V$ and $\Lambda_S$ are weighted measures of the strength of instruments in $V$ and $S$, respectively, and $\delta_S$ represents the extra uncertainty due to many weak and systematically pleiotropic instruments.

The variance components can be consistently estimated by

$$\hat{\eta}_V = \sum_{j \in V} \hat{\Omega}_j^{-1}(\hat{\beta}_X^2_j - \sigma^2_{X_j}), \quad \hat{\Lambda}_S = \sum_{j \in S} \hat{\Gamma}_j^{-1}(\hat{\beta}_X^2_j - \sigma^2_{X_j}), \quad \hat{\delta}_S = \sum_{j \in S} \hat{\Gamma}_j^{-1}(\sigma^2_{X_j} + \hat{\kappa}_S^2), \quad \hat{\Omega}_j = \sigma^2_{Y_j} + \hat{\theta}_V^2 \sigma^2_{X_j}, \quad \text{and} \quad \hat{\Gamma}_j = \hat{\Omega}_j + \hat{\kappa}_S^2.$$

Thus, we can compute consistent estimators of the asymptotic variances of causal effect estimates $\hat{\theta}_V, \hat{\theta}_S, \ldots, \hat{\theta}_{S_K}$. The post-selection estimator $\hat{\theta}_{PS}$ simply corresponds to the causal effect estimate with the estimated variance-minimising set of instruments.

The convergence result $n(\hat{\kappa}_S^2 - \kappa_S^2) \overset{P}{\to} 0$ means that the estimation error of $\hat{\kappa}_S^2$ is negligible in asymptotic variance calculations for the post-selection estimator. Furthermore, $\hat{\theta}_{PS}$ has the following oracle property: its asymptotic distribution is the same as the asymptotic distribution of the estimator which uses the true asymptotic variance-minimising instruments.

For our systematic pleiotropy model, a practical concern of this oracle property is that because the uncertainty from estimating the overdispersion parameters is considered negligible, the Valid estimator (which has an asymptotic variance $1/\eta_V$) is never selected on the basis that the there is too much pleiotropic variation in additional instruments. On efficiency grounds, adding any additional instrument set will always be favoured over using only the known valid instruments if the additional instruments are strong enough; Theorem 1 suggests we should include additional instruments $S$ if $\eta_V < \Lambda_S^2/(\delta_S - \Lambda_S)$.

### 3.2 The winner’s curse

Since the asymptotic results in Theorem 1 do not account for the uncertainty from model selection, the well-known problem of the winner’s curse remains an obstacle to making good inferences. In genetics, we often think of the winner’s curse as describing biases in estimation when weakly associated variants are filtered out (Göring et al., 2002). In our case, the ‘winner’ is defined in terms of minimising the asymptotic variance. Thus, although estimation remains asymptotically unbiased, on average, standard errors will be under-estimated, leading to inflated type I error rates.

As a brief illustration of this issue, consider the following simulation exercise. Suppose there are a total of $p = 36$ instruments, where $|V| = 4$ of them are known to be valid. We partition the remaining 32 instruments into two sets of $|S_1| = |S_2| = 16$ additional instruments. For the valid instruments $j \in V$ we set $\beta_{X_j} = 0.5$, and for all additional instruments, we set...
\[ \beta_{X_j} = \frac{3}{\sqrt{p}}, \ j \in S_1 \cup S_2. \]  
We also set \( \sigma_{X_j} = \frac{1}{\sqrt{n}} \) and \( \sigma_{Y_j} = \frac{2}{\sqrt{n}} \) for all \( j \in V \cup S_1 \cup S_2 \), where \( n = 100 \). Hence, all instruments are equally strong.

To focus on the performance under the null hypothesis, we set the true causal effect, \( \theta_0 = 0 \). We set the overdispersion parameter for \( S_1 \) as \( \kappa_{S_1}^2 = \frac{8}{n} \), and vary the parameter for \( S_2 \) in the range \( \kappa_{S_2}^2 \in [0, \frac{25}{n}] \). We simulate summary data according to the normal distribution in Assumptions 1 and 5.

The density plots in Figure 2 show the asymptotic approximations suggested by Theorem 1 are quite accurate even with \( n = 100 \) observations. However, for the post-selection estimators, the tails are heavier than the theorem predicts for two reasons: (i) the estimation error from estimating the overdispersion parameters \( \kappa_{S_1}^2 \) and \( \kappa_{S_2}^2 \) is not taken into account; (ii) the uncertainty in model selection is not taken into account. The first reason alone suggests perfect control of type 1 errors (when testing the hypothesis \( H_0 : \theta_0 = 0 \) against \( H_1 : \theta_0 \neq 0 \)) will be difficult in finite samples, but the problem is compounded by the second reason.

This is illustrated by the right panel in Figure 2. Since all instruments are equally strong, there is no reason to select the Valid estimator under our model assumptions. We then have a choice between whether the additional instrument set \( S_1 \) is included, or whether \( S_2 \) is included. When \( \kappa_{S_2}^2 \) is much smaller than \( \kappa_{S_1}^2 \), it is much easier to identify \( V \cup S_2 \) as the variance-minimising set of instruments, and so the type 1 errors are less inflated. Similarly, there is little uncertainty in model selection when \( \kappa_{S_2}^2 \) is much larger than \( \kappa_{S_1}^2 \). However, when \( \kappa_{S_1}^2 \) and \( \kappa_{S_2}^2 \) are similar in magnitude, it becomes less obvious which model should be selected, and this uncertainty manifests in higher type 1 error rates.
4 Conservative selection under idiosyncratic pleiotropy

For many applications, there is no good reason to assume pleiotropy balances out across variants, and we may not be comfortable imposing any group structure on unknown pleiotropic effects. We therefore consider estimation under a more general form of pleiotropy which allows variant-specific heterogeneity.

For each variant $j$, we denote $\tau_j$ as the unknown pleiotropic effect on the outcome which is not mediated by the exposure. These effects will be local-to-zero, following DiTraglia (2016)'s (henceforth DT’s) framework of locally invalid instruments. The specific rate restriction we will impose allows consistent estimation of the causal effect $\theta_0$ using any candidate instrument set, but making valid inferences requires us to account for the impact of unknown pleiotropy on the asymptotic bias.

**Assumption 6** (idiosyncratic pleiotropy). $\beta Y_j = \theta_0 \beta X_j + \tau_j$, where $\tau_j = 0$ for $j \in V$, $\|\tau\|^2_2 = O(1/n)$, and $\tau = (\tau_1, \ldots, \tau_p)'$.

By Assumptions 1 and 3, we can set $\tau_j = 0$ for $j \in V$. For the remaining effects $\tau_j$, $j \notin V$, the overall bias due to pleiotropy is decreasing in the sample size. This local-to-zero characterisation is not a substantive biological assumption, but rather a technical device which allows us to formulate a meaningful bias-variance trade-off.

Our rate restriction is slightly different to DT. Instead of requiring that pleiotropic effects are of order $1/\sqrt{n}$ for each genetic variant, we only limit the collective pleiotropy problem. Thus we allow some variants to have larger pleiotropic effects on the outcome than others, even in large samples.

4.1 Estimation under idiosyncratic pleiotropy

Under idiosyncratic pleiotropy, the causal effect estimator that uses the additional instrument set $S$ is given by

$$
\hat{\theta}_S = \arg \min_{\theta} \sum_{j \in V \cup S} \frac{(\hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j})^2}{\sigma_{Y_j}^2 + \theta^2 \sigma_{X_j}^2}.
$$

**Theorem 2** (Estimation under idiosyncratic pleiotropy). Under Assumptions 1-4 and 6, if $|V| = O(p)$,

$$
\frac{\eta_V + \eta_S}{\sqrt{\eta_V + \eta_S + \zeta_V + \zeta_S}} (\hat{\theta}_S - \theta_0 - \frac{b_S}{\eta_V + \eta_S}) \overset{D}{\underset{n \to \infty, p \to \infty}{\approx}} N(0, 1)
$$

as $n, p \to \infty$, where $b_S = \sum_{j \in S} \Omega_j^{-1} \beta_{X_j} \tau_j$, $\Omega_j = \sigma_{Y_j}^2 + \theta_0^2 \sigma_{X_j}^2$, and for any set of instruments $\tilde{S}$, $\eta_S = \sum_{j \in S} \Omega_j^{-1} \beta_{X_j}^2$ and $\zeta_S = \sum_{j \in S} \Omega_j^{-2} \sigma_{X_j}^2 \sigma_{Y_j}^2$. 


The variance components $\eta_V$ and $\eta_S$ are $\Theta(p)$, and $\zeta_V$ and $\zeta_S$ are $\Theta(p)$, so that, as in the systematic pleiotropy model, the convergence rate for $\hat{\theta}_S$ is $\sqrt{n}$ if $p = O(n)$. The key difference is that whereas systematic pleiotropy models the pleiotropic effects as random (around zero) according to a group-specific distribution, the idiosyncratic pleiotropy model considers them as unknown, fixed, and variant-specific. Therefore, the impact of idiosyncratic pleiotropy is to induce a bias in causal effect estimates.

In practice, the number of genetic variants taken to be valid instruments may be quite small. However, by allowing the set of valid instruments $V$ to increase proportionately with the total number of variants $p$, we are able to incorporate the extra uncertainty from weak instruments into the calculation of confidence intervals. In particular, we note that the asymptotic variance of $\hat{\theta}_S$ under fixed $p$ would be $(\eta_V + \eta_S)^{-1}$. Therefore, the additional variance term $(\eta_V + \eta_S)^{-2}(\zeta_V + \zeta_S)$, which also appears in ZWHBS’s analysis under systematic pleiotropy, can be regarded as a summary data version of Newey and Windmeijer (2009)’s variance correction term in the many weak instruments framework.

### 4.2 Estimating the asymptotic mean square error

To carry out the focused instrument selection approach, we first need to estimate the AMSE of each candidate instrument set. Constructing consistent estimators of the asymptotic variances is straightforward. If we were also able to construct consistent estimators of the asymptotic biases, we would have consistent model selection when using AMSE criteria. However, by design, the idiosyncratic pleiotropy model which allows for heterogenous pleiotropic effects prevents us from being as precise when estimating the asymptotic bias $(\eta_V + \eta_S)^{-1}b_S$ for any additional instrument set $S$. We are only able to pin down the asymptotic bias up to a distribution.

An asymptotically unbiased estimator for $b_S = \sum_{j \in S} \Omega_j^{-1} \beta_{X_j} \tau_j$ is $\hat{b}_S = \sum_{j \in S} \hat{\Omega}_j^{-1} \hat{\beta}_{X_j} \hat{\tau}_j + \hat{\theta}_V \sum_{j \in S} \hat{\Omega}_j^{-1} \sigma_{X_j}^2$, where $\hat{\tau}_j = \hat{\beta}_{Y_j} - \hat{\theta}_V \hat{\beta}_{X_j}$.

**Theorem 3** (Bias estimation under idiosyncratic pleiotropy). Under Assumptions 1-4 and 6, if $|V| = O(p)$, then

$$\Upsilon_S^{\frac{1}{2}}(\hat{b}_S - b_S) \overset{D}{\to} N(0, 1)$$

as $n, p \to \infty$, where $\Upsilon_S = \eta_S + \zeta_S + \xi_S + \eta_V^{-2} \eta_S^2 (\eta_V + \zeta_V)$, $\xi_S = 2\theta_0^2 \sum_{j \in S} \Omega_j^{-2} \sigma_{X_j}^4$, and $\eta_V$, $\eta_S$, $\zeta_V$, $\zeta_S$ are defined in Theorem 2.

The asymptotic variance $\Upsilon_S$ can be written as the sum of two terms: $\eta_S (1 + \eta_V^{-1} \eta_S)$ and $\zeta_S + \xi_S + \eta_V^{-2} \eta_S^2 \zeta_V$. The first variance term $\eta_S (1 + \eta_V^{-1} \eta_S)$ is of order $\Theta(n)$, and corresponds to the variance term obtained in DT’s framework in the case where $V$ and $S$ contain a finite
number of strong instruments. The second term \( \zeta_S + \xi_S + \eta_V^2 \eta_S^2 \zeta_V \) is of order \( \Theta(p) \) and represents the additional uncertainty due to many weak instruments.

From Theorem 2, the asymptotic variance of the causal effect estimate which uses the additional instrument set \( S \) satisfies \( (n_V + \eta_S)^{-1} + (n_V + \eta_S)^{-2} (\zeta_V + \zeta_S) \), where \( (n_V + \eta_S)^{-1} = \Theta(1/n) \), and \( (n_V + \eta_S)^{-2} (\zeta_V + \zeta_S) = \Theta(p/n^2) \). Thus, both components of the variance term are non-negligible in large samples when \( p/n = \Theta(1) \).

As noted by Newey and Windmeijer (2009), despite the knife-edge condition required to balance these variance components, it may be advisable to use the weak instrument variance correction \( (n_V + \eta_S)^{-2} (\zeta_V + \zeta_S) \) in general scenarios. Even when \( n \) is considerably larger than \( p \), and under relatively strong instruments, confidence intervals that do not involve the weak instrument variance correction can have poor coverage in finite samples. For example, the simulation study of Davies et al. (2015, pp. 457-9), which mimicked a MR design, showed that standard errors which did not correct for many weak instrument effects led to inflated type I error rates even for the case where \( n = 3000 \) and \( p = 9 \).

The expression for the asymptotic bias \( (n_V + \eta_S)^{-1} b_S \) is of magnitude \( O(1/\sqrt{n}) \), and therefore we have a meaningful bias-variance trade-off, where the AMSE is \( O(1/n) + \Theta(p/n^2) \).

For any set of instruments \( \hat{S} \), let \( \hat{n}_S = \sum_{j \in \hat{S}} \hat{\Omega}_j^{-1}(\hat{\beta}_{X_j}^2 - \sigma_{X_j}^2) \), \( \hat{\zeta}_S = \sum_{j \in \hat{S}} \hat{\Omega}_j^{-2} \sigma_{X_j}^2 \), \( \hat{\xi}_S = 2 \hat{\theta}_V \sum_{j \in \hat{S}} \hat{\Omega}_j^{-2} \sigma_{X_j}^2 \), \( \Omega_j = \sigma_{Y_j}^2 + \hat{\theta}_V^2 \sigma_{X_j}^2 \), and \( \hat{\Upsilon}_S = \hat{n}_S + \hat{\zeta}_S + \hat{\xi}_S + \hat{n}_V^2 \hat{\eta}_S^2 (\hat{n}_V + \hat{\zeta}_V) \). Let \( M(S, b_S) \) denote the AMSE of the causal effect estimator that uses the additional instrument set \( S \). Then, an asymptotically unbiased estimate of the AMSE of the estimator \( \hat{\theta}_S \) is given by

\[
\hat{M}(S, \hat{b}_S) = \frac{\hat{b}_S^2 - \hat{\Upsilon}_S}{(n_V + \eta_S)^2} + \frac{\hat{n}_V + \hat{n}_S + \hat{\zeta}_S + \hat{\xi}_S}{(n_V + \eta_S)^2},
\]

where the first term on the right hand side is an asymptotically unbiased estimate of \( b_S^2/(n_V + \eta_S)^2 \), and the second term on the right hand side is a consistent estimator of the asymptotic variance (see Lemma S.8 in Supplementary Material). By Theorem 3 and Slutsky’s lemma, we have \( \hat{M}(S, \hat{b}_S) \sim M(S, b_S) \) as \( n, p \to \infty \), where

\[
M(S, b_S) = \frac{(b_S + \sqrt{\Upsilon}_S \mathcal{Z})^2 - \Upsilon_S}{(n_V + \eta_S)^2} + \frac{\eta_V + \eta_S + \zeta_V + \zeta_S}{(n_V + \eta_S)^2},
\]

and \( \mathcal{Z} \sim N(0, 1) \).

### 4.3 Post-selection estimator

After estimating the AMSE of all possible instrument sets \( V, V \cup S_1, \ldots, V \cup S_K \), the post-selection estimator of \( \theta_0 \) simply chooses the estimated AMSE-minimising set of instruments.
A consistent estimate of the asymptotic variance of the Valid estimator is \( (\hat{\eta}_V + \hat{\zeta}_V)/\hat{\eta}_V^2 \). Therefore, we only use additional instruments if
\[
(\hat{\eta}_V + \hat{\zeta}_V)/\hat{\eta}_V^2 > \min_{k' \in [K]} \hat{M}(S_{k'}, \hat{b}_{S_{k'}}).
\]
Consider the weights \( \hat{\omega}_k = \mathbb{1}\{\hat{M}(S_k, \hat{b}_{S_k}) = \min_{k' \in [K]} \hat{M}(S_{k'}, \hat{b}_{S_{k'}})\} \), \( k \in [K] \), and \( \hat{\omega}_V = \mathbb{1}\{(\hat{\eta}_V + \hat{\zeta}_V)/\hat{\eta}_V^2 \leq \min_{k' \in [K]} \hat{M}(S_{k'}, \hat{b}_{S_{k'}})\} \). Then, the post-selection estimator can be written as a model averaging estimator (cf. DT, pp. 194-5),
\[
\hat{\theta}_{PS} = \hat{\omega}_V \hat{\theta}_V + \sum_{k=1}^{K} \hat{\omega}_k \hat{\theta}_{S_k}.
\]

### 4.4 Simulation study

This section studies the finite sample performance of the focused instrument selection approach under idiosyncratic pleiotropy. We simulated summary data on a total of \( p = 50 \) genetic variants, according to Assumptions 1-3, and 6. The sample size to generate all summary associations was set at \( n = 100 \), and the standard errors were fixed as \( \sigma_{X_j} = 1/\sqrt{n} \) and \( \sigma_{Y_j} = 1/\sqrt{n} \) for all \( j \in [50] \). A subset of \( |V| = 10 \) variants were set to be valid instruments. The remaining 40 variants were randomly partitioned into two groups of 20 variants, \( G_1 \) and \( G_2 \), where the pleiotropy parameters were generated as \( \tau_j \sim U[0, \bar{\tau}_{G_1}]/\sqrt{np} \) for \( j \in G_1 \), for values of \( \bar{\tau}_{G_1} \in \{0, 4, 8\} \), and \( \tau_j \sim U[0, 4]/\sqrt{np} \) for \( j \in G_2 \).

![Figure 3. RMSE performance. Square root of the estimated AMSE (RMSE) varying with the average instrument strength of \( G_1 \) and \( G_2 \), and invalidness of \( G_1 \).](image-url)
The concentration parameter is a measure of the average strength of instruments. For any set of instruments \( G \), it is defined as 
\[
\lambda_G = \sum_{j \in G} \frac{\hat{\beta}_{X_j}^2}{(\sigma_{X_j}^2 | G|)}.
\]
The exposure associations \( \hat{\beta}_{X_j} \) were chosen to achieve different levels of instrument strength, \( \lambda_G \). The true causal effect was set at \( \theta_0 = 0.2 \).

The selection problem was to choose between 4 instrument sets: \( V, V \cup S_1, V \cup S_2, \) and \( V \cup S_1 \cup S_2 \), where the additional instrument sets \( S_1 \) and \( S_2 \) were partitions formed by k-means clustering on the ratio of associations \( \hat{\beta}_{Y_j}/\hat{\beta}_{X_j}, j \notin V \). The post-selection estimator \( \hat{\theta}_{PS, Focused} \), only used the estimated AMSE-minimising instruments. We also note the estimates from the Valid estimator \( \hat{\theta}_V \) which only used the valid instruments \( V \).

Figure 3 shows that the inclusion of invalid instruments can improve estimation in an MSE sense. The extent of this improvement varies with both the instrument strength of all additional instruments, and degree of invalidness in \( G_1 \). Interestingly, when the valid instruments are relatively weak (\( \lambda_V \leq 12 \)), there are improvements in MSE from including even weaker, invalid instruments. When the valid instruments are strong, the flatter Focused RMSE curves indicate the strength of additional instruments matter less; it becomes harder to improve on an already precise Valid estimator.

Our results match the intuition that the focused selection approach is likely to be particularly valuable for MR studies which use weak instruments. For example, this could describe MR investigations of complex traits in which any nominated group of valid instruments may explain a very low proportion of exposure variation, or MR analyses on subgroups with smaller sample sizes.

5 Post-selection inference under idiosyncratic pleiotropy

While the primary focus of our work is to improve estimation, in this section we discuss the problem of post-selection inference. We first derive the asymptotic distribution of the post-selection estimator, which is used to construct honest confidence intervals using an analogous two-step approach suggested by DT.

Since these two-step confidence intervals are quite conservative, we propose a novel way to tighten intervals through the use of support restrictions on pleiotropic effects \( \tau_j, j \notin V \). These further restrictions are used to derive a weakly tighter asymptotic confidence interval for the asymptotic bias compared with Theorem 3, which can then be used to improve inference on the causal effect \( \theta_0 \).
5.1 Asymptotic distribution of the post-selection estimator

In Section 3, model selection was based on comparing only the asymptotic variances, which were consistently estimatable quantities. Here, model selection is based on the AMSE which we can only estimate up to its distribution. This creates non-ignorable uncertainty in the first-order asymptotic distribution of post-selection estimators.

To characterise the asymptotic distribution of \( \hat{\theta}_{PS} \), we first derive the joint asymptotic distribution of all candidate causal effect estimators \( \hat{\theta}_V, \hat{\theta}_{S_1}, \ldots, \hat{\theta}_{S_K} \), and their estimated biases \((\eta_V + \eta_{S_1})^{-1}\hat{b}_{S_1}, \ldots, (\eta_V + \eta_{S_K})^{-1}\hat{b}_{S_K}\).

**Lemma 1** (Joint convergence in distribution). Under Assumptions 1-4 and 6, if \(|V| = O(p)\),

\[
\begin{pmatrix}
\hat{\theta}_V - \theta_0 \\
\hat{\theta}_{S_1} - \theta_0 - \frac{b_{S_1}}{\eta_V + \eta_{S_1}} \\
\vdots \\
\hat{\theta}_{S_K} - \theta_0 - \frac{b_{S_K}}{\eta_V + \eta_{S_K}}
\end{pmatrix}
\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right),
\]

as \( n, p \to \infty \), where, denoting \( \Sigma_{21}^{(k)} \) the \( k \)-th element of \( \Sigma_{21} \), \( \Sigma_{31}^{(k)} \) the \( k \)-th element of \( \Sigma_{31} \), \( \Sigma_{22}^{(k,l)} \) the \( (k,l) \)-th element of \( \Sigma_{22} \), \( \Sigma_{32}^{(k,l)} \) the \( (k,l) \)-th element of \( \Sigma_{32} \), and \( \Sigma_{33}^{(k,l)} \) the \( (k,l) \)-th element of \( \Sigma_{33} \),

\[
\begin{align*}
\Sigma_{41} &= \frac{\eta_V + \zeta_V}{\eta_V^2} \\
\Sigma_{21}^{(k)} &= \frac{\eta_V + \zeta_V}{\eta_V(\eta_V + \eta_{S_k})} \\
\Sigma_{31}^{(k)} &= -\frac{\eta_{S_k}(\eta_V + \zeta_V)}{\eta_V^2(\eta_V + \eta_{S_k})} \\
\Sigma_{22}^{(k,l)} &= \frac{\eta_V + \zeta_V + \eta_{S_k}\cap S_l + \zeta_{S_k}\cap S_l}{(\eta_V + \eta_{S_k})(\eta_V + \eta_{S_l})}, \\
\Sigma_{32}^{(k,l)} &= \frac{\eta_V(\eta_{S_k}\cap S_l + \zeta_{S_k}\cap S_l) - \eta_{S_k}(\eta_V + \zeta_V)}{\eta_V(\eta_V + \eta_{S_k})(\eta_V + \eta_{S_l})}, \\
\Sigma_{33}^{(k,l)} &= \frac{\eta_V^2(\eta_{S_k}\cap S_l + \zeta_{S_k}\cap S_l + \zeta_{S_k}\cap S_l) + \eta_{S_k}\eta_{S_l}(\eta_V + \zeta_V)}{\eta_V^2(\eta_V + \eta_{S_k})(\eta_V + \eta_{S_l})}, \quad (k, l = 1, \ldots, K).
\end{align*}
\]

The asymptotic distribution of the post-selection estimator can then be derived using Lemma 1, consistent estimators of the variance components \( \hat{\eta}_V, \hat{\zeta}_V, \hat{\eta}_{S_k}, \hat{\zeta}_{S_k}, \hat{\xi}_{S_k} \), \( k \in [K] \) as defined in Section 4.2, and the continuous mapping theorem.
Theorem 4 (Asymptotic distribution of the post-selection estimator). Let $U = (U_1, \ldots, U_{2K+1})'$ be the normally distributed vector $U \sim N(0, \Sigma)$. Then, under Assumptions 1-4 and 6, if $|V| = O(p)$,

$$
\hat{\theta}_{PS} - \theta_0 \sim \omega_V U_1 + \sum_{k=1}^{K} \omega_k^* \left[ U_{k+1} + \frac{b_{S_k}}{\eta_V + \eta_{S_k}} \right],
$$

as $n, p \to \infty$, where \( \omega_V^* = 1\{ \Sigma_{11} \leq \min_{k' \in [K]} \left( [U_{K+k'+1} + (\eta_V + \eta_{S_k'})^{-2} b_{S_k'}] - \Sigma_{33}^{(k',k')} + \Sigma_{22}^{(k',k')} \right) \} \)

and \( \omega_k^* = 1\{ [U_{K+k+1} + (\eta_V + \eta_{S_k})^{-2} b_{S_k}]^2 - \Sigma_{33}^{(k,k)} + \Sigma_{22}^{(k,k)} = \min_{k' \in [K]} \left( [U_{K+k'+1} + (\eta_V + \eta_{S_k'})^{-2} b_{S_k'}] - \Sigma_{33}^{(k',k')} + \Sigma_{22}^{(k',k')} \right) \}, k \in [K]. \)

Since the AMSE is random even as $n, p \to \infty$, so are the population weights $\omega_V^*, \omega_1^*, \ldots, \omega_K^*$ which carry the model selection uncertainty. If we could estimate these population weights and unknown biases $b_{S_k}, k \in [K]$ consistently then we would be able to make exact inferences on $\theta_0$ which precisely account for model selection uncertainty in large samples. However, it is simply not possible to consistently estimate the distributions of post-selection estimators uniformly over the space of pleiotropic effects $\tau_j, j \in [p]$; see, for example, Leeb and Pötscher (2005, Section 2.3, pp. 38-40).

In practice, this leaves us with a choice between relatively narrow confidence intervals for $\theta_0$ which cannot guarantee correct coverage over all possible values of $\tau_j$, and conservative confidence intervals which may generally over-cover.

### 5.2 Confidence intervals

Here we apply the two-step approach proposed by DT to construct confidence intervals in our summary data setting with many weak instruments. From Theorem 2, the only non-consistently estimatable quantities in the asymptotic distributions of estimators $\hat{\theta}_{S_k}$ are the bias components $b_{S_k}, k \in [K]$. In the first step, the biases are left fixed, and confidence intervals are constructed using Theorem 4 by plugging in consistent estimators of the variance components. Therefore, in the first step, only the additional uncertainty due to model selection, but not bias estimation, is accounted for. For any given level of biases, these are the One-step intervals in DT. The natural choice is to plug-in the asymptotically unbiased estimates $\hat{b}_{S_k}, k \in [K]$, introduced in Section 4.2.

The second step aims to further incorporate the uncertainty due to bias estimation. Using Theorem 3, we can estimate a confidence region for the asymptotic bias of all candidate estimators. Then, the One-step intervals are calculated for each bias level in a corresponding grid of likely values. The widest calculated interval is retained as the Two-step interval (cf. DT, p. 195).
To construct One-step intervals, we start by taking \( R \) draws from a multivariate normal distribution

\[
\begin{pmatrix}
\tilde{U}_1 \\
\vdots \\
\tilde{U}_{2K+1}
\end{pmatrix} \sim N
\begin{pmatrix}
\begin{bmatrix} 0 & \hat{\Sigma}_{11} & \hat{\Sigma}_{21} & \hat{\Sigma}_{31} \\ 0 & \hat{\Sigma}_{22} & \hat{\Sigma}_{32} & \hat{\Sigma}_{33} \end{bmatrix} \\
\end{pmatrix},
\]

where estimators \( \hat{\Sigma}_{11}, \hat{\Sigma}_{22}, \hat{\Sigma}_{32}, \hat{\Sigma}_{33} \) are constructed by plugging-in the consistent estimates \( \hat{\eta}_V, \hat{\eta}_S, \hat{\xi}_V, \hat{\xi}_S, \hat{\bar{\eta}}_S, k \in [K] \) defined in Section 4.2. Consistency of these covariance estimators can be shown by identical arguments used in Lemma S.8 in Supplementary Material.

Let the \( r \)-th draw of \( (\tilde{U}_1, \ldots, \tilde{U}_{2K+1})' \) be denoted by \( (\tilde{U}^{(r)}_1, \ldots, \tilde{U}^{(r)}_{2K+1})' \), for \( r = 1, \ldots, R \). Also, for a fixed level of bias \( h_k = (\eta_V + \eta_S)^{-1}b_{S_k} \), \( k \in [K] \), let

\[
\tilde{A}^{(r)}(S_k, h_k, \tilde{U}^{(r)}_{K+k+1}) = (\tilde{U}^{(r)}_{K+k+1} + h_k)^2 - \frac{\hat{\bar{\eta}}_S}{(\hat{\eta}_V + \hat{\eta}_S)^2} + \frac{\hat{\eta}_V + \hat{\eta}_S + \hat{\xi}_V + \hat{\xi}_S}{(\hat{\eta}_V + \hat{\eta}_S)^2},
\]

and for \( h = (h_1, \ldots, h_K)' \), let

\[
\Lambda^{(r)}(h) = \sum_{k=1}^{K} \mathbb{I}\{\tilde{A}^{(r)}(S_k, h_k, \tilde{U}^{(r)}_{K+k+1}) = \min_{k' \in [K]} \tilde{A}^{(r)}(S_k, h_k, \tilde{U}^{(r)}_{K+k'+1})\} \left( \tilde{U}^{(r)}_{K+k+1} + \frac{b_{S_k}}{\hat{\eta}_V + \hat{\eta}_S} \right).
\]

Calculate \( \Lambda^{(r)}(h) \) for \( r \in [R] \), and, for any \( \alpha \in (0, 1) \), determine quantiles \( \hat{a}_1(h) \) and \( \hat{a}_2(h) \) such that \( \mathbb{P}(\hat{a}_1(h) \leq \Lambda(h) \leq \hat{a}_2(h)) = 1 - \alpha \). The One-step interval sets the bias \( h \) equal to the unbiased estimator \( \hat{h} = (\hat{b}_{S_1}/(\hat{\eta}_V + \hat{\eta}_S), \ldots, \hat{b}_{S_K}/(\hat{\eta}_V + \hat{\eta}_S))' \), so that the One-step interval is given by

\[
CI_{1,\alpha} = [\hat{\theta} - \hat{a}_1(\hat{h}), \hat{\theta} + \hat{a}_2(\hat{h})].
\]

The One-step intervals may have less than \((1 - \alpha)\)-level asymptotic coverage because they ignore errors from bias estimation.

### 5.2.2 Two-step intervals

We can estimate an asymptotic \((1 - \gamma)\)-level confidence region for \( b_{S_k}/(\eta_V + \eta_S) \), \( k \in [k] \) by using the \( R \) draws of \( (\tilde{U}_{K+1}, \ldots, \tilde{U}_{2K+1})' \). Denote this region by \( CI_{2,\gamma} \). For each \( c^* \in CI_{2,\gamma} \), we can calculate \( \hat{a}_1(c^*), \hat{a}_2(c^*) \), and set \( c_{\text{min}} = \arg\min_{c^* \in CI_{2,\gamma}} \hat{a}_1(c^*) \) and \( c_{\text{max}} = \arg\min_{c^* \in CI_{2,\gamma}} \hat{a}_2(c^*) \).
arg max_{c^* \in CI_2, \gamma} a_2(c^*). Then, the Two-step interval is given by

$$CI_{2,\alpha,\gamma} = [\hat{\theta} - \hat{a}_1(c_{min}), \hat{\theta} + \hat{a}_2(c_{max})].$$

By Theorem 4.4 of DT, $CI_{2,\alpha,\gamma}$ has asymptotic coverage of at least $1 - \alpha - \gamma$.

### 5.3 Finite-sample performance

This section aims to study how the performance of One-step and Two-step confidence intervals depend on the extent of pleiotropy and the strength of valid instruments. The same design introduced in Section 4.4 was used, but with the two following changes: first, we fixed $\bar{\lambda}_{G_1} = \bar{\lambda}_{G_2} = 8$, and varied the strength of the set of valid instruments $\lambda_V$. Second, the pleiotropy parameters were generated as $\tau_j = \bar{\tau}_{G_1} / \sqrt{n_p}$ for $j \in G_1$, for values $\bar{\tau}_{G_1} \in \{4, 8\}$, and $\tau_j = 4 / \sqrt{n_p}$ for $j \in G_2$. Intuitively, we may expect additional instruments to be more useful when valid instruments are weak. However, a weak set of valid instruments would limit our ability to precisely estimate the bias associated with including any additional set of instruments.

Table 1 shows that the One-step intervals are particularly useful when valid instruments are quite weak. They are able to achieve higher coverage probabilities than confidence intervals of the Valid estimator, and with shorter intervals. However, they offer no promise of correct coverage since the uncertainty from bias estimation is not taken into account. There is slight under-coverage of One-step intervals for the case where the valid instruments are strong ($\lambda_V = 16$) and there is a more severe pleiotropy problem ($\bar{\tau}_{G_1} = 8$). On the other hand, the

| $\bar{\tau}_{G_1} = 4$ | $\lambda_V = 4$ | $\lambda_V = 8$ | $\lambda_V = 16$ |
|------------------------|----------------|----------------|----------------|
|                        | Cov. | Wid. | Bias | RMSE | Cov. | Wid. | Bias | RMSE | Cov. | Wid. | Bias | RMSE |
| Valid                  | 96   | 100  | 0.01 | 0.39 | 96   | 100  | 0.01 | 0.24 | 96   | 100  | 0.00 | 0.16 |
| One-step              | 96   | 98   | 0.06 | 0.32 | 96   | 103  | 0.05 | 0.23 | 95   | 107  | 0.03 | 0.17 |
| Two-step             | 99   | 179  | 0.09 | 0.34 | 99   | 188  | 0.06 | 0.25 | 94   | 107  | 0.03 | 0.19 |

| $\bar{\tau}_{G_1} = 8$ | $\lambda_V = 4$ | $\lambda_V = 8$ | $\lambda_V = 16$ |
|------------------------|----------------|----------------|----------------|
|                        | Cov. | Wid. | Bias | RMSE | Cov. | Wid. | Bias | RMSE | Cov. | Wid. | Bias | RMSE |
| Valid                  | 97   | 100  | 0.00 | 0.38 | 96   | 100  | 0.00 | 0.24 | 95   | 100  | 0.00 | 0.17 |
| One-step              | 97   | 98   | 0.09 | 0.34 | 95   | 103  | 0.06 | 0.25 | 94   | 107  | 0.03 | 0.19 |
| Two-step             | 99   | 176  | 0.09 | 0.34 | 99   | 183  | 0.06 | 0.25 | 99   | 186  | 0.06 | 0.19 |

Table 1. Performance of 95% CIs. The Cov columns state percentage of 5000 experiments the true value $\theta_0 = 0.2$ was contained in the estimated CI. The Width columns are calculated as the average of (width of CI / width of the Valid CI)×100. The bias columns report the average absolute bias of the estimates.
Two-step intervals are generally very conservative; confidence intervals of the Valid estimator also achieve correct coverage but with a smaller width.

Overall, while the simulation exercise in Section 4.4 highlights that the focused selection approach is able to make careful use of additional pleiotropic variants to improve estimation, improving post-selection inferences is a very different challenge. By taking an analogous approach to that proposed in DT, our results here illustrate that we are able to obtain One-step and Two-step CIs that offer contrasting performances in terms of coverage and width.

5.3.1 Viability of One-step intervals

Given the competitive finite-sample performance of One-step intervals, we may be tempted to use them in practice even though they do not provide asymptotic guarantees of correct coverage. The natural question that arises is: how bad could the coverage be?

For the limited range of pleiotropic effects reported in Table 1, the coverage of the One-step intervals falls as the pleiotropy problem gets larger. However, if we were to keep increasing $\tau_j$ for all $j \not\in V$, eventually, the focused selection approach would simply choose to retain the Valid estimator, since including any additional instruments would lead to too much bias. In this case, from Theorem 4, $\omega^*_V \to 1$ and $\omega^*_k \to 0$, $k \in [K]$, so that the asymptotic distribution of the post-selection estimator should behave closely to the asymptotic distribution of the Valid estimator. Thus, the One-step interval should achieve near-correct coverage when additional instruments are very invalid; $\tau_j > \bar{\tau}$, $j \in [p]$, for some constant $\bar{\tau} > 0$.

Figure 3. Performance of 95% One-step intervals. The simulation design is the same as in Section 4.4, but with $\tau_j \sim U[\tau_0 - 1, \tau_0 + 1]$ for $j \not\in V$, $\lambda_V = \lambda_{G_1} = \lambda_{G_2} = 8$. The Width is calculated as the average of \((\text{width of One-step CI} / \text{width of the Valid CI}) \times 100\) over 5000 experiments.
Figure 3 supports this claim when searching over a restricted space of pleiotropy effects where the \( \tau_j \) behave similarly for all \( j \notin V \); as the pleiotropic effects increase in magnitude, the focused selection approach tends toward choosing the Valid estimator, and the coverage of One-step intervals improve.

The maximum coverage loss for this example is less than 7.5% for 95% asymptotic confidence intervals. Moreover, when the pleiotropic effects are small enough \( (\tau_0 \leq 5) \), the One-step intervals are shorter than intervals of the Valid estimator.

In practice, checking the worst-case coverage should involve a search through all possible combinations of \( \tau_j, j \notin V \), such that \( |\tau_j| \leq \bar{\tau} \). This could be computationally intensive, and the coverage performance would vary with the strength of instruments, along with the complexity of the model space.

### 5.4 Improving inference using a support restriction

It may be possible to obtain tight confidence intervals while also maintaining correct coverage by bounding the support of pleiotropic effects \( \tau_j, j \notin V \); see, for example, Conley et al. (2012). These bounds may be informed by results from existing observational studies. For example, if one were willing to assume that the pleiotropic effects were in the same direction as the bias of the OLS estimator from regressing the outcome on the exposure, then the results from Nevo and Rosen (2012) imply bounds on \( \tau_j, j \notin V \), which could then be used to bound the asymptotic biases \( (\eta_V + \eta_{S_k})^{-1}b_{S_k}, k \in [K] \). Alternatively, if it is known that genetics explains a low proportion of variation in a binary outcome, then there are restrictions implied on how large pleiotropic effects can be.

Recall that the Two-step interval for \( \theta_0 \) discussed in Section 5.2.2 searches over a confidence region for the asymptotic biases. If bounds on the pleiotropic effects allow us to tighten these confidence regions, then the width of the Two-step intervals could also shorten significantly.

This is demonstrated in this section by imposing the condition that \( |\tau_j| \leq \bar{\tau}, j \notin V \), for some known \( \bar{\tau} \geq 0 \) such that \( \bar{\tau} = O(1/\sqrt{np}) \). With this additional restriction, we derive bounds on the asymptotic bias expressions, with a view to highlighting the improved performance of Two-step intervals under a support restriction. For each additional instrument set \( S \), since the asymptotic bias \( (\eta_V + \eta_s)^{-1}b_S \) cannot be consistently estimated, we seek a bound on \( (\eta_V + \eta_s)^{-1}|b_S| \), denoted \( \hat{B}_S \), which we can construct using summary data only.

**Theorem 5** (Bounding the asymptotic bias under a support restriction). Under Assumptions 1-4, 6, if \( |\tau_j| \leq \bar{\tau}, j \notin V \), for some known \( \bar{\tau} \geq 0 \) such that \( \bar{\tau} = O(1/\sqrt{np}) \), and \( |V| = O(p) \), then up to order \( o_p(1/\sqrt{m}) \) terms, \( (\eta_V + \eta_S)^{-1}|b_S| \leq \hat{B}_S \), where \( \hat{B}_S \) =...
(\hat{\eta}_\nu + \hat{\eta}_s)^{-1} \tau \left( \sum_{j \in S} \hat{\Omega}_j^{-1}(|\hat{\beta}_{X_j}| + \sigma_{X_j} \sqrt{2/\pi}) \right), \pi \text{ is the mathematical constant } \pi = 3.1415\ldots, \hat{\Omega}_j = \sigma_{X_j}^2 + \hat{\theta}_V^2 \sigma_{X_j}^2, \text{ and } \hat{\eta}_\nu, \hat{\eta}_s \text{ are consistent estimators of } \eta_\nu, \eta_s \text{ introduced in Section 4.2.}

Using Theorem 5, we have the two-sided bound, up to asymptotically ignorable \( o_P(1/\sqrt{n}) \) terms, \(-\hat{B}_S \leq (\hat{\eta}_\nu + \hat{\eta}_s)^{-1} b_S \leq \hat{B}_S \) for every additional instrument set \( S \). This can be combined with the asymptotic confidence interval implied by Theorem 3, so that an asymptotic \((1 - \alpha)\)-level confidence interval for \((\hat{\eta}_\nu + \hat{\eta}_s)^{-1} b_S \) is given by \([\hat{B}_S^{(L)}, \hat{B}_S^{(U)}] \) where

\[
\hat{B}_S^{(L)} = \max \left\{ -\hat{B}_S, (\hat{\eta}_\nu + \hat{\eta}_s)^{-1}(\hat{b}_S - q_\alpha/2 \sqrt{\hat{\pi}_S}) \right\}
\]

\[
\hat{B}_S^{(U)} = \min \left\{ \hat{B}_S, (\hat{\eta}_\nu + \hat{\eta}_s)^{-1}(\hat{b}_S + q_\alpha/2 \sqrt{\hat{\pi}_S}) \right\},
\]

where \( q_\alpha \) is \((1 - \alpha)\)-th quantile of the standard normal distribution. By repeating the procedure for the Two-step interval described in Section 5.2.2, but now limiting the search of the asymptotic bias in the interval \([\hat{B}_S^{(L)}, \hat{B}_S^{(U)}] \), we can construct a Two-step confidence interval for \( \theta_0 \) under a Support Restriction, Two-step SR.

We now re-visit the simulation exercise in Section 5.3 to compare the performance of Two-step SR with the One-step interval, and Two-step interval without the support restriction. The simulation design is the same as described in Section 4.4, but now the pleiotropy effects are generated as \( \tau_j \sim U[0, \bar{\tau}_0/\sqrt{n_p}] \) for all \( j \notin V \), and we vary the instrument strength of groups \( \lambda_\nu, \lambda_{G_1}, \lambda_{G_2} \in \{4,8\} \). The model selection problem is the same as before; we choose from candidate sets \( V, V \cup S_1, V \cup S_2, V \cup S_1 \cup S_2 \), where \( S_1 \) and \( S_2 \) are partitions of the additional instruments formed by \( k \)-means clustering \((k = 2)\) on the ratio estimates \( \hat{\beta}_{Y_j}/\hat{\beta}_{X_j} \), \( j \notin V \).

| \( (\bar{\tau}_0, \lambda_\nu) \) | \( (2, 4) \) | \( (2, 8) \) | \( (2, 16) \) | \( (4, 4) \) | \( (4, 8) \) | \( (4, 16) \) |
|---|---|---|---|---|---|---|
| \( \lambda_{G_1} = \lambda_{G_2} = 8 \) | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. |
| One-step | 96 | 98 | 95 | 105 | 95 | 108 | 96 | 98 | 95 | 104 | 96 | 108 |
| Two-step | 99 | 182 | 99 | 194 | 99 | 195 | 99 | 182 | 99 | 192 | 100 | 193 |
| Two-step SR | 97 | 82 | 96 | 91 | 96 | 103 | 98 | 90 | 97 | 103 | 98 | 120 |
| \( (\bar{\tau}_0, \lambda_\nu) \) | \( (6, 4) \) | \( (6, 8) \) | \( (6, 16) \) | \( (8, 4) \) | \( (8, 8) \) | \( (8, 16) \) |
| \( \lambda_{G_1} = \lambda_{G_2} = 8 \) | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. | Cov. Wid. |
| One-step | 96 | 98 | 95 | 104 | 95 | 107 | 97 | 98 | 96 | 103 | 95 | 107 |
| Two-step | 99 | 181 | 99 | 190 | 99 | 191 | 99 | 180 | 99 | 188 | 99 | 189 |
| Two-step SR | 98 | 99 | 98 | 117 | 99 | 138 | 99 | 107 | 99 | 131 | 99 | 153 |

Table 2. Performance of 95% Two-step SR intervals. The Width columns are calculated as the average of \((width \text{ of CI} / width \text{ of the Valid CI}) \times 100 \).
Table 2 shows that if we have a bound on the maximum pleiotropic effect over $\tau_j, j \not\in V$, then the Two-step SR intervals can be a substantial improvement over the Two-step intervals which do not use that information. Interestingly, despite remaining conservative in coverage, when the valid instruments are weak ($\lambda_V = 4$) or if the pleiotropic effects are known to be small in magnitude ($\bar{\tau}_0 \leq 4$), Two-step SR intervals can be much tighter than the confidence intervals of the Valid estimator. We also note that, as in Section 5.3, the One-step intervals are quite competitive in terms of both coverage and width for this setting of weak valid instruments.

In summary, our results suggest that by including many slightly pleiotropic variants into MR analyses, as well as estimation, it is possible to improve inference (with asymptotic guarantees) if additional restrictions are placed on the magnitude of pleiotropic effects.

6 Real data examples

In this section we demonstrate how the focused instrument selection method can be applied to an MR study in practice. We consider several examples which comprise estimating: (i) the effect of vitamin D supplementation on Coronary Heart Disease (CHD); (ii) the effect of vitamin D supplementation on Multiple Sclerosis (MS); (iii) the effect of Systolic Blood Pressure (SBP) on CHD; (iv) the effect of SBP on Alzheimer’s Disease (AD).

Examples (i) and (ii) involve vitamin D as the exposure of interest. Previous MR studies instrumenting vitamin D have used variants from genes implicated in the modulation of 25OHD levels through known mechanisms. In particular, the GC, DHC7R, CYP2R1 and CYP24A1 genes have known functions in Vitamin D transport, synthesis, or metabolism. Following Berry et al. (2012) and Mokry et al. (2015), we take genetic variants from these gene regions as our valid instruments.

Evidence from observational studies suggests that low serum vitamin D levels are associated with an increased risk of cardiovascular disease (CVD); see, for example, Dobnig et al. (2008). However, these reported associations may be due to unmeasured confounding, as evidence from randomized clinical trials suggests no causal link (Barbarawi et al., 2019). Meanwhile, there is greater existing evidence of an association between decreased vitamin D level and risk of MS; see, for example, the MR study of Mokry et al. (2015).

Examples (iii) and (iv) have SBP as the exposure of interest. Genetic variants from genes encoding antihypertensive drug targets have been used to study the genetically-predicted effects of blood pressure lowering drugs (Gill et al., 2019; Georgakis et al., 2020). In particular, Gill et al. (2019) identify genes corresponding to drug targets for the following antihypertensive drugs: angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, $\beta$-
blockers, calcium channel blockers, and thiazide diuretic agents. We selected genetic variants from those gene regions as our set of valid instruments for SBP.

A recent network meta-analysis of 42 randomized clinical trials by Bundy et al. (2017) suggests a positive and linear effect of mean-achieved SBP on risk of CVD. Hypertension is also a risk factor for AD, with MR studies suggesting a causal association between genetically-predicted SBP and AD risk (Ostergaard et al., 2015). However, this does not imply that antihypertensive drugs could be repurposed for the prevention or treatment of AD; see Walker et al. (2020).

The twosampleMR R package (Hemani et al., 2018) was used to find uncorrelated genetic variants for genetic–exposure associations, that were then matched with data on genetic associations with the outcome from a non-overlapping data source. The package also contained all the GWAS summary data that we used, involving studies by Revez et al. (2020), Sudlow et al. (2015), Nikpay et al. (2015), Evangelou et al. (2018), Patsopoulos et al. (2019), and Lambert et al. (2013). More details on the summary data can be found in Supplementary Material.

The average strength of a set of instruments $\bar{S}$ was measured using an unbiased estimator of the concentration parameter, $\hat{\lambda}_{\bar{S}} = \sum_{j \in \bar{S}} \left( \frac{\hat{\beta}_X^2 j - \sigma^2 X j}{\sigma^2 X j | \bar{S}} \right)$.

To form additional instrument sets from the additional genetic variants, we again used $k$-means clustering on the ratio estimates $\hat{\beta}_Y j / \hat{\beta}_X j$, $j \not\in V$. For each example, the number of clusters $k^* \geq 2$ was determined by the majority rule from the NbClust R package; see Charrad et al. (2014). Therefore, we partitioned the additional genetic variants into $k^*$ groups by $k$-means clustering, and considered all possible combinations of those $k^*$ partitions, which resulted in $2^{k^*} - 1$ additional instrument sets, $S_1, \ldots, S_{(2^{k^*} - 1)}$.

| $X \rightarrow Y$ | $|V|$ | $|S^*|$ | $\hat{\lambda}_V$ | $\hat{\lambda}_{S^*}$ | $\hat{\theta}_V$ | $\hat{\theta}_{PS}$ | Valid | One-step | Two-step |
|------------------|------|--------|--------------|----------------|----------------|----------|------|---------|---------|
| VitD $\rightarrow$ CHD | 7 | 47 | 2078 | 117 | -0.011 | -0.027 | [-0.09, 0.07] | [-0.12, 0.06] | [-0.18, 0.12] |
| VitD $\rightarrow$ MS | 10 | 88 | 1410 | 65 | -0.267 | -0.241 | [-0.38, -0.09] | [-0.49, -0.04] |
| SBP $\rightarrow$ CHD | 11 | 413 | 108 | 73 | 0.045 | 0.042 | [0.01, 0.09] | [0.03, 0.06] | [0.03, 0.06] |
| SBP $\rightarrow$ AD | 11 | 352 | 108 | 78 | -0.033 | -0.024 | [-0.09, 0.05] | [-0.09, 0.05] | [-0.09, 0.05] |

Table 3. Results from the data examples. For each example, the AMSE-minimising set of instruments included an additional set of genetic variants $S^*$. The units for the estimated parameter $\theta_0$ are log odds per 25OHD nmol/L for vitamin D effects on CHD and MS, and log odds per mmHg for SBP effects on CHD and AD. Asymptotic 95% confidence intervals are reported in the last three columns.

Table 3 shows that for all examples, at most 11 genetic variants were chosen as valid instruments. However, the AMSE-minimising combination of instruments $V \cup S^*$ involved incorporating information from many more genetic variants ($|S^*|$ is much larger than $|V|$).
Interestingly, for the examples with vitamin D as the exposure, the focused selection method led the use of many additional genetic variants even though the valid instruments were much stronger ($\hat{\lambda}_V$ is much larger than $\hat{\lambda}_{S^*}$).

The post-selection estimates $\hat{\theta}_{PS}$ were quite similar to the valid estimates $\hat{\theta}_V$. For the effect of SBP on CHD, the post-selection estimate $\hat{\theta}_{PS}$ was almost identical to the Valid estimate $\hat{\theta}_V$ despite including 413 of the 446 additional variants available. The Two-step interval for this example also rejects the null of no effect, indicating a robust association. Similarly, the focused selection method also suggests a robust finding of the effect of vitamin D on MS. Our null result of the effect of vitamin D on CHD is consistent with the findings of Barbarawi et al. (2019)’s meta-analysis of 21 randomized clinical trials.

![Figure 4. SBP and AD study. Genetic associations with SBP (mmHg) and AD (log odds ratio) (left). Two-step SR intervals as a function of the support restriction $\tau_j \in [-\bar{\tau}, \bar{\tau}]$ (right). The dashed lines on the right panel indicate the 90% asymptotic confidence interval of the Valid estimator, $[-0.058, -0.008]$. The additional instrument sets here considered all combinations of 4 partitions of 446 additional variants.](image)

One notable difference between the valid and focused approach was in their estimates of the effect SBP on AD. While the Valid estimator indicated a significant negative association between SBP and AD ($\hat{\theta} = -0.033$, p-value 0.03), the focused selection approach suggested a point estimate closer to 0, and could not reject a null finding with either the One-step or Two-step intervals at a 95% asymptotic confidence level. The right panel of Figure 4 illustrates sensitivity of Two-step SR intervals to the support restriction on pleiotropy ($\tau_j \leq \bar{\tau}$, $j \notin V$). For small enough values of pleiotropy ($\bar{\tau} \leq 0.004$), the Two-step SR intervals suggest a significant association at a 90% asymptotic confidence level.

We conclude with a note of caution regarding our choice of valid instruments for the examples with SBP as the exposure. The left panel of Figure 4 shows some heterogeneity in the ratio...
estimates $\hat{\beta}_{Y_j} / \hat{\beta}_{X_j}, j \in \mathcal{V}$ in our set of valid instruments $\mathcal{V}$. We grouped together variants from gene regions aiming to mimic the effects of different SBP-lowering medication. However, more thorough investigations by Gill et al. (2019) and Walker et al. (2020) suggest that genetic variants corresponding to different antihypertensive drugs do not necessarily suggest the same causal effect estimates. It is possible that simply taking genetic variants corresponding to one drug target, as opposed to several different ones, could be a better choice for $\mathcal{V}$.

7 Conclusion

Compared with usual instrumental variable analyses in the social sciences, MR studies must usually select from very many potential instruments; there are often hundreds of genetic variants that could be considered viable instruments. Popular existing methods tend to rely on strong assumptions which restrict the behaviour of pleiotropy to systematic effects, or assumptions that a large proportion of candidate instruments are valid.

However, if there is a reason for believing, for example, that 50% of a large number of genetic variants are valid instruments, then this usually stems from some knowledge of the biology a few genes which suggests that variants from those regions are more likely to satisfy the exclusion restriction. As such, our assumption that a relatively few variants are known to be valid, while the vast majority of variants may have heterogeneous pleiotropic effects, could be considered a more palatable assumption.

Our goal in this paper was to minimise the asymptotic risk of estimation. This is in stark contrast to the current MR methods which focus on achieving unbiased estimation. By building on DT’s focused selection approach under local misspecification in ZWHBS’s two-sample summary data framework, we were able to formulate and tackle an asymptotic bias-variance trade-off which results from the inclusion of many weak and locally invalid instruments. Through several real data examples, we have demonstrated how the optimal choice of instruments to minimise estimation risk may involve many more genetic variants than just those that are biologically-justified.

Our idiosyncratic pleiotropy model for additional genetic variants only allows conservative model selection. Therefore, although we can construct uniformly valid confidence intervals, the intervals will in general over-cover. To improve on DT’s two-step approach for honest confidence intervals, we have proposed a way to tighten post-selection inferences using support restrictions on pleiotropic effects. This paves a path for the focused selection method to potentially improve inference, as well as estimation, in MR analyses.
References

Anderson, T. W. and H. Rubin (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics* 20(1), 46–63.

Barbarawi, M., B. Kheiri, Y. Zayed, O. Barbarawi, H. Dhillon, B. Swaid, A. Yelangi, S. Sundus, G. Bachuwa, M. L. Alkotob, and J. E. Manson (2019). Vitamin D Supplementation and Cardiovascular Disease Risks in More Than 83 000 Individuals in 21 Randomized Clinical Trials: A Meta-analysis. *JAMA Cardiology* 4(8), 765–776.

Bentkus, V. (2005). A Lyapunov-type bound in $R^d$. *Theory of Probability and Its Applications* 49(2), 400–410.

Berry, D. J., K. S. Vmaleswaran, J. C. Whittaker, A. D. Hingorani, and E. Hyppönen (2012). Evaluation of genetic markers as instruments for Mendelian randomization studies on vitamin D. *PLoS ONE* 7(5), 1–10.

Bi, N., H. Kang, and J. Taylor (2019). Inference after selecting plausibly valid instruments with application to mendelian randomization. *arXiv:1911.03985*, 1–59.

Bowden, J., G. Davey Smith, P. C. Haycock, and S. Burgess (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology* 40(4), 304–314.

Bowden, J., F. Del Greco M, C. Minelli, Q. Zhao, D. A. Lawlor, N. A. Sheehan, J. Thompson, and G. Davey Smith (2019). Improving the accuracy of two-sample summary-data Mendelian randomization: Moving beyond the NOME assumption. *International Journal of Epidemiology* 48(3), 728–742.

Bundy, J. D., C. Li, P. Stuchlik, X. Bu, T. N. Kelly, K. T. Mills, H. He, J. Chen, P. K. Whelton, and J. He (2017). Systolic blood pressure reduction and risk of cardiovascular disease and mortality a systematic review and network meta-analysis. *JAMA Cardiology* 2(7), 775–781.

Burgess, S., C. N. Foley, E. Allara, J. R. Staley, and J. M. Howson (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* 11(1).

Burgess, S., R. A. Scott, N. J. Timpson, G. Davey Smith, and S. G. Thompson (2015). Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology* 30, 543–552.

Burgess, S. and S. G. Thompson (2011). Avoiding bias from weak instruments in mendelian randomization studies. *International Journal of Epidemiology* 40(3), 755–764.
Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs (2014). NbClust: An R Package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* 61(6), 1–36.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly Exogenous. *The Review of Economics and Statistics* 94(1), 260–272.

Davey Smith, G. and S. Ebrahim (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 32(1), 1–22.

Davies, N. M., S. von Hinke Kessler Scholder, H. Farbmacher, S. Burgess, F. Windmeijer, and G. D. Smith (2015). The many weak instruments problem and mendelian randomization. *Statistics in Medicine* 34(3), 454–468.

DiTraglia, F. J. (2016). Using invalid instruments on purpose: Focused moment selection and averaging for GMM. *Journal of Econometrics* 195(2), 187–208.

Dobnig, H., S. Pilz, H. Scharnagl, W. Renner, U. Seelhorst, B. Wellnitz, J. Kinkeldei, B. O. Boehm, G. Weihrauch, and W. Maerz (2008). Independent association of low serum 25-hydroxyvitamin d and 1,25-dihydroxyvitamin d levels with all-cause and cardiovascular mortality. *JAMA - The Archives of Internal Medicine* 168(12), 1340–1349.

Evangelou et al., E. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics* 50(10), 1412–1425.

Foley, C. N., A. M. Mason, P. D. Kirk, and S. Burgess (2021). MR-Clust: Clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* 37(4), 531–541.

Georgakis, M. K., D. Gill, A. J. Webb, E. Evangelou, P. Elliott, C. L. Sudlow, A. Dehghan, R. Malik, I. Tzoulaki, and M. Dichgans (2020). Genetically determined blood pressure, antihypertensive drug classes, and risk of stroke subtypes. *Neurology* 95(4), E353–E361.

Gill et al., D. (2019). Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. *Circulation* 140(4), 270–279.

Göring, H. H., J. D. Terwilliger, and J. Blangero (2002). Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. *The American Journal of Human Genetics* 69(6), 1357–1369.

Guo, Z., H. Kang, T. Tony Cai, and D. S. Small (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 80(4), 793–815.
Hemani, G., J. Bowden, and G. Davey Smith (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics* 27(R2), R195–R208.

Hemani et al., G. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 7, e34408.

Iong, D., Q. Zhao, and Y. Chen (2020). A Latent Mixture Model for Heterogeneous Causal Mechanisms in Mendelian Randomization. *arXiv:2007.06476v2*.

Kang, H., A. Zhang, T. T. Cai, and D. S. Small (2016). Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization. *Journal of the American Statistical Association* 111(513), 132–144.

Lambert et al., J. C. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nature Genetics* 45(12), 1452–1458.

Lawlor, D. A. (2016). Commentary: Two-sample Mendelian randomization: Opportunities and challenges. *International Journal of Epidemiology* 45(3), 908–915.

Lawlor, D. A., R. M. Harbord, J. A. Sterne, N. Timpson, and G. Davey Smith (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* 27, 1133–1163.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory* 21(1), 21–59.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics* 30(1), 67–80.

Liu, Z., T. Ye, B. Sun, M. Schooling, and E. J. Tchetgen Tchetgen (2020). On Mendelian Randomization Mixed-Scale Treatment Effect Robust Identification (MR MiSTERI) and Estimation for Causal Inference. *arXiv:2009.14484*.

Mokry, L. E., S. Ross, O. S. Ahmad, V. Forgetta, G. D. Smith, A. Leong, C. M. Greenwood, G. Thanassoulis, and J. B. Richards (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. *PLoS Medicine* 12(8), 1–20.

Morrison, J., N. Knoblauch, J. H. Marcus, M. Stephens, and X. He (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics* 52(7), 740–747.

Nevo, A. and A. M. Rosen (2012). Identification with imperfect instruments. *Review of Economics and Statistics* 94(3), 659–671.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica* 77(3), 687–719.
Nikpay et al., M. (2015). A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics 47*(10), 1121–30.

Ostergaard et al., S. D. (2015). Associations between Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study. *PLoS Medicine 12*(6), 1–16.

Patsopoulos et al., N. A. (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science 365*(6460).

Revez, J. A., T. Lin, Z. Qiao, A. Xue, Y. Holtz, Z. Zhu, J. Zeng, H. Wang, J. Sidorenko, K. E. Kemper, A. A. Vinkhuyzen, J. Frater, D. Eyles, T. H. Burne, B. Mitchell, N. G. Martin, G. Zhu, P. M. Visscher, J. Yang, N. R. Wray, and J. J. McGrath (2020). Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nature Communications 11*(1), 1–12.

Sanderson, E., W. Spiller, and J. Bowden (2020). Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomisation. *BioRxiv: 10.1101/2020.04.02.021980*.

Schmidt et al., A. F. (2020). Genetic drug target validation using Mendelian randomization. *Nature Communications 11*(3255).

Solovieff, N., C. Cotsapas, P. H. Lee, S. M. Purcell, and J. W. Smoller (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics 14*(7), 483–495.

Stelzer et al., G. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*.

Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, and J. Danesh et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine 12*(3).

Tchetgen Tchetgen, E. J., B. L. Sun, and S. Walter (2019). The GENIUS Approach to Robust Mendelian Randomization Inference. *arXiv:1709.07779*, 1–32.

Verbanck, M., C. Y. Chen, B. Neale, and R. Do (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics 50*(5), 693–698.

Walker, V. M., P. G. Kehoe, R. M. Martin, and N. M. Davies (2020). Repurposing antihypertensive drugs for the prevention of Alzheimer’s disease: A Mendelian randomization study. *International Journal of Epidemiology 49*(4), 1132–1140.

Windmeijer, F., H. Farbmacher, N. Davies, and G. Davey Smith (2019). On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. *Journal of the American Statistical Association 114*(527), 1339–1350.
Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika 92*(4), 937–950.

Yavorska, O., J. Staley, and S. Burgess (2020). MendelianRandomization (R package).

Ye, B. T., J. Shao, and H. Kang (2020). Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Annals of Statistics (forthcoming)*.

Yuan, Z., H. Zhu, P. Zeng, S. Yang, S. Sun, C. Yang, J. Liu, and X. Zhou (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nature Communications 11*(1), 1–14.

Zhao, Q., J. Wang, G. Hemani, J. Bowden, and D. S. Small (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics 48*(3), 1742 – 1769.
SUPPLEMENTARY MATERIAL: THEORETICAL RESULTS

Ashish Patel, Francis J. DiTraglia, Verena Zuber, and Stephen Burgess

Notation

Let CS denote ‘the Cauchy-Schwarz inequality’, CH denote ‘Chebyshev’s inequality’, and T denote ‘the triangle inequality’. For all $j \in [p]$, let $e_{Y_j} = \hat{\beta}_{Y_j} - \beta_{Y_j}$, $e_{X_j} = \hat{\beta}_{X_j} - \beta_{X_j}$, $\Omega_j = \sigma_{Y_j}^2 + \theta_0^2\sigma_{X_j}^2$, and $J_{1j}^* = \Omega_j^{-1}\beta_{X_j}(e_{Y_j} - \theta_0e_{X_j})$. For the systematic pleiotropy model, let $u_j = \varepsilon_{Y_j} - \theta_0e_{X_j} \sim N(0, \Gamma_{jS})$, where $\Gamma_{jS} = \Omega_j + \kappa_S^2$ and $\varepsilon_{Y_j} = \varepsilon_{Y_j} + \alpha_j \sim N(0, \sigma_{Y_j}^2 + \kappa_S^2)$ for $j \in S$ by Assumptions 1 and 5.

Systematic pleiotropy model

Lemma S.1. (Valid estimator under $|V| = \Theta(1)$). Under Assumptions 1-5 and $|V| = \Theta(1)$, $\hat{\theta}_V - \theta_0 = \eta_V^{-1}\sum_{j \in V} J_{1j}^* + o_P(1/\sqrt{n})$, where $J_{1j}^* = \Omega_j^{-1}\beta_{X_j}(e_{Y_j} - \theta_0e_{X_j})$.

Proof. It can be shown by Lemmas B.1-B.3 of ZWHBS, pp. 42-5 that $\eta_V = \sum_{j \in V} J_{1j}^* + o_P(1/\sqrt{n})$, where $J_{1j} = J_{1j}^* + J_{1j}^{**}$, and $J_{1j}^{**} = \Omega_j^{-2}(e_{Y_j} - \theta_0e_{X_j})(\sigma_{Y_j}^2e_{X_j} + \theta_0\sigma_{X_j}^2e_{Y_j})$. Note that $E[\sum_{j \in V} J_{1j}^{**}] = 0$ and $Var(\sum_{j \in V} J_{1j}^{**}) = \sum_{j \in V} \Omega_j^{-2}\sigma_{Y_j}^2\sigma_{X_j}^2 = \Theta(1)$, by Assumption 1. Therefore, by CH, $\sum_{j \in V} J_{1j}^{**} = O_P(1)$. Since $\eta_V = \sum_{j \in V} \Omega_j^{-1}\beta_{X_j}^2 = \Theta(n)$, it follows that $\eta_V^{-1}\sum_{j \in V} J_{1j}^{**} = o_P(1/\sqrt{n})$. 

Lemma S.2. (Overdispersion parameters). Under Assumptions 1-5, and $|V| = \Theta(1)$, for any additional instrument set $S$, $n(\kappa_S^2 - \kappa_S^2) \xrightarrow{P} 0$.

Proof. Let $m = |S|$, $\tilde{\tau}_S = \sum_{j \in S}(\hat{\beta}_{Y_j} - \hat{\theta}_V\hat{\beta}_{X_j})/m$, and

\[
\hat{\kappa}_S^2 = \frac{1}{m-1}\sum_{j \in S}(\hat{\beta}_{Y_j} - \hat{\theta}_V\hat{\beta}_{X_j} - \tilde{\tau}_S)^2 - \frac{1}{m}\sum_{j \in S}(\sigma_{Y_j}^2 + \hat{\theta}_V^2\sigma_{X_j}^2) - \frac{1}{\eta_V} \cdot \frac{1}{m}\sum_{j \in S} \hat{\beta}_{X_j}^2 \\
+ \frac{1}{\eta_V} \cdot \frac{1}{m(m-1)} \sum_{j \in S} \sum_{k \neq j} \hat{\beta}_{X_j}\hat{\beta}_{X_k} \\
:= R_1^{(\kappa)} - R_2^{(\kappa)} - \frac{1}{\eta_V} R_3^{(\kappa)} + \frac{1}{\eta_V} R_4^{(\kappa)}.
\]

First, we can write

\[
(m-1)R_1^{(\kappa)} = \sum_{j \in S}(\hat{\beta}_{Y_j} - \hat{\theta}_V\hat{\beta}_{X_j})^2 - m\tilde{\tau}_S^2.
\]

For any $j \in S$, $\hat{\beta}_{Y_j} - \hat{\theta}_V\hat{\beta}_{X_j} = u_j - (\hat{\theta}_V - \theta_0)(\beta_{X_j} + e_{X_j})$, where $u_j = \varepsilon_{Y_j} - \theta_0e_{X_j} \sim N(0, \Gamma_{jS})$, and $\varepsilon_{Y_j} = \varepsilon_{Y_j} + \alpha_j \sim N(0, \sigma_{Y_j}^2 + \kappa_S^2)$. 

[S-1]
We construct $\tilde{R}_1^{(\kappa)}$ by replacing $\hat{\theta}_V - \theta_0$ in the above expansion for $R_1^{(\kappa)}$ with $\sum_{j \in V} J_{1j}^* / \eta_V$ from Lemma S.1,

$$
\tilde{R}_1^{(\kappa)} = \frac{1}{m} \sum_{j \in S} u_j^2 - 2 \left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right) \frac{1}{m} \sum_{j \in S} (\beta X_j + e_{X_j}) u_j + \left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right)^2 \frac{1}{m} \sum_{j \in S} (\beta X_j + e_{X_j})^2
$$

and define an infeasible estimator $\tilde{\kappa}_S^2$ of $\kappa_S^2$ such that

$$
\tilde{\kappa}_S^2 = \tilde{R}_1^{(\kappa)} - R_2^{(\kappa)} - \frac{1}{\eta_V} R_3^{(\kappa)} + \frac{1}{\eta_V} R_4^{(\kappa)}.
$$

We prove the following:

(i) $n(\tilde{\kappa}_S^2 - \kappa_S^2) = o_P(1)$;

(ii) $n(\tilde{\kappa}_S^2 - \tilde{\kappa}_S^2) = o_P(1)$.

Given (i) and (ii), the statement of Lemma S.2 follows by T.

Part (i)

Using the above expansions, and noting that $\sum_{j \in S} \Gamma_{js} = m \kappa_S^2 + \sum_{j \in S} \Omega_j$,

$$
m(\kappa_S^2 - \kappa_S^2) = \sum_{j \in S} (u_j^2 - \Gamma_{js}) - 2 \left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right) \sum_{j \in S} \beta X_j u_j - 2 \left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right)^2 \sum_{j \in S} e_{X_j} u_j
$$

and

$$
\frac{1}{m-1} \sum_{j \in S} \sum_{k \neq j} u_j u_k - \frac{1}{m-1} \sum_{j \in S} u_j u_k
$$

and

$$
\left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right) \frac{1}{m-1} \sum_{j \in S} \sum_{k \neq j} ((\beta X_j + e_{X_j}) u_k + (\beta X_k + e_{X_k}) u_j)
$$

and

$$
\left[ \left( \frac{1}{\eta_V} \sum_{j \in V} J_{1j}^* \right)^2 - \frac{1}{\eta_V} \right] \frac{1}{m-1} \sum_{j \in S} \sum_{k \neq j} (\beta X_j + e_{X_j})(\beta X_k + e_{X_k}) - (\hat{\theta}_V - \theta_0^2) \sum_{j \in S} \sigma_{X_j}^2.
$$

[S-2]
First, since $u_j \sim N(0, \Gamma_{j|s}^2)$, we have $E[(u_j^2 - \Gamma_{j|s}^2)^2] = 2\Gamma_{j|s}^4$. Hence, by CH, $R_1^{(k)} = O_P(\sqrt{m/n^2}) = O_P(1/\sqrt{n})$. Similarly, $\sum_{j \in S} \beta_{X_j} u_j = O_P(\|\beta_X\|_2/\sqrt{n})$, so that, by $\eta_V^{-1} \sum_{j \in V} J^*_{ij} = O_P(1/\sqrt{n})$, we have $R_2^{(k)} = O_P(1/n)$. For $R_3^{(k)}$, note that $\sum_{j \in S} e_{X_j} u_j = \sum_{j \in S} e_{X_j} \varepsilon_{Y_j} - \theta_0 \sum_{j \in S} (e_{X_j}^2 - \sigma_{X_j}^2) - \theta_0 \sum_{j \in S} \sigma_{X_j}^2 = O_P(\sqrt{m/n^2}) + O_P(m/n)$. Hence, $R_3^{(k)} = O_P(1/\sqrt{n})$. By CH, $R_5^{(k)} = O_P(1/n)$ noting that $\text{Var}(\sum_{j \in S} \sum_{k \neq j} u_j u_k) = \sum_{j \in S} \sum_{k \neq j} \Gamma_j^2 \Gamma_k^2 = O(m^2/n^2)$. By similar arguments, $R_6^{(k)} = O_P(1/n)$. For $R_8^{(k)}$, note that $\hat{\theta}_V - \theta_0 = (\hat{\theta}_V - \theta_0)(\hat{\theta}_V - \theta_0) = O_P(1/\sqrt{n})$. Thus, $R_8^{(k)} = O_P(1/\sqrt{n})$.

Terms $R_4^{(k)}$ and $R_7^{(k)}$ involve the zero-mean random variable $\eta_V^{-1} \sum_{j \in V} J^*_{ij} - \eta_V^{-1}$. Since $\sum_{j \in V} J^*_{ij} \sim N(0, \eta_V)$ under finite $|V|$, $(\eta_V^{-1} \sum_{j \in V} J^*_{ij})^2 - \eta_V^{-1} = O_P(1/n)$ as $\eta_V = \Theta(n)$. Similarly, $\sum_{j \in S} (\beta_{X_j} + e_{X_j})^2 = O_P(\|\beta_X\|_2^2) = O_P(1)$ and $\sum_{j \in V} \sum_{k \neq j} \beta_{X_j} \beta_{X_k}^2 = \sum_{j \in V} \beta_{X_j}^2 - (\sum_{j \in V} \beta_{X_j}^2)^2 = O(\|\beta_X\|_4^4 + \|\beta_X\|_2^4) = O(1)$ by Assumption 2. Hence, by CS, it follows that $R_4^{(k)} = O_P(1/n)$ and $R_7^{(k)} = O_P(1/n)$.

Overall, we have shown that $m(\hat{\kappa}_S^2 - \kappa_S^2) = o_P(1)$. The statement from Part (i) follows by $m = \Theta(n)$.

Part (ii)

By definition,

$$
\hat{\kappa}_S^2 - \kappa_S^2 = (R_1^{(k)} - \hat{R}_1^{(k)}) + \left(\frac{1}{\eta_V} - \frac{1}{\eta_V}ight)R_3^{(k)} + \left(\frac{1}{\eta_V} - \frac{1}{\eta_V}ight)R_4^{(k)}.
$$

First,

$$
m(R_1^{(k)} - \hat{R}_1^{(k)}) = -2[\hat{\theta}_V - \theta_0 - \frac{1}{\eta_V} \sum_{j \in V} J^*_{ij}] \sum_{j \in S} (\beta_{X_j} + e_{X_j}) u_j
$$

$$
+ \left[ (\hat{\theta}_V - \theta_0)^2 - \left( \frac{1}{\eta_V} \sum_{j \in V} J^*_{ij} \right)^2 \right] \sum_{j \in S} (\beta_{X_j} + e_{X_j})^2
$$

$$
+ \left[ \hat{\theta}_V - \theta_0 - \frac{1}{\eta_V} \sum_{j \in V} J^*_{ij} \right] \frac{1}{m - 1} \sum_{j \in S} \sum_{k \neq j} [ (\beta_{X_j} + e_{X_j}) u_k + (\beta_{X_k} + e_{X_k}) u_j ]
$$

$$
- \left[ \hat{\theta}_V - \theta_0 \right] \frac{1}{\eta_V} \sum_{j \in V} J^*_{ij} \frac{1}{m - 1} \sum_{j \in S} \sum_{k \neq j} (\beta_{X_j} + e_{X_j})(\beta_{X_k} + e_{X_k}).
$$

Calculations used in Part (i) show that $\sum_{j \in S} \sum_{k \neq j} (\beta_{X_j} + e_{X_j}) u_j = O_P(1/\sqrt{n})$, $\sum_{j \in S} (\beta_{X_j} + e_{X_j})^2 = O_P(1)$, $\sum_{j \in S} \sum_{k \neq j} [ (\beta_{X_j} + e_{X_j}) u_k + (\beta_{X_k} + e_{X_k}) u_j ] / (m - 1) = O_P(1)$, $\sum_{j \in V} \sum_{k \neq j} (\beta_{X_j} + e_{X_j})(\beta_{X_k} + e_{X_k})/(m - 1) = O_P(1)$.

Also, by Lemma S.1, $\hat{\theta}_V - \theta_0 - \eta_V^{-1} \sum_{j \in V} J^*_{ij} = o_P(1/\sqrt{n})$, so that $(\hat{\theta}_V - \theta_0)^2 - (\eta_V^{-1} \sum_{j \in V} J^*_{ij})^2 = o_P(1/n)$ since $\eta_V^{-1} \sum_{j \in V} J^*_{ij} = O_P(1/\sqrt{n})$. Therefore, $n(R_1^{(k)} - \hat{R}_1^{(k)}) = o_P(1)$ by $m = \Theta(n)$. 

[S-3]
For terms \( \hat{\eta}_V^{-1} - \eta_V^{-1} \) and \( \hat{\eta}_V^{-1} - \eta_V^{-1} \), we first show \( \hat{\eta}_V^{-1} - \eta_V^{-1} = o_P(1/n) \). Note that

\[
\hat{\eta}_V - \eta_V = \sum_{j \in V} \hat{\Omega}_j^{-1}(\Omega_j - \hat{\Omega}_j)\hat{\Omega}_j^{-1}((\beta X_j + e X_j)^2 - \sigma^2_{X_j}) + \sum_{j \in V} \Omega_j^{-1}(\beta^2_{X_j} - \beta_{X_j}^2 - \sigma^2_{X_j})
\]

\[
= (\theta_0^2 - \hat{\theta}_V^2)\Theta(n \sum_{j \in V}((\beta X_j + e X_j)^2 - \sigma^2_{X_j})) + \Theta(n \sum_{j \in V} 2\beta_{X_j}e X_j + e^2_{X_j} - \sigma^2_{X_j})
\]

\[
= O_P(\sqrt{n}),
\]

where the second equality follows by \( \hat{\Omega}_j = \Theta(n), \Omega_j = \Theta(n), \Omega_j - \hat{\Omega}_j = (\theta_0^2 - \hat{\theta}_V^2)\sigma^2_{X_j}, \) and \( \beta^2_{X_j} = (\beta X_j + e X_j)^2 \). The third equality follows by \( \theta_0^2 - \hat{\theta}_V^2 = (\theta_0 - \hat{\theta}_V)(\theta_0 + \hat{\theta}_V) = O_P(1/\sqrt{n}), \sum_{j \in V}((\beta X_j + e X_j)^2 - \sigma^2_{X_j}) = O_P(1), \) and \( \sum_{j \in V}(2\beta_{X_j}e X_j + e^2_{X_j} - \sigma^2_{X_j}) = O_P(1/\sqrt{n}) \).

Therefore,

\[
\frac{\hat{\eta}_V - \eta_V}{\eta_V} = 1 + \frac{\hat{\eta}_V - \eta_V}{\eta_V} = 1 + o_P(1/\sqrt{n}),
\]

and by identical arguments, \( n \hat{\eta}_V^{-1} = 1 + O_P(1/\sqrt{n}) \), so that

\[
\frac{1}{\hat{\eta}_V} - \frac{1}{\eta_V} = \frac{\eta_V - \hat{\eta}_V}{\eta_V} \cdot \frac{1}{\eta_V} = O_P(1/\sqrt{n})(1 + O_P(1/\sqrt{n}))o_P(1/n) = o_P(1/n).
\]

Finally, \( R_3^{(\kappa)} = m^{-1}\sum_{j \in S}(\beta X_j + e X_j)^2 = O_P(1/n) \) by CH, and it has been shown that \( mR_4^{(\kappa)} = \sum_{j \in S}\sum_{k \neq j}(\beta X_j + e X_k)(\beta X_k + e X_k)/(m - 1) = O_P(1) \). Thus, \( m(\hat{\kappa}_S - \kappa_S^2) = o_P(1) \), and Part (ii) follows by \( m = \Theta(n) \).

\[\square\]

**Lemma S.3. (Consistency of \( \hat{\theta}_S \)).** Under Assumptions 1-5, and \( |V| = \Theta(1) \), for any additional instrument set \( S, \hat{\theta}_S - \theta_0 \xrightarrow{p} 0 \).

**Proof.** Let \( \hat{Q}_V(\theta) = -2^{-1}\sum_{j \in V}(\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j})^{-1}(\hat{\beta}_Y_j - \theta \hat{\beta}_X_j)^2 \) and \( \hat{Q}_S(\theta, \kappa^2) = -2^{-1}\sum_{j \in V}(\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2)^{-1}(\hat{\beta}_Y_j - \theta \hat{\beta}_X_j)^2 \) so that \( \hat{\theta}_S = \arg \max \hat{Q}_V(\theta) + \hat{Q}_S(\theta, \kappa^2) \).

First, we focus on \( \hat{Q}_S(\theta, \kappa^2) \). For any \( j \in S, \hat{\beta}_Y_j - \theta \hat{\beta}_X_j = (\varepsilon_{Y_j} - \theta e_{X_j}) - \beta X_j(\theta - \theta_0) \), so that \( (\hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j})^2 = (\varepsilon_{Y_j} - \theta e_{X_j})^2 + (\theta - \theta_0)^2\beta^2_{X_j} - 2(\theta - \theta_0)\beta_{X_j}(\varepsilon_{Y_j} - \theta e_{X_j}) \).

\[\begin{align*}
-2\hat{Q}_S(\theta, \kappa^2) = (\theta - \theta_0)^2 \sum_{j \in S} \frac{\beta_{X_j}^2}{\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2_S} + \sum_{j \in S} \frac{(\varepsilon_{Y_j} - \theta e_{X_j})^2 - (\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2_S)}{\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2_S} \\
- (\kappa^2_S - \kappa^2_S) \sum_{j \in S} \frac{1}{\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2_S} + |S| - 2(\theta - \theta_0) \sum_{j \in S} \frac{\beta_{X_j}(\varepsilon_{Y_j} - \theta e_{X_j})}{\sigma^2_{Y_j} + \theta^2\sigma^2_{X_j} + \kappa^2_S} \\
:= (\theta - \theta_0)^2 R_4^{(Q)} + R_2^{(Q)} - (\kappa^2_S - \kappa^2_S) R_3^{(Q)} + |S| - 2(\theta - \theta_0) R_4^{(Q)}.
\end{align*}\]
First, 

\[ R_1^{(Q)} = \sum_{j \in S} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} + (\kappa^2_S - \kappa^2_S) \sum_{j \in S} \frac{\beta^2_{X_j}}{(\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)(\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)} \]

\[ = \sum_{j \in S} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} + o_P(1/n)O_P(n^2 \| \beta_X \|_2^2) \]

\[ = \sum_{j \in S} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} + o_P(n \| \beta_X \|_2^2), \]

where the second and third equalities follow by Assumption 1 and Lemma S.2.

Similarly,

\[ R_2^{(Q)} = \sum_{j \in S} \frac{(\varepsilon_{Y_j} - \theta e_{X_j})^2 - (\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} \]

\[ + (\kappa^2_S - \kappa^2_S) \sum_{j \in S} \frac{(\varepsilon_{Y_j} - \theta e_{X_j})^2 - (\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)}{(\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)(\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)} \]

\[ = O_P(p/n) + o_P(p/n), \]

by Assumption 1, Lemma S.2, and CH, noting that \( \text{Var}[(\varepsilon_{Y_j} - \theta e_{X_j})^2] = 2(\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S)^2 = \Theta(1/n^2). \)

By similar arguments it can be shown that \( R_3^{(Q)} = O_P(n) \) and \( R_4^{(Q)} = O_P(\sqrt{n} \| \beta_X \|_2). \) Thus, by Lemma S.2, \( p = O(n), \) and \( \| \beta_X \|_2 = \Theta(1), \)

\[ -2 \tilde{Q}_S(\theta, \kappa^2_S) = (\theta - \theta_0)^2 \sum_{j \in S} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} + |S| + o_P(n \| \beta_X \|_2^2) + O_P(p/n) + O_P(\sqrt{n} \| \beta_X \|_2) \]

\[ = (\theta - \theta_0)^2 \sum_{j \in S} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j} + \kappa^2_S} + |S| + o_P(n). \]

By identical arguments for \( \tilde{Q}_V(\theta), \) we have

\[ -2 \tilde{Q}_V(\theta) = (\theta - \theta_0)^2 \sum_{j \in V} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j}} + \sum_{j \in V} \frac{(e_{Y_j} - \theta e_{X_j})^2 - (\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j})}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j}} \]

\[ + |V| - 2(\theta - \theta_0) \sum_{j \in V} \frac{\beta_{X_j}(e_{Y_j} - \theta e_{X_j})}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j}} \]

\[ := (\theta - \theta_0)^2 \sum_{j \in V} \frac{\beta^2_{X_j}}{\sigma^2_{Y_j} + \theta^2 \sigma^2_{X_j}} + o_P(n). \]
Therefore,

\[-2[\bar{Q}_S(\theta, \hat{\kappa}^2_S) + \bar{Q}_V(\theta)] = (\theta - \theta_0)^2 \sum_{j \in V \cup S} \frac{\beta_{X_j}^2}{\hat{\sigma}_{Y_j}^2 + \hat{\theta}^2 \sigma_{X_j}^2} + |S| + o_p(n),\]

where \(\hat{\sigma}_{Y_j}^2 = \sigma_{Y_j}^2 + I\{j \in S\} \kappa_j^2\). The rest of the proof then follows by identical arguments used in the ZWHBS's Proof of Theorem 3.1, p.41. □

**Proof of Theorem 1 (Asymptotic distribution of \(\hat{\theta}_S\)).**

For any additional instrument set \(S\), we have shown that \(n(\hat{\kappa}_S - \kappa_S^2) = o_p(1)\) in Lemma S.2. To derive the asymptotic distribution of \(\hat{\theta}_S\), note that the first-order condition for \(\hat{\theta}_S\) is \(\sum_{j \in V} \hat{\psi}^{(V)}_j(\hat{\theta}_S) + \sum_{j \in S} \hat{\psi}^{(S)}_j(\hat{\theta}_S, \kappa_S^2) = 0\) where \(\hat{\psi}^{(V)}_j(\theta) = \Omega_j(\theta)^{-2}(\hat{\beta}_Y_j - \theta \hat{\beta}_X_j)(\hat{\beta}_X_j \sigma_{Y_j}^2 + \theta \hat{\beta}_Y_j \sigma_{X_j}^2), \hat{\psi}^{(S)}_j(\theta, \kappa^2) = \Gamma_j(\theta, \kappa^2)^{-2}(\hat{\beta}_Y_j - \theta \hat{\beta}_X_j)(\hat{\beta}_X_j (\sigma_{Y_j}^2 + \kappa^2) + \theta \hat{\beta}_Y_j \sigma_{X_j}^2)\) and \(\Gamma_j(\theta, \kappa^2) = \sigma_{Y_j}^2 + \theta^2 \sigma_{X_j}^2 + \kappa^2\).

We first focus on \(\sum_{j \in S} \hat{\psi}^{(S)}_j(\theta, \kappa_S^2)\). Note that \(\hat{\beta}_Y_j - \theta \hat{\beta}_X_j = \varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}\), so that \(\sum_{j \in S} \hat{\psi}^{(S)}_j(\theta, \kappa_S^2) = \sum_{j \in S} (J_{1j}^{(S)} + J_{2j}^{(S)})\), where \(J_{1j}^{(S)} = \Gamma(\theta_0, \kappa_S^2)^{-1}\beta_X_j (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j})\) and \(J_{2j}^{(S)} = \Gamma(\theta_0, \kappa_S^2)^{-2}(\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\varepsilon_{X_j} (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2)\).

For the first term, we have

\[
\sum_{j \in S} J_{1j}^{(S)} = \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-1}\beta_X_j (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) + (\kappa_S^2 - \kappa_S^2) \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-1} (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) = \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-1}\beta_X_j (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) + o_p(\sqrt{n}) \beta_X_2,
\]

where the second equality follows by Lemma S.2, Assumptions 1, 2, and 5, and CH.

Similarly, we can write

\[
\sum_{j \in S} J_{2j}^{(S)} = \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-2}(\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\varepsilon_{X_j} (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2)
\]

\[
+2(\kappa_S^2 - \kappa_S^2) \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-1} \Omega_j (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\varepsilon_{X_j} (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2)
\]

\[
+2(\kappa_S^2 - \kappa_S^2)(\kappa_S^2 + \kappa_S^2) \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-1}(\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\sigma_{X_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2).
\]

Note that \(E[\sum_{j \in S} (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\varepsilon_{X_j} (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2)] = 0\) and \(Var(\sum_{j \in S} (\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2) = \sum_{j \in S} \Gamma_j^2 \sigma_{X_j}^4 (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2\) where \(\Gamma_j^2 = \Theta(1/n)\). Therefore, by similar arguments used on \(\sum_{j \in S} J_{1j}^{(S)}\) we have

\[
\sum_{j \in S} J_{2j}^{(S)} = \sum_{j \in S} \Gamma(\theta_0, \kappa_S^2)^{-2}(\varepsilon_{Y_j} - \theta_0 \varepsilon_{X_j}) (\varepsilon_{X_j} (\sigma_{Y_j}^2 + \kappa_S^2) + \theta_0 \varepsilon_{Y_j} \sigma_{X_j}^2) + o_p(\sqrt{p}).
\]
By repeated use of CS,

\[
\mathbb{E}[|J_{1j}^{(S)} + J_{2j}^{(S)}|^3] \leq O(n^3|\beta_{X_j}|^3\mathbb{E}[|\varepsilon_{Y_j} - \theta_0 e_{X_j}|^3])
\]
\[
+ O\left(n^3|\beta_{X_j}|^2\mathbb{E}[|\varepsilon_{Y_j} - \theta_0 e_{X_j}|^6]^{\frac{1}{2}} \left(\mathbb{E}[|e_{X_j}|^2]^{\frac{1}{2}} + \mathbb{E}[|\varepsilon_{Y_j}|^2]^{\frac{1}{2}}\right)\right)
\]
\[
+ O\left(n^3|\beta_{X_j}|\mathbb{E}[|\varepsilon_{Y_j} - \theta_0 e_{X_j}|^6]^{\frac{1}{2}}\left(\mathbb{E}[|e_{X_j}|^4]^{\frac{1}{2}} + 2\mathbb{E}[|e_{X_j}|^2]^{\frac{1}{2}}\mathbb{E}[|\varepsilon_{Y_j}|^2]^{\frac{1}{2}} + \mathbb{E}[|\varepsilon_{Y_j}|^4]^{\frac{1}{2}}\right)\right)
\]
\[
+ O\left(n^3\mathbb{E}[|\varepsilon_{Y_j} - \theta_0 e_{X_j}|^6]^{\frac{1}{2}}\left(\mathbb{E}[|e_{X_j}|^6]^{\frac{1}{2}} + 3\mathbb{E}[|e_{X_j}|^4]^{\frac{1}{2}}\mathbb{E}[|\varepsilon_{Y_j}|^2]^{\frac{1}{2}} + 3\mathbb{E}[|e_{X_j}|^2]^{\frac{1}{2}}\mathbb{E}[|\varepsilon_{Y_j}|^4]^{\frac{1}{2}} + \mathbb{E}[|\varepsilon_{Y_j}|^6]^{\frac{1}{2}}\right)\right).
\]

Since \(\varepsilon_{Y_j} - \theta_0 e_{X_j}, e_{X_j}, \) and \(e_{Y_j}\) are zero-mean normally distributed random variables, \(|\varepsilon_{Y_j} - \theta_0 e_{X_j}|, |e_{X_j}|, \) and \(|e_{Y_j}|\) follow a half-normal distribution. Thus, for any \(\ell > 0, \mathbb{E}[|e_{X_j}|^{2\ell}] = Var(e_{X_j})^{\frac{\ell}{2}}(\ell! 2^\ell), \) so that \(\mathbb{E}[|e_{X_j}|^2] = \Theta(1/n), \mathbb{E}[|e_{X_j}|^4] = \Theta(1/n^2), \) and \(\mathbb{E}[|e_{X_j}|^6] = \Theta(1/n^3). \) Identical arguments apply for the moments of \(|\varepsilon_{Y_j}|\) and \(|\varepsilon_{Y_j} - \theta_0 e_{X_j}|. \) Thus,

\[
\mathbb{E}[|J_{1j}^{(S)} + J_{2j}^{(S)}|] \leq O(n^\frac{3}{2}|\beta_{X_j}|^3) + O(n|\beta_{X_j}|^2) + O(n^{\frac{3}{2}}|\beta_{X_j}|) + O(1),
\]

cf. ZWHBS, proof of Theorem 3.2.

We note that \(\sum_{j \in S} Var(J_{1j}^{(S)} + J_{2j}^{(S)}) = \sum_{j \in S} \frac{1}{\sigma_j^2} + \sum_{j \in S} \frac{1}{\beta_j^2} + \frac{\mathbb{E}[(\theta_j - \bar{\theta}_S)\sigma_j^2]}{\mathbb{E}[\hat{\beta}_j^2]} = \Theta(1/p)\). Therefore the Lyapunov condition holds: \(\sum_{j \in S} Var(J_{1j}^{(S)} + J_{2j}^{(S)}) = \frac{1}{\sigma_j^2} \mathbb{E}[|J_{1j}^{(S)} + J_{2j}^{(S)}|^3] \not \rightarrow 0, \) by Assumptions 1, 2, and 4, noting that \(\|\beta_X\|_3 = o(1)\). Then, by Lyapunov’s CLT,

\[
\sum_{j \in S} \tilde{\psi}_j^{(S)}(\theta_0, \hat{\kappa}_S^2) \overset{D}{\rightarrow} N\left(0, \sum_{j \in S} \frac{1}{\sigma_j^2} + \sum_{j \in S} \frac{1}{\beta_j^2} + \sum_{j \in S} \frac{1}{\beta_j^2} \mathbb{E}[|\hat{\beta}_j|^2] + \sum_{j \in S} \frac{1}{\beta_j^2} \mathbb{E}[|\hat{\beta}_j|^2] \right).
\]

The random components in \(\tilde{\psi}_j^{(V)}(\theta_0)\) are functions of errors \(e_{X_j} \) and \(e_{Y_j}, j \in V, \) and so are independent from \(\tilde{\psi}_j^{(S)}(\theta_0, \hat{\kappa}_S^2), j \in S, \) by Assumption 1. Thus,

\[
\sum_{j \in V} \tilde{\psi}_j^{(V)}(\theta_0) + \sum_{j \in S} \tilde{\psi}_j^{(S)}(\theta_0, \hat{\kappa}_S^2) \overset{D}{\rightarrow} N\left(0, \eta_V + \Lambda_S + \delta_S\right).
\]

By direct calculation, for any \(\bar{\theta}_S\) in the line joining \(\hat{\theta}_S\) and \(\theta_0,\)

\[
\sum_{j \in S} \nabla_{\theta_0} \tilde{\psi}_j^{(S)}(\bar{\theta}_S, \hat{\kappa}_S^2) = \sum_{j \in S} \gamma_j(\bar{\theta}_S, \hat{\kappa}_S^2)^{-2} \tilde{\beta}_j (\tilde{\beta}_j - \bar{\theta}_S^2 \tilde{\beta}_j) - \sum_{j \in S} \gamma_j(\bar{\theta}_S, \hat{\kappa}_S^2)^{-2} \tilde{\beta}_j (\tilde{\beta}_j - \bar{\theta}_S^2 \tilde{\beta}_j) - \sum_{j \in S} \gamma_j(\bar{\theta}_S, \hat{\kappa}_S^2)^{-1} \bar{\theta}_S \tilde{\beta}_j (\tilde{\beta}_j - \bar{\theta}_S^2 \tilde{\beta}_j) + \bar{\theta}_S \tilde{\beta}_j \sigma_j^2.
\]
By similar arguments to handle the terms of $\sum_{j \in S} \eta_j^{(S)}(\theta, \hat{\kappa}_S^2)$, it can be shown that:

$$
\sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \hat{\kappa}_S^2 \beta_{X_j} - \hat{\theta}_S \beta_{X_j} = \sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \hat{\kappa}_S^2 \beta_{X_j} + o_p(n\|\beta_X\|^2) + O_p(\sqrt{n}\|\beta_X\|_2) + O_p(\sqrt{p});
$$

$$
\sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \hat{\kappa}_S^2 \beta_{X_j} - \hat{\theta}_S \beta_{X_j} = \sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \hat{\kappa}_S^2 \beta_{X_j} + o_p(n\|\beta_X\|^2) + O_p(\sqrt{n}\|\beta_X\|_2) + o_p(p);
$$

$$
\hat{\theta}_S \sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \beta_{X_j} \hat{\kappa}_S^2 \beta_{X_j} = \sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2} \beta_{Y_j} \beta_{X_j} \hat{\kappa}_S^2 \beta_{X_j} + o_p(n\|\beta_X\|^2) + O_p(\sqrt{n}\|\beta_X\|_2) + o_p(p).
$$

Also, using the first-order condition, we showed above that $\sum_{j \in S} \Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-2}(\beta_{Y_j} - \hat{\theta}_S \beta_{X_j})(\hat{\beta}_{Y_j} \hat{\kappa}_S^2) + \hat{\theta}_S \beta_{Y_j} \sigma_{X_j}^2) = O_p(\sqrt{n}\|\beta_X\|_2) + O_p(\sqrt{p})$. Under Assumptions 1 and 5, Lemma S.2, and given consistency of $\hat{\theta}_S$ for $\theta_0$, we have $\Gamma_j(\hat{\theta}_S, \hat{\kappa}_S^2)^{-1} \hat{\theta}_S \sigma_{X_j}^2 = \Theta_p(1)$. Then, overall, we have

$$
\sum_{j \in S} \nabla_{\theta} \eta_j^{(S)}(\hat{\theta}_S, \hat{\kappa}_S^2) = -\sum_{j \in S} \Gamma_j^{-1} \beta_{X_j}^2 + o_p(n\|\beta_X\|^2) + o_p(p) + O_p(\sqrt{n}\|\beta_X\|_2) + O_p(\sqrt{p}).
$$

By identical arguments,

$$
\sum_{j \in V} \nabla_{\theta} \eta_j^{(V)}(\hat{\theta}_S) = -\sum_{j \in V} \Omega_j^{-1} \beta_{X_j}^2 + o_p(n) + O_p(\sqrt{n}).
$$

Therefore, under $\|\beta_X\|_2 = \Theta(1)$,

$$
\sum_{j \in V} \nabla_{\theta} \eta_j^{(V)}(\hat{\theta}_S) + \sum_{j \in S} \nabla_{\theta} \eta_j^{(S)}(\hat{\theta}_S, \hat{\kappa}_S^2) = -(\eta_V + \Lambda_S) + o_p(n) + o_p(p) + O_p(\sqrt{n}) + O_p(\sqrt{p}). \quad (S.2)
$$

By the mean value theorem expanding the first-order conditions $\sum_{j \in V} \eta_j^{(V)}(\hat{\theta}_S) + \sum_{j \in S} \eta_j^{(S)}(\hat{\theta}_S, \hat{\kappa}_S^2) = 0$ around $\hat{\theta}_S = \theta_0$, the result follows by (S.1) and (S.2). \qed

**Idiosyncratic pleiotropy model**

**Lemma S.4. (Consistency of $\hat{\theta}_S$).** Under Assumptions 1-4 and 6, for any additional instrument set $S$, $\hat{\theta}_S - \theta_0 \xrightarrow{P} 0$.

**Proof.** Let $\hat{g}_j(\theta) = \hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j}$. Some simple algebra shows that $\hat{g}_j(\theta) = (\epsilon_{Y_j} - \theta \epsilon_{X_j}) + \beta_{X_j}(\theta_0 - \theta) + \tau_j$. 

[S-8]
so that for $\dot{Q}(\theta) = -\sum_{j=1}^{p} \dot{g}_j(\theta^2)/2(\sigma_{Y_j}^2 + \theta^2\sigma_{X_j}^2)$,

$$-2\ddot{Q}(\theta) = (\theta_0 - \theta)^2 \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} + p + \sum_{j \in V \cup S} \Omega_j(\theta)^{-1}[(e_{Y_j} - \theta e_{X_j})^2 - (\sigma_{Y_j}^2 + \theta^2\sigma_{X_j}^2)]$$

$$+ \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} \tau_j^2 + 2(\theta_0 - \theta) \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} \beta_{X_j}(e_{Y_j} - \theta e_{X_j})$$

$$+ 2 \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} \tau_j(e_{Y_j} - \theta e_{X_j}) + 2(\theta_0 - \theta) \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} \beta_{X_j} \tau_j$$

$$:= (\theta_0 - \theta)^2 \sum_{j \in V \cup S} \Omega_j(\theta)^{-1} \beta_{X_j}^2 + p + R_{1p} + R_{2p} + R_{3p} + R_{4p} + R_{5p}.$$ 

Note that $E[R_{1p}] = 0$ and $Var(R_{1p}) = 2p$. Hence, $R_1 = O_P(\sqrt{p})$ by CH. Similarly, $R_{3p} = O_P(\sqrt{n}\|\beta_X\|_2 \cdot |\theta_0 - \theta|)$, and $R_{4p} = O_P(1)$. By CS, $R_{2p} = O(1)$ and $R_{5p} = O_P(\sqrt{n}\|\beta_X\|_2 \cdot |\theta_0 - \theta|)$. Thus,

$$-2\ddot{Q}(\theta) = (\theta_0 - \theta)^2 \sum_{j \in V \cup S} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2 + \theta^2\sigma_{X_j}^2} + p + O_P(\sqrt{n} + \sqrt{n}\|\beta_X\|_2 \cdot |\theta_0 - \theta|).$$

The rest of the proof is then identical to the Proof of Theorem 3.1 of ZWHBS, p.41. □

For Lemma S.5-S.7, let $\psi_j(\theta) = \Omega_j(\theta)^{-2}(\dot{\beta}_{Y_j} - \theta \dot{\beta}_{X_j})/3(\dot{\beta}_{X_j}\sigma_{Y_j}^2 + \theta \dot{\beta}_{Y_j}\sigma_{X_j}^2).$

**Lemma S.5.** Under Assumptions 1-4 and 6, $\sum_{j \in V \cup S} \nabla_{\theta} \psi_j(\theta_0)/(\eta V + \eta S) \xrightarrow{P} -1$.

**Proof.** The first-order condition is given by $\sum_{j \in V \cup S} \psi_j(\theta) = 0$. Also,

$$\nabla_{\theta} \psi_j(\theta) = -\dot{\beta}_{X_j}(\dot{\beta}_{Y_j} - \beta_{Y_j} - \beta_{X_j}) \Omega_j(\theta)^{-2} + \dot{\beta}_{X_j}\sigma_{Y_j}^2(\dot{\beta}_{Y_j} - \theta \dot{\beta}_{X_j}) \Omega_j(\theta)^{-2}$$

$$-4\theta \sigma_{X_j}^2(\dot{\beta}_{Y_j} - \theta \dot{\beta}_{X_j})/3(\dot{\beta}_{X_j}\sigma_{Y_j}^2 + \theta \dot{\beta}_{Y_j}\sigma_{X_j}^2) \Omega_j(\theta)^{-3}.$$

Let $e_{Y_j} = \dot{\beta}_{Y_j} - \beta_{Y_j} - \tau_j$ and $e_{X_j} = \dot{\beta}_{X_j} - \beta_{X_j}$. Also, let $\Omega_j = \sigma_{Y_j}^2 + \theta^2\sigma_{X_j}^2$. Then,

$$E[\nabla_{\theta} \psi_j(\theta_0)] = -\Omega_j^{-1} \beta_{X_j}^2 + \Omega_j^{-2} \sigma_{X_j}^2 \tau_j^2 - 4\theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} \tau_j - 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j^2.$$

Let $\phi_j(\theta) = \nabla_{\theta} \psi_j(\theta) - E[\nabla_{\theta} \psi_j(\theta)]$. Then,

$$\phi_j(\theta_0) = -\Omega_j^{-1} \beta_{X_j} e_{X_j} - \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - \Omega_j^{-2} \sigma_{X_j}^2 (e_{X_j}^2 - \sigma_{X_j}^2)$$

$$- \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \tau_j e_{X_j} - \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 e_{X_j} e_{X_j} + \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j}$$

$$+ \Omega_j^{-2} \sigma_{X_j}^2 \tau_j e_{X_j} - \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j}$$

$$+ \Omega_j^{-2} \sigma_{X_j}^2 \tau_j e_{X_j} - 4\theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} - 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j e_{X_j}$$

$$- 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 (e_{X_j}^2 - \sigma_{X_j}^2) + 4\theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \beta_{X_j} e_{X_j} + 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^2 \beta_{X_j} e_{X_j}$$

$$+ 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j e_{X_j} + 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j e_{X_j} - 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j e_{X_j} - 4\theta_0 \Omega_j^{-3} \sigma_{X_j}^4 \tau_j e_{X_j}$$

$$\equiv \sum_{l=1}^{23} \phi^{(l)}_j(\theta_0).$$
First, \( \text{Var}(\phi_j^{(1)}(\theta_0)) = \Omega_j^{-2} \beta_{X_j}^2 \sigma_{X_j}^2 \), so that \( \sum_{j \in V \cup S} \text{Var}(\phi_j^{(1)}(\theta_0)) \leq \Theta(n \| \beta_X \|_2^2) \). Similarly, for each \( \ell \in [23] \), we have that \( \sum_{j \in V \cup S} \text{Var}(\phi_j^{(1)}(\theta_0)) = O(n \| \beta_X \|_2^2) + O(p) \). Therefore, by CH, we have

\[
\mathbb{P}
\left[
\frac{1}{\eta_V + \eta_S} \sum_{j \in V \cup S} \phi_j(\theta_0) \right] > \kappa 
\leq \frac{1}{\kappa^2 (\eta_V + \eta_S)^2} \left( \sum_{j \in V \cup S} \text{Var}(\phi_j(\theta_0)) + 2 \sum_{j \in V \cup S} \sum_{k \neq j} \text{Cov}(\phi_j(\theta_0), \phi_k(\theta_0)) \right)
\leq O(1/n \| \beta_X \|_2^2) + O(p/n^2 \| \beta_X \|_2^2)
\leq o(1),
\]

where the second inequality uses the fact that \( \text{Cov}(\phi_j(\theta_0), \phi_k(\theta_0)) \leq \text{Var}(\phi_j(\theta_0))^{1/2} \text{Var}(\phi_k(\theta_0))^{1/2} \) and \( \eta_V + \eta_S = \Theta(n \| \beta_X \|_2^2) \), and equality follows by Assumption 2. Therefore,

\[
\frac{1}{\eta_V + \eta_S} \sum_{j \in V \cup S} \nabla_\theta \psi_j(\theta_0) = \frac{1}{\eta_V + \eta_S} \sum_{j \in V \cup S} \mathbb{E}[\nabla_\theta \psi_j(\theta_0)] + o_P(1).
\]

and

\[
\frac{1}{\eta_V + \eta_S} \sum_{j \in V \cup S} \mathbb{E}[\nabla_\theta \psi_j(\theta_0)] = -1 + \frac{1}{\eta_V + \eta_S} \left[ \sum_{j \in V \cup S} \Omega_j^{-2} \sigma_{X_j}^2 \tau_j^2 - 4 \theta_0 \sum_{j \in V \cup S} \Omega_j^{-2} \sigma_{X_j}^2 \beta_X \tau_j - 4 \theta_0^2 \sum_{j \in V \cup S} \Omega_j^{-3} \sigma_{X_j}^4 \tau_j^2 \right]
\leq -1 + \frac{1}{\eta_V + \eta_S} \left[ O(n \| \tau \|_2^2) + O(n \| \beta_X \|_2 \| \tau \|_2) \right]
\leq -1 + O(1/n \| \beta_X \|_2^2) + O(1/n^2 \| \beta_X \|_2)
= -1 + o_P(1).
\]

Hence, by T, \( \sum_{j \in V \cup S} \nabla_\theta \psi_j(\theta_0)/(\eta_V + \eta_S) = -1 + o_P(1) \). \hfill \square

**Lemma S.6.** Under Assumptions 1-4 and 6, \( \sum_{j \in V \cup S} \psi_j(\theta_0)/\sqrt{\eta_V + \eta_S + \zeta_V + \zeta_S} \overset{D}{\rightarrow} N(b_S/(\eta_V + \eta_S), 1) \).

**Proof.** We can decompose \( \psi_j(\theta_0) \) into a fixed bias term \( b_j \), and stochastic terms \( J_{1j} \) and \( J_{2j} \),

\[
\psi_j(\theta_0) = b_j + J_{1j} + J_{2j},
\]

where \( b_j = \Omega_j^{-1} \beta_{X_j} \tau_j + \theta_0 \Omega_j^{-2} \sigma_{X_j}^2 \tau_j^2, J_{1j} = \Omega_j^{-1} \beta_{X_j} (e_{X_j} - \theta_0 e_{X_j}) + \Omega_j^{-2} (e_{Y_j} - \theta_0 e_{X_j}) (\sigma_{Y_j}^2 e_{X_j} + \theta_0 \sigma_{X_j}^2 e_{Y_j}), J_{2j} = \Omega_j^{-2} (\sigma_{Y_j}^2 e_{X_j} + 2 \theta_0 \sigma_{X_j}^2 e_{Y_j} - \theta_0^2 \sigma_{X_j}^2 e_{X_j}) \tau_j). \)

By CS,

\[
\sum_{j \in V \cup S} \Omega_j^{-2} \sigma_{X_j}^2 \tau_j^2 = \Theta(n \| \tau \|_2^2) = O(1).
\]

[S-10]
Also,

\[
\text{Var} \left( \sum_{j \in V \cup S} J_{2j} \right) = \sum_{j \in V \cup S} \Omega_j^{-4} \tau_j^2 \sigma_X^2 \left[ (\sigma_{Y_j}^2 - \theta_0^2 \sigma_X^2) + 4\theta_0^2 \sigma_{X_j}^2 \right] \\
= \sum_{j \in V \cup S} \Omega_j^{-2} \tau_j^2 \sigma_X^2 \\
= O(1),
\]

where the last equality follows by \( n \| \mu \|_2^2 = O(1) \). Therefore, by CH and \( \mathbb{E}[J_{2j}] = 0 \) for all \( j \), \( \sum_{j \in V \cup S} J_{2j} = O_P(1) \).

Then, since \( \eta_V + \eta_S + \zeta_V + \zeta_S = \Theta(n \| \beta_X \|_2^2 + p) \),

\[
\frac{1}{\sqrt{\eta_V + \eta_S + \zeta_V + \zeta_S}} \sum_{j \in V \cup S} \left[ \psi_j(\theta_0) - \Omega_j^{-1} \beta_{X_j} \tau_j \right] = \left( \frac{1}{\sqrt{\eta_V + \eta_S + \zeta_V + \zeta_S}} \sum_{j \in V \cup S} J_{1j} \right) + o_P(1).
\]

By identical arguments used in ZWHBS, \( V_{1/2} \sum_{j \in V \cup S} J_{1j} \overset{D}{\to} N(0, 1) \). The result (ii) then follows by Slutsky’s lemma. \( \square \)

**Lemma S.7.** Under Assumptions 1-4 and 6, for any \( \tilde{\theta} \overset{P}{\to} \theta_0 \), \( \sum_{j \in V \cup S} \nabla_{\theta \theta} \psi_j(\tilde{\theta}) / (\eta_V + \eta_S) = O_P(1) \).

Proof. Let \( \Omega_j(\theta) = \sigma_{Y_j}^2 + \theta^2 \sigma_X^2 \), so that \( \nabla_{\theta} \Omega_j(\theta) = 2\theta \sigma_X^2 \). For any \( \theta \), we can write

\[
\nabla_{\theta \theta} \psi_j(\theta) = -2\hat{\beta}_{X_j} \hat{\beta}_{Y_j} \sigma_X^2 \Omega_j(\theta)^{-2} + 2\hat{\beta}_{X_j} \hat{\beta}_{X_j} \sigma_Y^2 (\nabla_{\theta} \Omega_j(\theta)) \Omega_j(\theta)^{-3} \\
-2\hat{\beta}_{Y_j} \sigma_X^2 \hat{\beta}_{X_j} (\nabla_{\theta} \Omega_j(\theta)) \Omega_j(\theta)^{-3} \\
-4\sigma_X^2 (\beta_{Y_j} - \theta \hat{\beta}_{X_j}) (\beta_{X_j} \sigma_Y^2 + \theta \hat{\beta}_{Y_j} \sigma_X^2) \Omega_j(\theta)^{-3} \\
+4\theta \sigma_X^2 \hat{\beta}_{X_j} \sigma_Y^2 (\beta_{X_j} \sigma_X^2 + \theta \hat{\beta}_{Y_j} \sigma_Y^2) \Omega_j(\theta)^{-3} \\
+12\theta \sigma_X^2 (\beta_{Y_j} - \theta \hat{\beta}_{X_j}) (\beta_{X_j} \sigma_Y^2 + \theta \hat{\beta}_{Y_j} \sigma_X^2) (\nabla_{\theta} \Omega_j(\theta)) \Omega_j(\theta)^{-4}
\]

\[= \sum_{l=1}^{8} H_{1j}(\theta).\]

We can expand \( H_{1j}(\theta) \) so that

\[
\sum_{j \in V \cup S} H_{1j}(\theta) = -\sum_{j \in V \cup S} (\beta_{X_j} + e_{X_j})(\theta_0 \beta_{X_j} + e_{Y_j} + \tau_j) \sigma_X^2 \Omega_j(\theta)^{-2} \\
= -\theta_0 \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} \beta_{X_j} - \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} \beta_{X_j} e_{Y_j} - \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} \beta_{X_j} \tau_j \\
-\theta_0 \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} \beta_{X_j} e_{X_j} - \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} e_{X_j} e_{Y_j} - \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} e_{X_j} \tau_j.
\]

For the first term on the right hand side, note that \( \sum_{j \in V \cup S} \sigma_X^2 \Omega_j(\theta)^{-2} \beta_{X_j}^2 = \Theta(n \| \beta_X \|_2^2) \) since \( \Omega_j(\theta) = \Theta(n^{-1}) \). For the second term,
Taylor expansion of the first-order condition so that by CH, 
\[
\sum_{j \in V \cup S} \sigma_{X_j}^2 \Omega_j(\theta)^{-2} \beta_{X_j} e_{Y_j} = \Theta(n\|\beta_X\|_2^2),
\]
so that by CH, 
\[
\sum_{j \in V \cup S} \sigma_{X_j}^2 \Omega_j(\theta)^{-2} \beta_{X_j} e_{Y_j} = \Theta(n^{\frac{1}{2}}\|\beta_X\|_2).
\]

Similarly, 
\[
\left| \sum_{j \in V \cup S} \sigma_{X_j}^2 \Omega_j(\theta)^{-2} \beta_{X_j} \tau_j \right| \leq \Theta(n\|\beta_X\|_2\|\tau\|_2) = O(n^{\frac{1}{2}}\|\beta_X\|).
\]

For the fifth term on the right hand side, 
\[
\text{Var} \left( \sum_{j \in V \cup S} \sigma_{X_j}^2 \Omega_j(\theta)^{-2} e_{X_j} e_{Y_j} \right) = \Theta \left( n^2 \sum_{j \in V \cup S} \sigma_{X_j}^2 \sigma_{Y_j}^2 \right) = \Theta(p),
\]
so 
\[
\sum_{j \in V \cup S} \sigma_{X_j}^2 \Omega_j(\theta)^{-2} e_{X_j} e_{Y_j} = O(p^{\frac{1}{2}}) \text{ by CH.}
\]

Using similar arguments for the remaining terms of \(\sum_{j \in V \cup S} H_{ij}(\theta)\), and for \(\sum_{j=1}^p H_{ij}(\theta)\), \(l \in [8]\), we have that 
\[
\sum_{l=1}^8 \sum_{j=1}^p H_{lj}(\theta) = O(n\|\beta_X\|_2^2 + p).
\]

Proof of Theorem 2 (Asymptotic distribution of \(\hat{\theta}_S\)).

The proof strategy follows ZWHBS’s Proof of Theorem 3.2, but with Lemmas S.5-S.7 involving additional steps required to control the non-negligible bias term from our model.

Let \(\psi_j(\theta) = (\hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j}) (\hat{\beta}_{X_j} \sigma_{Y_j}^2 + \theta \hat{\beta}_{Y_j} \sigma_{X_j}^2)(\sigma_{Y_j}^2 + \theta^2 \sigma_{X_j}^2)^{-2}\). Given consistency of \(\hat{\theta}\), a second-order Taylor expansion of the first-order condition \(\sum_{j=1}^8 \psi_j(\theta) = 0\) around \(\hat{\theta} = \theta_0\) implies that there exists \(\hat{\theta}\) on the line segment joining \(\hat{\theta}\) and \(\theta_0\) such that 
\[
\frac{\eta_\nu + \eta_S}{\sqrt{\eta_\nu + \eta_S + \zeta_\nu + \zeta_S}} (\hat{\theta} - \theta_0) = - \left( \frac{1}{\eta_\nu + \eta_S} \sum_{j \in V \cup S} \nabla_{\theta} \psi_j(\theta_0) + o_P \left( \frac{1}{\eta_\nu + \eta_S} \sum_{j \in V \cup S} \nabla_{\theta} \psi_j(\theta) \right) \right)^{-1} \times \frac{1}{\sqrt{\eta_\nu + \eta_S + \zeta_\nu + \zeta_S}} \sum_{j \in V \cup S} \psi_j(\theta_0).
\]

The result then follows by Slutsky’s lemma, and Lemmas S.5-S.7, which show:

(i) \(\sum_{j \in V \cup S} \nabla_{\theta} \psi_j(\theta_0)/(\eta_\nu + \eta_S) \overset{P}{\rightarrow} -1\);

(ii) \(\sum_{j \in V \cup S} \psi_j(\theta_0)/\sqrt{\eta_\nu + \eta_S + \zeta_\nu + \zeta_S} \overset{D}{\rightarrow} N(b_S/(\eta_\nu + \eta_S), 1)\);

(iii) for any \(\tilde{\theta} \overset{P}{\rightarrow} \theta_0\), \(\sum_{j \in V \cup S} \nabla_{\theta} \psi_j(\tilde{\theta})/(\eta_\nu + \eta_S) = O_P(1)\).

Proof of Theorem 3 (Asymptotic distribution of the bias \(b_S\)).

[S-12]
Let $\hat{B}_j(\theta) = \hat{\Omega}_j^{-1} \beta_X (\hat{\beta}_{ij} - \theta \hat{X}_j) + \theta \hat{\Omega}_j^{-1} \sigma^2_{x_j}$. Note that $\hat{\Omega}_j - \Omega_j = (\hat{\theta} - \theta_0)(\theta + \theta_0) \sigma^2_{x_j} = \Theta(n^{-1}|\hat{\theta} - \theta_0|)$. We can write

$$\sum_{j \in S} [\hat{B}_j(\theta_0) - \Omega_j^{-1} \beta_X \tau_j] = \sum_{j \in S} \left[ \Omega_j^{-1} \beta_X (e_{Y_j} - \theta_0 e_{X_j}) + \Omega_j^{-1} e_{X_j} e_{Y_j} - \theta_0 \Omega_j^{-1} (e_{X_j}^2 - \sigma^2_{x_j}) 
+ (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \beta_X (e_{Y_j} - \theta_0 e_{X_j}) + (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) e_{X_j} e_{Y_j}
- \theta_0 (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) (e_{X_j}^2 - \sigma^2_{x_j}) + (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \beta_X \tau_j
+ \Omega_j^{-1} \tau_j e_{X_j} + (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \tau_j e_{X_j} \right].$$

Using $\hat{\Omega}_j^{-1} - \Omega_j^{-1} = \Theta(n|\hat{\theta} - \theta_0|)$, note that $|\sum_{j \in S} (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \beta_X (e_{Y_j} - \theta_0 e_{X_j})| = O_P(n^{3/2}||\beta_x||_2)$ by CS, CH, and consistency of $\hat{\theta}$ for $\theta_0$. Similarly, the last five terms on the right hand side are $o_P(n^{3/2}||\beta_x||_2^2 + p^{3/2})$. Therefore, for $\hat{B}_j(\theta_0) = \Omega_j^{-1} \beta_X (e_{Y_j} - \theta_0 e_{X_j}) + \Omega_j^{-1} e_{X_j} e_{Y_j} - \theta_0 \Omega_j^{-1} (e_{X_j}^2 - \sigma^2_{x_j})$,

$$\sum_{j \in S} [\hat{B}_j(\theta_0) - \Omega_j^{-1} \beta_X \tau_j] = \sum_{j \in S} \hat{B}_j(\theta_0) + o_P(n^{3/2}||\beta_x||_2^2 + p^{3/2}).$$

By CH, $|e_{X_j}| = O_P(n^{-1/2})$ and $|e_{Y_j}| = O_P(n^{-3/2})$, so that $E[|\hat{B}_j(\theta_0)|^3] = O(n^{3/2}||\beta_x||^3) + O(n||\beta_x||^2) + O(n^{1/2}||\beta_x||) + O(1)$, and

$$\sum_{j \in S} E[|\hat{B}_j(\theta_0)|^3] = O(n^{3/2}||\beta_x||^3) + O(n||\beta_x||^2) + O(n^{1/2}||\beta_x||) + O(p)$$

$$= O(n^{3/2}||\beta_x||^3) + O(n||\beta_x||^2) + O(p),$$

where the last equality follows by $n^{1/2}||\beta_x||_1 \leq n^{1/2} p^{1/2}||\beta_x||_2 \leq (n||\beta_x||^2_2 + p)/2$.

Let $\xi_5 = 2q_0^2 \sum_{j \in S} \Omega_j^{-2} \sigma^4_{x_j}$. The variance of $\sum_{j \in S} \hat{B}_j(\theta_0)$ is given by

$$Var\left( \sum_{j \in S} \hat{B}_j(\theta_0) \right) = \eta_S + \zeta_S + \xi_S,$$

where $\eta_S + \zeta_S = \Theta(n||\beta_x||^2_2 + p)$ and $\xi_S = \Theta(p)$. Therefore, the following Lyapunov condition holds,

$$\frac{1}{(\eta_S + \zeta_S + \xi_S)^{3/2}} \sum_{j \in S} E[|\hat{B}_j(\theta_0)|^3] = O\left( \frac{||\beta_x||^3}{||\beta_x||^2} \right) + O\left( \frac{1}{n^{1/2}||\beta_x||_2} \right) + O\left( \frac{1}{p^{3/2}} \right) = o(1),$$

by $||\beta_x||^3/||\beta_x||^2 \rightarrow 0$. Thus, by Lyapunov’s CLT,

$$\frac{1}{\sqrt{\eta_S + \zeta_S + \xi_S}} \sum_{j \in S} \hat{B}_j(\theta_0) \overset{D}{\rightarrow} N(0,1).$$
Note that $\hat{B}_j(\theta_V) - \hat{B}_j(\theta_0) = -(\hat{\theta}_V - \theta_0)\hat{\Omega}_j^{-1}(\hat{\beta}_{X_j} - \sigma^2_{X_j})$, and

$$\sum_{j \in S} \hat{\Omega}_j^{-1}(\hat{\beta}_{X_j} - \sigma^2_{X_j}) = \sum_{j \in S} \left[ \Omega_j^{-1}\hat{\beta}_{X_j}^2 + (\hat{\Omega}_j^{-1} - \Omega_j^{-1})\beta_{X_j}^2 + 2\hat{\Omega}_j^{-1}\beta_{X_j}e_{X_j} + 2(\hat{\Omega}_j^{-1} - \Omega_j^{-1})\beta_{X_j}e_{X_j} + \Omega_j^{-1}(e_{X_j}^2 - \sigma_{X_j}^2) + (\hat{\Omega}_j^{-1} - \Omega_j^{-1})(e_{X_j}^2 - \sigma_{X_j}^2) \right].$$

By similar arguments used above, and since $p/n^2\|\beta_X\|^2 \to 0$, the last five terms on the right hand side are $o_P(n\|\beta_X\|^2)$. Therefore, since $\hat{\theta}_V - \theta_0 = O_P(1/n^{\frac{3}{2}}\|\beta_X\|_2) + O_P(p^{\frac{1}{2}}/n\|\beta_X\|^2)$, by the above results,

$$\sum_{j \in S} [\hat{B}_j(\theta_V) - \Omega_j^{-1}\beta_{X_j}\tau_j] = \sum_{j \in S} \hat{B}_j(\theta_0) - \eta_S(\hat{\theta}_V - \theta_0) + o_P(n^{\frac{1}{2}}\|\beta_X\|_2 + p^{\frac{1}{2}}).$$

Let $V_B = \eta_S + \zeta_S + (\eta_S^2/\eta_V^2)(\eta_V + \zeta_V)$. Note that $V_B = \Theta(n\|\beta_X\|^2 + p)$ so that

$$V_B^{-\frac{1}{2}} \sum_{j \in S} [\hat{B}_j(\theta_V) - \Omega_j^{-1}\beta_{X_j}\tau_j] = V_B^{-\frac{1}{2}} \left( \sum_{j \in S} \hat{B}_j(\theta_0) - \eta_S(\hat{\theta}_V - \theta_0) \right) + o_P(1).$$

From Proposition 1, we have the expansion

$$\hat{\theta}_V - \theta_0 = \frac{1}{\eta_V} \sum_{j \in V} J_{1,j} + o_P\left( \frac{\sqrt{\eta_V + \zeta_V}}{\eta_V} \right),$$

Then,

$$V_B^{-\frac{1}{2}} \sum_{j \in S} [\hat{B}_j(\hat{\theta}_V) - \Omega_j^{-1}\beta_{X_j}\tau_j] = V_B^{-\frac{1}{2}} \left( \sum_{j \in S} \hat{B}_j(\theta_0) - \eta_S \sum_{j \in V} J_{1,j} \right) + o_P(1),$$

as $o_P(V_B^{-\frac{1}{2}}(\eta_S/\eta_V^2)\sqrt{\eta_V + \zeta_V}) = o_P(1)$. From the above arguments, Proposition 1, and noting that $\sum_{j \in S} \hat{B}_j(\theta_0)$ and $\sum_{j \in V} J_{1,j}$ are mutually independent, the result follows by Slutsky’s lemma.

\[ \square \]

**Lemma S.8 (Consistent variance estimation).**

Let $\hat{\Omega}_j = \sigma_{X_j}^2 + \hat{\theta}_V^2\sigma_{X_j}^2$, $\hat{V}_B = \hat{\eta}_S + \hat{\zeta}_S + (\hat{\eta}_S^2/\hat{\eta}_V^2)(\hat{\eta}_V + \hat{\zeta}_V)$, where $\hat{\zeta}_S = 2\hat{\theta}_V^2 \sum_{j \in S} \hat{\Omega}_j^{-2}\sigma_{X_j}^4$.

We show that, under Assumptions 1-4 and 6,

(i) $(\eta_V + \eta_S)^2(\hat{\eta}_V + \hat{\eta}_S + \hat{\zeta}_V + \hat{\zeta}_S)/(\hat{\eta}_V + \hat{\eta}_S)^2(\eta_V + \eta_S + \zeta_V + \zeta_S) \xrightarrow{P} 1$;

(ii) $\hat{V}_B/V_B \xrightarrow{P} 1$.

**Proof.**

**Part (i)**

First, note that $\hat{\Omega}_j - \Omega_j = 2\sigma_{X_j}^2(\hat{\theta}_S - \theta_0) = \Theta(|\hat{\theta}_S - \theta_0|/n)$, so that $\hat{\Omega}_j^{-1} - \Omega_j^{-1} = \hat{\Omega}_j^{-1}(\Omega_j - \hat{\Omega}_j)\Omega_j^{-1} = \Theta(n|\hat{\theta}_S - \theta_0|)$. 

[S-14]
Then,

\[
\hat{\eta}V + \hat{\eta}S - \eta V - \eta S = 2 \sum_{j \in V \cup S} \Omega_j^{-1} \beta_{X_j} e_{X_j} + \sum_{j \in V \cup S} \Omega_j^{-1}(\epsilon_{X_j}^2 - \sigma_{X_j}^2) + \sum_{j \in V \cup S} (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \beta_{X_j}^2 \\
+ 2 \sum_{j \in V \cup S} (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \beta_{X_j} e_{X_j} + \sum_{j \in V \cup S} (\hat{\Omega}_j^{-1} - \Omega_j^{-1})(\epsilon_{X_j}^2 - \sigma_{X_j}^2)
\]

\[
= O_P(n^{\frac{1}{2}} \|\beta_X\|_2) + O_P(p^{\frac{1}{2}}) + o_P(n \|\beta_X\|_2^2) \\
= o_P(n \|\beta_X\|_2^2) + O_P(p^{\frac{1}{2}}),
\]

by similar arguments used in the proof of Propositions 1 and 2. By T, since \(\eta V + \eta S = \Theta(n \|\beta_X\|_2^2)\) and \(p/n^2 \|\beta_X\|_2^4 \to 0\), we have \((\hat{\eta}V + \hat{\eta}S)^{-1} = O_P(1/n \|\beta_X\|_2^2)\).

Therefore,

\[
\frac{(\eta V + \eta S)^2}{(\hat{\eta}V + \hat{\eta}S)^2} - 1 = \frac{(\eta V + \eta S + \hat{\eta}V + \hat{\eta}S)(\eta V + \eta S - \hat{\eta}V - \hat{\eta}S)}{(\hat{\eta}V + \hat{\eta}S)^2} \\
= O_P(n \|\beta_X\|_2^2)[o_P(n \|\beta_X\|_2^2) + O_P(p^{\frac{1}{2}})]O_P(1/n^2 \|\beta_X\|_2^4) \\
= o_P(1) + O_P(p^{\frac{1}{2}}/n \|\beta_X\|_2^2) \\
= o_P(1),
\]

where the last line follows from \(p/n^2 \|\beta_X\|_2^4 \to 0\).

Similarly, noting that \(\hat{\Omega}_j^{-2} - \Omega_j^{-2} = \hat{\Omega}_j^{-2}\Omega_j^{-2}(\Omega_j + \hat{\Omega}_j)(\Omega_j - \hat{\Omega}_j) = \Theta(n^2 |\hat{\theta}_S - \theta_0|)\), we have

\[
\hat{\zeta}V + \hat{\zeta}S - \zeta V - \zeta S = \sum_{j \in V \cup S} (\hat{\Omega}_j^{-2} - \Omega_j^{-2}) \sigma_{X_j}^2 \sigma_{\epsilon_j}^2 \\
= \Theta(p |\hat{\theta}_S - \theta_0|) \\
= o_P(p).
\]

Using the above results,

\[
\frac{(\eta V + \eta S)^2}{(\hat{\eta}V + \hat{\eta}S)^2} \cdot \frac{\hat{\eta}V + \hat{\eta}S + \hat{\zeta}V + \hat{\zeta}S}{\eta V + \eta S + \zeta V + \zeta S} = (1 + o_P(1)) \left(1 + \frac{\hat{\eta}V + \hat{\eta}S - \eta V - \eta S}{\eta V + \eta S + \zeta V + \zeta S} + \frac{\hat{\zeta}V + \hat{\zeta}S - \zeta V - \zeta S}{\eta V + \eta S + \zeta V + \zeta S}\right) \\
= (1 + o_P(1)) \left(1 + \frac{O_P(n \|\beta_X\|_2^2) + O_P(p^{\frac{1}{2}}) + o_P(p)}{\Theta(n \|\beta_X\|_2^2) + \Theta(p)}\right) \\
= (1 + o_P(1))^2 \\
= 1 + o_P(1).
\]

Part (ii)
First, note that

\[
\hat{s}_i - \xi_s = 2(\hat{\theta}_S + \theta_0)(\hat{\theta}_S - \theta_0) \sum_{j \in S} \Omega_j^{-2} \sigma_{X_j}^2 + 2(\hat{\theta}_S + \theta_0)(\hat{\theta}_S - \theta_0) \sum_{j \in S} (\Omega_j^{-2} - \Omega_j^{-2}) \sigma_{X_j}^4 \\
+ 2\theta_0^2 \sum_{j \in S} (\Omega_j^{-2} - \Omega_j^{-2}) \sigma_{X_j}^4
\]

\[
= \Theta(p|\hat{\theta}_S - \theta_0|) + \Theta(p|\hat{\theta}_S - \theta_0|^2) \\
= o_P(p).
\]

Therefore,

\[
\hat{V}_B - V_B = (\hat{\eta}_S - \eta_S) + (\hat{\xi}_S - \xi_S) + (\hat{\xi}_S - \xi_S) + (\hat{\eta}_S - \eta_S)(\hat{\eta}_S + \eta_S)(\hat{\eta}_V + \hat{\xi}_V) \\
+ \eta_S^2 \left( \frac{\hat{\eta}_V + \hat{\xi}_V}{\eta_V} - \frac{\eta_V + \xi_V}{\eta_V} \right) \\
= o_P(n\|\beta_X\|^2) + o_P(1) + o_P(p) + \left[ O_P(n\|\beta_X\|^2) + O_P(p^2) \right] O_P(n\|\beta_X\|^2) \left[ O_P(1/n\|\beta_X\|^2) + o_P(1/n\|\beta_X\|^2) \right] \\
+ o_P(1) + O(n^2\|\beta_X\|^2) \left[ o_P(1/n\|\beta_X\|^2) + o_P(p/n\|\beta_X\|^2) \right]
\]

\[
= o_P(n\|\beta_X\|^2) + o_P(p) + o_P(p^2/n\|\beta_X\|^2)
\]

where the second equality follows by (i) and T, since \(\hat{\eta}_V^{-2}(\hat{\eta}_V + \hat{\xi}_V) = O_P(\eta_V^{-2}(\eta_V + \xi_V)) = O_P(1/n\|\beta_X\|^2) + O_P(p/n^2\|\beta_X\|^2)\), and \(\hat{\eta}_V^{-2}(\hat{\eta}_V + \hat{\xi}_V) - \eta_V^{-2}(\eta_V + \xi_V) = o_P(\eta_V^{-2}(\eta_V + \xi_V)) = o_P(1/n\|\beta_X\|^2) + o_P(p/n^2\|\beta_X\|^2)\).

Therefore, since \(V_B = \Theta(n\|\beta_X\|^2) + \Theta(p)\),

\[
\frac{\hat{V}_B - V_B}{V_B} = o_P(1) + o_P \left( \frac{p}{n\|\beta_X\|^2} \right)
\]

\[
= o_P(1),
\]

as \(p = O(n)\) and \(\|\beta_X\|^2 = \Theta(1)\). Finally, we note \(\hat{V}_B/V_B = 1 + (\hat{V}_B - V_B)/V_B = 1 + o_P(1)\). \(\square\)

**Proof of Lemma 1 (Convergence in distribution of effect and bias estimates).**

As shown in the proofs of Theorem 2 and 3, for any \(S\), ignoring \(o_P(1/n^2\|\beta_X\|^2)\) and \(o_P(p^2/n\|\beta_X\|^2)\) terms,

\[
\hat{\theta}_V - \theta_0 = \frac{1}{\eta_V} \sum_{j \in V} J_{1j}
\]

\[
\hat{\theta}_j - \theta_0 - \frac{b_{S_k}}{\eta_V + \eta_{S_k}} = \frac{1}{\eta_V + \eta_{S_k}} \sum_{j \in V \cup S_k} J_{1j}
\]

\[
\hat{b}_{S_k} - b_{S_k} = \frac{1}{\eta_V + \eta_{S_k}} \sum_{j \in S_k} B_j - \frac{\eta_{S_k}}{\eta_V(\eta_V + \eta_{S_k})} \sum_{j \in V} J_{1j}
\]
where \( J_{1j} = \Omega_j^{-1} \beta_X (e_Y - \theta_0 e_X) + \Omega_j^{-2} (e_Y - \theta_0 e_X) (\sigma^2 \beta_j e_X + \theta_0 \sigma^2 \beta_j e_Y) \), and \( B_j = \Omega_j^{-1} \beta_X (e_Y - \theta_0 e_X) + \Omega_j^{-1} e_X e_Y - \theta_0 \Omega_j^{-1} (e^2 \beta_j - \sigma^2 \beta_j) \).

We can partition the the \( K \) additional instrument sets into \( L \leq 2^K - 1 \) distinct sets which span the additional instrument sets \( S_1, ..., S_K \). For example, for \( K = 3 \), each instrument must belong to one, and only one, of the following sets: \( M_1 = S_1 \cap S_2 \cap S_3, M_2 = S_1 \cap S_2 \cap S_3^C, M_3 = S_1 \cap S_2^C \cap S_3, M_4 = S_2 \cap S_2 \cap S_3, M_5 = S_1 \cap S_2^C \cap S_3^C, M_6 = S_1^C \cap S_2 \cap S_3^C, \) and \( M_7 = S_1^C \cap S_2^C \cap S_3^C \). Then, for each \( j \in [3] \), we can construct selection indicators \( \alpha_{\ell} \in \{0, 1\} \), \( \ell \in [7] \), such that \( S_j = \bigcup_{\ell=1}^{7} \alpha_\ell M_\ell \).

For \( L \leq 2^{[K]} - 1 \), let \( M_1, ..., M_L \) be distinct sets of the additional instruments which span the additional instrument sets \( S_1, ..., S_K \).

We can therefore write

\[
\begin{pmatrix}
\hat{\theta}_V - \theta_0 \\
\hat{\theta}_{S_1} - \theta_0 - \frac{b_{S_1}}{\eta_V + \eta_{S_1}} \\
\vdots \\
\hat{\theta}_{S_K} - \theta_0 - \frac{b_{S_K}}{\eta_V + \eta_{S_K}} \\
\end{pmatrix}
= \pi_V \left( \sum_{j \in V} J_{1j} \right) + \sum_{\ell=1}^{L} \pi_{M_\ell} \left( \sum_{j \in M_\ell} \mu_j \right),
\]

where

\[
\mu_j = \begin{pmatrix} J_{1j} \\ R_j \end{pmatrix}, \quad \pi_V = \begin{pmatrix} 1 \\ \frac{1}{\eta_V + \eta_{S_1}} \\ \vdots \\ \frac{1}{\eta_V + \eta_{S_K}} \end{pmatrix}, \quad \pi_{M_\ell} = \begin{pmatrix} 0 \\ \frac{1}{\eta_V + \eta_{S_1}} \chi_{M_\ell \subseteq S_1} \\ \vdots \\ \frac{1}{\eta_V + \eta_{S_K}} \chi_{M_\ell \subseteq S_K} \end{pmatrix}, \quad (1 \ 0) \text{, } (\ell = 1, ..., L).
\]

and \( R_j = \Omega_j^{-2} \theta_0 (\sigma^2 \beta_j e_Y - \sigma^2 \beta_j e_X) - 2 \theta_0 \sigma^2 \beta_j e_X e_Y + \theta_0 \Omega_j^{-1} (e^2 \beta_j - \sigma^2 \beta_j) \). For any set \( M_\ell \), we will show that

\[
\sum_{j \in M_\ell} \mu_j \sim N \begin{pmatrix} 0 \\ \eta_{M_\ell} + \zeta_{M_\ell} \end{pmatrix}, \quad (\ell = 1, ..., L). \tag{S.3}
\]

Also, as in the Proof of Lemma S.6, we have

\[
\sum_{j \in V} J_{1j} \sim N(0, \eta_V + \zeta_V).
\]

Then, since (i) the random components in \( J_{1j} \) and \( \mu_j \) are functions of the error terms \( e_X \) and \( e_Y \),

[S-17]
(ii) for any $j \neq k$, $e_{X_j}$ and $e_{X_k}$ are jointly normal and uncorrelated, and hence mutually independent (likewise for $e_{Y_j}$ and $e_{X_j}$), we have that $\sum_{j \in V} J_{1j}$, $\sum_{j \in M_1} \mu_j$, ..., $\sum_{j \in M_K} \mu_j$ are mutually independent sums. Therefore,

$$\pi_V \left( \sum_{j \in V} J_{1j} \right) + \sum_{\ell=1}^L \pi_{M_\ell} \left( \sum_{j \in M_\ell} \mu_j \right) \sim N \left( 0_{(2K+1) \times 1}, V + W \right),$$

where $V = \pi_V (\eta_V + \zeta_V)\pi_V^T$ and $W = \sum_{\ell=1}^L \pi_{M_\ell} \begin{bmatrix} \eta_{M_\ell} + \zeta_{M_\ell} & 0 \\ 0 & \xi_{M_\ell} \end{bmatrix} \pi_{M_\ell}^T$.

Some straight-forward calculations show that

$$V = \begin{bmatrix} V_{11} & V_{12}^T & V_{13}^T \\ V_{21} & V_{22} & V_{23}^T \\ V_{31} & V_{32} & V_{33} \end{bmatrix},$$

where $V_{11} = \eta_V^{-2}(\eta_V + \zeta_V)$, the $k$-th element of $V_{21}$ is given by $V_{21}^{(k)} = \eta_V^{-1}(\eta_V + \eta_S) - 1(\eta_V + \zeta_V)$, the $k$-th element of $V_{31}$ is given by $V_{31}^{(k)} = -\eta_V^{-2}(\eta_V + \eta_S) - 1(\eta_V + \zeta_V)$, the $(k, l)$-th element of $V_{22}$ is given by $V_{22}^{(k, l)} = (\eta_V + \eta_S) - 1(\eta_V + \zeta_V)$, the $(k, l)$-th element of $V_{32}$ is given by $V_{32}^{(k, l)} = -\eta_V^{-1}(\eta_V + \eta_S) - 1(\eta_V + \zeta_V)$, and the $(k, l)$-th element of $V_{33}$ is given by $V_{33}^{(k, l)} = \eta_V^{-2}(\eta_V + \eta_S) - 1(\eta_V + \eta_S) - 1(\eta_V + \zeta_V)$.

Similarly, we have

$$W = \begin{bmatrix} 0 & 0 & 0 \\ 0 & W_1 & W_1 \\ 0 & W_1 & W_2 \end{bmatrix},$$

where the $(k, l)$-th element of $W_1$ is given by $W_{1}^{(k, l)} = (\eta_V + \eta_S) - 1(\eta_V + \eta_S) - 1(\eta_S \cap S) + \zeta_S \cap S)$, and the $(k, l)$-th element of $W_2$ is given by $W_{2}^{(k, l)} = (\eta_V + \eta_S) - 1(\eta_V + \eta_S) - 1(\eta_S \cap S) + \zeta_S \cap S + \xi_S \cap S)$.

The expression for the covariance matrix in Lemma 1 is then $\Sigma = V + W$.

**Proof of Equation S.3.**

If $|M_\ell| = o(p)$, then the following asymptotic distribution result for $\sum_{j \in M_\ell} \mu_j$ still applies but variance components $\zeta_{M_\ell}$ and $\xi_{M_\ell}$ from Equation S.3 would be negligible. Therefore, we focus on the case $|M_\ell| = O(p)$ for $\ell \in [L]$. We wish to apply the following multivariate Berry-Esseen result by Bentkus (2005). For a subset of instruments $M_\ell$ and the 2-dimensional vector of independent variables $\mu_1, ..., \mu_{|M_\ell|}$, let $\Sigma_{\ell} = Var \left( \sum_{j \in M_\ell} \mu_j \right)$. Then for $U \sim N(0, \Sigma_{\ell})$, for any convex set $S$,

$$|\mathbb{P} \left( \sum_{j \in M_\ell} \mu_j \in S \right) - \mathbb{P}(U \in S)| \leq O \left( \sum_{j \in M_\ell} \mathbb{E} \left[ \| \Sigma_{\ell}^{-1/2} \mu_j \|^2 \right] \right).$$

[S-18]
Thus, joint convergence of effect and bias estimates follows by the following equations

\[ \Sigma_{\ell} = \Theta(n\|\beta_X\|_2^2 + p) \]  \hspace{1cm} (S.4)

\[ \sum_{j \in M_{\ell}} \mathbb{E}[\|\mu_j\|_2^3] = O(n^{3/2}\|\beta_X\|_2^2) + O(p^{3/2}n\|\beta_X\|_2^2) + O(n\|\beta_X\|_2^2) + O(p) \]  \hspace{1cm} (S.5)

since (4) implies \( \Sigma_{\ell}^{-1/2} = \Theta(1/n^{1/2}\|\beta_X\|_2) + \Theta(1/p^{1/2}) \), and by (5),

\[ \sum_{j \in M_{\ell}} \mathbb{E}[\|\Sigma_{\ell}^{-1/2}\mu_j\|_2^3] = \sum_{j \in M_{\ell}} \mathbb{E}[\|\mu_j\|_2^3] \]

\[ = O\left(\|\beta_X\|_2^4\right) + O\left(\frac{p^{1/2}}{n^{1/2}\|\beta_X\|_2^2}\right) + O\left(\frac{1}{n^{1/2}\|\beta_X\|_2^2}\right) + O\left(\frac{1}{p^{1/2}}\right) \]

\[ = o(1), \]

as \( \|\beta_X\|_4/\|\beta_X\|_2 \to 0 \) by Assumption 2, and the many weak instruments set-up \( n\|\beta_X\|_2^2/p = \Theta(1) \) (see discussion in Section 4.2).

To show (4), we calculate the covariance matrix \( \Sigma_{\ell} = \text{Var}(\sum_{j \in M_{\ell}} \mu_j) \) as

\[ \Sigma_{\ell} = \sum_{j \in M_{\ell}} \text{Var}(\mu_j) \]

\[ = \begin{bmatrix} \eta_{M_{\ell}} + \zeta_{M_{\ell}} & 0 \\ 0 & \xi_{M_{\ell}} \end{bmatrix} \]

\[ = \Theta(n\|\beta_X\|_2^2 + p), \]

since \( \text{Var}(J_{1j}) = \Omega_{1j}^{-1}\beta_{Xj}^2 + \Omega_{1j}^{-2}\sigma_{Xj}^2\sigma_{Yj}^2 \), \( \text{Cov}(J_{1j}, R_j) = 0 \), and \( \text{Var}(R_j) = 2\theta_0\Omega_{1j}^{-2}\sigma_{Xj}^4 \).

To show (5), note that \( \|\mu_j\|_2^2 = J_{1j}^2 + (J_{1j} - B_j)^2 \leq 3J_{1j}^2 + 2B_j^2 \), and \( \|\mu_j\|_2^4 \leq 9J_{1j}^4 + 12J_{1j}^2 B_j^2 + 4B_j^4 \).

Thus, \( \mathbb{E}[\|\mu_j\|_2^3] = 3\mathbb{E}[J_{1j}^3] + 2\mathbb{E}[B_j^3] \), and \( \mathbb{E}[\|\mu_j\|_2^4] \leq 15\mathbb{E}[J_{1j}^4] + 10\mathbb{E}[B_j^4] \).

Then,

\[ \sum_{j \in M_{\ell}} \mathbb{E}[\|\mu_j\|_2^3] = 3 \sum_{j \in M_{\ell}} \mathbb{E}[J_{1j}^3] + 2 \sum_{j \in M_{\ell}} \mathbb{E}[B_j^3] \]

\[ = 5\eta_{M_{\ell}} + 5\zeta_{M_{\ell}} + 2\xi_{M_{\ell}} \]

\[ = \Theta(n\|\beta_X\|_2^3 + p), \]

and

\[ \sum_{j \in M_{\ell}} \mathbb{E}[\|\mu_j\|_2^4] \leq 15 \sum_{j \in M_{\ell}} \mathbb{E}[J_{1j}^4] + 10 \sum_{j \in M_{\ell}} \mathbb{E}[B_j^4] \]

\[ = O(n^2\|\beta_X\|_4^4) + O(n\|\beta_X\|_2^4) + O(p), \]
since by direct calculation, it can be shown that there exist constants $C_l$, $l \in [10]$ such that

$$
\sum_{j \in M_k} \mathbb{E}[J_{1j}^4] = C_1 \sum_{j \in M_k} \Omega_j^{-2} \beta_j^4 + C_2 \sum_{j \in M_k} \Omega_j^{-3} \sigma_j^2 \beta_j^2 + C_3 \sum_{j \in M_k} \Omega_j^{-4} \sigma_j^4 \beta_j^4 + C_4 \sum_{j \in M_k} \Omega_j^{-5} \sigma_j^3 \beta_j^3 + C_5 \sum_{j \in M_k} \Omega_j^{-6} \sigma_j^6 \beta_j^6 + \cdots
$$

= \Theta(n^2 \|\beta_X\|_4^4) + \Theta(n \|\beta_X\|_2^2) + \Theta(p),

and

$$
\sum_{j \in M_k} \mathbb{E}[^{\hat B}_j^4] = C_4 \sum_{j \in M_k} \Omega_j^{-2} \beta_j^4 + C_5 \sum_{j \in M_k} \Omega_j^{-3} \sigma_j^2 \beta_j^2 + C_6 \theta_0^2 \sum_{j \in M_k} \Omega_j^{-3} \sigma_j^6 \beta_j^6 + C_7 \theta_0^2 \sum_{j \in M_k} \Omega_j^{-4} \sigma_j^4 \beta_j^4 + C_8 \sum_{j \in M_k} \Omega_j^{-4} \sigma_j^2 \beta_j^2 + C_9 \theta_0^2 \sum_{j \in M_k} \Omega_j^{-3} \sigma_j^6 \beta_j^6 + C_{10} \theta_0^2 \sum_{j \in M_k} \Omega_j^{-4} \sigma_j^4 \beta_j^4
$$

$$
= \Theta(n^2 \|\beta_X\|_4^4) + \Theta(n \|\beta_X\|_2^2) + \Theta(p).
$$

Thus, by CS and the above,

$$
\sum_{j \in M_k} \mathbb{E}[\|\mu_j\|^2_2^2] \leq \left( \sum_{j \in M_k} \mathbb{E}[\|\mu_j\|^2_2] \right)^{\frac{1}{2}} \left( \sum_{j \in M_k} \mathbb{E}[\|\mu_j\|^2_2] \right)^{\frac{1}{2}}
$$

$$
\leq (O(n^2 \|\beta_X\|_4^4) + O(n \|\beta_X\|_2^2) + O(p)) \frac{1}{2} (O(n \|\beta_X\|_2^2) + O(p)) \frac{1}{2}
$$

$$
= O(n\|\beta_X\|_4^4 \|\beta_X\|_2^2) + O(p^2 \|\beta_X\|_2^2) + O(n \|\beta_X\|_2^2) + O(p).
$$

Proof of Theorem 4 (Asymptotic distribution of the post-selection estimator $\hat \theta$).

Since $\sum_{k=1}^K \hat \omega_k = 1$, we have

$$
\hat \theta - \theta_0 = \sum_{k=1}^K \hat \omega_k (\hat \theta_{S_k} - \theta_0).
$$

For all $k \in [K]$, as $\hat \omega_k$ is a function of the estimated AMSE, which consists of consistent estimators of constants (see Lemma S.8) and $\hat \theta_{S_k}$, the result then follows by Slutsky’s lemma and Lemma 1 which shows the joint convergence in distribution of $\hat b_{S_k}$ and $\hat \theta_{S_k}$ over the whole model space $S_k$, $k \in [K]$.

Proof of Theorem 5 (Bound implied by support restriction).

Let $\bar b_S = \bar \tau \sum_{j \in S} \Omega_j^{-1} \|\beta_j\|$. By CS,

$$
- \bar b_S \leq b_S \leq \bar b_S.
$$

(S.6)
Let $\hat{b}_S = \bar{\tau} \sum_{j \in S} \Omega_j^{-1} |\hat{\beta}_{X_j}|$. By the reverse triangle inequality,

$$|\tilde{b}_S - \hat{b}_S| \leq \bar{\tau} \sum_{j \in S} \Omega_j^{-1} \left| |\hat{\beta}_{X_j}| - |\beta_{X_j}| \right| \leq \bar{\tau} \sum_{j \in S} \Omega_j^{-1} |e_{X_j}|.$$ 

Let $\dot{b}_S = \bar{\tau} \sum_{j \in S} \Omega_j^{-1} |e_{X_j}|$, so that by the above, 

$$\tilde{b}_S - \dot{b}_S \leq \tilde{b}_S \leq \dot{b}_S + \ddot{b}_S. \quad (S.7)$$

Note that

$$E[\dot{b}_S] = \bar{\tau} \sum_{j \in S} \Omega_j^{-1} E[|e_{X_j}|] = \sqrt{\frac{2}{\pi}} \bar{\tau} \sum_{j \in S} \Omega_j^{-1} \sigma_{X_j} = O(\sqrt{p}),$$

by Assumption 1, $|S| = O(p)$, and $\bar{\tau} = O(1/\sqrt{np})$. Similarly,

$$\text{Var}(\dot{b}_S) = \bar{\tau}^2 \sum_{j \in S} \Omega_j^{-2} \text{Var}(|e_{X_j}|) = (1 - \frac{2}{\pi}) \bar{\tau}^2 \sum_{j \in S} \Omega_j^{-2} \sigma_{X_j}^2 = O(1),$$

by Assumption 1, $|S| = O(p)$, and $\bar{\tau} = O(1/\sqrt{np})$. Thus, by CH, $|\hat{b}_S - E[\hat{b}_S]| = O_P(1)$. Using (S.7),

$$\frac{1}{\sqrt{p}} (\tilde{b}_S - E[\tilde{b}_S]) + o_P(1) \leq \frac{1}{\sqrt{p}} \ddot{b}_S \leq \frac{1}{\sqrt{p}} (\tilde{b}_S + E[\tilde{b}_S]) + o_P(1). \quad (S.8)$$

Let $\ddot{b}_S = \sqrt{\frac{2}{\pi}} \bar{\tau} \sum_{j \in S} \hat{\Omega}_j^{-1} \sigma_{X_j}$. Then,

$$\frac{1}{\sqrt{p}} (\ddot{b}_S - E[\ddot{b}_S]) = \frac{1}{\sqrt{p}} \bar{\tau} \sqrt{\frac{2}{\pi}} \sum_{j \in S} (\hat{\Omega}_j^{-1} - \Omega_j^{-1}) \sigma_{X_j} = o_P(1),$$

by Assumption 1, $|S| = O(p)$, $\bar{\tau} = O(1/\sqrt{np})$, and Lemma S.8. Therefore, by (S.8),

$$\frac{1}{\sqrt{p}} (\ddot{b}_S - \ddot{b}_S) + o_P(1) \leq \frac{1}{\sqrt{p}} \ddot{b}_S \leq \frac{1}{\sqrt{p}} (\ddot{b}_S + \ddot{b}_S) + o_P(1). \quad (S.9)$$
By (S.6) and (S.9),
\[-\frac{1}{\sqrt{p}}(\hat{b}_S + \tilde{b}_S) + o_P(1) \leq \frac{1}{\sqrt{p}}b_S \leq \frac{1}{\sqrt{p}}(\tilde{b}_S + \tilde{b}_S) + o_P(1).\]

Finally, let \(\tilde{b}_S^* = \tilde{\tau} \sum_{j \in S} \hat{\Omega}_j^{-1} |\hat{\beta}_{X_j}|.\) Then, by CS and T,
\[
\frac{1}{\sqrt{p}}(\tilde{b}_S^* - \tilde{b}_S) \leq \frac{1}{\sqrt{p}} \tilde{\tau} \sum_{j \in S} |\hat{\Omega}_j^{-1} - \Omega_j^{-1}| |\hat{\beta}_{X_j}| + \frac{1}{\sqrt{p}} \tilde{\tau} \sum_{j \in S} |\hat{\Omega}_j^{-1} - \Omega_j^{-1}| |e_{X_j}| \\
= o_P(1),
\]
by CS, Lemma S.8, and similar arguments used to establish (S.8). The result then follows by (S.9), (S.10), and noting that \(\eta_V + \eta_S = \Theta(n \|\beta_X\|_2^2 + p),\) so that by Lemma S.8, \(\sqrt{p}(\hat{\eta}_V + \hat{\eta}_S)^{-1} = O(1/\sqrt{n})\) if \(p = \Theta(n)\) and \(\|\beta_X\|_2 = \Theta(1).\)

\[\Box\]

**Data**

We took genetic associations with vitamin D (serum 25-hydroxyvitamin D; 25OHD) concentration from a GWAS by Revez et al. (2020) based on a large UK Biobank sample of 417,580 European individuals (Sudlow et al., 2015). Genetic associations with CHD were taken from a meta-GWAS of 48 studies with a total of 60,801 cases and 123,504 controls from a majority European population, conducted by the CARDIoGRAMplusC4D consortium (Nikpay et al., 2015). Genetic associations with SBP were taken from a meta-GWAS of 757,601 individuals from a European population drawn from the UK Biobank and the International Consortium of Blood Pressure (Evangelou et al., 2018). Genetic associations with MS were taken from a GWAS of 47,429 cases and 68,374 controls from a European population conducted by the International Multiple Sclerosis Genetics Consortium (Pat-sopoulos et al., 2019). Finally, genetic associations with AD were taken from a GWAS of 17,008 cases and 37,154 controls from a European population conducted by International Genomics of Alzheimer’s Project (Lambert et al., 2013). To define gene regions, we considered variant positions ±500 mb from gene positions indicated on GeneCards, Stelzer et al. (2016).