Identification- and many instrument-robust tests via invariant moment conditions*

Tom Boot
University of Groningen
t.boot@rug.nl

Johannes W. Ligtenberg
University of Groningen
j.w.ligtenberg@rug.nl

February 28, 2023

Abstract

Identification-robust hypothesis tests are commonly based on the continuous updating objective function or its score. When the number of moment conditions grows proportionally with the sample size, the large-dimensional weighting matrix prohibits the use of conventional asymptotic approximations and the behavior of these tests remains unknown. We show that the structure of the weighting matrix opens up an alternative route to asymptotic results when the distribution of the moment conditions evaluated at the true parameters is reflection invariant. In a heteroskedastic linear instrumental variables model, this allows us to establish joint asymptotic normality of conventional identification-robust tests statistics under many-instrument sequences. The additional variance terms that appear are negative, indicating that the conventional approximation leads to conservative tests. We revisit a study on the elasticity of substitution between immigrant and native workers where the number of instruments is over a quarter of the sample size. As the theory predicts, the many-instrument robust approximation leads to substantially narrower confidence intervals.

Keywords: robust inference, many instruments, heteroskedasticity.
JEL codes: C12, C26,

*We thank Federico Crudu, Paul Elhorst, Jinyong Hahn, Bruce Hansen, Artūras Juodis, Lingwei Kong, Nick Koning, Zhipeng Liao, Xinwei Ma, Jack Porter, Andres Santos, Xiaoxia Shi, Yixao Sun, Aico van Vuuren, and Tom Wansbeek for many insightful comments.

1 Introduction

Identification-robust inference procedures guard researchers against spurious conclusions that arise due to a lack of instrument relevance. These procedures are commonly based on the continuous updating (CU) objective function of Hansen, Heaton, and Yaron (1996), either directly through Anderson and Rubin's (1949) AR statistic, via its score as in Kleibergen's (2005) K statistic, or via a combination thereof (Moreira, 2003; Kleibergen, 2005; Andrews, 2016). While conventional tools yield the limiting distribution of these statistics when the number of instruments is small, their asymptotic behavior is unknown under Bekker's (1994) many instrument sequences. This limits their use in recent applications with large instrument sets such as the judge design dummies from Kling (2006), genetic variants instrumenting socioeconomic variables (Davies, Hemani, Timpson, Windmeijer, and Davey Smith, 2015), Bartik instruments that interact local industry shares with national industry growth rates (Goldsmith-Pinkham, Sorkin, and Swift, 2020), and saturated specifications allowing nonparametric conditioning (Angrist and Imbens, 1995; Blandhol, Bonney, Mogstad, and Torgovitsky, 2022).

The main obstacle in establishing the limiting distribution of statistics based on the CU objective function is that under many-instrument sequences the weighting matrix for the moment conditions is large-dimensional and does not converge to a well-defined object. As Newey and Windmeijer (2009) write: "If the number of instruments grows as fast as the sample size, the number of elements of the weight matrix grows as fast as the square of the sample size. It seems difficult to simultaneously control the estimation error for all these elements." In this paper, we show how this difficulty can be circumvented when the distribution of the moment conditions evaluated at the true parameters satisfies a suitable invariance condition. We focus throughout on the linear instrumental variables (IV) model with heteroskedasticity, although our results on the AR statistic apply in a generic generalized methods of moments set-up.

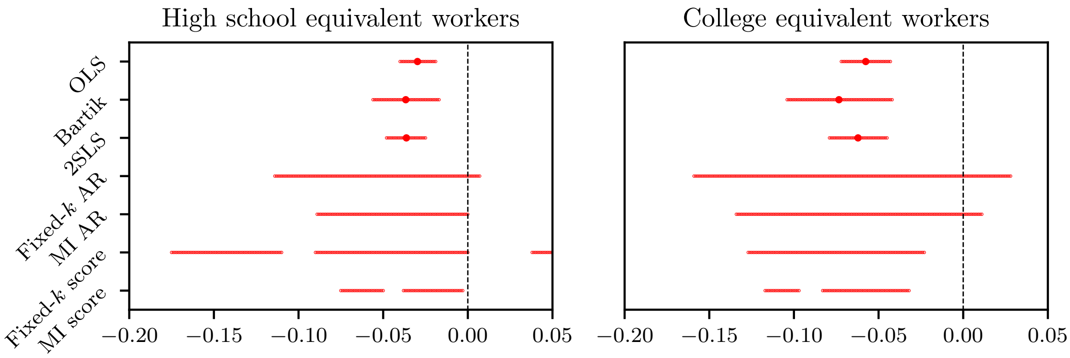
Our approach is motivated by the observation that if the distribution of the moment conditions, evaluated at the true parameter vector, is orthogonally invariant, the finite sample distribution of the AR statistic is available in closed form. This sidesteps the issue that the dimension of the weighting matrix is nonneglible compared to the sample size, but the scope of application is limited: when the moment conditions are independent, orthogonal invariance implies that they are normally distributed. However, under a substantially weaker invariance property, known as orthant symmetry (Efron, 1969) or reflection invariance (Bekker and

Lawford, 2008), the obstruction posed by the weighting matrix can be circumvented as well. This type of invariance is in particular suitable for heteroskedastic models as it allows the distribution of the moment conditions to differ across observations. Under reflection invariance, the finite sample distribution of the CU objective function is no longer tractable, but its limiting distribution, and hence, that of the Anderson-Rubin statistic, follows from known results on the limiting behavior of bilinear forms by Chao, Swanson, Hausman, Newey, and Woutersen (2012). We propose a procedure that in finite samples produces narrower confidence intervals relative to those based on an asymptotic approximation that assumes the number of instruments to be fixed. Conservativeness of the usual AR statistic is also studied by Bun, Farbmacher, and Poldermans (2020). They show that even under the null and with a fixed number of moment conditions the test based on the AR statistic is undersized in finite samples.

A well-known downside of the AR statistic is that it lacks power in overidentified models. This problem is particularly severe under many instrument sequences. We therefore turn to a derivation of the asymptotic distribution of the score function. We establish a new central limit theorem for cubic forms that implies joint asymptotic normality of the score and the appropriately centered and scaled AR statistic under many instrument sequences. The variance of the score function contains several terms that do not appear when the number of instruments grows slower than proportionally with the sample size. Importantly, we show that the sum of these terms is negative, indicating that the conventional asymptotic approximation leads to unnecessarily wide confidence intervals. Furthermore, as the score is the derivative of the CU objective function it has a low value in any region where the objective function is flat. This is the case for tested parameters around the true value, but also for distant alternatives. We combine the many instrument robust AR and score statistics in a two step test to overcome this power loss. The first step is based on the AR statistic and ensures that alternatives far from the true value are rejected. The second step uses the score to improve the overall power.

To highlight the practical relevance of the many-instrument robust approximations, we consider data from Card (2009) that was recently revisited in Goldsmith-Pinkham et al. (2020). The goal is to conduct inference on the negative inverse elasticity of substitution between immigrant and native workers by using the share of immigrants from a particular country of origin in various US cities as instruments. Many-instrument analysis is relevant here as there are 38 instruments (equal to the number of countries of origin) and 124 observations (equal to the

Figure 1: Illustration: 95% CIs for the negative inverse elasticity of substitution.



number of cities). Figure 1 shows 95% confidence intervals for the negative inverse elasticity of substitution for high-school equivalent workers (left panel) and college equivalent workers (right panel). The intervals are based on (i) ordinary least squares (OLS), (ii) Bartik instruments that use a particular accounting identity to weight the set of available instruments, (iii) two-stage least squares (2SLS), (iv) and (v) inverting the AR test assuming a fixed number of instruments (k) or many instruments (MI), (vi) and (vii) inverting the score test assuming a fixed number of instruments (k) or many instruments (MI). As expected the identification-robust confidence intervals are wider relative to their non-robust counterparts. The confidence intervals based on the score are disjoint, which is not uncommon for this type of test. Most importantly for our purposes is the narrowing of the confidence intervals when using the developed many-instrument approximations. In the left panel, the width of the confidence interval based on inverting the AR test is reduced by 25%. For the score based interval, even if we only focus on the central part of the confidence interval of the fixed-k approximation, the width is reduced by 19%. In the right panel, the width is reduced with 28% for the AR based confidence intervals, and 18% for the score based interval. Further details are given in Section 7.

As a second application, we briefly revisit the canonical example in the many-instrument literature that studies the return of education using the quarter-of-birth dummy proposed by Angrist and Krueger (1991). As suggested by Mikusheva and Sun (2022), in this application the dimension of the instrument set can be increased up to 1530 instruments by interacting the quarter-of-birth instrument with various fixed effects. For the AR test we find a small reduction in the width of the confidence intervals when moving from the assumption that the number of instruments is fixed to the many-instrument robust approximation. For the score test on the other hand we observe no narrowing of the confidence interval.

We further assess the finite sample performance of the tests in a simulation that highlights the effect of (i) the identification strength, (ii) the number of instruments, (iii) conditional heteroskedasticity and (iv) robustness to deviations from the invariance assumption. The simulation shows that unlike conventional asymptotic approximations, the many instrument identification robust tests have close to nominal size control regardless of the instrument strength and regardless of the number of instruments. As the theory suggests, the procedures developed for a fixed number of instruments get progressively more conservative when the number of instruments increases relative to the sample size. Moreover, we find that the developed procedures are robust to small deviations from the assumed reflection invariance. This raises the important question whether the results continue to be valid under a well-defined approximate reflection invariance condition, similar in spirit to the ideas of Canay, Romano, and Shaikh (2017) in randomization tests, but we have no theoretical results in this regard.

In the appendix we also compare the power properties with an alternative approach that alters the weighting matrix to no longer depend on the regression errors. This approach is recently proposed by Mikusheva and Sun (2022) and Crudu, Mellace, and Sándor (2021) for the AR statistic and Matsushita and Otsu (2022) for the score-based statistic. For the AR statistic, the difference in power appears to depend on the sign of the true coefficient on the endogenous variable. For the score statistic, we find that the alternative weighting matrix is generally suboptimal in terms of power when identification is weak, while differences are small when identification is strong.

Related literature Many instrument sequences can be traced back to Kunitomo (1980) and Morimune (1983). Bekker (1994) shows that in a homoskedastic IV model with normally distributed errors and strong instruments, the two-stage least squares estimator is inconsistent under many instruments. The limited information maximum likelihood estimator remains consistent, but the presence of many instruments changes the asymptotic variance. Hansen, Hausman, and Newey (2008) extend these results by removing the normality assumption. Anatolyev (2019) provides a survey of the literature on many instruments.

The consistency of the limited information maximum likelihood estimator is lost under heteroskedasticty, with the exception of balanced group structures as in Bekker and van der Ploeg (2005). Estimators and specification tests that remain consistent under many instruments and heteroskedasticity were developed by Hausman, Newey, Woutersen, Chao, and Swanson (2012), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Chao, Hausman, Newey, Swanson, and Woutersen (2014) and Bekker and Crudu (2015). The key idea is to explicitly remove the terms in the LIML objective function that cause the inconsistency under heteroskedasticity, leading to various jackknife estimators. In this sense it is not surprising that continuous updating is useful under heteroskedasticity given the jackknife interpretation by Donald and Newey (2000).

When instruments are weak or even irrelevant, the parameters of interest cannot be estimated consistently, and the focus shifts to inference procedures that guarantee size control uniformly over the strength of the instruments. In homoskedastic linear IV models, such identification-robust inference is commonly based on (i) the Anderson-Rubin statistic (Anderson and Rubin, 1949) that is a scaled version of the LIML objective function, (ii) statistics based on the score of this objective function (Kleibergen, 2002), or (iii) a combination of (i) and (ii) as in the conditional likelihood-ratio (CLR) test (Moreira, 2003). The CLR test is particularly attractive as it provides near optimal power (Andrews, Marmer, and Yu, 2019). Under heteroskedasticity, inference can be based on the continuous updating objective function, its score (Kleibergen, 2005) or generally more powerful conditional test statistics (Andrews, 2016).

Allowing many instruments to be potentially weak can be done through what is called many weak instrument sequences developed by Chao and Swanson (2005) and Stock and Yogo (2005). Such sequences are crucially different from many instrument sequences as they restrict the number of instruments to increase at a slower rate relative to the sample size. A combination of many instrument sequences with identification robust statistics was considered by Bekker and Kleibergen (2003) in the context of the homoskedastic Gaussian IV model. This shows that under many instrument sequences the score-based statistic by Kleibergen (2002) needs to be scaled to obtain the familiar χ^2 limiting distribution.

Finally, the combination of robust inference in heteroskedastic linear IV models with many instruments that is considered in this paper has been studied recently by Crudu, Mellace, and Sándor (2021), Mikusheva and Sun (2022), Matsushita and Otsu (2022) and Lim, Wang, and Zhang (2022). Instead of using the continuous updating objective function, these papers change the objective function by using the weighting matrix from the homoskedastic set-up. Critical values for the resulting AR statistic can then be derived that yield a valid test even under heteroskedasticity. Using the homoskedastic weighting matrix, Matsushita and Otsu (2022) propose a score-based statistic that is also identification and many instrument robust under heteroskedasticity. Lim et al. (2022) consider a conditional linear combination of the squared jackknife AR statistic and an orthogonalized LM statistic with critical values from a minimum regret approach as in Andrews (2016). A formal comparison of the results under various weighting matrices is an important avenue for further research.

Invariance properties can open up a route to exact finite sample inference via randomization tests (Lehmann and Romano, 2005; Bekker and Lawford, 2008; Canay et al., 2017). In special cases, invariance can even be used to derive the exact finite sample distribution, e.g. the t-statistic has a Student's t-distribution under rotational invariance (Fisher, 1925). In other cases, the distribution must be simulated by drawing transformations from the invariance group. For a recent example of such randomization inference in economics, see Young (2019). In our setting, one could indeed simulate the exact finite sample distribution of the AR statistic, be it at substantial computational costs. However, this does not appear to be case for the score, which depends on the first stage errors and the covariance between the first and second stage errors, both of which are unknown. The invariance of the second stage errors does allow us to find the asymptotic distribution of the AR and score. Invariance properties to find limiting distributions have been used before in the symmetrization of empirical processes (van der Vaart and Wellner, 1996). In special cases the results for the invariant process extend to more general processes that do not satisfy the invariance property. We leave it for further research whether this is the case for the model considered in this paper.

Structure In Section 2 we discuss the heteroskedastic IV model and the CU objective function. Two invariance conditions and their implications for the distribution of the AR statistic are discussed in Section 3. Section 4 focuses on results for the score. Section 5 provides the variance estimators required to implement the tests and discusses their consistency. Section 6 contains the Monte Carlo results. The empirical applications are given in Section 7. Section 8 concludes.

Notation For a vector $\boldsymbol{v}$, denote by $\boldsymbol{D}_v$ the diagonal matrix with $\boldsymbol{v}$ on its diagonal. Moreover, for a square matrix $\boldsymbol{A}$, let $\boldsymbol{D}_A = \boldsymbol{A} \odot \boldsymbol{I}$, where $\odot$ is the Hadamard product. We use $\dot{\boldsymbol{A}} = \boldsymbol{A} - \boldsymbol{D}_A$ for a matrix with all diagonal elements equal to zero. Projection matrices are denoted as $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. $\boldsymbol{\iota}$ indicates a vector of ones and $\boldsymbol{e}_i$ a vector with its i^{th} entry equal to one and the remaining entries equal to zero. Let $\boldsymbol{a}_{(h)} = \boldsymbol{A}\boldsymbol{e}_h$ denote the h^{th} column of a matrix $\boldsymbol{A}$. For random variables A and B, $A \overset{(d)}{=} B$ means that A is distributionally equivalent to B. $A \overset{(E)}{=} B$ means that $\mathrm{E}[A] = \mathrm{E}[B]$. $\mathrm{E}_A[\cdot]$ is the expectation over the distribution of the random variable A. $\to_d$ denotes convergence in distribution, $\to_p$ convergence in probability and $\to_{a.s.}$ almost sure convergence. $a.s.n.$ is short for with probability 1 for all n sufficiently large. For a symmetric $n \times n$ matrix $\boldsymbol{A}$, $\lambda_{\min}(\boldsymbol{A}) = \lambda_1(\boldsymbol{A}) \leq \cdots \leq \lambda_n(\boldsymbol{A}) = \lambda_{\max}(\boldsymbol{A})$ denote its eigenvalues. C denotes a generic finite positive constant that can differ between appearances. We tacitly assume $C > 1/C$ if necessary.

2 Continuous updating and the heteroskedastic linear IV model

While our results on the AR statistic apply to any setting with reflection invariant moment conditions, our main focus is the heteroskedastic linear IV model. The model has p endogenous regressors with both the first and second stage exactly linear,

$$
\begin{aligned}
y_i &= \boldsymbol{x}_i'\boldsymbol{\beta}_0 + \varepsilon_i, \\
\boldsymbol{x}_i &= \boldsymbol{\Pi}'\boldsymbol{z}_i + \boldsymbol{\eta}_i \\
&= \bar{\boldsymbol{z}}_i + \boldsymbol{\eta}_i,
\end{aligned} \tag{1}
$$

with $\boldsymbol{\beta}_0$ a $p \times 1$ vector, $\boldsymbol{\Pi}$ a $k \times p$ matrix, and $i = 1, \ldots, n$. We denote $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, and $\bar{\boldsymbol{Z}} = (\bar{\boldsymbol{z}}_1, \ldots, \bar{\boldsymbol{z}}_n)'$. We also introduce the following notation that will be convenient below: for some $\boldsymbol{\beta}$, not necessarily equal to $\boldsymbol{\beta}_0$, $\varepsilon_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = (\varepsilon_1(\boldsymbol{\beta}), \ldots, \varepsilon_n(\boldsymbol{\beta}))'$.

The model (1) is accompanied by the following assumptions.

Assumption A1. *(a) Conditional on $\boldsymbol{Z}$, $\{\varepsilon_i, \boldsymbol{\eta}_i'\}_{i=1}^n$ is independent, with mean zero and $\mathrm{E}[(\varepsilon_i, \boldsymbol{\eta}_i')'(\varepsilon_i, \boldsymbol{\eta}_i')|\boldsymbol{Z}] = \boldsymbol{\Sigma}_i = (\sigma_i^2 \;\; \boldsymbol{\sigma}_{12i}'; \;\; \boldsymbol{\sigma}_{12i} \;\; \boldsymbol{\Sigma}_{22i})$, (b) $0 < C^{-1} \leq \lambda_{\min}(\boldsymbol{\Sigma}_i) \leq \lambda_{\max}(\boldsymbol{\Sigma}_i) \leq C < \infty$ a.s., (c) For all i, $\mathrm{E}[\varepsilon_i^4|\boldsymbol{Z}] \leq C < \infty$ a.s. and $\mathrm{E}[\|\boldsymbol{\eta}_i\|^4|\boldsymbol{Z}] \leq C < \infty$ a.s.*

To conduct inference on $\boldsymbol{\beta}_0$, we have k moment conditions $\boldsymbol{g}_i(\boldsymbol{\beta})$ that conditionally on $\boldsymbol{Z}$ are independent across i and satisfy $\mathrm{E}[\boldsymbol{g}_i(\boldsymbol{\beta}_0)] = \mathrm{E}[\boldsymbol{z}_i\varepsilon_i] = \boldsymbol{0}$. We stack the moment conditions in the $n \times k$ matrix $\boldsymbol{G}(\boldsymbol{\beta}) = [\boldsymbol{g}_1(\boldsymbol{\beta}), \ldots, \boldsymbol{g}_n(\boldsymbol{\beta})]'$, with $\mathrm{rank}(\boldsymbol{G}(\boldsymbol{\beta}_0)) = k$. Define the projector $\boldsymbol{P}(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta})(\boldsymbol{G}(\boldsymbol{\beta})'\boldsymbol{G}(\boldsymbol{\beta}))^{-1}\boldsymbol{G}(\boldsymbol{\beta})'$. The continuous updating (CU) objective function introduced by Hansen et al. (1996)

can be written as

$$Q(\boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{\iota}'\boldsymbol{P}(\boldsymbol{\beta})\boldsymbol{\iota} = \frac{k}{n} + \frac{1}{n}\sum_{i\neq j}[\boldsymbol{P}(\boldsymbol{\beta})]_{ij}. \tag{2}$$

The minimizer of (2) is the continuous updating estimator (CUE), which Donald and Newey (2000) show has a jackknife interpretation. Newey and Windmeijer (2009) show that the estimator is asymptotically normal when the number of instruments grows slowly in the sense that $k^3/n \to 0$.

The CU objective function is closely related to the Anderson-Rubin GMM (abbreviated as AR) statistic, defined as

$$\mathrm{AR}(\boldsymbol{\beta}) = nQ(\boldsymbol{\beta}). \tag{3}$$

For a fixed number of instruments k, the AR statistic is asymptotically $\chi^2(k)$ distributed when evaluated at $\boldsymbol{\beta}_0$. Extending this result to the case where the number of moment conditions grows proportionally with the sample size is challenging. Specializing to the linear IV model (1), the CU objective function is

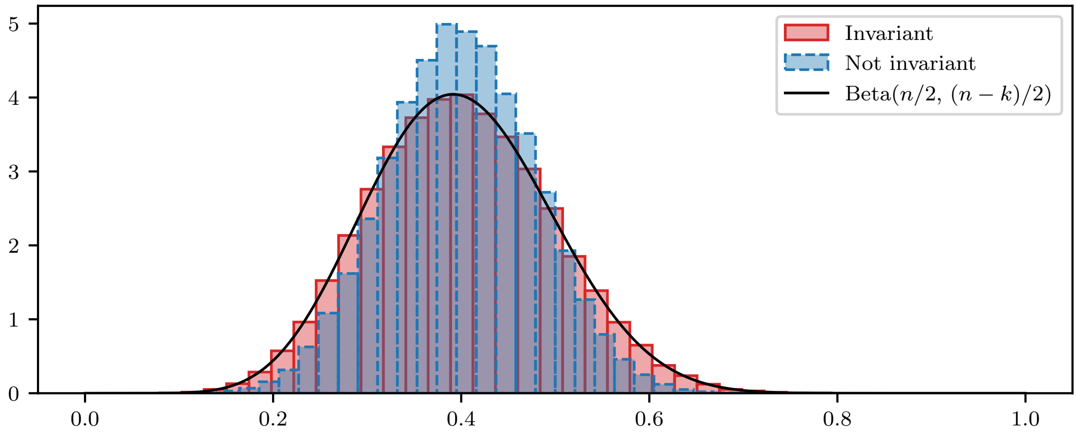
$$Q(\boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{\iota}'\boldsymbol{D}_{\varepsilon(\boldsymbol{\beta})}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}^2_{\varepsilon(\boldsymbol{\beta})}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon(\boldsymbol{\beta})}\boldsymbol{\iota}. \tag{4}$$

The weighting matrix $\boldsymbol{Z}'\boldsymbol{D}^2_{\varepsilon(\boldsymbol{\beta})}\boldsymbol{Z}$ is $k \times k$ dimensional and contains the second stage regression errors $\boldsymbol{\varepsilon}$. This combination makes the behavior of this weighting matrix challenging to control when k is a non-negligible fraction of the sample size.

3 Invariant moment conditions

To motivate the use of invariance conditions to obtain the distribution of test statistics, consider the setting where the moment conditions are orthogonally invariant, i.e. $\boldsymbol{G}(\boldsymbol{\beta}_0) \overset{(d)}{=} \boldsymbol{G}(\boldsymbol{\beta}_0)\boldsymbol{Q}$ for any orthogonal matrix $\boldsymbol{Q}$. In this case, the finite sample distribution of $Q(\boldsymbol{\beta}_0)$ can be obtained as follows. Orthogonal invariance implies that $\boldsymbol{\iota}'\boldsymbol{P}(\boldsymbol{\beta}_0)\boldsymbol{\iota}/n \overset{(d)}{=} \boldsymbol{z}'\boldsymbol{P}(\boldsymbol{\beta}_0)\boldsymbol{z}$ where $\boldsymbol{z}$ is uniformly distributed over the $(n-1)$-dimensional unit sphere and $\boldsymbol{P}(\boldsymbol{\beta}_0)$ can be regarded as fixed, see e.g. Vershynin (2018, Chapter 5). As a result, $Q(\boldsymbol{\beta}_0) \overset{(d)}{=} (Z_1 + Z_2)^{-1}Z_1$ where $Z_1 \sim \chi^2(k)$ independently of $Z_2 \sim \chi^2(n-k)$, and hence, $Q(\boldsymbol{\beta}_0) \sim \mathrm{Beta}(k/2, (n-k)/2)$ or equivalently $(n-k)Q(\boldsymbol{\beta}_0)/[k(1-Q(\boldsymbol{\beta}_0))] \sim \mathrm{F}(k, n-k)$. Orthogonal invariance thus allows us to bypass the fact that the dimensions of the weighting matrix

Figure 2: Empirical PDF of $Q(\beta_0)$ with and without orthogonal invariance.



Note: empirical PDF obtained by simulating 100,000 draws of $Q(\beta_0) = \boldsymbol{\iota}'\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{\iota}/n$ with $\boldsymbol{G}$ an $n \times k$ matrix with $n = 50$ and $k = 20$. The elements $[\boldsymbol{G}]_{ij} \sim N(0,1)$ and $[\boldsymbol{G}]_{ij} = \varepsilon_i(\beta_0)Z_{ij}$ with $\varepsilon_i(\beta_0) \sim N(0,1)$ and independently $Z_{ij} \sim N(0,1)$. Dark solid line indicates the PDF of the Beta$(k/2,(n-k)/2)$ distribution.

$\boldsymbol{G}(\boldsymbol{\beta}_0)'\boldsymbol{G}(\boldsymbol{\beta}_0)$ are nonnegligible relative to the sample size. Figure 2 illustrates this result by showing the empirical PDF of $Q(\boldsymbol{\beta}_0)$ over 100,000 draws with $n = 50$, $k = 20$ and we ensure orthogonal invariance by drawing the elements of $\boldsymbol{G}$ as independent standard normal random variables.

Unfortunately, orthogonal invariance is restrictive. In particular, combined with independence of the moment conditions, it implies that the moment conditions are normally distributed. Even in a highly stylized set-up where we have an IV model with second stage errors ε_i and the instruments Z_{ij} generated as independent standard normals, orthogonal invariance breaks down as $[\boldsymbol{G}]_{ij} = \varepsilon_i Z_{ij}$. The blue dashed histogram in Figure 2 shows that in this case the empirical PDF differs from the Beta PDF.

The class of allowed distributions can be substantially enlarged by the following invariance assumption, referred to as orthant symmetry by Efron (1969) and reflection invariance by Bekker and Lawford (2008). In the context of the IV model in (1), we impose the invariance on the second stage regression errors ε_i.

Assumption A2. *Let $\{r_i\}_{i=1}^n$ be a sequence of independent Rademacher random variables and $\boldsymbol{r} = (r_1, \ldots, r_n)'$. Then, conditional on $\boldsymbol{Z}$, $\boldsymbol{\varepsilon} \stackrel{(d)}{=} \boldsymbol{D}_r\boldsymbol{\varepsilon}$.*

This assumption implies reflection invariance in the distribution of the moment conditions as conditional on $\boldsymbol{Z}$, $\boldsymbol{G}(\boldsymbol{\beta}_0) \stackrel{(d)}{=} \boldsymbol{D}_r\boldsymbol{G}(\boldsymbol{\beta}_0)$. The results in this section therefore apply to any application where the distribution of the moment conditions is reflection invariant. The assumption is substantially weaker than assumption that the errors are homoskedastic and Gaussian under which many of the previous results for identification robust tests with many moment conditions have been derived (Bekker, 1994; Bekker and Kleibergen, 2003). The key observation is that Assumption A2 allows the distribution of the moment conditions to differ across observations. This makes it particularly suitable to use in the context of heteroskedastic models.

Under Assumption A2 we can relate the distribution of the CU objective function with a similar function written in terms of Rademacher random variables. Conditional on $\boldsymbol{Z}$,

$$Q(\boldsymbol{\beta}_0) \overset{(d)}{=} Q_r(\boldsymbol{\beta}_0) = \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}(\boldsymbol{\beta}_0)\boldsymbol{r}.$$

While the exact finite sample distribution of $Q(\boldsymbol{\beta}_0)$ is no longer tractable, the asymptotic distribution under many instrument sequences is. As conditional on $\boldsymbol{Z}$, $Q(\boldsymbol{\beta}_0)$ and $Q_r(\boldsymbol{\beta}_0)$ are distributionally equivalent, it suffices to analyze the asymptotic distribution $Q_r(\boldsymbol{\beta}_0)$. Likewise, we analyze $\text{AR}_r(\boldsymbol{\beta}_0) = nQ_r(\boldsymbol{\beta}_0)$ to establish the asymptotic distribution of the AR statistic defined in (3).

Conditioning now on $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$, the only randomness in $\text{AR}_r(\boldsymbol{\beta}_0)$ comes from the Rademacher random variables. Under the following assumptions, we can directly apply the CLT for bilinear forms by Chao et al. (2012) to obtain the asymptotic distribution of the AR statistic under many instrument sequences.

Assumption A3. *Conditional on $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$ we have with probability 1 for all sufficiently large n: (a)* $\text{rank}[\boldsymbol{P}(\boldsymbol{\beta}_0)] = k$, *(b)* $P_{ii}(\boldsymbol{\beta}_0) \leq C < 1$ *for* $i = 1, \ldots, n$, *(c)* $\sigma_n^2 > 1/C$ *where*

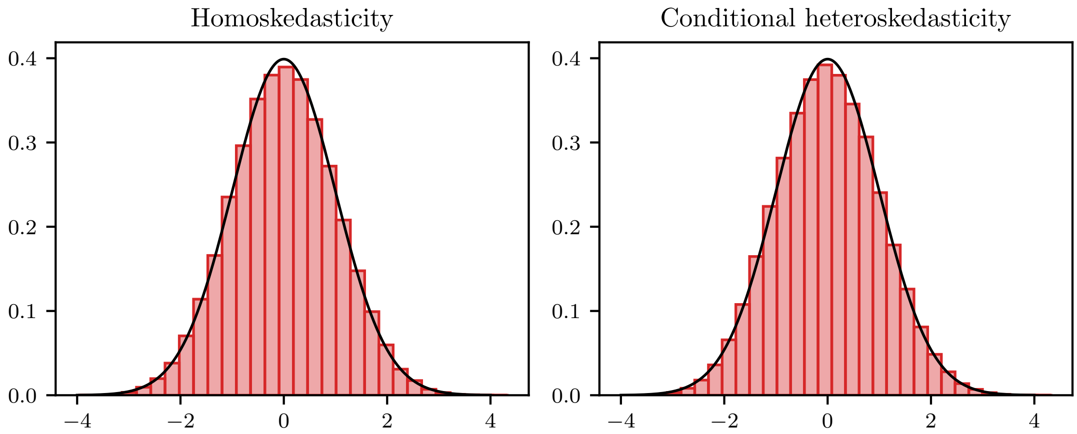
$$\sigma_n^2 = \frac{2}{k}\sum_{i \neq j} P_{ij}(\boldsymbol{\beta}_0)^2. \tag{5}$$

Part (a) excludes any redundant moment conditions. Part (b) is common in the many instruments literature, see e.g. Hausman et al. (2012), Bekker and Crudu (2015) and Anatolyev (2019). Part (c) bounds the variance of the scaled AR statistic away from zero and is required to apply the central limit theorem provided in Lemma A2 by Chao et al. (2012). The following result follows directly from that Lemma and its proof.

Theorem 1. *Under Assumptions A1 to A3 and conditional on* $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$, *when* $k \to \infty$ *as* $n \to \infty$, $(k\sigma_n^2)^{-1/2}(\text{AR}_r(\boldsymbol{\beta}_0) - k) \to_d N(0,1)$ *a.s. This implies that* $(k\sigma_n^2)^{-1/2}(\text{AR}(\boldsymbol{\beta}_0) - k) \to_d N(0,1)$.

The assumption of finite moments for the errors and the eigenvalue bounds for their variance in Assumption A1 and the technical assumption on $\boldsymbol{P}(\boldsymbol{\beta}_0)$ in

Figure 3: Empirical PDF of the centered and scaled AR statistic.



Note: empirical PDF of $(k\sigma_n^2)^{-1/2}(\text{AR}(\beta_0)-k)$ as defined in Theorem 1 obtained by simulating 100,000 draws of the $n \times k$ moment condition matrix $\boldsymbol{G}$ with $n = 50$ and $k = 20$. The elements $[\boldsymbol{G}]_{ij} = \varepsilon_i(\beta_0)Z_{ij}$ with $\varepsilon_i(\beta_0) \sim N(0,1)$ and independently $Z_{ij} \sim N(0,1)$ (left panel) and $\varepsilon_i \sim N(0, Z_{1i}^2)$ (right panel). Dark solid line indicates the PDF of the standard normal distribution.

Assumption A3 made in Theorem 1 are comparable to the assumptions Crudu et al. (2021) and Mikusheva and Sun (2022) need to obtain the distribution of the jackknife AR statistic.

Theorem 1 shows that the AR statistic needs to be shifted and scaled to have a well-defined asymptotic distribution. A similar result is obtained by Anatolyev and Gospodinov (2011) for the AR statistic in a homoskedastic IV model with many instruments. We note that Theorem 1 applies in a general GMM set-up where the moment conditions are reflection invariant, as we make no use of the particulars of the linear IV model (1), but only exploit the invariance in the moment conditions. While Theorem 1 requires $k \to \infty$, we can achieve uniform inference across k by testing based on the quantiles of the distribution of $Z = (2k)^{-1/2}(Z_1 - k)$ where $Z_1 \sim \chi^2(k)$. When k is fixed, $\sigma_n^2 \to_p 2$, and hence, we compare $\text{AR}(\boldsymbol{\beta}_0)$ against the quantiles of a $\chi^2(k)$ distribution. When k increases, the quantiles of Z approach that of the standard normal distribution and Theorem 1 applies.

To illustrate Theorem 1, we revisit the setting from Figure 2 with $n = 50$ observations, $k = 20$ instruments and $[\boldsymbol{G}]_{ij} = \varepsilon_i Z_{ij}$ with ε_i and Z_{ij} independent standard normal random variables. The left panel of Figure 3 shows that the asymptotic distribution derived in Theorem 1 provides an accurate finite sample realization. In the right panel, we introduce conditional heteroskedasticity through $\varepsilon_i \sim N(0, Z_{i1}^2)$. We hardly observe any effect on the finite sample distribution.

4 Inference based on the score

Since the AR test is known not to be efficient in overidentified models, we now consider the application of Assumption A2 in the linear IV model to analyze a test statistic based on the score of the CU objective function given in (4). To obtain the limiting distribution of the first order conditions of the CU objective function, we make the following assumption on the IV model in (1).

Assumption A4. *Consider* $\boldsymbol{\eta}_i$ *and* ε_i *as in* (1)*. Then,* $\boldsymbol{\eta}_i = \varepsilon_i \boldsymbol{a}_i + \boldsymbol{u}_i$*, where* $\boldsymbol{a}_i = \boldsymbol{\sigma}_{21i}/\sigma_i^2$*, and, conditional on* $\boldsymbol{Z}$*,* $\{\boldsymbol{u}_i, \varepsilon_i\}$ *are mutually independent.*

This assumption parameterizes the relation between the first and second stage errors and also appears in Bekker and Kleibergen (2003). It is for example satisfied if $(\varepsilon_i, \boldsymbol{\eta}_i')$ is multivariate normal. We use Assumption A4 to write

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}}_i + \varepsilon_i \boldsymbol{a}_i, \quad \bar{\boldsymbol{x}}_i = \bar{\boldsymbol{z}}_i + \boldsymbol{u}_i. \tag{6}$$

The fact that $\bar{\boldsymbol{x}}_i$ does not depend on ε_i is useful when applying Assumption A2. The flexibility of the model can be increased by including higher-order polynomials of the errors ε_i in Assumption A4 at the expense of more elaborate notation. The assumption on the eigenvalues of the second moment matrix of $(\varepsilon_i, \boldsymbol{\eta}_i')$ in Assumption A1 implies that the eigenvalues of $\boldsymbol{\Sigma}_i^U = \mathrm{E}[\boldsymbol{u}_i \boldsymbol{u}_i' | \boldsymbol{Z}]$ are bounded from above and below by positive constants and that $\mathrm{E}[\|\boldsymbol{u}_i\|^4 | \boldsymbol{Z}] \leq C < \infty$ *a.s.*

Denote $\boldsymbol{V}(\boldsymbol{\beta}) = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon(\beta)^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. To simplify the notation, we write $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\beta}_0)$ and likewise $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{\beta}_0)$. The score of the CU objective function is

$$S_{(i)}(\boldsymbol{\beta}) = \frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i} = -\frac{1}{n}\boldsymbol{x}_{(i)}'(\boldsymbol{I} - \boldsymbol{D}_{P(\boldsymbol{\beta})\iota})\boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{\varepsilon}(\boldsymbol{\beta}).$$

Under Assumption A2, and using the decomposition of $\boldsymbol{\eta}_i$ in Assumption A4 we find that, conditional on $\boldsymbol{Z}$, $S_{(i)}(\boldsymbol{\beta}_0) \overset{(d)}{=} S_{(i),r}(\boldsymbol{\beta}_0)$, where with $\bar{\boldsymbol{x}}_{(i)}$ as in (6),

$$\begin{aligned} S_{(i),r}(\boldsymbol{\beta}_0) = -\frac{1}{n}\bar{\boldsymbol{x}}_{(i)}'\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{r} \\ -\frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}. \end{aligned} \tag{7}$$

The score consists of one linear term, two quadratic terms and one cubic term. Our strategy is to derive the asymptotic distribution of $S_{(i),r}(\boldsymbol{\beta}_0)$ to obtain the distribution of $S_{(i)}(\boldsymbol{\beta}_0)$. As for the AR statistic, we derive the limiting distribution conditional on $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$. For the score, this implies that we do need to take into account the randomness that enters via $\boldsymbol{u}_i$.

4.1 Conditional variance of the score

We start by finding the conditional variance of (7) that is decomposed in two parts. First, the conditional expectation of the estimator for the variance of the score as in Kleibergen (2005) after applying Assumption A2,

$$\Omega^L_{ij}(\boldsymbol{\beta}_0) = \frac{1}{n}\,\mathrm{E}[(\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_r\boldsymbol{\varepsilon})'(\boldsymbol{I} - \boldsymbol{D}_r\boldsymbol{D}_{Pr})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{D}_r\boldsymbol{D}_{Pr})(\bar{\boldsymbol{x}}_{(j)} + \boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_r\boldsymbol{\varepsilon})|\mathcal{J}].$$

Second, we find correction terms that are relevant under many-instrument sequences. The following matrix appears under nonzero identification strength. Define $S_{ij} = P_{ii} + P_{jj}$. Then for $i \neq j$,

$$\begin{aligned} [\boldsymbol{V} \odot \boldsymbol{W}]_{ij} &= V_{ij}\big[(P_{ii}P_{jj} + P_{ij}^2)(3 - 4S_{ij}) - 2S_{ij} + 2S_{ij}^2\big], \\ [\boldsymbol{V} \odot \boldsymbol{W}]_{ii} &= -2V_{ii}P_{ii}(1 - 2P_{ii}) - 2\sum_{k=1}^{n} V_{ik}^2\varepsilon_k^2 P_{ik}^2. \end{aligned} \tag{8}$$

Using this notation, we have the following theorem.

Theorem 2. *For* $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$ *and under Assumption A2,* $\mathrm{E}[S_{(i),r}(\boldsymbol{\beta}_0)|\mathcal{J}] = 0$*. The* (i,j)*-th element of the conditional variance matrix is*

$$\Omega_{ij}(\boldsymbol{\beta}_0) = \mathrm{E}\left[n \cdot S_{(i),r}(\boldsymbol{\beta}_0)S_{(j),r}(\boldsymbol{\beta}_0)\big|\mathcal{J}\right] = \Omega^L_{ij}(\boldsymbol{\beta}_0) + \Omega^H_{ij}(\boldsymbol{\beta}_0), \tag{9}$$

where $\Omega^L_{ij}(\boldsymbol{\beta}_0) = \Omega^{L,z}_{ij}(\boldsymbol{\beta}_0) + \Omega^{L,a}_{ij}(\boldsymbol{\beta}_0) + \Omega^{L,u}_{ij}(\boldsymbol{\beta}_0)$ *with*

$$\begin{aligned} \Omega^{L,z}_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'[(\boldsymbol{I} - \boldsymbol{D}_P)\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{D}_P) + \boldsymbol{D}_P\boldsymbol{D}_V(\boldsymbol{I} - 2\boldsymbol{D}_P) + \boldsymbol{V} \odot \boldsymbol{P} \odot \boldsymbol{P}]\bar{\boldsymbol{z}}_{(j)}, \\ \Omega^{L,a}_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{a}_{(i)}'(\boldsymbol{D}_P - \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{a}_{(j)}, \\ \Omega^{L,u}_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_V(\boldsymbol{I} - \boldsymbol{D}_P)), \end{aligned}$$

and $\Omega^H_{ij}(\boldsymbol{\beta}_0) = \Omega^{H,z}_{ij}(\boldsymbol{\beta}_0) + \Omega^{H,a}_{ij}(\boldsymbol{\beta}_0) + \Omega^{H,u}_{ij}(\boldsymbol{\beta}_0)$ *with*

$$\begin{aligned} \Omega^{H,z}_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'(\boldsymbol{V} \odot \boldsymbol{W})\bar{\boldsymbol{z}}_{(j)}, \\ \Omega^{H,a}_{ij}(\boldsymbol{\beta}_0) &= -\frac{2}{n}\boldsymbol{a}_{(i)}'(\boldsymbol{D}_P - \boldsymbol{P} \odot \boldsymbol{P})^2\boldsymbol{a}_{(j)}, \\ \Omega^{H,u}_{ij}(\boldsymbol{\beta}_0) &= -\frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{I} - 2\boldsymbol{D}_P + \boldsymbol{P} \odot \boldsymbol{P})), \end{aligned}$$

with $(\boldsymbol{W} \odot \boldsymbol{V})$ *defined in (8) and* $\boldsymbol{D}_{\Sigma^U(i,j)}$ *an* $n \times n$ *diagonal matrix with the* k*-th diagonal element equal to* $\mathrm{cov}(u_{ki}, u_{kj}|\mathcal{J})$*.*

Proof. See Appendix A.2. □

Theorem 2 does not involve any asymptotic approximation, but gives the exact conditional variance of the score under Assumption A2. The components $\Omega_{ij}^{L}(\boldsymbol{\beta}_0)$ and $\Omega_{ij}^{H}(\boldsymbol{\beta}_0)$ are split into contributions due to the instruments, due to heteroskedasticity in the relation between the first and second stage errors, and finally due to the second stage errors themselves. The terms involving $\boldsymbol{a}_{(i)}$ and $\boldsymbol{a}_{(j)}$ cancel when $\boldsymbol{a}_{(i)} = a_{(i)} \cdot \boldsymbol{\iota}$ for all $i = 1, \ldots, p$. This is true under homoskedasticity, but holds more generally. From Assumption A4 we see that even if ε_i and $\boldsymbol{u}_i$ are (conditionally) heteroskedastic, we can still have $\boldsymbol{a}_{(i)} = a_{(i)} \cdot \boldsymbol{\iota}$. The most important property of the correction $\boldsymbol{\Omega}^H(\boldsymbol{\beta}_0)$ is given by the following result.

Corollary 1. *Let (i)* $\min_{i=1,\ldots,n} \lambda_{\min}(\boldsymbol{\Sigma}_i^U) \geq C > 0$*, (ii)* $\max\limits_{i=1,\ldots,n} P_{ii} \leq 0.9$ *and (iii)* $n^{-1}\sum_{i=1}^n V_{ii}P_{ii} > 0$*. Then,* $\boldsymbol{\Omega}^H(\boldsymbol{\beta}_0)$ *from Theorem 2 is negative definite.*

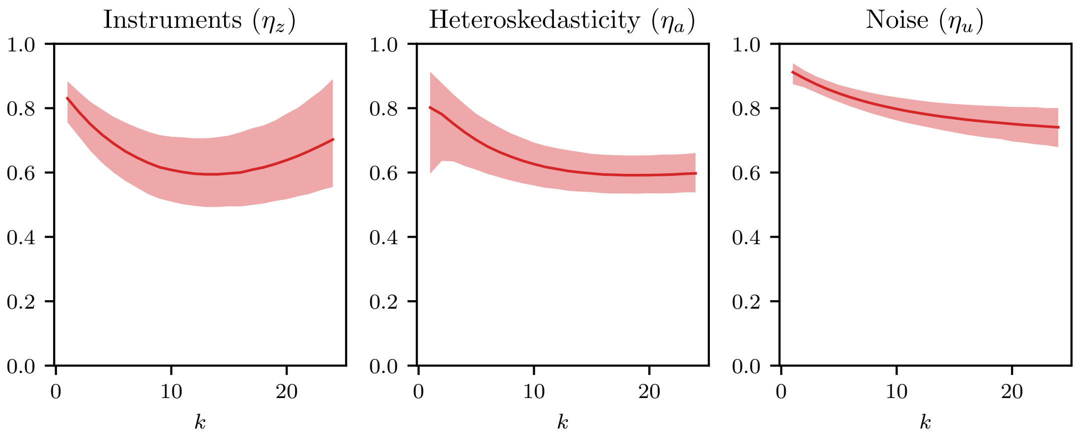
Proof. See Appendix A.3. □

Under Assumption A4, condition (i) is implied by Assumption A1. Condition (ii) ensures that in Theorem 2 $[\boldsymbol{W}]_{ij} \leq 0$ for $i \neq j$, which in turn implies that $\boldsymbol{\Omega}^{H,z}(\boldsymbol{\beta}_0) \preceq \boldsymbol{O}$. Moreover, together with the assumption (iii), it ensures that $\boldsymbol{\Omega}^{H,u}(\boldsymbol{\beta}_0) \prec \boldsymbol{O}$. If we drop conditions (i) and (iii), the result holds with negative definite replaced by negative semidefinite.

Corollary 1 shows that the variance correction is negative definite. This implies that the use of the conventional inference procedures based on the score will be conservative. The condition that $n^{-1}\sum_{i=1}^n V_{ii}P_{ii} > 0$ makes it clear that this is more likely to occur, and asymptotically only occurs, when the number of instruments is a non-negligible fraction of the sample size. For instance, if $\{\varepsilon_i\}_{i=1}^n$ is itself a sequence of Rademacher variables and $\boldsymbol{Z}$ has independent elements with finite fourth moment. Bai et al. (2007) show that $\frac{1}{n}\sum_{i=1}^n V_{ii}P_{ii} - \frac{k^2}{n^2} \to_{a.s.} 0$, and hence asymptotically (iii) requires that $k/n \to \lambda > 0$.

We revisit the numerical example from Section 3 to quantify the reduction in the conditional variance when using the conditional variance from Theorem 2, versus the conditional expectation of the conventional variance estimator. We consider the ratios $\eta_z = \left[\Omega_{ij}^{L,z}(\boldsymbol{\beta}_0) + \Omega_{ij}^{H,z}(\boldsymbol{\beta}_0)\right]/\Omega_{ij}^{L,z}(\boldsymbol{\beta}_0)$, that contains the terms due to the signal in the instruments, as well as the analogously defined ratios η_a that is due to heteroskedasticity, and η_u that is due to the independent (noise) component in the second stage errors. We now specify there is $p = 1$ endogenous variable, the number of instruments ranges from $k = 1, \ldots, n/2$, and the instrument slope coefficients are $\boldsymbol{\Pi} = \boldsymbol{e}_1$. Since we study the ratio η_z, this can be scaled

Figure 4: Reduction in the components of the conditional variance.



Note: depicted is the ratio $\eta_z = \hat{\mathrm{E}}[(\Omega^{L,z}(\beta_0) + \Omega^{H,z}(\beta_0))/\Omega^{L,z}(\beta_0)]$ as a function of the number of instruments k (left panel), and similarly for η_a (middle panel) and η_u (right panel) with the components of $\Omega^L(\beta_0)$ and $\Omega^H(\beta_0)$ defined in Theorem 2 and $\hat{\mathrm{E}}$ indicates the average over the 10,000 simulation draws. The band is between the 5th percentile and the 95th percentile.

with any nonzero constant without changing the results. We set $\boldsymbol{D}_{\Sigma^U_{(1,1)}} = \boldsymbol{I}$ and $a_{(1),i} = |Z_{i1}|$. Results are based on 10,000 draws.

Figure 4 shows the average of the ratios $\{\eta_z, \eta_a, \eta_u\}$ as well as the 5th percentile and the 95th percentile. We find a substantial reduction in the different components of the conditional variance when taking into account the correction terms contained in $\Omega^H_{ij}(\boldsymbol{\beta}_0)$. This reduction even occurs with a single instrument, although it generally increases when the number of instruments grows.

4.2 Asymptotic results for the AR and score statistics

To describe the joint limiting distribution of the AR statistic and the score, we need the following assumptions.

Assumption A5. *(a)* $\frac{1}{n}\sum_{i=1}^n \|\bar{\boldsymbol{z}}_i\|^2 \leq C < \infty$ *a.s.n., (b)* $\frac{1}{n}\max_{i=1,\dots,n} \|\bar{\boldsymbol{z}}_i\|^2 \to_{a.s.} 0$, *(c)* $\frac{1}{n}\max_{i=1,\dots,n} \|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_\varepsilon \boldsymbol{e}_i\|^2 \to_{a.s.} 0$, *(d)* $0 < C^{-1} \leq \lambda_{\min}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}) \leq \lambda_{\max}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}) \leq C < \infty$ *a.s.n.*, $0 < C^{-1} \leq \lambda_{\min}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}^2_\varepsilon \boldsymbol{Z}) \leq \lambda_{\max}(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}^2_\varepsilon \boldsymbol{Z}) \leq C < \infty$ *a.s.n.*

Part (a) and (b) are standard assumptions under many instruments. Part (a) also appears in Chao et al. (2012) and Hausman et al. (2012), who instead of (b) require $n^{-2}\sum_{i=1}^n \|\bar{\boldsymbol{z}}_i\|^4 \to_{a.s.} 0$. We see that this condition is implied by Assumption A5 parts (a) and (b). In particular, (b) is a Lyapunov condition needed for the central limit theorem we employ. Part (c) is another Lyapunov condition needed for the CLT under heteroskedasticity. Part (d) ensures that $\boldsymbol{V} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ has bounded eigenvalues $a.s.n.$

The joint limiting distribution of the AR statistic and the score evaluated at the true parameter $\boldsymbol{\beta}_0$ is given in the following theorem.

Theorem 3. *Under Assumptions A1 to A5, when $n \to \infty$ and $k/n \to \lambda \in (0, 1)$,*

$$\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)^{-1/2} \begin{pmatrix} \frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}_0) - k) \\ \sqrt{n} \cdot \boldsymbol{S}(\boldsymbol{\beta}_0) \end{pmatrix} \to_d N(\boldsymbol{0}, \boldsymbol{I}_{p+1}).$$

Here $[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{1,1} = \sigma_n^2$ from (5), $[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{2:p+1,2:p+1}$ is given by $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$ in Theorem 2, and the covariance between the rescaled AR statistic and the score is

$$[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{1,j+1} = [\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{j+1,1} = \frac{2}{\sqrt{nk}} \mathrm{tr}(\boldsymbol{\Psi}^{(j)} \odot \boldsymbol{P}), \quad j = 1, \ldots p, \tag{10}$$

with $\boldsymbol{\Psi}^{(j)} = \boldsymbol{M}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}$ and $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$.

Proof. See Appendix B. □

Our proof uses that the eigenvalues of $\boldsymbol{\Sigma}_n$ are bounded away from zero under many instrument sequences. In sharp contrast to the analysis under a fixed number of instruments, this holds even in the unidentified case where $\boldsymbol{\Pi} = \boldsymbol{O}$. This is the reason that under many instruments Theorem 3 does not need to be analyzed separately for different identification strengths as is the case under a fixed number of instruments. We further observe that the covariance between the objective function and the score is only nonzero when the number of instruments increases *and* when there is heteroskedasticity in the sense that $\boldsymbol{a}_{(j)}$ varies across observations. When $\boldsymbol{a}_{(j)}$ is constant across observations, we have $\boldsymbol{D}_{a_{(j)}} = a_{(j)}\boldsymbol{I}_n$ and hence $\boldsymbol{\Psi}^{(j)} = \boldsymbol{O}$.

5 Implementation

5.1 An unbiased and consistent variance estimator

To use Theorems 1 and 3 for hypothesis testing and the construction of asymptotically valid confidence intervals, we require a consistent estimator for $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$. Define,

$$\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}) = \begin{pmatrix} \hat{\sigma}_n^2(\boldsymbol{\beta}) & \left[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta})\right]_{2:p,1}' \\ \left[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta})\right]_{2:p,1} & \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) \end{pmatrix}, \tag{11}$$

with for the variance of the AR statistic $\hat{\sigma}_n^2(\boldsymbol{\beta}) = 2k^{-1}(k - \boldsymbol{\iota}'\boldsymbol{D}_P^2\boldsymbol{\iota})$ and to lighten the notation $\boldsymbol{D}_P$ is short for $\boldsymbol{D}_{P(\beta)}$. To estimate the variance matrix of the score vector $\boldsymbol{\Omega}(\boldsymbol{\beta})$ we follow the decomposition in Theorem 2,

$$\begin{aligned}
\hat{\Omega}_{ij}^L(\boldsymbol{\beta}) &= \frac{1}{n}\boldsymbol{x}_{(i)}'(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{x}_{(j)}, \\
\hat{\Omega}_{ij}^H(\boldsymbol{\beta}) &= \frac{1}{n}\boldsymbol{x}_{(i)}'[7\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2 \\
&\qquad + 3\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P} - 4\boldsymbol{D}_P(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}) - 4(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{D}_P \\
&\qquad - 2\boldsymbol{D}_P\dot{\boldsymbol{V}} - 2\dot{\boldsymbol{V}}\boldsymbol{D}_P + 2\boldsymbol{D}_P^2\dot{\boldsymbol{V}} + 2\dot{\boldsymbol{V}}\boldsymbol{D}_P^2]\boldsymbol{x}_{(j)} \\
&\quad - \frac{2}{n}\boldsymbol{x}_{(i)}'\big[\big((\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P} \odot \boldsymbol{P})\big) \odot \boldsymbol{I}\big]\boldsymbol{x}_{(j)}.
\end{aligned} \tag{12}$$

When $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the final line estimates the terms $\Omega_{ij}^{H,u}(\boldsymbol{\beta}_0)$ and $\Omega_{ij}^{H,a}(\boldsymbol{\beta}_0)$ from Theorem 2, as well as the terms corresponding to $[\boldsymbol{V} \odot \boldsymbol{W}]_{ii}$ in $\Omega_{ij}^{H,z}(\boldsymbol{\beta}_0)$. The preceeding terms in $\hat{\Omega}_{ij}^H(\boldsymbol{\beta})$ estimate the terms corresponding to $[\boldsymbol{V} \odot \boldsymbol{W}]_{ij}$ for $i \neq j$ in $\Omega_{ij}^{H,z}(\boldsymbol{\beta}_0)$. Finally, the covariance between the AR statistic and the jth component of the score given in (10) can be estimated by

$$[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta})]_{1,j+1} = [\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta})]_{j+1,1} = \frac{2}{\sqrt{nk}}\boldsymbol{x}_{(j)}'(\boldsymbol{D}_V - (\boldsymbol{V} \odot \boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon}. \tag{13}$$

The consistency of (11) for $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ is stated in the following result.

Theorem 4. *Under Assumption A2,* $\mathrm{E}[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)|\mathcal{J}] = \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$. *Also, under Assumptions A1 to A5,* $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0) \to_p \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$.

Proof. See Appendix A.5. □

While the estimator is conditionally unbiased and consistent, it is not guaranteed to be positive definite. In a setting where there is only a single endogenous regressor, a crude way of dealing with negative variances is to set the variance equal to zero when this occurs, such that tests using this variance always reject. This is the solution we employ here.

5.2 AR statistic: power relative to fixed-k approximations

With the estimator for the variance, we can use Theorem 3 to perform identification robust inference. To conduct inference based on the AR statistic that is valid regardless of the number of instruments, we obtain a confidence region for

$\boldsymbol{\beta}_0$ with asymptotic coverage rate $1-\alpha$ by including all values for $\boldsymbol{\beta}$ for which

$$(k\sigma_n^2)^{-1/2}(\mathrm{AR}(\boldsymbol{\beta})-k) \leq (2k)^{-1/2}(\chi^2(k)_{1-\alpha}-k), \tag{14}$$

where $\chi^2(k)_{1-\alpha}$ is the $1-\alpha$ quantile of a $\chi^2(k)$ distribution. Using (14) ensures that asymptotically we compare the AR statistic with $\chi^2(k)$ critical values when k is fixed, and that we compare the centered and scaled AR statistic with standard normal critical values when k is large. We have the following result on the probability of rejecting based on the many-instrument approximation versus the fixed-k approximation. Note that this finite-sample result does not contradict the fact that both procedures are asymptotically size correct under a fixed number of instruments.

Corollary 2. *Let* $\phi_1(\boldsymbol{\beta})=1$ *if* $(k\sigma_n^2)^{-1/2}(AR(\boldsymbol{\beta})-k) > (2k)^{-1/2}(\chi^2(k)_{1-\alpha}-k)$, *and let* $\phi_2(\boldsymbol{\beta})=1$ *if* $\mathrm{AR}(\boldsymbol{\beta}) > \chi^2(k)_{1-\alpha}$. *Then, if* $\alpha < 0.3$, $\mathrm{P}(\phi_1(\boldsymbol{\beta})=1) > \mathrm{P}(\phi_2(\boldsymbol{\beta})=1)$.

Proof. After some rewriting we see that $\phi_1(\boldsymbol{\beta})=1$ if

$$AR(\boldsymbol{\beta}) > \chi^2(k)_{1-\alpha} - [\chi^2(1-\alpha)-k]\cdot\left[1-\left(1-k^{-1}\sum_{i=1}^n P_{ii}(\boldsymbol{\beta})^2\right)^{1/2}\right].$$

If $\alpha<0.3$, we have that $\chi^2(k)_{1-\alpha} > k$ for all k. The result follows. □

5.3 Combining the score and AR statistic

Given that the score-based test lacks power in regions away from the true value where the objective function is flat, we also combine the AR and score test. From Theorems 3 and 4 we have

$$\begin{pmatrix}\mathrm{AR}^o(\boldsymbol{\beta}_0)\\ \boldsymbol{S}^o(\boldsymbol{\beta}_0)\end{pmatrix} = \hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)^{-1/2}\begin{pmatrix}\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}_0)-k)\\ \sqrt{n}\cdot\boldsymbol{S}(\boldsymbol{\beta}_0)\end{pmatrix} \to_d N(\boldsymbol{0}, \boldsymbol{I}_{p+1}).$$

Therefore, for a given size α and $\alpha_{\mathrm{AR}}<\alpha$ the test that rejects if either $\mathrm{AR}^o(\boldsymbol{\beta}) > (2k)^{-1/2}(\chi^2(k)_{1-\alpha_{\mathrm{AR}}}-k)$ or if both $\mathrm{AR}^o(\boldsymbol{\beta}) \leq (2k)^{-1/2}(\chi^2(k)_{1-\alpha_{\mathrm{AR}}}-k)$ and $\boldsymbol{S}^o(\boldsymbol{\beta})'\boldsymbol{S}^o(\boldsymbol{\beta}) > \chi^2(k)_{1-\alpha_S}$, with $\alpha_S=(\alpha-\alpha_{\mathrm{AR}})/(1-\alpha_{\mathrm{AR}})$, is size correct under the assumptions of Theorem 3.

In finite samples $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)$ is not guaranteed to be invertible. We implement the combination test by rejecting the test whenever $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)$ is singular.

For the fixed-k linear IV model with homoskedastic normal errors it is known that the CLR test of Moreira (2003) has excellent power properties regardless of instrument strength (Andrews et al., 2006, 2019). Since the CLR statistic can be written as particular combination of the fixed-k AR and score statistic, it is desirable to also combine the many instrument AR and score into a many instrument CLR statistic. Such an extension is non-trivial however, so we leave it for further research.

6 Simulation results

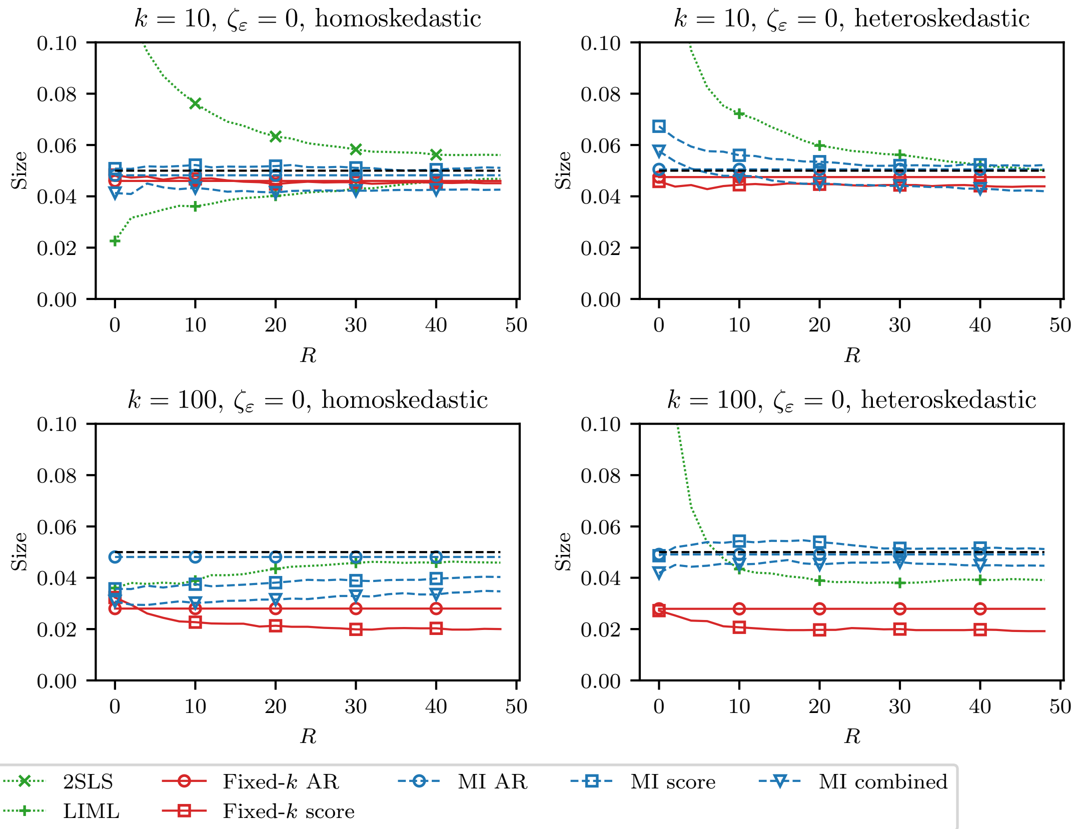
We test the finite sample performance of the proposed tests by generating $n = 800$ observations from the model in (1), with $p = 1$ endogenous regressor. The elements of the instrument matrix $\boldsymbol{Z}$ are independent standard normal distributed. We set $\boldsymbol{\Pi} = (R\sqrt{k}/n)^{1/2}\boldsymbol{e}_1$, such that the first instrument is relevant for the endogenous regressor. The strength of the instrument is governed by the parameter R.

To allow for violations of the invariance assumption we draw the errors ε_i independently from a skew normal distribution with skewness parameter ζ_ε. That is, the PDF of ε_i is $f(x) = (2/\omega_\varepsilon)\phi\left((x-\xi_\varepsilon)/\omega_\varepsilon\right)\Phi\left(\zeta_\varepsilon(x-\xi_\varepsilon)/\omega_\varepsilon\right)$, with ϕ and Φ the PDF and CDF of the standard normal distribution and $\omega_\varepsilon = \sigma_{\varepsilon i}/(1 - \frac{2\delta_\varepsilon^2}{\pi})$ and $\xi_\varepsilon = -\omega_\varepsilon\delta_\varepsilon\sqrt{2/\pi}$ for $\delta_\varepsilon = \zeta_\varepsilon/\sqrt{1+\zeta_\varepsilon^2}$ and π the mathematical constant. We consider both a homoskedastic setting and a heteroskedastic setting. In the former case we generate $\eta_i = \rho\varepsilon_i + \sqrt{1-\rho^2}w_i$ for $\rho = 0.3$ and $w_i \sim N(0,1)$. In the latter case we set $\eta_i = a_i\varepsilon_i + \frac{1}{2}w_i$ where $a_i = |Z_{i1}|$ and w_i again standard normal. The results are based on 10,000 draws from this data generating process (DGP).

6.1 Size

To analyze the size properties of the proposed tests, we set $\beta_0 = 0$ and a nominal size of $\alpha = 0.05$. Figure 5 shows the size of 2SLS, LIML, the fixed-k and the many instrument robust tests, when $k \in \{10, 100\}$ and R ranges from 0 to 50. In the upper left panel we see that for this relatively small instrument set and homoskedasticity 2SLS is oversized and especially so for weak instruments. LIML and the fixed-k tests on the other hand are slightly conservative. The many instrument robust tests are size correct. When k increases to 100, which is depicted in the lower left panel, the size distortion of 2SLS increase, but LIML remains relatively unaffected. The many instrument robust AR test maintains the five percent rejection rate. The size of the other many instrument robust tests drops

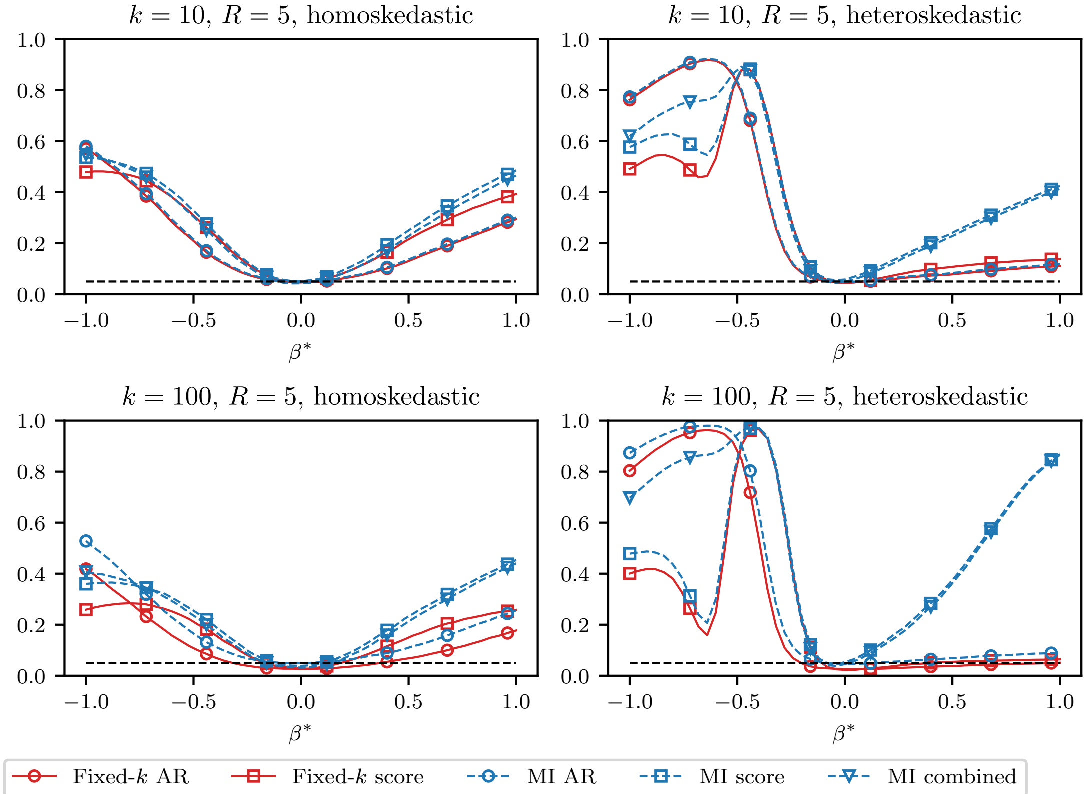
Figure 5: Size under identification robust inference.



Note: size when testing $H_0\colon \beta = 0$ at $\alpha = 0.05$ based on the fixed-k Anderson-Rubin test, the fixed-k score, the tests developed here, 2SLS and LIML. k denotes the number of instruments and R their strength. ζ_ε is the skewness parameter for ε. The combined test uses $\alpha_{\text{AR}} = 0.01$. The Monte Carlo is described in Section 6.

slightly, but not as much as the size of the fixed-k tests. These observations hold uniformly over the instrument strength and largely extend to the heteroskedastic case, with the exception that LIML becomes oversized for weak instruments and that the many instrument robust score, and with it the combination test, becomes slightly oversized when there is a small number of weak instruments. However, the many instrument robust score remains size correct for large instrument sets. The many instrument robust score's size distortion for $k = 10$ and weak instruments is due to negative variance estimates that occasionally prevail in this setting. By default the many instrument robust score test rejects for negative variance estimates, thus explaining the higher rejection rates. We do not observe negative variance estimates under the null for higher values of k or stronger instruments, and hence the test is size correct in these cases.

Figure 6: Power under identification robust inference with weak instruments.



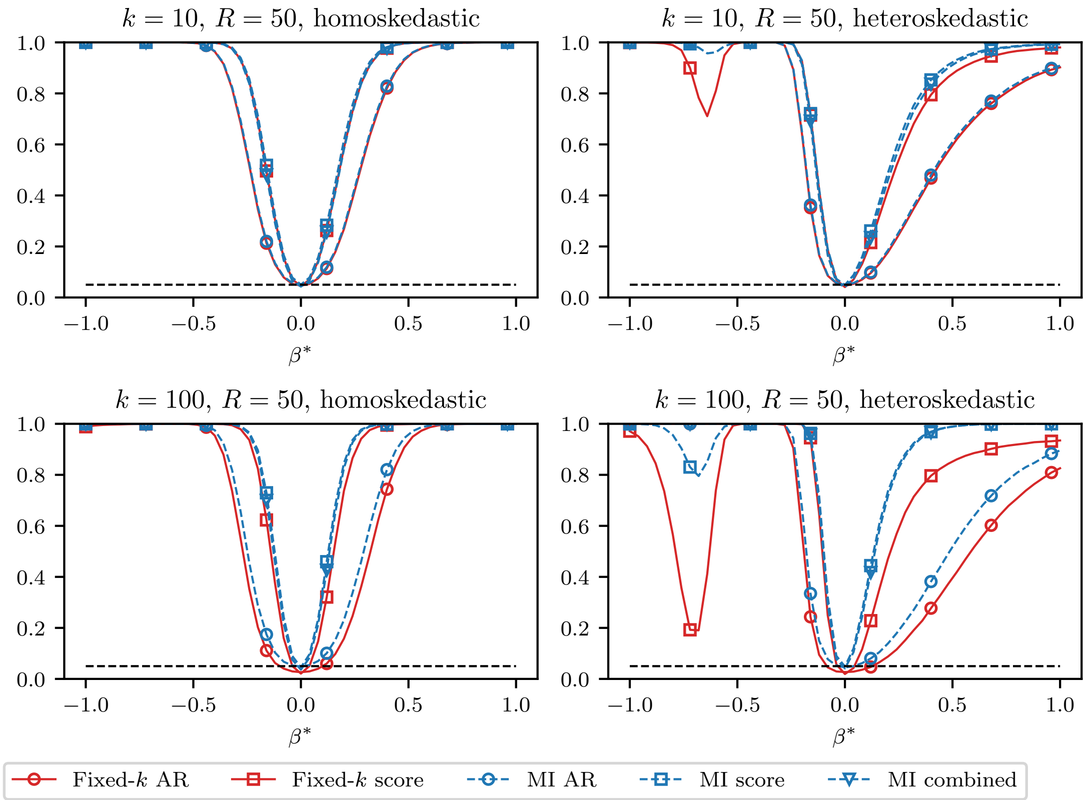
Note: power when testing $H_0 : \beta = 0$ when the true $\beta = \beta^*$ at $\alpha = 0.05$ based on the fixed-k Anderson-Rubin test, the fixed-k score test and the tests developed here. k denotes the number of instruments, R their strength and the invariance assumption is satisfied. The combined test uses $\alpha_{\text{AR}} = 0.01$. The Monte Carlo is described in Section 6.

6.2 Power

We now analyze the power against $H_0 \colon \beta = 0$ when the true β_0 equals β^* in the interval $[-1; 1]$ for $\alpha = 0.05$. We vary the identification strength by setting $R = 5$ and $R = 50$. Throughout the invariance assumption is satisfied. Figure 6 shows the power of the fixed-k and many instrument robust tests in different settings when the instruments are weak. If we first focus on the homoskedastic DGP depicted in the left panels, we observe that for small k the power of the fixed-k tests is close to their many instrument robust counterparts. Nevertheless, even for this small instrument set the many instrument correction improves the power. The power differences increase when $k = 100$ as shown in the lower left panel. Furthermore, using the score rather than the AR increases power.

These observations extend to the heteroskedastic case as given in the right panels. For small k the power of the fixed-k and many instrument robust tests

Figure 7: Power under identification robust inference with strong instruments.



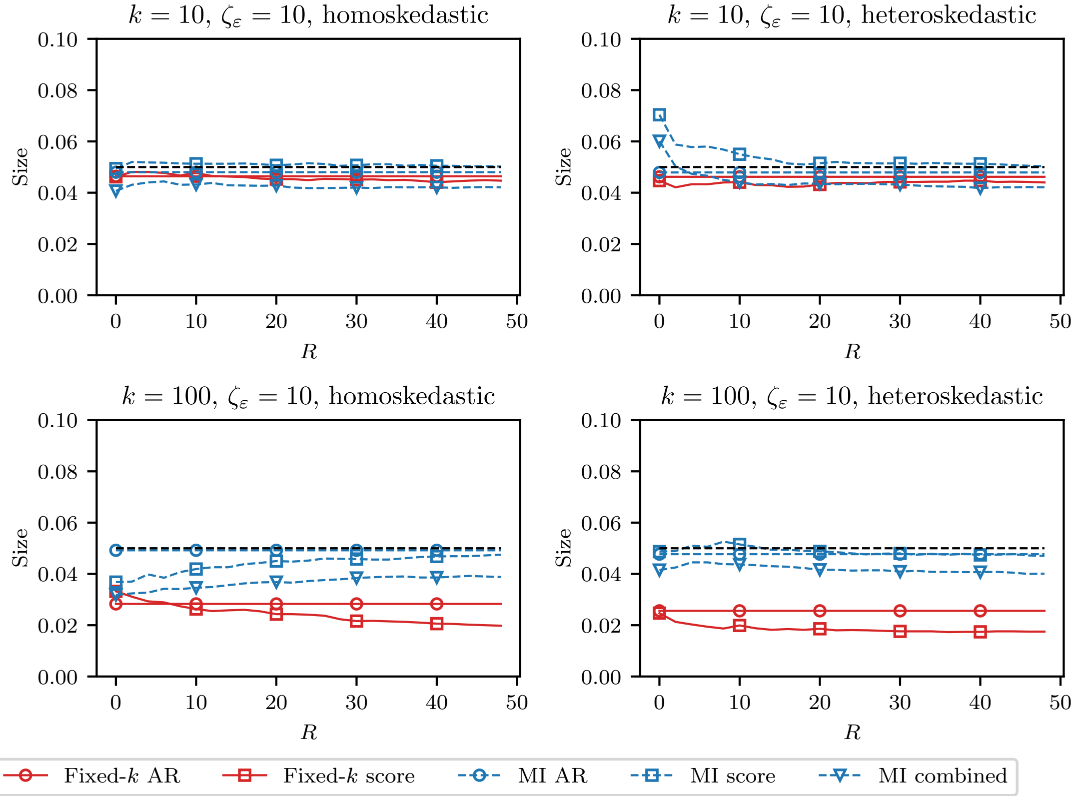
Note: see the note to Figure 6.

are close, but the power difference increases in the number of instruments, and the score generally tests have higher power than the AR tests. A major difference however, is the right tail of the power curve. For positive values of the true β only the many instrument robust score has good power. Part of this power comes from negative variance estimates for which the many instrument robust score test always rejects. Such estimates are common for alternatives far away from the true value of β. We illustrate this point in Appendix C.2.

For stronger instruments, as shown in Figure 7, we see similar behavior of the tests. A notable difference is the power in in the right tail for the heteroskedastic DGP. Although power remains lower than in the left tail, it increases towards one. The advantage of the many instrument robust tests over the fixed-k tests remains.

In both figures the combination of the many instrument robust AR and score test closely follows the power curve of the many instrument robust score. Only when the power of the score drops because the tested value is far from the true parameter, the power curves separate and we see that the combination successfully overcomes the loss in power.

Figure 8: Size under identification robust inference without invariance.



Note: see the note to Figure 5.

6.3 Moment conditions without invariance

An important issue is what happens to the size of the many instrument robust tests when the invariance assumption on the second stage regression errors is violated. Figure 8 shows the size when the errors ε_i are generated from a skewed normal distribution with $\zeta_\varepsilon = 10$. All panels show rejection rates that are very close to those in Figure 5. This suggests some robustness to departures from the invariance imposed in Assumption A2.

7 Empirical applications

7.1 Card (2009)

We illustrate the tests by estimating the negative inverse elasticity of substitution between immigrant and native workers using the model and data by Card (2009) and Goldsmith-Pinkham et al. (2020). Card (2009) uses the wage gap between natives and immigrants and their respective labor supplies in 124 cities in the US to determine how easily workers from one of these groups can be replaced by workers from the other group. Since this may depend on the skill group the employee is in, Card (2009) subdivides the sample in high school and college equivalent workers. Denote y_{jl} the residual log wage gap between immigrant and native men in skill group j and location l, x_{lj} the ratio of immigrant to native hours worked in skill group j and location l of both men and women and $\boldsymbol{X}_l$ the vector of location specific controls including an intercept. Then we can estimate the negative inverse elasticity of substitution, β, from

$$y_{lj} = \beta \log x_{lj} + \boldsymbol{\gamma}' \boldsymbol{X}_l + \varepsilon_{lj},$$

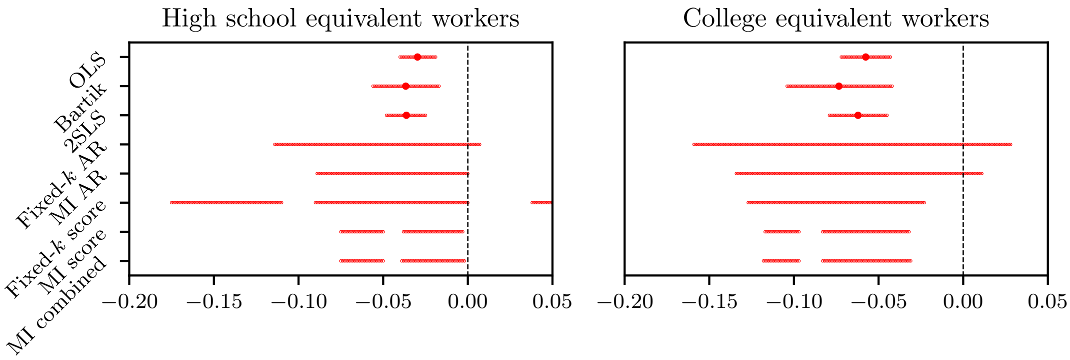
where $\boldsymbol{\gamma}$ are other coefficients and ε_{lj} is an error term.

It is possible though, that a local labor demand shock both increases relative earnings and the number of immigrants that settles in that location, thus increasing the relative labor supply. Consequently x_{lj} is an endogenous regressor. Card (2009) therefore instruments labor supply with the predicted number of immigrants in skill group j that settles in l arguing that immigrants tend to settle in cities with a large population of immigrants from their country of origin. Therefore settlement patterns in an initial period together with arrival rates of immigrants from specific country groups in subsequent periods, can be used to predict the inflow of immigrants in each city. To be precise, let $N_{lk,1980}$ be the number of immigrants from $k = 1, \ldots, 38$ country (group) of origin in l in 1980, $N_{k,1980}$ be the total number of immigrants from k in the US in 1980, $P_{l,2000}$ be the population of l in 2000 and g_{kj} be the number of immigrants from k in skill group j arriving in the US from 1990 to 2000. Then x_{lj} is instrumented by $B_{lj} = \sum_{k=1}^{38} z_{lk,1980} g_{kj}$ where $z_{lk,1980} = N_{lk,1980}/(N_{k,1980} \cdot P_{l,2000})$. B_{lj} is also known as a Bartik instrument.

Goldsmith-Pinkham et al. (2020) note that instead of combining the $z_{lk,1980}$ into a single instrument via the g_{kj}, they can also be used separately as instruments, leaving the coefficients on the $z_{lk,1980}$ unrestricted. In this case the number of instruments amounts to 38, which is large compared to the 124 observations.

In this section we compare confidence intervals for the elasticity of substitution obtained by ignoring the endogeneity and using ordinary least squares (OLS); 2SLS; the fixed-k AR and score statistic; and the tests developed in this paper. We apply 2SLS twice. Once with the Bartik instrument and once with the $z_{lk,1980}$ as instruments. The other instrumental variable approaches use the $z_{lk,1980}$ as instruments. To estimate the standard errors for the 2SLS estimate with B_{lj} as an instrument, we bootstrap over the 124 cities with 1000 replications. We prefer

Figure 9: 95% CIs for the negative inverse elasticity of substitution.



Note: 95% confidence intervals for the negative inverse elasticity of substitution between native and immigrant workers in the indicated skill. Point estimates, when available, are indicated with a dot. Confidence intervals are constructed using (i) OLS (ii) 2SLS with the Bartik instrument (iii) 2SLS (iv) the fixed-k AR statistic (v) the fixed-k score statistic and (vi-viii) the tests developed here.

bootstrap standard errors over analytical standard errors since, as Goldsmith-Pinkham et al. (2020) note, the 2SLS estimator is identical to a GMM estimator using a weighting matrix based on the g_{jk}. This weighting matrix is singular however, which invalidates standard results for GMM estimation and makes that conventional standard errors may be far off.

We repeat Figure 1 and add the many instrument robust combination test in Figure 9. From the results for high school equivalent workers, we observe the following. Firstly, OLS, 2SLS with the Bartik instrument and 2SLS yield narrower confidence intervals than the identification and many instrument robust methods. However, in this application there are many instruments compared to the sample size. These instruments are, moreover, likely to be weak as indicated by the disjoint confidence interval of the many instrument robust score statistic. 2SLS is oversized when there are many and/or weak instruments and thus these confidence intervals may be unreliably small. Secondly, the confidence intervals based on the fixed-k AR and score test are wider than their many instrument robust counterparts. Thirdly, the upper bounds of the fixed-k score and the many instrument robust confidence intervals are close to zero. The table in Appendix D which reports the exact values, indicates that according to the fixed-k score and the many instrument robust AR the negative inverse elasticity of substitution is not significantly different from zero. The confidence intervals from the score and the combination of the many instrument robust AR and score test on the other hand excludes zero. Fourthly, the confidence interval of the combined test is close to those of the many instrument robust score, albeit slightly wider.

The estimates for college equivalent workers show that OLS, and 2SLS again give the smallest confidence intervals. Furthermore, also for this skill group the intervals based on the fixed-k tests are the widest, hinting again at heteroskedasticity. Besides, the two AR tests yield wide intervals which includes zero, whereas the intervals from the score tests and the combined test do not. This thus shows the benefit of considering the score of the objective function, rather than only the objective function itself.

7.2 Angrist and Krueger (1991)

The article by Angrist and Krueger (1991), that studies the return to education, is a motivating study for the many and weak instrument literature. We revisit this study using the extended instrument set of all year-of-birth, quarter-of-birth and place-of-birth interactions as suggested by Mikusheva and Sun (2022), which contains up to 1530 instruments. Table 1 shows the 95% confidence intervals for the return to education for the fixed-k and many instrument tests. We note that the confidence intervals by the many instrument AR are slightly narrower than those by the fixed-k AR test for $k = 1530$. For the score and the combined test we do not observe any difference between the fixed-k and many instrument test however.

The widths of the confidence intervals for the many instrument robust AR are similar to those found by Mikusheva and Sun (2022) for the jackknife AR with crossfit variance. They find $[0.008, 0.20]$ and $[-0.047, 0.20]$ for 180 and 1530 instruments respectively. The confidence intervals for the many instrument score are also comparable to what Matsushita and Otsu (2022) find for the jackknife score: $[0.067, 0.133]$ for $k = 180$ and $[0.025, 0.123]$ for $k = 1530$.

8 Conclusion

We develop a new approach for identification robust inference under many instruments and heteroskedasticity using reflection invariance in the moment conditions to derive the joint limiting distributions of the AR and score statistic. We find that the many instrument corrections to the variance of both statistics are negative, suggesting that conventional approximations lead to conservative tests and a potential loss of power under many instruments. Monte Carlo simulations show

Table 1: 95% confidence intervals for the return to education

	$k = 30$		$k = 180$		$k = 1530$	
	Lower	Upper	Lower	Upper	Lower	Upper
Fixed-k AR	0.00	0.18	0.01	0.18	-0.01	0.23
MI AR	0.00	0.18	0.01	0.18	0.00	0.23
Fixed-k score	0.05	0.12	0.07	0.13	0.04	0.15
MI score	0.05	0.12	0.07	0.13	0.04	0.15
MI combined	0.05	0.12	0.07	0.13	0.04	0.15

Note: 95% confidence intervals for the return to education. k denotes the number of instruments. Estimates are based on people born between 1930 and 1940.

close to nominal size of the developed procedures regardless of the strength of the instruments and the number of instruments, as well as a substantial increase in power under many instruments. We apply our new tests to the elasticity of substitution study by Card (2009) and to the return to education study by Angrist and Krueger (1991). The applications show that the reduction in the length of the confidence intervals by using the many instrument approximation can be sizable.

References

Anatolyev, S. (2019): "Many instruments and/or regressors: a friendly guide," *Journal of Economic Surveys*, 33, 689–726.

Anatolyev, S. and N. Gospodinov (2011): "Specification testing in models with many instruments," *Econometric Theory*, 27, 427–441.

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, 46 – 63.

Andrews, D. W. K., V. Marmer, and Z. Yu (2019): "On optimal inference in the linear IV model," *Quantitative Economics*, 10, 457–485.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal two-sided invariant similar tests for instrumental variables regression," *Econometrica*, 74, 715–752.

Andrews, I. (2016): "Conditional linear combination tests for weakly identified models," *Econometrica*, 84, 2155–2182.

Angrist, J. D. and G. W. Imbens (1995): "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American statistical Association*, 90, 431–442.

Angrist, J. D. and A. B. Krueger (1991): "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

Bai, Z. D., B. Q. Miao, and G. M. Pan (2007): "On asymptotics of eigenvectors of large sample covariance matrix," *Annals of Probability*, 35, 1532 – 1572.

Bao, Y. and A. Ullah (2010): "Expectation of quadratic forms in normal and nonnormal variables with applications," *Journal of Statistical Planning and Inference*, 140, 1193–1205.

Bekker, P. and F. Kleibergen (2003): "Finite-sample instrumental variables inference using an asymptotically pivotal statistic," *Econometric Theory*, 19, 744–753.

Bekker, P. A. (1994): "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica*, 62, 657–681.

Bekker, P. A. and F. Crudu (2015): "Jackknife instrumental variable estimation with heteroskedasticity," *Journal of Econometrics*, 185, 332–342.

Bekker, P. A. and S. Lawford (2008): "Symmetry-based inference in an instrumental variable setting," *Journal of Econometrics*, 142, 28–49.

Bekker, P. A. and J. van der Ploeg (2005): "Instrumental variable estimation based on grouped data," *Statistica Neerlandica*, 59, 239–267.

Billingsley, P. (1995): *Probability and Measure*, John Wiley & Sons Ltd, Third ed.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is TSLS actually LATE?" Working Paper 29709, National Bureau of Economic Research.

Bun, M. J. G., H. Farbmacher, and R. W. Poldermans (2020): "Finite sample properties of the GMM Anderson–Rubin test," *Econometric Reviews*, 39, 1042–1056.

Canay, I. A., J. P. Romano, and A. M. Shaikh (2017): "Randomization tests under an approximate symmetry assumption," *Econometrica*, 85, 1013–1030.

Card, D. (2009): "Immigration and inequality," *American Economic Review*, 99, 1–21.

Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen (2014): "Testing overidentifying restrictions with many instruments and heteroskedasticity," *Journal of Econometrics*, 178, 15–21.

Chao, J. C. and N. R. Swanson (2005): "Consistent estimation with a large number of weak instruments," *Econometrica*, 73, 1673–1692.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen

(2012): “Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments,” *Econometric Theory*, 28, 42–86.

Crudu, F., G. Mellace, and Z. Sándor (2021): “Inference in instrumental variable models with heteroskedasticity and many instruments,” *Econometric Theory*, 37, 281–310.

Davies, N. M., G. Hemani, N. J. Timpson, F. Windmeijer, and G. Davey Smith (2015): “The role of common genetic variation in educational attainment and income: evidence from the National Child Development Study,” *Scientific Reports*, 5, 1–12.

Donald, S. G. and W. K. Newey (2000): “A jackknife interpretation of the continuous updating estimator,” *Economics Letters*, 67, 239–243.

Efron, B. (1969): “Student's t-test under symmetry conditions,” *Journal of the American Statistical Association*, 64, 1278–1302.

Fisher, R. A. (1925): “Applications of “Student's” Distribution.” *Metron*, 5, 90–104.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): “Bartik instruments: What, when, why, and how,” *American Economic Review*, 110, 2586–2624.

Hansen, C., J. Hausman, and W. Newey (2008): “Estimation with many instrumental variables,” *Journal of Business & Economic Statistics*, 26, 398–422.

Hansen, L. P., J. Heaton, and A. Yaron (1996): “Finite-sample properties of some alternative GMM estimators,” *Journal of Business & Economic Statistics*, 14, 262–280.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): “Instrumental variable estimation with heteroskedasticity and many instruments,” *Quantitative Economics*, 3, 211–255.

Kleibergen, F. (2002): “Pivotal statistics for testing structural parameters in instrumental variables regression,” *Econometrica*, 70, 1781–1803.

——— (2005): “Testing parameters in GMM without assuming that they are identified,” *Econometrica*, 73, 1103–1123.

Kling, J. R. (2006): “Incarceration length, employment, and earnings,” *American Economic Review*, 96, 863–876.

Kunitomo, N. (1980): “Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations,” *Journal of the American Statistical Association*, 75, 693–700.

Lehmann, E. L. and J. P. Romano (2005): *Testing statistical hypotheses*, Springer, Third ed.

Lim, D., W. Wang, and Y. Zhang (2022): "A Conditional Linear Combination Test with Many Weak Instruments," *arXiv preprint arXiv:2207.11137*.

Matsushita, Y. and T. Otsu (2022): "Jackknife Lagrange multiplier test with many weak instruments," *Econometric Theory (in press)*.

Mikusheva, A. and L. Sun (2022): "Inference with Many Weak Instruments," *Review of Economic Studies*, 89, 2663–2686.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Morimune, K. (1983): "Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size," *Econometrica*, 51, 821–841.

Newey, W. K. and F. Windmeijer (2009): "Generalized method of moments with many weak moment conditions," *Econometrica*, 77, 687–719.

Stock, J. H. and M. Yogo (2005): "Asymptotic distributions of instrumental variables statistics with many instruments," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 109–120.

Ullah, A. (2004): *Finite sample econometrics*, Oxford University Press.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes With Applications to Statistics*, Springer Nature, chap. Symmetrization and Measurability.

Vershynin, R. (2018): "High-dimensional probability: an introduction with applications in data science," in *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, vol. 47.

Whittle, P. (1960): "Bounds for the moments of linear and quadratic forms in independent variables," *Theory of Probability & Its Applications*, 5, 302–305.

Young, A. (2019): "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *Quarterly Journal of Economics*, 134, 557–598.

Appendix A Proofs

A.1 Preliminary results

In the proofs of our theorems we make use of the following results.

A.1.1 Expectations over Rademacher random variables

Theorem A.1. *Consider a $n \times 1$ vector $\boldsymbol{r}$ with independent Rademacher entries. Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_4$ denote generic $n \times n$ matrices and $\boldsymbol{v}$ an $n \times 1$ vector. Then,*

1. $\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r}] = \mathrm{tr}(\boldsymbol{A}_1)$.
2. Ullah (2004):
$\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{r}] = -2\,\mathrm{tr}(\boldsymbol{D}_{A_1}\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1)\,\mathrm{tr}(\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1'\boldsymbol{A}_2)$.
3. $\mathrm{E}[\boldsymbol{v}'\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{D}_r\boldsymbol{A}_1\boldsymbol{r}] = \boldsymbol{v}'\boldsymbol{A}_2\boldsymbol{D}_{A_1}\boldsymbol{\iota} + \boldsymbol{\iota}'(\boldsymbol{A}_2 \odot \boldsymbol{A}_1')\boldsymbol{v} + \boldsymbol{\iota}'\boldsymbol{D}_{A_2}\boldsymbol{A}_1\boldsymbol{v} - 2\boldsymbol{\iota}'\boldsymbol{D}_{A_2}\boldsymbol{D}_{A_1}\boldsymbol{v}$.
4. $$\begin{aligned}
\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}] &= \mathrm{tr}(\boldsymbol{D}_{A_4A_1}\boldsymbol{D}_{A_2A_3}) \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_3}\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} - 2\,\mathrm{tr}\left(\boldsymbol{D}_{A_3}\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{D}_{A_2}\right) + \boldsymbol{\iota}'\boldsymbol{A}_2 \odot \boldsymbol{A}_3 \odot (\boldsymbol{A}_1'\boldsymbol{A}_4')\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}\boldsymbol{\iota} - 2\,\mathrm{tr}\left(\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}\right) + \boldsymbol{\iota}'\boldsymbol{A}_4 \odot \boldsymbol{A}_1 \odot (\boldsymbol{A}_3'\boldsymbol{A}_2')\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'(\boldsymbol{A}_1 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_4)\boldsymbol{\iota} + \mathrm{tr}(\boldsymbol{A}_1'\boldsymbol{A}_2' \odot \boldsymbol{A}_3'\boldsymbol{A}_4') - 2\,\mathrm{tr}((\boldsymbol{A}_1 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_4)) \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} - 2\,\mathrm{tr}(\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}) \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_2A_1}\boldsymbol{A}_3\boldsymbol{D}_{A_4}\boldsymbol{\iota} - 2\,\mathrm{tr}(\boldsymbol{D}_{A_2A_1}\boldsymbol{A}_3\boldsymbol{D}_{A_4}) \\
&- 2\boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{D}_{A_2}\boldsymbol{A}_3\boldsymbol{D}_{A_4}\boldsymbol{\iota} - 2\boldsymbol{\iota}'\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} + 16\,\mathrm{tr}(\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}) \\
&- 2\boldsymbol{\iota}'\boldsymbol{A}_1\boldsymbol{D}_{A_2} \odot \boldsymbol{A}_4 \odot \boldsymbol{A}_3'\boldsymbol{\iota} - 2\boldsymbol{\iota}'\boldsymbol{A}_3\boldsymbol{D}_{A_4} \odot \boldsymbol{A}_2 \odot \boldsymbol{A}_1'\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_4}\boldsymbol{A}_3'\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} - \mathrm{tr}(\boldsymbol{D}_{A_4}\boldsymbol{A}_3'\boldsymbol{A}_1\boldsymbol{D}_{A_2}) - \mathrm{tr}(\boldsymbol{D}_{A_2}\boldsymbol{A}_1'\boldsymbol{A}_3\boldsymbol{D}_{A_4}) \\
&+ \boldsymbol{\iota}'((\boldsymbol{A}_3 \odot \boldsymbol{A}_1')\boldsymbol{A}_4) \odot \boldsymbol{A}_2\boldsymbol{\iota} - 2\boldsymbol{\iota}'((\boldsymbol{A}_1 \odot \boldsymbol{A}_3' \odot \boldsymbol{I})\boldsymbol{A}_2) \odot \boldsymbol{A}_4\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'(\boldsymbol{A}_1 \odot (\boldsymbol{A}_3(\boldsymbol{A}_2' \odot \boldsymbol{A}_4)))\boldsymbol{\iota} - 2\,\mathrm{tr}((\boldsymbol{A}_1 \odot (\boldsymbol{A}_3(\boldsymbol{A}_2' \odot \boldsymbol{A}_4)))) \\
&- 2\,\mathrm{tr}((\boldsymbol{A}_3 \odot (\boldsymbol{A}_1(\boldsymbol{A}_4' \odot \boldsymbol{A}_2)))) \\
&+ \boldsymbol{\iota}'(\boldsymbol{A}_1 \odot \boldsymbol{A}_2')\boldsymbol{A}_4'\boldsymbol{D}_{A_3}\boldsymbol{\iota} - 2\boldsymbol{\iota}'(\boldsymbol{A}_1 \odot \boldsymbol{A}_2' \odot \boldsymbol{I})\boldsymbol{A}_4'\boldsymbol{D}_{A_3}\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'(\boldsymbol{A}_3 \odot \boldsymbol{A}_4')\boldsymbol{A}_2'\boldsymbol{D}_{A_1}\boldsymbol{\iota} - 2\boldsymbol{\iota}'(\boldsymbol{A}_3 \odot \boldsymbol{A}_4' \odot \boldsymbol{I})\boldsymbol{A}_2'\boldsymbol{D}_{A_1}\boldsymbol{\iota} \\
&+ \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_4'\boldsymbol{D}_{A_3}\boldsymbol{\iota}.
\end{aligned}$$
5. $$\begin{aligned}
\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}] &= \mathrm{tr}(\boldsymbol{D}_{A_4A_1}\boldsymbol{D}_{A_2\boldsymbol{A}_3}) + \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}\boldsymbol{\iota} \\
&- 2\,\mathrm{tr}(\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}) + \boldsymbol{\iota}'(\boldsymbol{A}_4 \odot \boldsymbol{A}_1 \odot (\boldsymbol{A}_3'\boldsymbol{A}_2'))\boldsymbol{\iota}.
\end{aligned}$$

Suppose now that $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are symmetric matrices with all diagonal elements equal to zero. Then,

6. *Bao and Ullah (2010):*

$$\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r})^2(\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{r})^2] = 4\,\mathrm{tr}(\boldsymbol{A}_1^2)\,\mathrm{tr}(\boldsymbol{A}_2^2) + 8\,\mathrm{tr}^2(\boldsymbol{A}_1\boldsymbol{A}_2) + 32\,\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_1\boldsymbol{A}_2) + 16\,\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_2\boldsymbol{A}_1) - 32\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_1^2)(\boldsymbol{I}\odot\boldsymbol{A}_2^2)\boldsymbol{\iota} - 64\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_1\boldsymbol{A}_2)(\boldsymbol{I}\odot\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{\iota} + 32\boldsymbol{\iota}'(\boldsymbol{A}_1\odot\boldsymbol{A}_1\odot\boldsymbol{A}_2\odot\boldsymbol{A}_2)\boldsymbol{\iota}.$$

7. *Ullah (2004):*

$$\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_3\boldsymbol{r}] = \mathrm{tr}(\boldsymbol{A}_3[20(\boldsymbol{A}_1\odot\boldsymbol{A}_2) - 3\boldsymbol{I}\odot(2\boldsymbol{A}_1\boldsymbol{A}_2 + 2\boldsymbol{A}_2\boldsymbol{A}_1) + 4\boldsymbol{A}_1\boldsymbol{A}_2 + 4\boldsymbol{A}_2\boldsymbol{A}_1 + 2\,\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{I}]).$$

Proof. 1. $\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r}] = \mathrm{tr}(\boldsymbol{A}_1\,\mathrm{E}[\boldsymbol{r}\boldsymbol{r}']) = \mathrm{tr}(\boldsymbol{A}_1)$.

2. See Ullah (2004), Appendix A5.

3. Denote $\boldsymbol{\Delta} = \boldsymbol{r}\boldsymbol{r}' - \boldsymbol{I}$. We split the expectation into two parts,

$$\mathrm{E}[\boldsymbol{v}'\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}] = \underbrace{\mathrm{E}[\boldsymbol{v}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}]}_{(I)} + \underbrace{\mathrm{E}[\boldsymbol{v}'\boldsymbol{\Delta}\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}]}_{(II)}.$$

For the first part, using independence of the Rademacher random variables,

$$(I) = \mathrm{E}\left[\sum_{i,j,k=1}^{n} v_i a_{1,ij} a_{2,jk} r_j r_k\right] = \sum_{i,j=1}^{n} v_i a_{1,ij} a_{2,jj} = \boldsymbol{v}'\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota}.$$

For (II) we write $(II) = \mathrm{E}\left[\sum_{i,j,k,l=1}^{n} v_i\delta_{ij}a_{1,jk}a_{2,kl}r_k r_l\right]$. There are two cases where the expectation is nonzero. In case $(II.a)$ $i=k, j=l, i\neq j$, and

$$(II.a) = \sum_{i\neq j} v_i a_{1,ji} a_{2,ij} = \boldsymbol{\iota}\boldsymbol{A}_1\odot\boldsymbol{A}_2'\boldsymbol{v} - \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{D}_{A_2}\boldsymbol{v}.$$

In case $(II.b)$ $i=l, j=k, i\neq j$, such that

$$(II.b) = \sum_{i\neq j} v_i a_{1,jj} a_{2,ji} = \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{v} - \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{D}_{A_2}\boldsymbol{v}.$$

4. We decompose the expectation as

$$\begin{aligned}&\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}]\\ &= \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4)]}_{(I)} + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I}))]}_{(III)}\\ &\quad + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4)]}_{(II)} + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I}))]}_{(II')}.\end{aligned} \tag{A.1}$$

Starting with (I), we have that

$$
\begin{aligned}
(I) &= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_r\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{D}_r\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{A}_4\boldsymbol{e}_i\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_4\boldsymbol{e}_i = \mathrm{tr}(\boldsymbol{D}_{A_4A_1}\boldsymbol{D}_{A_2A_3}).
\end{aligned}
$$

For (II), define $\delta_{kl} = [\boldsymbol{r}\boldsymbol{r}' - \boldsymbol{I}]_{kl}$ and note that $\delta_{kk} = 0$, and $\mathrm{E}[\delta_{kl}] = \mathrm{E}[r_k r_l] = 0$ if $k \neq l$ and $\mathrm{E}[\delta_{kl}^2] = \mathrm{E}[r_k^2 r_l^2] = 1$.

$$
\begin{aligned}
(II) &= \mathrm{E}\left[\sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}' - \boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{e}_i\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j,m,k,l=1}^{n} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_r\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{e}_m\delta_{mk}\boldsymbol{e}_k'\boldsymbol{A}_3\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{D}_r\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{A}_4\boldsymbol{e}_i\right].
\end{aligned}
$$

There are two cases when the expectation is nonzero: (a) $j = m, l = k, j \neq l$ and (b) $j = k, l = m, j \neq l$. Starting with case (a),

$$
\begin{aligned}
(II.a) &= \sum_{i=1}^{n}\sum_{j\neq l} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{e}_j\boldsymbol{e}_l'\boldsymbol{A}_3\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{A}_4\boldsymbol{e}_i \\
&= \boldsymbol{\iota}'\boldsymbol{D}_{A_3}\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota}\underbrace{-\,\mathrm{tr}\left(\boldsymbol{D}_{A_3}\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{D}_{A_2}\right)}_{(II.a.2)}.
\end{aligned}
$$

For case (b), we have

$$
\begin{aligned}
(II.b) &= \sum_{i=1}^{n}\sum_{j\neq l} \boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{e}_l\boldsymbol{e}_j'\boldsymbol{A}_3\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{A}_4\boldsymbol{e}_i \\
&= \boldsymbol{\iota}'\boldsymbol{A}_2 \odot \boldsymbol{A}_3 \odot (\boldsymbol{A}_1'\boldsymbol{A}_4')\boldsymbol{\iota} + (II.a.2).
\end{aligned}
$$

By rotation invariance, the expressions for (II') can be obtained by changing $\boldsymbol{A}_2 \to \boldsymbol{A}_4$, $\boldsymbol{A}_3 \to \boldsymbol{A}_1$, $\boldsymbol{A}_4 \to \boldsymbol{A}_2$, $\boldsymbol{A}_1 \to \boldsymbol{A}_3$.

The most difficult term to deal with in (A.1) is

$$
(III) = \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j,k,m,l,s} a_{1,ij}a_{2,jk}a_{3,ml}a_{4,ls}r_j r_l \delta_{km}\delta_{si}\right].
$$

There are now 10 cases to consider, which we label $(III.a) - (III.j)$. All of them satisfy $k \neq m$, $s \neq i$

a.	$j=l$	$k=s$	$m=i$	$k\neq i$	f.		$k=i$	$l=s$
b.		$k=i$	$m=s$	$i\neq m$	g.	$j=s$	$k=l$	$m=i$
c.	$j\neq l$	$j=k$	$m=s$	$l=i$	h.		$k=i$	$l=m$
d.			$m=i$	$l=s$	i.	$j=i$	$k=s$	$m=l$
e.		$j=m$	$k=s$	$l=i$	j.		$k=l$	$m=s$

We work out $(III.a)-(III.c)$ explicitly. The remaining cases follow by analogous calculations.

$$
\begin{aligned}
(III.a) &= \sum_{i=1}^{n}\sum_{j,k\neq i} a_{1,ij}a_{2,jk}a_{3,ij}a_{4,jk} \\
&= \sum_{i=1}^{n}\sum_{j,k\neq i} \boldsymbol{e}_i'(\boldsymbol{A}_1\odot\boldsymbol{A}_3)\boldsymbol{e}_j\boldsymbol{e}_j'(\boldsymbol{A}_2\odot\boldsymbol{A}_4)\boldsymbol{e}_k \\
&= \boldsymbol{\iota}'(\boldsymbol{A}_1\odot\boldsymbol{A}_3)(\boldsymbol{A}_2\odot\boldsymbol{A}_4)\boldsymbol{\iota}\underbrace{-\operatorname{tr}((\boldsymbol{A}_1\odot\boldsymbol{A}_3)(\boldsymbol{A}_2\odot\boldsymbol{A}_4))}_{(III.a.2)}.
\end{aligned}
$$

$$
\begin{aligned}
(III.b) &= \sum_{i=1}^{n}\sum_{j,m\neq i} a_{1,ij}a_{2,ji}a_{3,mj}a_{4,jm} \\
&= \operatorname{tr}(\boldsymbol{A}_1'\boldsymbol{A}_2'\odot\boldsymbol{A}_3'\boldsymbol{A}_4') + (III.a.2). \\
(III.c) &= \sum_{i=1}^{n}\sum_{j\neq i, j\neq m, m\neq i} a_{1,ij}a_{2,jj}a_{4,im}a_{3,mi} \\
&= \sum_{i=1}^{n}\sum_{j\neq i} \boldsymbol{e}_i\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{e}_j - \boldsymbol{e}_i'\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{e}_j \\
&\quad - \boldsymbol{e}_i'(\boldsymbol{A}_1\boldsymbol{D}_{A_2})\odot\boldsymbol{A}_4\odot\boldsymbol{A}_1'\boldsymbol{e}_j \\
&= \boldsymbol{\iota}'\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} - \operatorname{tr}(\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}) - \boldsymbol{\iota}'\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota} \\
&\quad + 2\operatorname{tr}(\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}) - \boldsymbol{\iota}'(\boldsymbol{A}_1\boldsymbol{D}_{A_2})\odot\boldsymbol{A}_4\odot\boldsymbol{A}_3'\boldsymbol{\iota}.
\end{aligned}
$$

There are many repeated elements in the expressions for $(III.d)-(III.j)$. We introduce the following notation

$$
\begin{aligned}
&(c.1) = \boldsymbol{\iota}'\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota}, && (c.2) = -\operatorname{tr}(\boldsymbol{D}_{A_4A_3}\boldsymbol{A}_1\boldsymbol{D}_{A_2}), \\
&(c.3) = -\boldsymbol{\iota}'\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota}, && (c.4) = \operatorname{tr}(\boldsymbol{D}_{A_3}\boldsymbol{D}_{A_4}\boldsymbol{A}_1\boldsymbol{D}_{A_2}), \\
&(c.5) = -\boldsymbol{\iota}'(\boldsymbol{A}_1\boldsymbol{D}_{A_2})\odot\boldsymbol{A}_4\odot\boldsymbol{A}_3'\boldsymbol{\iota}, && (d.1) = \boldsymbol{\iota}'\boldsymbol{D}_{A_4}\boldsymbol{A}_3'\boldsymbol{A}_1(\boldsymbol{I}\odot\boldsymbol{A}_2)\boldsymbol{\iota}, \\
&(d.2) = -\operatorname{tr}(\boldsymbol{D}_{A_4}\boldsymbol{A}_3'\boldsymbol{A}_1\boldsymbol{D}_{A_2}), && (e.1) = \boldsymbol{\iota}'((\boldsymbol{A}_3\odot\boldsymbol{A}_1')\boldsymbol{A}_4)\odot\boldsymbol{A}_2\boldsymbol{\iota}, \\
&(e.3) = -\boldsymbol{\iota}'((\boldsymbol{A}_1\odot\boldsymbol{A}_3'\odot\boldsymbol{I})\boldsymbol{A}_2)\odot\boldsymbol{A}_4\boldsymbol{\iota}, && (g.1) = \boldsymbol{\iota}'(\boldsymbol{A}_1\odot(\boldsymbol{A}_3(\boldsymbol{A}_2'\odot\boldsymbol{A}_4)))\boldsymbol{\iota}, \\
&(g.2) = -\operatorname{tr}((\boldsymbol{A}_1\odot(\boldsymbol{A}_3(\boldsymbol{A}_2'\odot\boldsymbol{A}_4)))), && (h.1) = \boldsymbol{\iota}'(\boldsymbol{A}_1\odot\boldsymbol{A}_2')\boldsymbol{A}_4'\boldsymbol{D}_{A_3}\boldsymbol{\iota}, \\
&(h.5) = -\boldsymbol{\iota}'((\boldsymbol{A}_1\odot\boldsymbol{A}_2')\odot\boldsymbol{I})'A_4(\boldsymbol{A}_3'\odot\boldsymbol{I})\boldsymbol{\iota}, && (i.1) = \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_4'\boldsymbol{D}_{A_3}\boldsymbol{\iota}.
\end{aligned}
$$

Furthermore, let any of these with a asterisk denote the same term but with $\boldsymbol{A}_2 \to \boldsymbol{A}_4$, $\boldsymbol{A}_3 \to \boldsymbol{A}_1$, $\boldsymbol{A}_4 \to \boldsymbol{A}_2$, $\boldsymbol{A}_1 \to \boldsymbol{A}_3$. Then

$$
\begin{aligned}
(III.c) &= (c.1) + (c.2) + (c.3) + (c.4) + (c.5) + (c.4),\\
(III.d) &= (d.1) + (d.2) + (c.3)^* + (c.4) + (c.3) + (c.4),\\
(III.e) &= (e.1) + (c.5)^* + (e.3) + (c.4) + (c.5)^* + (c.4),\\
(III.f) &= (c.1)^* + (c.5)^* + (c.3)^* + (c.4) + (c.2)^* + (c.4),\\
(III.g) &= (g.1) + (g.2)^* + (g.2) + (c.4) + (d.2)^* + (c.4),\\
(III.h) &= (h.1) + (g.2)^* + (c.2)^* + (c.4) + (h.5) + (c.4),\\
(III.i) &= (i.1) + (e.3) + (h.5)^* + (c.4) + (h.5) + (c.4),\\
(III.j) &= (h.1)^* + (g.2) + (h.5)^* + (c.4) + (c.2) + (c.4).
\end{aligned}
$$

Putting everything together, we obtain the desired result.

5. Can be obtained from Item 4 by only considering the terms (I) and $(II)'$ in the proof.
6. See Bao and Ullah (2010), Theorem 2.
7. See Ullah (2004), Appendix A5.

□

A.1.2 Eigenvalues of Hadamard products

Theorem A.2. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ real symmetric matrices. Then*

$$\lambda_{\max}(\boldsymbol{A} \odot \boldsymbol{B}) \leq \lambda_{\max}(\boldsymbol{A} \otimes \boldsymbol{B}) \leq \max\{\lambda_{\max}(\boldsymbol{A})\, \lambda_{\max}(\boldsymbol{B}), \lambda_{\min}(\boldsymbol{A})\, \lambda_{\min}(\boldsymbol{B})\},$$

and

$$
\begin{aligned}
\lambda_{\min}(\boldsymbol{A} \odot \boldsymbol{B}) &\geq \lambda_{\min}(\boldsymbol{A} \otimes \boldsymbol{B})\\
&\geq \min\{\lambda_{\min}(\boldsymbol{A})\, \lambda_{\min}(\boldsymbol{B}), \lambda_{\min}(\boldsymbol{A})\, \lambda_{\max}(\boldsymbol{B}), \lambda_{\max}(\boldsymbol{A})\, \lambda_{\min}(\boldsymbol{B})\}.
\end{aligned}
$$

Proof. Let $\boldsymbol{v} \in \mathbb{R}^n$ be given and define $\boldsymbol{u} \in \mathbb{R}^{n^2}$ with $u_{(i-1)\cdot n+i} = v_i$ for $i = 1, \ldots, n$ and zeroes elsewhere. Then $\boldsymbol{v}'(\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{v} = \boldsymbol{u}'(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{u}$.

Now since $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric, so are $\boldsymbol{A} \odot \boldsymbol{B}$ and $\boldsymbol{A} \otimes \boldsymbol{B}$. Consequently both have real eigenvalues of which the maximum and minimum can be written as

$$\lambda_{\max}(\boldsymbol{A}\odot\boldsymbol{B}) = \max_{\boldsymbol{v}:\boldsymbol{v}'\boldsymbol{v}=1} \boldsymbol{v}'(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{v} = \max_{\boldsymbol{u}:\boldsymbol{u}'\boldsymbol{u}=1} \boldsymbol{u}'(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{u}$$
$$\leq \max_{\boldsymbol{w}:\boldsymbol{w}'\boldsymbol{w}=1} \boldsymbol{w}'(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{w} = \lambda_{\max}(\boldsymbol{A}\otimes\boldsymbol{B}),$$

and

$$\lambda_{\min}(\boldsymbol{A}\odot\boldsymbol{B}) = \min_{\boldsymbol{v}:\boldsymbol{v}'\boldsymbol{v}=1} \boldsymbol{v}'(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{v} = \min_{\boldsymbol{u}:\boldsymbol{u}'\boldsymbol{u}=1} \boldsymbol{u}'(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{u}$$
$$\geq \min_{\boldsymbol{w}:\boldsymbol{w}'\boldsymbol{w}=1} \boldsymbol{w}'(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{w} = \lambda_{\min}(\boldsymbol{A}\otimes\boldsymbol{B}),$$

where $\boldsymbol{u}$ is restricted to follow the structure above and $\boldsymbol{w}$ is any vector in $\mathbb{R}^{n^2}$. We therefore go from a restricted extremum to an unrestricted extremum, which explains the inequality.

The last set of equalities in the theorem then follows because the n^2 eigenvalues of $\boldsymbol{A}\otimes\boldsymbol{B}$ equal $\lambda_i(\boldsymbol{A})\lambda_j(\boldsymbol{B})$ for $i,j=1,\dots,n$. □

Corollary 3. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n\times n$ real symmetric matrices. If $\lambda_{\min}(\boldsymbol{A})\geq 0$ then $\lambda_{\max}(\boldsymbol{A}\odot\boldsymbol{B})\leq\lambda_{\max}(\boldsymbol{A})\,\lambda_{\max}(\boldsymbol{B})$. If in addition $\lambda_{\min}(\boldsymbol{B})\geq 0$, then $\lambda_{\min}(\boldsymbol{A}\odot\boldsymbol{B})\geq\lambda_{\min}(\boldsymbol{A})\,\lambda_{\min}(\boldsymbol{B})$.*

A.2 Proof of Theorem 2

Proof. The score function from the continuous updating objective function is given by $\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_i} = -\frac{1}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{I}-\boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon}$. Under Assumption A2, the score satisfies

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_i} &\overset{(d)}{=} -\frac{1}{n}(\bar{\boldsymbol{x}}_{(i)}+\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)})'(\boldsymbol{I}-\boldsymbol{D}_r\boldsymbol{D}_{Pr})\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{\varepsilon}\\
&= -\frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}+\frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}-\frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}+\frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\\
&\overset{(\mathrm{E}_r)}{=} 0.
\end{aligned}$$

This proves the first statement of Theorem 2.

The $(i,j)^{\text{th}}$ element of the conditional variance is given by

$$\mathrm{E}\left[n\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_i}\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_j}|\mathcal{J}\right] = \mathrm{E}\Bigg[\underbrace{\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{x}_{(j)}}_{(I)}+\underbrace{\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{D}_{P\iota}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{D}_{P\iota}\boldsymbol{x}_{(j)}}_{(II)}$$
$$\underbrace{-\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{D}_{P\iota}\boldsymbol{x}_{(j)}}_{(III)}\underbrace{-\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{D}_{P\iota}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{x}_{(j)}}_{(IV)}|\mathcal{J}\Bigg].$$

We first write $(I)-(IV)$ in terms of the Rademacher random variables from Assumption A2, take the expectation over these random variables by applying the results from Section A.1.1, and then take an additional expectation over the first stage errors $\boldsymbol{U}$. The marker "fixed-k approximation" indicates terms that appear when we take the estimator for the variance of the score as in Kleibergen (2005) and Newey and Windmeijer (2009), and then take the expectation over $\boldsymbol{r}$ and $\boldsymbol{U}$.

Using that $\boldsymbol{x}_{(i)} = \bar{\boldsymbol{x}}_{(i)} + \boldsymbol{D}_\varepsilon \boldsymbol{a}_{(i)}$ we get,

$$
\begin{aligned}
(I) &\overset{(d)}{=} \frac{1}{n}\bar{\boldsymbol{x}}_{(i)}'\boldsymbol{V}\boldsymbol{D}_\varepsilon \boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} + \boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} \\
&\quad + \frac{1}{n}\bar{\boldsymbol{x}}_{(i)}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} \\
&\overset{(\mathrm{E}_r)}{=} \frac{1}{n}\Big[\bar{\boldsymbol{x}}_{(i)}'\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} - 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{a_{(j)}}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\,\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) \\
&\quad + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})\Big] \\
&\overset{(\mathrm{E}_U)}{=} \frac{1}{n}\Big[\underbrace{\bar{\boldsymbol{z}}_{(i)}'\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{V}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})}_{\text{fixed-}k\text{ approximation}} \\
&\quad - 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{a_{(j)}}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\,\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})\Big].
\end{aligned}
$$

$$
\begin{aligned}
(II) &\overset{(d)}{=} \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r} \\
&\quad + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} \\
&\overset{(\mathrm{E}_{r,U})}{=} \underbrace{\frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'[\boldsymbol{D}_P\boldsymbol{D}_V + \boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P + \boldsymbol{P}\odot\boldsymbol{P}\odot\boldsymbol{V} - 2\boldsymbol{D}_P^2\boldsymbol{D}_V]\bar{\boldsymbol{z}}_{(j)}}_{\text{fixed-}k\text{ approximation (I)}} \\
&\quad + \underbrace{\frac{1}{n}\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) + \frac{1}{n}\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{fixed-}k\text{ approximation (II)}} \\
&\quad + \frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'\{2\boldsymbol{D}_P\boldsymbol{D}_V + 7\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P - 10\boldsymbol{D}_P\boldsymbol{D}_V\boldsymbol{D}_P + 3(\boldsymbol{V}\odot\boldsymbol{P}\odot\boldsymbol{P}) \\
&\quad - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot\boldsymbol{I} - 4\boldsymbol{D}_P^2\boldsymbol{V}\boldsymbol{D}_P - 4\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P^2 + 16\boldsymbol{D}_P^3\boldsymbol{D}_V \\
&\quad -4(\boldsymbol{V}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P - 4\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P}\odot\boldsymbol{P})\}\,\bar{\boldsymbol{z}}_{(j)} \\
&\quad + \frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_P\boldsymbol{D}_V - (\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}))) \\
&\quad - \frac{2}{n}\,\mathrm{tr}((\boldsymbol{I}\odot\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + \frac{1}{n}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) + \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}).
\end{aligned}
$$

Note that $\mathrm{tr}((\boldsymbol{I}\odot \boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) = \boldsymbol{a}'_{(i)}(\boldsymbol{P}\odot\boldsymbol{P})^2\boldsymbol{a}_{(j)}$.

$$\begin{aligned}
(III) &\overset{(d)}{=} -\frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_\varepsilon \boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon \boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} - \frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} \\
&\quad - \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} - \frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r} \\
&\overset{(\mathrm{E}_{r,U})}{=} \underbrace{-\frac{1}{n}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(i)} - \frac{1}{n}\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{fixed-}k\text{ approximation}} \\
&\quad - \frac{2}{n}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} - \frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{D}_V) \\
&\quad + \frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}).
\end{aligned}$$

Note that $\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) = \boldsymbol{a}'_{(i)}(\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{a}_{(j)}$.

$$\begin{aligned}
(IV) &\overset{(\mathrm{E}_{r,U})}{=} \underbrace{-\frac{1}{n}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)} - \frac{1}{n}\mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{fixed-}k\text{ approximation}} \\
&\quad - \frac{2}{n}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} - \frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{D}_V) \\
&\quad + \frac{2}{n}\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\,\mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}).
\end{aligned}$$

Rearranging, we find

$$\Omega_{ij}(\boldsymbol{\beta}_0) = \mathrm{E}_r\left[n\cdot S_{(i),r}(\boldsymbol{\beta}_0)S_{(j),r}(\boldsymbol{\beta}_0)\middle|\mathcal{J}\right] = \Omega^L_{ij}(\boldsymbol{\beta}_0) + \Omega^H_{ij}(\boldsymbol{\beta}_0),$$

where $\Omega^L_{ij}(\boldsymbol{\beta}_0)$ includes all terms labeled "fixed-k approximation", and $\Omega^H_{ij}(\boldsymbol{\beta}_0)$ includes the remaining terms. Importantly, the products $n^{-1}\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})$ cancel when adding (I)-(IV). Some further algebraic manipulations give the result in Theorem 2.

□

A.3 Proof of Corollary 1

Proof. It is clear that in Theorem 2, $\boldsymbol{\Omega}^{H,a}(\boldsymbol{\beta}_0)$ is negative semidefinite. To show that $\boldsymbol{\Omega}^{H,u}(\boldsymbol{\beta}_0) \prec \boldsymbol{O}$, observe that $\boldsymbol{\Omega}^{H,u}(\boldsymbol{\beta}_0) = -\frac{2}{n}\sum_{i=1}^n c_i\boldsymbol{\Sigma}^U_i$ where

$$\begin{aligned}
c_i &= \boldsymbol{e}'_i(\boldsymbol{D}_V\boldsymbol{D}_P(1-2\boldsymbol{D}_P) + (\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}))\boldsymbol{e}_i \\
&= V_{ii}P_{ii}(1-2P_{ii}+P_{ii}^2) + \sum_{k\neq i}^n V_{ik}^2\varepsilon_k^2P_{ki}^2 \geq V_{ii}P_{ii}(1-P_{ii})^2.
\end{aligned}$$

Since by assumption $P_{ii} \le C < 1$, $\lambda_{\min}(\boldsymbol{\Sigma}_i^U) \ge C$, and $\sum_{i=1}^n V_{ii} P_{ii} > 0$, we have that $\boldsymbol{\Omega}^{H,u}(\boldsymbol{\beta}_0) \le -(C \sum_{i=1}^n V_{ii} P_{ii}) \cdot \boldsymbol{I}_p \prec \boldsymbol{O}$.

It now suffices to prove that $\bar{\boldsymbol{Z}}'(\boldsymbol{V} \odot \boldsymbol{W})\bar{\boldsymbol{Z}} \preceq \boldsymbol{O}$. We first consider the diagonal terms of $\boldsymbol{V} \odot \boldsymbol{W}$. If $V_{ii} = 0$, the result is trivial. We therefore assume $V_{ii} > 0$. First, we make the observation that $\boldsymbol{\Omega}_{ii}(\boldsymbol{\beta}_0) \ge 0$, since it is the expectation of the squared score. This also holds in a model with $\boldsymbol{\Pi} = \boldsymbol{O}$, and $\boldsymbol{a}_{(i)} = 0$ for all (i) and $\boldsymbol{\Sigma}^U(i,i) = 1$ for some i and 0 for all $j \ne i$. We then deduce that for all i,

$$V_{ii} - 3P_{ii}V_{ii} + 4V_{ii}P_{ii}^2 - 2\sum_{j=1}^n V_{ij}^2 \varepsilon_j^2 P_{ij}^2 \ge 0. \tag{A.2}$$

Now, $w_{ii}V_{ii} = -2(V_{ii}P_{ii} - 2V_{ii}P_{ii}^2 + \sum_{j=1}^n V_{ij}^2\varepsilon_j^2 P_{ij}^2)$. To show that $w_{ii} \le 0$, we need to show that $-\frac{1}{2}w_{ii}V_{ii} = V_{ii}P_{ii} - 2V_{ii}P_{ii}^2 + \sum_{j=1}^n V_{ij}^2\varepsilon_j^2 P_{ij}^2 \ge 0$. Using (A.2), we have that

$$-\frac{1}{2}w_{ii}V_{ii} \ge V_{ii}\left(P_{ii} - 2P_{ii}^2 + \frac{1}{2} - \frac{3}{2}P_{ii} + 2P_{ii}^2\right) = \frac{1}{2}V_{ii}(1 - P_{ii}).$$

Since $V_{ii} > 0$, we can now conclude that $w_{ii} < 0$ if $P_{ii} \le C < 1$, which holds by Assumption A3.

Consider now the off-diagonal terms of $\boldsymbol{V} \odot \boldsymbol{W}$. For $i \ne j$, we have

$$w_{ij} = (P_{ii}P_{jj} + P_{ij}^2)(3 - 4(P_{ii} + P_{jj})) - 2(P_{ii} + P_{jj}) + 2(P_{ii} + P_{jj})^2$$

Suppose that $P_{ii} + P_{jj} < 3/4$. Then, $P_{ii}P_{jj} + P_{ij}^2 \le (P_{ii} + P_{jj})^2$ by using that $P_{ij}^2 \le P_{ii}P_{jj}$ and $2P_{ii}P_{jj} \le (P_{ii} + P_{jj})^2$. Defining $x_{ij} = P_{ii} + P_{jj}$, we then have that $w_{ij} \le -(4x_{ij}^2 - 5x_{ij} + 2)x_{ij} < 0$. Now suppose that $P_{ii} + P_{jj} \ge 3/4$, then

$$w_{ij} \le P_{ii}P_{jj}(3 - 4(P_{ii} + P_{jj})) - 2(P_{ii} + P_{jj}) + 2(P_{ii} + P_{jj})^2.$$

We can verify that $w_{ij} \le 0$ if $\max_{i=1,\dots,n} P_{ii} \le \frac{1}{8}[3 + 2\sqrt{2} + \sqrt{3}(4\sqrt{2} - 5)^{1/2}] \approx 0.904$.

Define $\tilde{\boldsymbol{\Pi}} = (\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})^{1/2}\boldsymbol{\Pi}$ and $\tilde{\boldsymbol{z}}_i = (\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})^{-1/2}\boldsymbol{z}_i$. Then,

$$\begin{aligned}\boldsymbol{\Pi}'\sum_{i,j}\boldsymbol{z}_i V_{ij}\boldsymbol{z}_j'\boldsymbol{\Pi}\cdot w_{ij} &= \boldsymbol{\Pi}'\sum_{i,j}\boldsymbol{z}_i\boldsymbol{z}_i'(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})^{-1}\boldsymbol{z}_j\boldsymbol{z}_j'\boldsymbol{\Pi}\cdot w_{ij} \\ &= \tilde{\boldsymbol{\Pi}}'\left(\sum_{i=1}^n \tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i' w_{ii} + \sum_{i>j}(\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j' + \tilde{\boldsymbol{z}}_j\tilde{\boldsymbol{z}}_j'\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i')\cdot w_{ij}\right)\tilde{\boldsymbol{\Pi}}.\end{aligned}$$

The first term within the brackets is a negatively weighted sum of symmetric positive semidefinite matrices and hence, negative semidefinite. The second term is a nonpositively weighted sum of symmetric positive semidefinite matrices, so that the resulting matrix is symmetric and negative semidefinite. We conclude that $\bar{\boldsymbol{Z}}'(\boldsymbol{W} \odot \boldsymbol{V})\bar{\boldsymbol{Z}} \preceq \boldsymbol{O}$. □

A.4 Proof of Theorem 3

We defer the proof of Theorem 3 to Appendix B, due to its length.

A.5 Proof of Theorem 4

A.5.1 Unbiasedness

Proof. To show that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_0)$ is unbiased conditionally on $\mathcal{J} = \{\boldsymbol{Z}_i, \varepsilon_i\}_{i=1}^n$, we start by analyzing the variance of the score. The variance estimator given in (12), evaluated at the true parameter vector $\boldsymbol{\beta}_0$, consists of the following components.

$$\begin{aligned}
\hat{\Omega}_{ij}^L(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{I} - \boldsymbol{D}_{P_\iota})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{D}_{P_\iota})\boldsymbol{x}_{(j)},\\
\hat{\Omega}_{ij}^H(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}[7\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2\\
&\qquad + 3\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P} - 4\boldsymbol{D}_P(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}) - 4(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{D}_P\\
&\qquad - 2\boldsymbol{D}_P\dot{\boldsymbol{V}} - 2\dot{\boldsymbol{V}}\boldsymbol{D}_P + 2\boldsymbol{D}_P^2\dot{\boldsymbol{V}} + 2\dot{\boldsymbol{V}}\boldsymbol{D}_{P_\beta}^2]\boldsymbol{x}_{(j)}\\
&\quad - \frac{2}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{D}_V - \boldsymbol{V} \odot \boldsymbol{P})\boldsymbol{D}_\varepsilon^2(\boldsymbol{D}_V - \boldsymbol{V} \odot \boldsymbol{P})\boldsymbol{x}_{(j)}
\end{aligned}$$

For $\hat{\Omega}_{ij}^L(\boldsymbol{\beta}_0)$, we use Assumption A4 and then the distributional equivalence from Assumption A2 to obtain

$$\begin{aligned}
\boldsymbol{x}'_{(j)}\boldsymbol{V}\boldsymbol{x}_{(i)} &= \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{a}'_{(j)}\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)} + \boldsymbol{a}'_{(j)}\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)}\\
&\overset{(d)}{=} \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{r}'\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(i)}}\boldsymbol{r}\\
&\overset{(\mathrm{E}_{r,U})}{=} \mathrm{E}[\boldsymbol{x}'_{(j)}\boldsymbol{V}\boldsymbol{x}_{(i)}|\mathcal{J}] = \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{V}) + \mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}).
\end{aligned} \tag{A.3}$$

Similarly, we obtain

$$\begin{aligned}
\boldsymbol{x}'_{(j)}\boldsymbol{D}_{P_\iota}\boldsymbol{V}\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P\boldsymbol{D}_V) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}),\\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_{P_\iota}\boldsymbol{V}\boldsymbol{D}_{P_\iota}\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}_{(j)}[\boldsymbol{D}_P\boldsymbol{D}_V + \boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P - 2\boldsymbol{D}_P^2\boldsymbol{D}_V + (\boldsymbol{P} \odot \boldsymbol{P} \odot \boldsymbol{V})]\bar{\boldsymbol{z}}_{(i)}\\
&\qquad + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_V\boldsymbol{D}_P) + \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}).
\end{aligned} \tag{A.4}$$

Aggregating these results, we see that $\mathrm{E}[\hat{\Omega}^L_{ij}(\boldsymbol{\beta}_0)|\mathcal{J}] = \Omega^L_{ij}(\boldsymbol{\beta}_0)$.

For $\hat{\Omega}^H_{ij}(\boldsymbol{\beta}_0)$, we use the following results

$$\begin{aligned}
\boldsymbol{x}'_{(j)}\boldsymbol{D}^k_P\dot{\boldsymbol{V}}\boldsymbol{D}^l_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}^k_P\dot{\boldsymbol{V}}\boldsymbol{D}^l_P\bar{\boldsymbol{z}}_{(i)}, \quad l,k=0,1,2,\\
\boldsymbol{x}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}^k_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}^k_P\bar{\boldsymbol{z}}_{(i)}, \quad k=0,1,\\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P\boldsymbol{D}_V\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P\boldsymbol{D}_V\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P\boldsymbol{D}_V)\\
&\quad + \mathrm{tr}(\boldsymbol{D}^2_P\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}),\\
\boldsymbol{x}'_{(j)}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}^2_P\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_V\boldsymbol{D}^2_P)\\
&\quad + \boldsymbol{a}'_{(j)}(\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{a}_{(i)},\\
\boldsymbol{x}'_{(j)}\big[\big((\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\big)\odot\boldsymbol{I}\big]\boldsymbol{x}_{(i)} &\overset{(\mathrm{E}_{r,U})}{=} \bar{\boldsymbol{z}}'_{(j)}[((\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}))\odot\boldsymbol{I}]\bar{\boldsymbol{z}}_{(i)}\\
&\quad + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}))\\
&\quad + \boldsymbol{a}'_{(j)}(\boldsymbol{P}\odot\boldsymbol{P})^2\boldsymbol{a}_{(i)}.
\end{aligned} \tag{A.5}$$

Aggregating these results and using symmetry shows that $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}_0)$ is a conditionally unbiased estimator for $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$.

Similarly under the null we have $\mathrm{E}[\hat{\sigma}^2_n(\boldsymbol{\beta}_0)] = \mathrm{E}[\frac{2}{k}(k-\boldsymbol{\iota}'\boldsymbol{D}^2_P\boldsymbol{\iota})] = \frac{2}{k}(\sum_{i=1}^n P_{ii} - \sum_{i=1}^n P^2_{ii}) = \frac{2}{k}(\sum_{i,j=1}^n P^2_{ij} - \sum_{i=1}^n P^2_{ii}) = \frac{2}{k}(\sum_{i\neq j} P^2_{ij}) = \sigma^2_n$ and

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{1,j+1}(\boldsymbol{\beta}_0) &= \frac{2}{\sqrt{n\cdot k}}\boldsymbol{x}'_{(j)}(\boldsymbol{D}_V-(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon}\\
&= \frac{2}{\sqrt{n\cdot k}}\bar{\boldsymbol{x}}'_{(j)}(\boldsymbol{D}_V-(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{D}_{a_{(j)}}(\boldsymbol{D}_V-(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon}\\
&\overset{(d)}{=} \frac{2}{\sqrt{n\cdot k}}[\bar{\boldsymbol{x}}'_{(j)}(\boldsymbol{D}_V-(\boldsymbol{V}\odot\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r))\boldsymbol{D}^2_r\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{r}\\
&\quad + \boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}(\boldsymbol{D}_V-(\boldsymbol{V}\odot\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r))\boldsymbol{D}^2_r\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{r}]\\
&\overset{(\mathrm{E}_{r,U})}{=} \frac{2}{\sqrt{n\cdot k}}[\mathrm{tr}(\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_\varepsilon) - \boldsymbol{\iota}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{\iota}]\\
&= \frac{2}{\sqrt{n\cdot k}}[\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P) - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P)]\\
&= \frac{2}{\sqrt{n\cdot k}}\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P)\\
&= \frac{2}{\sqrt{n\cdot k}}\,\mathrm{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}).
\end{aligned}$$

We conclude that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_0)$ is a conditionally unbiased estimator for $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$.

A.5.2 Consistency

We first show consistency of the variance estimator of the AR statistic. Under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, σ_n^2 and $\hat{\sigma}_n^2$ are identical, hence under H_0 the estimator is consistent.

Next, we consider the variance estimator of the score statistic. Define $\boldsymbol{x}_{(i),r} = \bar{\boldsymbol{z}}_{(i)} + \boldsymbol{u}_{(i)} + \boldsymbol{D}_r \boldsymbol{D}_\varepsilon \boldsymbol{a}_{(i)}$. We first observe that under many instrument sequences the variance of the score is bounded away from zero as established in Section B.2.3. Then, to show consistency of the variance estimator, we need to show for some matrix $\boldsymbol{A}_r$ that possibly depends on the vector of Rademacher random variables $\boldsymbol{r}$, that

$$n^{-2}\,\mathrm{E}[(\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r} - \mathrm{E}[\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r}|\mathcal{J}])^2|\mathcal{J}] \to_p 0. \tag{A.6}$$

For $\boldsymbol{A}_r$ we consider the general cases (a) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{A}\boldsymbol{D}_r$ and (b) $\boldsymbol{A}_r = \boldsymbol{A}$, and the specific cases (c) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}$, and (d) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r$. Cases (a) and (b) cover the consistency of the terms listed in (A.3) and (A.5) that are all of the form $\boldsymbol{x}'_{(i)}\boldsymbol{A}_r\boldsymbol{x}_{(j)}$. For all these terms $\lambda_{\max}(\boldsymbol{A}\odot\boldsymbol{A}) \le C$ *a.s.n.* and $\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}') \le C$ *a.s.n.*, which we will use repeatedly below. We frequently invoke the bound that for a random vector $\boldsymbol{w}$ with independent elements that have bounded fourth moment, we have $\mathrm{E}[(\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w} - \mathrm{E}[\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w}])^2|\boldsymbol{A}] \le C\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}')$, see for instance Whittle (1960). Cases (c) and (d) will cover the consistency of the terms in (A.4).

For (a)–(d), we decompose (A.6) into three parts that will be treated separately,

$$\begin{aligned}
&n^{-2}\,\mathrm{E}[(\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r} - \mathrm{E}[\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r}|\mathcal{J}])^2|\mathcal{J}] \\
&\quad\le 4\underbrace{n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(I)} \\
&\qquad+ 4\underbrace{n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j),r} - \mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(II)} \\
&\qquad+ 4\underbrace{n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(III)}.
\end{aligned}$$

We start with $(a.I) - (a.III)$.

$$\begin{aligned}
(a.I) &= n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}r)^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}) + n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(j)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(i)}}) \\
&\le 2\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}})\left(\frac{1}{n^2}\sum_{k=1}^n \bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^n \bar{z}^4_{(j),k}\right)^{1/2} \to_{a.s.} 0,
\end{aligned}$$

by Assumption A5. Similarly, for $(a.II)$

$$
\begin{aligned}
(a.II) &= n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)} - \mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&\le 2n^{-1}\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}})\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}[u^4_{(i),k}]^{1/2}\,\mathrm{E}[u^4_{(j),k}]^{1/2} \to_{a.s.} 0,
\end{aligned}
$$

since Assumptions A1 and A4 imply that $u_{(i),k}$ has bounded fourth moment. Finally, $(a.III) = \mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}] = 0$.

For $(b.I)$, conditional on $\mathcal{J}$ there is no randomness, so we get $\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}] = 0$. For $(b.II)$ we have

$$
\begin{aligned}
(b.II) &= n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}\boldsymbol{u}_{(j)} - \mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{E}[\sum_{k,k',l,l'} u_{(i),k}u_{(i),k'}u_{(j),l}u_{(j),l'}A_{kl}A_{k'l'}] - n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_A)^2 \\
&\le n^{-2}\boldsymbol{u}'_{(i)}\boldsymbol{D}_{u_{(i)}}(\boldsymbol{A}\odot\boldsymbol{A})\boldsymbol{D}_{u_{(j)}}\boldsymbol{u}_{(j)} + n^{-2}\boldsymbol{u}'_{(j)}\boldsymbol{D}_{u_{(i)}}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{D}_{u_{(i)}}\boldsymbol{u}_{(j)} \\
&\le n^{-1}(\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}}) + \lambda_{\max}(\boldsymbol{A}\boldsymbol{A}'))\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}[u^4_{(i),k}]^{1/2}\,\mathrm{E}[u^4_{(j),k}]^{1/2} \to_{a.s.} 0.
\end{aligned}
$$

Finally, $(b.III)$ satisfies

$$
\begin{aligned}
(b.III) &= n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{r})^2|\mathcal{J}] - \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}})^2 \\
&\le Cn^{-2}\,\mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(i)}}) \\
&\le Cn^{-2}\,\mathrm{tr}(\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}^2_\varepsilon\boldsymbol{A}'\boldsymbol{D}_\varepsilon)
\end{aligned}
$$

Using the expressions for $\boldsymbol{A}$ as in (A.3) and (A.5), we see that $(b.III) \to_{a.s.} 0$.

We continue with $(c.I) - (c.III)$.

$$
\begin{aligned}
(c.I) &= n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}^2_{V\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(i)}}) - 2n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P) \\
&\quad + n^{-2}\boldsymbol{\iota}'(\boldsymbol{P}\odot\boldsymbol{P}\odot(\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}))\boldsymbol{\iota} \\
&= n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - 2n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} \\
&\quad + n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}(\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} \\
&\le \left(\frac{1}{n^2}\sum_{k=1}^{n}\bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^{n}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{e}_k)^4\right)^{1/2} \to_{a.s.} 0,
\end{aligned}
$$

with the convergence implied by Assumption A5.

$(c.II)$ follows by analogous arguments. For $(c.III)$, we have

$$\begin{aligned}(c.III) &= n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\ &= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}]\\ &\le n^{-2}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}^2_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}) \to_{a.s.} 0.\end{aligned}$$

Proceeding with $(d.I) - (d.III)$, we have

$$(d.I) = n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}].$$

Notice that

$$\begin{aligned}n^{-1}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)} &= n^{-1}\boldsymbol{\iota}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\iota}\\ &= n^{-1}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)} + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{\iota}\\ &\quad + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_P\boldsymbol{\iota}\\ &\quad + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{\iota}.\end{aligned}$$

The second and third term after the final equality sign have expectation equal to zero. The difference of these terms from their expectation converges almost surely to zero by the same arguments as used in showing convergence of parts $(a) - (c)$. The final term has expectation $\bar{\boldsymbol{z}}'_{(i)}(\dot{\boldsymbol{V}}\odot\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})\bar{\boldsymbol{z}}_{(j)}$. Subtracting this expectation, and defining $\boldsymbol{r}_{-ij}$ as the vector $\boldsymbol{r}$ with the ith and jth element set to zero, the final term can be written as

$$\mathrm{tr}(\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}) + n^{-1}\sum_{k=1}^{n}\sum_{l\neq k} r_k r_l \boldsymbol{r}'_{-kl}\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{r}_{-kl}.$$

Squaring and taking the expectation, we get the bound

$$\begin{aligned}&\frac{2}{n^2}\,\mathrm{E}[\mathrm{tr}(\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}})^2|\mathcal{J}] + \frac{4}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{E}[(\boldsymbol{r}'_{-kl}\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{r}_{-kl})^2|\mathcal{J}]\\ &\le \frac{2}{n^2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(i)}}(\dot{\boldsymbol{V}}\odot\dot{\boldsymbol{P}}^2)\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{r})^2]\\ &\quad + \frac{4}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{tr}(\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{D}_{Pe_l}\boldsymbol{D}_{\bar{z}_{(j)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_{Pe_k})\\ &\le C\left(\frac{1}{n^2}\sum_{k=1}^{n}\bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^{n}\bar{z}^4_{(j),k}\right)^{1/2} \to_{a.s.} 0.\end{aligned}$$

$(d.II)$ follows from analogous arguments. Finally,

$$\begin{aligned}
(d.III) &= n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_{\varepsilon}\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_{\varepsilon}\boldsymbol{a}_{(j)} - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}] \\
&= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r})^2|\mathcal{J}] - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}]^2 \\
&\leq n^{-2}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}^2_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}) \to_{a.s.} 0.
\end{aligned}$$

Lastly, we consider the estimator of the covariance between the AR and the score statistic. From (10) and (13) we can bound the variance of $[\hat{\boldsymbol{\Sigma}}_{n,r}(\boldsymbol{\beta}_0)]_{1,j}$ as

$$\begin{aligned}
&\mathrm{E}[([\hat{\boldsymbol{\Sigma}}_{n,r}(\boldsymbol{\beta}_0)]_{1,j})^2|\mathcal{J}] \\
&\leq \frac{4}{nk}\,\mathrm{E}[(\mathrm{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}) - (\bar{\boldsymbol{z}}_{(j)} + \boldsymbol{D}_r\boldsymbol{D}_{\varepsilon}\boldsymbol{a}_{(j)} + \boldsymbol{u}_{(j)})'(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] \\
&\leq \frac{C}{nk}(\mathrm{E}[(\mathrm{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}) - \boldsymbol{a}'_{(j)}\boldsymbol{D}_r\boldsymbol{D}_{\varepsilon}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] \\
&\quad + \mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \mathrm{E}[(\boldsymbol{u}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]) \\
&= \frac{C}{nk}(\mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \mathrm{E}[(\boldsymbol{u}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]).
\end{aligned}$$

The first term becomes, by using the law of iterated expectations, Assumption A2 and Theorem A.1,

$$\begin{aligned}
&\frac{C}{nk}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] \\
&\leq \frac{C}{nk}(\mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_r\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]) \\
&= \frac{C}{nk}(\mathrm{E}[\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_{\varepsilon}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_{\varepsilon}\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}] \\
&\quad + \mathrm{E}[\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{r}|\mathcal{J}]) \\
&= \frac{C}{nk}(\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}^3_P\bar{\boldsymbol{z}}_{(j)}) + \boldsymbol{\varepsilon}'\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon}) \\
&= \frac{C}{nk}(\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}^3_P\bar{\boldsymbol{z}}_{(j)}) + \boldsymbol{\iota}'\boldsymbol{D}_{\varepsilon}\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{D}_{\varepsilon}\boldsymbol{\iota}) \\
&\leq \frac{C}{nk}(\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}^3_P\bar{\boldsymbol{z}}_{(j)}) + \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)}) \to_{a.s.} 0,
\end{aligned}$$

by Assumption A5. The last inequality uses that $\boldsymbol{e}'_j(\boldsymbol{V}\odot\boldsymbol{V})\boldsymbol{D}_P\boldsymbol{D}^2_{\varepsilon}\boldsymbol{\iota} = \sum_{i=1}^n V^2_{ji}P_{ii}\varepsilon^2_i \leq \sum_{i=1}^n V^2_{ji}\varepsilon^2_i = V_{jj}$, and hence,

$$\boldsymbol{\iota}'\boldsymbol{D}_{\varepsilon}\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{D}_{\varepsilon}\boldsymbol{\iota} \leq \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_{\varepsilon}\boldsymbol{D}^2_V\boldsymbol{D}_{\varepsilon}\bar{\boldsymbol{z}}_{(j)} = \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)}.$$

We conclude that, under $H_0\colon \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\hat{\boldsymbol{\Sigma}}_n$ is consistent for $\boldsymbol{\Sigma}_n$. □

Appendix B Central limit theorem

The proof of Theorem 3 is similar to the proof of Lemma A2 in Chao et al. (2012) and consists of the following steps. First, in Appendix B.1 we rewrite the statistic

$$\begin{pmatrix} \frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}) - k) \\ \sqrt{n}\cdot \boldsymbol{S} \end{pmatrix} = \boldsymbol{Y}_n \overset{(d)}{=} \boldsymbol{Y}_{nr},$$

such that it is a martingale difference array. Note that $\overset{(d)}{=}$ is defined as distributional equivalence conditional on $\boldsymbol{Z}$.

Second, in Appendix B.2 we show that, conditional on $\mathcal{J} = \{\boldsymbol{Z}_i, \varepsilon_i\}_{i=1}^n$, any linear combination of the elements in $\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr}$ converges to the same linear combination of a multivariate normally distributed random vector. That is, conditional on $\mathcal{J}$ $\boldsymbol{t}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} \to_d \boldsymbol{t}'\boldsymbol{Z}$ for any $\boldsymbol{t} \in \mathbb{R}^{p+1}$ and $\boldsymbol{Z}$ a multivariate normally distributed random variable with identity covariance matrix.

Third, in Appendix B.3 we use a version of Lebesgue's dominated convergence theorem to show that $\boldsymbol{t}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d \boldsymbol{t}'\boldsymbol{Z}$ unconditionally.

Fourth, in Appendix B.4 we invoke the Cramér-Wold theorem to conclude that $\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d \boldsymbol{Z}$ and thus that $\boldsymbol{Y}_n$ is multivariate normally distributed.

B.1 Rewriting the statistic

First we rewrite the AR statistic. In Section 2 we showed that $\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}_0) - k) \overset{(d)}{=} \frac{1}{\sqrt{k}}(\mathrm{AR}_r(\boldsymbol{\beta}_0) - k)$. Then defining

$$w_{1n,\mathrm{AR}} = \frac{2}{\sqrt{k}}P_{12}, \quad y_{in,\mathrm{AR}} = \frac{2}{\sqrt{k}}\left[\sum_{j<i} P_{ij}r_j\right]\cdot r_i,$$

we have $\frac{1}{\sqrt{k}}(\mathrm{AR}_r(\boldsymbol{\beta}_0) - k) = w_{1n,\mathrm{AR}} + \sum_{i=3}^n y_{in,\mathrm{AR}}$.

Next, we consider the score. We rewrite the first order conditions as

$$\begin{aligned}
\left.\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_h}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} &= -\frac{1}{n}\boldsymbol{x}'_{(h)}(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon} \\
&= -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{D}_{a_{(h)}}(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon}\right] \\
&\overset{(d)}{=} -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{r}\right] \\
&= -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{\Psi}^{(h)}\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\right],
\end{aligned}$$

where

$$\boldsymbol{\Psi}^{(h)} \equiv \boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}, \quad \boldsymbol{\Phi}^{(h)} \equiv \boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_{\varepsilon}.$$

We rewrite the final term as

$$\begin{aligned}
\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{r} &= \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}) + \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{\Delta}) \qquad \boldsymbol{\Delta} \equiv \boldsymbol{r}\boldsymbol{r}' - \boldsymbol{I}_n \\
&= \operatorname{tr}(\boldsymbol{\Phi}^{(h)}\boldsymbol{D}_r) + \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{\Delta}) \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{\substack{i,j,k\\ i\neq k}} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j, k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} P_{ij}\Phi^{(h)}_{jj}r_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} \varepsilon_iV_{ij}\varepsilon_j\bar{x}_{h,j}V_{jj}\varepsilon_jr_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} P_{ii}\Phi^{(h)}_{ij}r_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + 2\sum_{j\neq i} \Phi^{(h)}_{ji}P_{jj}r_i + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k.
\end{aligned}$$

Notice that $\boldsymbol{\Phi}^{(h)}\boldsymbol{P} = \boldsymbol{\Phi}^{(h)}$ and therefore $\boldsymbol{\Phi}^{(h)}(\boldsymbol{\Psi}^{(h)})' = \boldsymbol{\Phi}^{(h)}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}$. Furthermore, $\operatorname{tr}(\boldsymbol{\Psi}^{(h)}) = 0$. We conclude that,

$$\begin{aligned}
\sqrt{n}\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_h} &\overset{(d)}{=} -\frac{1}{\sqrt{n}}\sum_{j\neq i}\Phi^{(h)}_{ji}(1-2P_{jj})r_i - \frac{1}{\sqrt{n}}\sum_{j\neq i}\Psi^{(h)}_{ji}r_jr_i + \frac{1}{\sqrt{n}}\sum_{i\neq j\neq k\neq i}P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k \\
&= w^{(h)}_{1n,S} + \sum_{i=3}^{n} y^{(h)}_{in,S},
\end{aligned}$$

where we defined

$$\begin{aligned}
w^{(h)}_{1n,S} &= -\frac{1}{\sqrt{n}}\sum_{j\neq 1}\Phi^{(h)}_{j1}(1-2P_{jj})r_1 - \frac{1}{\sqrt{n}}\sum_{j\neq 2}\Phi^{(h)}_{j2}(1-2P_{jj})r_2 - \frac{1}{\sqrt{n}}\Psi^{(h)}_{[21]}r_2r_1, \\
y^{(h)}_{in,S} &= \left[-\frac{1}{\sqrt{n}}\sum_{j\neq i}\Phi^{(h)}_{ji}(1-2P_{jj}) - \frac{1}{\sqrt{n}}\sum_{j<i}\Psi^{(h)}_{[ij]}r_j + \frac{1}{\sqrt{n}}\sum_{l<j<i}A^{(h)}_{[ijl]}r_jr_l\right]\cdot r_i, \\
\Psi^{(h)}_{[ij]} &= \Psi^{(h)}_{ij} + \Psi^{(h)}_{ji}, \\
A^{(h)}_{[ijk]} &= A^{(h)}_{ijk} + A^{(h)}_{ikj} + A^{(h)}_{jik} + A^{(h)}_{jki} + A^{(h)}_{kij} + A^{(h)}_{kji}, \quad A^{(h)}_{ijk} = P_{ij}\Phi^{(h)}_{jk}
\end{aligned}$$

We have now shown the following distributional equivalence,

$$\boldsymbol{Y}_n = \begin{pmatrix} \frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}) - k) \\ \sqrt{n} \cdot \boldsymbol{S} \end{pmatrix} \overset{(d)}{=} \boldsymbol{Y}_{nr} = \begin{pmatrix} w_{1n,\mathrm{AR}} \\ \boldsymbol{w}_{1n,S} \end{pmatrix} + \sum_{i=3}^{n} \begin{pmatrix} y_{in,\mathrm{AR}} \\ \boldsymbol{y}_{in,S} \end{pmatrix}.$$

B.2 Conditional distribution of $\boldsymbol{t}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr}$

To use the Cramér-Wold theorem in Section B.4 we need to show that for any $\boldsymbol{t} \in \mathbb{R}^{p+1}$ $\boldsymbol{t}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} \to_d \boldsymbol{t}'\boldsymbol{Z}$. When $\boldsymbol{t} = \boldsymbol{0}$ the condition is trivially satisfied. Therefore, let $\boldsymbol{t} \in \mathbb{R}^{p+1} \setminus \boldsymbol{0}$ and write $\boldsymbol{t} = C\boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}$ for $\boldsymbol{\alpha} \in \mathbb{R}^{p+1} \setminus \boldsymbol{0}$. Consider $(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr}$ and define $\Xi_n = \mathrm{var}(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr}|\mathcal{J})$. Then,

$$(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} = w_{1n} + \sum_{i=3}^{n} y_{in},$$

where we define

$$\begin{aligned} w_{1n} &= \Xi_n^{-1/2}\left[c_{1n}w_{1n,\mathrm{AR}} + \boldsymbol{c}_{2n}'\boldsymbol{w}_{1n,S}\right], \\ y_{in} &= \Xi_n^{-1/2}\Bigg[-\frac{1}{\sqrt{n}}\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii}) - \frac{1}{\sqrt{n}}\sum_{j<i}(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ij]} - 2c_{1n}\gamma_n P_{ij})r_j \\ &\qquad + \frac{1}{\sqrt{n}}\sum_{l<j<i}\boldsymbol{c}_{2n}'\boldsymbol{a}_{[ijl]}r_l r_j\Bigg] \cdot r_i, \end{aligned} \tag{B.1}$$

where $\boldsymbol{c}_n = (c_{1n}, \boldsymbol{c}_{2n}')' = \boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\alpha}$, $0 < \boldsymbol{\alpha}'\boldsymbol{\alpha} \leq C$, $\boldsymbol{\phi}_{ji} = (\Phi_{ji}^{(1)}, \ldots, \Phi_{ji}^{(p)})'$, $\boldsymbol{\psi}_{[ji]} = (\Psi_{[ji]}^{(1)}, \ldots, \Psi_{[ji]}^{(p)})'$, $\boldsymbol{a}_{[ijk]} = (A_{[ijk]}^{(1)}, \ldots, A_{[ijk]}^{(p)})$ and $\gamma_n = \frac{\sqrt{n}}{\sqrt{k}}$. Notice that $\boldsymbol{c}_n'\boldsymbol{c}_n \leq C$, which implies $c_{1n}^2 \leq C$ and $\boldsymbol{c}_{2n}'\boldsymbol{c}_{2n} \leq C$.

For later purposes, it will be useful to write the bracketed term in y_{in} in matrix notation. Define $\boldsymbol{S}_{i-1}$ as the $n \times n$ matrix with in the left-upper $i-1 \times i-1$ block the identity matrix and zeroes elsewhere. Define

$$\boldsymbol{\Psi} = \boldsymbol{M}\boldsymbol{D}_{\sum_{h=1}^{p} c_{2n,h}a_{(h)}}\boldsymbol{P}, \quad \boldsymbol{\Phi} = \boldsymbol{D}_{\sum_{h=1}^{p} c_{2n,h}\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_{\varepsilon}$$

then we can write

$$\begin{aligned} y_{in} &= \Xi_n^{-1/2}\Bigg\{-\frac{1}{\sqrt{n}}\boldsymbol{c}_{2n}'\bar{\boldsymbol{X}}'(\boldsymbol{I}_n - 2\boldsymbol{D}_P)\dot{\boldsymbol{V}}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i \\ &\qquad - \frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{S}_{i-1}\left[(\boldsymbol{\Psi} + \boldsymbol{\Psi}' - 2\boldsymbol{D}_{\Psi}) - 2c_{1n}\gamma_n\dot{\boldsymbol{P}}\right]\boldsymbol{e}_i + \frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r}\Bigg\} \cdot r_i, \end{aligned} \tag{B.2}$$

$$\boldsymbol{A}_{-i} = \boldsymbol{S}_{i-1}\boldsymbol{A}_i\boldsymbol{S}_{i-1} = \boldsymbol{S}_{i-1}[\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i} + \boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}} + \boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi} - \boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}]\boldsymbol{S}_{i-1}.$$

To see that the last term in (B.2) equals the last term of y_{in} in (B.1) note that $\boldsymbol{A}_{-i}$ consists of the sum of three matrices with zero diagonal. Furthermore, the quadratic form with $\boldsymbol{r}'\boldsymbol{S}_{i-1}$ selects only the upper left block of the matrix. By splitting the sums into the part stemming from the upper and lower triangular parts we get for the first term

$$\begin{aligned}\boldsymbol{r}'\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i}\boldsymbol{S}_{i-1}\boldsymbol{r} &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} P_{jl}\bar{x}_{(h),l}V_{li}\varepsilon_i r_j r_l + \sum_{l<j<i} P_{lj}\bar{x}_{(h),l}V_{ji}\varepsilon_i r_j r_l] \\ &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} A^{(h)}_{jli} r_j r_l + \sum_{l<j<i} A^{(h)}_{lji} r_j r_l],\end{aligned}$$

the second term

$$\begin{aligned}\boldsymbol{r}'\boldsymbol{S}_{i-1}\boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}\boldsymbol{r} &= \sum_{h=1}^{p} c_{2n,h}\boldsymbol{r}'\boldsymbol{S}_{i-1}\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{D}_{Ve_i}\boldsymbol{\varepsilon}_i(\boldsymbol{D}_{\varepsilon}\boldsymbol{V}\boldsymbol{D}_{\varepsilon} - \boldsymbol{D}_{\varepsilon}\boldsymbol{D}_{V}\boldsymbol{D}_{\varepsilon})\boldsymbol{S}_{i-1}\boldsymbol{r} \\ &= \sum_{h=1}^{p} c_{2n,h}\boldsymbol{r}'\boldsymbol{S}_{i-1}\boldsymbol{D}_{\varepsilon}\boldsymbol{D}_{Ve_i}\boldsymbol{\varepsilon}_i(\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_{\varepsilon} - \boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{D}_{V}\boldsymbol{D}_{\varepsilon})\boldsymbol{S}_{i-1}\boldsymbol{r} \\ &= \sum_{h=1}^{p} c_{2n,h}\boldsymbol{r}'\boldsymbol{S}_{i-1}(\boldsymbol{D}_{e_i'P}\boldsymbol{\Phi}^{(h)} - \boldsymbol{D}_{e_i'P}\boldsymbol{D}_{\Phi^{(h)}})\boldsymbol{S}_{i-1}\boldsymbol{r} \\ &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} P_{ij}\bar{x}_{(h),j}V_{jl}\varepsilon_l r_j r_l + \sum_{l<j<i} P_{il}\bar{x}_{(h),l}V_{lj}\varepsilon_j r_j r_l] \\ &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} A^{(h)}_{ijl} r_j r_l + \sum_{l<j<i} A^{(h)}_{ilj} r_j r_l],\end{aligned}$$

and third term

$$\begin{aligned}&\boldsymbol{r}'\boldsymbol{S}_{i-1}(\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}^{(h)} - \boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi^{(h)}})\boldsymbol{S}_{i-1}\boldsymbol{r} \\ &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} P_{ji}\bar{x}_{(h),i}V_{il}\varepsilon_l r_j r_l + \sum_{l<j<i} P_{li}\bar{x}_{(h),i}V_{ij}\varepsilon_j r_j r_l] \\ &= \sum_{h=1}^{p} c_{2n,h}[\sum_{l<j<i} A^{(h)}_{jil} r_j r_l + \sum_{l<j<i} A^{(h)}_{lij} r_j r_l].\end{aligned}$$

Furthermore note that $\boldsymbol{A}_{-i}$ is a symmetric matrix.

We will now show that $(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}_n$ converges to a standard normally distributed random variable. As in Chao et al. (2012) we first show that $w_{1n} = o_p(1)$ such that we can focus on $\sum_{i=3}^{n} y_{in}$. Next, we check conditions of the martingale difference array CLT.

B.2.1 $w_{1n} = o_p(1)$ unconditionally

Consider w_{1n} as defined in (B.1). Following Chao et al. (2012), we show that $w_{1n} = o_p(1)$ by showing that $\mathrm{E}[\|w_{1n}\|^4|\mathcal{J}] \to_{a.s.} 0$. To bound the terms from $\boldsymbol{w}_{1n,S}$ we need the following three bounds. First, using Assumption A5, we have that

$$
\begin{aligned}
\frac{1}{n^2}\sum_{j=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_j\varepsilon_j\|^4 &\leq \frac{1}{n^2}\max_{l=1,\ldots,n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_l\varepsilon_l\|^2\sum_{j=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_j\varepsilon_j\|^2 \\
&\leq \frac{o_{a.s.}(1)}{n}\sum_{h=1}^{p}\sum_{j=1}^{n}(\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_j\varepsilon_j)^2 \\
&= \frac{o_{a.s.}(1)}{n}\sum_{h=1}^{p}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{V}\bar{\boldsymbol{Z}}\boldsymbol{e}_h \\
&\leq \frac{o_{a.s.}(1)}{n}\lambda_{\max}(\boldsymbol{V})\sum_{h=1}^{p}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\bar{\boldsymbol{Z}}\boldsymbol{e}_h \to_{a.s.} 0,
\end{aligned}
\tag{B.3}
$$

by Assumption A5 and where $o_{a.s.}(1)$ is a term converging to zero $a.s.$

Second, under Assumption A5 and by the finite fourth moment of the elements of $\boldsymbol{U}$ following from Assumption A1, we have

$$
\begin{aligned}
\mathrm{E}\left[\frac{1}{n^2}\sum_{i,j=1}^{n}\|\phi_{ij}\|^4\middle|\mathcal{J}\right] &= \mathrm{E}\left[\frac{1}{n^2}\sum_{i,j=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_iV_{ij}\varepsilon_j\|^4\middle|\mathcal{J}\right] \\
&\leq \mathrm{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4V_{ii}^2\middle|\mathcal{J}\right] \\
&\leq \mathrm{E}\left[\frac{C}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4\middle|\mathcal{J}\right] \\
&\leq \frac{C}{n^2}\sum_{i=1}^{n}\left(\|\bar{\boldsymbol{Z}}'\boldsymbol{e}_i\|^4 + \mathrm{E}[\|\boldsymbol{U}'\boldsymbol{e}_i\|^4|\mathcal{J}]\right) \to_{a.s.} 0.
\end{aligned}
\tag{B.4}
$$

Third, as the rows of $\boldsymbol{U}$ are independent and by Theorem 2 in Whittle (1960),

$$
\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\|^4|\mathcal{J}] &\leq \frac{C}{n^2}\sum_{h=1}^{p}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{u}_{(h)}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i)^4|\mathcal{J}] \\
&\leq \frac{C}{n^2}\sum_{h=1}^{p}\mathrm{E}[u_{(h),j}^4|\mathcal{J}]\sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}^2\boldsymbol{D}_\varepsilon\boldsymbol{e}_i)^2 \\
&\leq \frac{C}{n^2}\sum_{h=1}^{p}\mathrm{E}[u_{(h),j}^4|\mathcal{J}]\mathrm{tr}(\boldsymbol{V}^2) \to_{a.s.} 0.
\end{aligned}
\tag{B.5}
$$

Now we have that, since $\boldsymbol{c}_{2n}'\boldsymbol{c}_{2n} \leq C$ and using the definition of $\boldsymbol{w}_{1n,S}$,

$$
\begin{aligned}
&\mathrm{E}\left[\|\boldsymbol{c}_{2n}'\boldsymbol{w}_{1n,S}\|^4\middle|\mathcal{J}\right] \leq C\cdot\mathrm{E}\left[\|\boldsymbol{w}_{1n,S}\|^4\middle|\mathcal{J}\right]\\
&= C\,\mathrm{E}\left[\|\frac{-1}{\sqrt{n}}\sum_{j\neq 1}\boldsymbol{\phi}_{j1}(1-2P_{jj})r_1 - \frac{1}{\sqrt{n}}\sum_{j\neq 2}\boldsymbol{\phi}_{j2}(1-2P_{jj})r_2 - \frac{1}{\sqrt{n}}\boldsymbol{\psi}_{[21]}r_2r_1\|^4\middle|\mathcal{J}\right]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\left[\|\sum_{j\neq 1}\boldsymbol{\phi}_{j1}(1-2P_{jj})\|^4 + \|\sum_{j\neq 2}\boldsymbol{\phi}_{j2}(1-2P_{jj})\|^4 + \|\boldsymbol{\psi}_{[21]}\|^4\middle|\mathcal{J}\right]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\Big[\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_1\varepsilon_1\|^4 + \|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{e}_1\varepsilon_1\|^4 + \|\boldsymbol{\phi}_{11}(1-2P_{11})\|^4\\
&\quad + \|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_2\varepsilon_2\|^4 + \|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{e}_2\varepsilon_2\|^4 + \|\boldsymbol{\phi}_{22}(1-2P_{22})\|^4 + p\max_{h=1,\ldots,p}(\Psi_{[21]}^{(h)})^4\Big|\mathcal{J}\Big]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\left[\sum_{j=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{e}_j\varepsilon_j\|^4 + \|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{e}_j\varepsilon_j\|^4 + \|\boldsymbol{\phi}_{11}\|^4 + \|\boldsymbol{\phi}_{22}\|^4 + C\middle|\mathcal{J}\right]\\
&\to_{a.s.} 0,
\end{aligned}
$$

where for the final line we use (B.3), (B.4), (B.5), Assumption A5 and that

$$
\Psi_{jk}^{(h)} = \boldsymbol{e}_j'\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{e}_k \leq \boldsymbol{e}_j\boldsymbol{M}\boldsymbol{e}_j\boldsymbol{e}_k'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}^2}\boldsymbol{P}\boldsymbol{e}_k \leq \max_{i=1,\ldots,n} a_{(h),i}^2 \leq C \quad a.s.n. \quad \text{(B.6)}
$$

with the second inequality by $P_{ii} < 1$ $a.s.n.$

For the part of w_{1n} due to the AR statistic, we have

$$
\mathrm{E}[\|c_{1n}w_{1n,\mathrm{AR}}\|^4|\mathcal{J}] = \frac{16\cdot c_{1n}^4}{k^2}P_{12}^4 \leq \frac{C}{k^2}(\sum_{i=1}^{n}P_{1i}^2)^2 \leq \frac{C}{k^2}P_{11}^2 \to_{a.s.} 0.
$$

As in the proof of Lemma A2 in Chao et al. (2012), the above results imply that $w_{1n} = c_{1n}w_{1n,\mathrm{AR}} + \boldsymbol{c}_{2n}'\boldsymbol{w}_{1n,S} \to_p 0$ unconditionally, and hence,

$$
(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n = \sum_{i=3}^{n} y_{in} + o_p(1).
$$

B.2.2 Martingale difference sequence

Define the σ-fields $\mathcal{F}_{i,n} = \sigma(r_1,\ldots,r_i)$ such that $\mathcal{F}_{i-1,n} \subset \mathcal{F}_{i,n}$. It is clear that, $\mathrm{E}[y_{in}|\mathcal{J},\mathcal{F}_{i-1,n}] = 0$, due to the r_i that multiplies all the terms. Hence, conditional on $\mathcal{J}$, $\{y_{in},\mathcal{F}_{i,n}, 1\leq i\leq n, n\geq 3\}$ is a martingale difference array.

B.2.3 Variance bounded away from zero

For our statistic to be well defined we require the existence of $\boldsymbol{\Sigma}_n^{-1}$ almost surely. We start by considering a quadratic form of $\boldsymbol{\Omega}$ defined in (9) in Theorem 2. Let $\boldsymbol{v}$ be any p dimensional vector satisfying $\boldsymbol{v}'\boldsymbol{v} = 1$. Then,

$$\boldsymbol{v}'\boldsymbol{\Omega}\boldsymbol{v} \geq \sum_{j=1}^{n} c_j \boldsymbol{v}'\boldsymbol{\Sigma}_j^U \boldsymbol{v},$$

where from (9) we have $c_i = V_{jj} - 3V_{jj}P_{jj} + 4V_{jj}P_{jj}^2 - 2\sum_{k=1}^n V_{jk}^2\varepsilon_k^2 P_{kj}^2$.

Then since $\boldsymbol{\Sigma}_j^U$ is positive definite and using that for $k \neq j$, $P_{kj}^2 \leq P_{jj}(1-P_{jj})$

$$\begin{aligned} c_j &= V_{jj}(1-3P_{jj}+4P_{jj}^2) - 2\left(V_{jj}P_{jj}^3 + \sum_{k\neq j} V_{jk}^2\varepsilon_k^2 P_{kj}^2\right) \\ &\geq V_{jj}(1-3P_{jj}+4P_{jj}^2) - 2\left(V_{jj}P_{jj}^3 + P_{jj}(1-P_{jj})\sum_{k\neq j} V_{jk}^2\varepsilon_k^2\right) \\ &= V_{jj}(1-3P_{jj}+4P_{jj}^2) - 2\left(V_{jj}P_{jj}^3 + P_{jj}(1-P_{jj})(V_{jj}-V_{jj}P_{jj})\right) \\ &= V_{jj}(1-5P_{jj}+8P_{jj}^2-4P_{jj}^3) \\ &= V_{jj}(1-P_{jj})(1-2P_{jj})^2. \end{aligned} \tag{B.7}$$

Consider the case where the inequality holds with equality, such for every k either $P_{kj}^2 = P_{jj}(1-P_{jj})$ or $V_{jk}^2\varepsilon_k^2 = 0$. The last condition cannot hold for every $k \neq j$, because $\sum_{k\neq j} V_{jk}^2\varepsilon_k^2 = \sum_{k=1}^n V_{jk}^2\varepsilon_k^2 - V_{jj}P_{jj} = \boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{D}_\varepsilon^2\boldsymbol{V}\boldsymbol{e}_j - V_{jj}P_{jj} = V_{jj}(1-P_{jj})$. Now since $P_{jj} < 1$ by Assumption A3 and $V_{jj} > 0$ by Assumption A5, we conclude that there must be at least one k such that $V_{jk}^2\varepsilon_k^2 \neq 0$. For such a k, equality therefore only obtains if $P_{kj}^2 = P_{jj}(1-P_{jj})$. Assume moreover that $P_{jj} = 1/2$, so that $P_{kj} = \pm\frac{1}{2}$. Using that $\boldsymbol{P}$ is a projection matrix, this implies that $P_{kk} = 1/2$, $P_{k'j}^2 = 0$ for $k' \neq k$ and $P_{j'k} = 0$ for $j' \neq j$. However, this simply means that two columns of $\boldsymbol{D}_\varepsilon\boldsymbol{Z}$ coincide up to their sign. This case is excluded by Assumption A5 that states that $\lambda_{\min}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z}/n) \geq C$ $a.s.n.$ We then have

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^n c_j \boldsymbol{v}'\boldsymbol{\Sigma}_j^U\boldsymbol{v} &\geq \frac{\lambda_{\min}(\boldsymbol{\Sigma}_i^U)}{n}\sum_{j=1}^n c_j \geq \frac{C}{n}\operatorname{tr}(\boldsymbol{V}) \geq \frac{C\operatorname{tr}(\boldsymbol{Z}\boldsymbol{Z}')}{n\,\lambda_{\max}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})} \\ &\geq \frac{Ck\,\lambda_{\min}(\boldsymbol{Z}'\boldsymbol{Z})}{n\,\lambda_{\max}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z})} > 0, \quad a.s.n. \end{aligned}$$

by Assumption A5 and because $k/n > 0$.

Now let $\boldsymbol{b} = [\boldsymbol{\Sigma}]_{2:p+1,1}$ the covariance between the AR statistic and the score.

Then $\det(\boldsymbol{\Sigma}_n) = \det(\boldsymbol{\Omega})\det(\boldsymbol{\Omega} - \boldsymbol{b}\boldsymbol{b}'\sigma_n^{-2})$ by Schur complements. The $(i,j)^{\text{th}}$ element in $\boldsymbol{b}\boldsymbol{b}'\sigma_n^{-2}$ is the covariance of the AR statistic with i^{th} and j^{th} element of the score divided by the variance of the AR statistic. Hence this is equal to the correlation of the AR statistic with the i^{th} and j^{th} element of the score statistic times the standard deviations of the i^{th} and j^{th} element of the score statistic. Let $\boldsymbol{\rho}$ be the vector of correlations between the AR statistic and the score. That is, $\rho_i = \text{corr}(1/\sqrt{k}(\text{AR}(\boldsymbol{\beta}) - k), 1/\sqrt{n}S_{(i)}|\mathcal{J})$. Then

$$\begin{aligned}\det(\boldsymbol{\Sigma}_n) &= \det(\boldsymbol{\Omega})\det(\boldsymbol{\Omega} - \boldsymbol{b}\boldsymbol{b}'\sigma_n^{-2}) \\ &= \det(\boldsymbol{\Omega})\det(\boldsymbol{I} + \boldsymbol{D}_\rho)\det(\boldsymbol{\Omega})\det(\boldsymbol{I} - \boldsymbol{D}_\rho) > 0,\end{aligned}$$

if $\rho_i \neq \pm 1$ for all i. We now prove that this is indeed the case. Consider first the variance of the score. We have with $[\boldsymbol{D}_c]_{ii} = c_i$ from (B.7)

$$\Omega_{ii} \geq \frac{1}{n}\left[\text{tr}(\boldsymbol{\Sigma}^U_{(i,i)}\boldsymbol{D}_c) + \boldsymbol{a}'_{(j)}(\boldsymbol{D}_P - \boldsymbol{P}\odot\boldsymbol{P})(\boldsymbol{I} - 2(\boldsymbol{D}_P - (\boldsymbol{P}\odot\boldsymbol{P})))\boldsymbol{a}_{(j)}\right].$$

Define $\boldsymbol{\Delta} = \boldsymbol{D}_P - \boldsymbol{P}\odot\boldsymbol{P}$. Then, for the squared correlation coefficient, we get

$$\rho_i^2 = 2\frac{(n^{-1}\boldsymbol{a}'_{(i)}\boldsymbol{\Delta}\boldsymbol{D}_P\boldsymbol{\iota})^2}{n^{-1}\text{tr}(\boldsymbol{\Delta})[n^{-1}\text{tr}(\boldsymbol{\Sigma}^U_{(i,i)}\boldsymbol{D}_c) + n^{-1}\boldsymbol{a}'_{(i)}\boldsymbol{\Delta}(\boldsymbol{I} - 2\boldsymbol{\Delta})\boldsymbol{a}_{(i)}]}.$$

As this is a correlation coefficient, we have $|\rho_i^2| \leq 1$ and this holds even if $\text{tr}(\boldsymbol{\Sigma}^U_{(i,i)}\boldsymbol{D}_c)$ is arbitrarily small. However, as $n^{-1}\text{tr}(\boldsymbol{\Sigma}^U_{(i,i)}\boldsymbol{D}_c) \geq C > 0$ *a.s.n.*, we have $|\rho_i^2| \leq C < 1$ *a.s.n.* We conclude that $\boldsymbol{\Sigma}_n^{-1}$ exists *a.s.n.*

B.2.4 Lyapunov condition

In this section we show that the martingale difference array $\{y_{in}, \mathcal{F}_{i,n}, 1 \leq i \leq n, n \geq 3\}$ satisfies the following Lyapunov condition

$$\begin{aligned}\sum_{i=3}^n \text{E}[y_{in}^4|\mathcal{J}] \leq C\Xi_n^{-2}\sum_{i=3}^n &\underbrace{\text{E}\left[\left(-\frac{1}{\sqrt{n}}\sum_{j\neq i}\boldsymbol{c}'_{2n}\boldsymbol{\phi}_{ji}(1-2P_{ii})r_i\right)^4\middle|\mathcal{J}\right]}_{\text{linear}} \\ &+ \underbrace{\text{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{j<i}(\boldsymbol{c}'_{2n}\boldsymbol{\psi}_{[ij]} - 2c_{1n}\gamma_n P_{ij})r_j r_i\right)^4\middle|\mathcal{J}\right]}_{\text{quadratic}} \\ &+ \underbrace{\text{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{l<j<i}\boldsymbol{c}'_{2n}\boldsymbol{a}_{[ijk]}r_l r_j r_i\right)^4\middle|\mathcal{J}\right]}_{\text{cubic}} \to_{a.s.} 0.\end{aligned} \tag{B.8}$$

Since the variance $\boldsymbol{\Sigma}_n$ was shown in Section B.2.3 to be bounded away from zero, Ξ_n^{-2} is finite, as $\Xi_n = \operatorname{var}(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\boldsymbol{Y}_{nr}|\mathcal{J}) = (\boldsymbol{\alpha}'\boldsymbol{\alpha})\operatorname{var}(w_{1n} + \sum_{i=3}^n y_{in}|\mathcal{J}) = (\boldsymbol{\alpha}'\boldsymbol{\alpha})(1 + o_{a.s.}(1)) > 0$. We now subsequently consider the linear, quadratic and cubic terms in (B.8).

Linear term For the term linear in $\boldsymbol{r}$, we have that

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{n^2}\sum_{i=3}^n\left(\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii})\right)^4\middle|\mathcal{J}\right] \\
&\leq \mathrm{E}\left[\frac{C}{n^2}\sum_{i=3}^n(1-2P_{ii})^4\|\sum_{j=1}^n\boldsymbol{\phi}_{ji}-\boldsymbol{\phi}_{ii}\|^4\middle|\mathcal{J}\right] \\
&\leq \mathrm{E}\left[\frac{C}{n^2}\sum_{i=3}^n\left(\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\|^4+\|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\|^4+\|\boldsymbol{\phi}_{ii}\|^4\right)\middle|\mathcal{J}\right]\to_{a.s.}0,
\end{aligned}$$

since $(1-2P_{ii})^2 < 1$ and by Assumption A5, (B.3), (B.4) and (B.5).

Quadratic term For the term quadratic in $\boldsymbol{r}$ in (B.8), we first notice that

$$\frac{1}{n^2}\sum_{i=3}^n\mathrm{E}\left[\|\sum_{j<i}\gamma_nP_{ij}r_ir_j\|^4\middle|\mathcal{J}\right]\leq\frac{\gamma_n^2}{n^2}\sum_{i=3}^n\left(\sum_{j<i}P_{ij}^4+3\sum_{\substack{(j,m)<i\\ j\neq m}}P_{ij}^2P_{im}^2\right)\leq C\frac{k}{nk}\to 0.$$

Similarly,

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=3}^n\mathrm{E}\left[\|\sum_{k<i}\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}r_ir_k\|^4\middle|\mathcal{J}\right] \\
&\leq\frac{C}{n^2}\sum_{i=3}^n\sum_{k<i}\sum_{l<i}\sum_{m<i}\sum_{s<i}|\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[il]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[im]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[is]}|\,\mathrm{E}[r_kr_lr_mr_s|\mathcal{J}] \\
&\leq\frac{C}{n^2}\sum_{i=3}^n\left(\sum_{k<i}\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}\right)^4+3\sum_{\substack{(k,m)<i\\ k\neq m}}\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}\right)^2\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[im]}\right)^2\right) \\
&\leq\frac{C}{n^2}\sum_{i=3}^n\left(\sum_{k<i}\sum_{h=1}^p\left(\Psi_{[ik]}^{(h)}\right)^4+3\sum_{\substack{(k,m)<i\\ k\neq m}}\sum_{h=1}^p\left(\Psi_{[ik]}^{(h)}\right)^2\sum_{h=1}^p\left(\Psi_{[im]}^{(h)}\right)^2\right).
\end{aligned} \tag{B.9}$$

To bound this expression, note that by (B.6) $\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{e}_i \leq C\ a.s.n.$ Also, for any vector $\boldsymbol{v}$, $(\boldsymbol{\Psi}^{(h)}\boldsymbol{v})^2 = \boldsymbol{v}'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{v} \leq \max_{i=1,\dots,n} a_{(h),i}^2 \cdot \boldsymbol{v}'\boldsymbol{P}\boldsymbol{v}$. This implies that $\sum_{i=1}^{n}(\Psi_{ij}^{(h)})^2 \leq \max_{i=1,\dots,n} a_{(h),i}^2 \cdot P_{jj} \leq C \quad a.s.n.$ Then,

$$\begin{aligned}\frac{1}{n^2}\sum_{i,k=1}^{n}(\Psi_{ik}^{(h)})^4 &\leq \frac{1}{n^2}\sum_{i,k=1}^{n}\left(\sum_{j=1}^{n}(\Psi_{jk}^{(h)})^2\right)(\Psi_{ik}^{(h)})^2 \\ &\leq \frac{1}{n^2}\max_{j=1,\dots,n} a_{(h),j}\sum_{k=1}^{n}P_{kk}\sum_{i=1}^{n}(\Psi_{ik}^{(h)})^2 \\ &\leq \frac{1}{n^2}\max_{j=1,\dots,n} a_{(h),j}^2\sum_{k=1}^{n}P_{kk}^2 \leq \frac{Ck}{n^2} \to_{a.s.} 0.\end{aligned}$$

Using this result, we have for the first term on the final line of (B.9)

$$\begin{aligned}\frac{1}{n^2}\sum_{i=3}^{n}\sum_{k<i}\left(\Psi_{[ik]}^{(h)}\right)^4 &\leq \frac{1}{n^2}\sum_{k,i=1}^{n}\left(\Psi_{ik}^{(h)}+\Psi_{ki}^{(h)}\right)^4 \\ &\leq \frac{C}{n^2}\sum_{k,i=1}^{n}\left[(\Psi_{ik}^{(h)})^4+(\Psi_{ki}^{(h)})^4\right] \leq \frac{Ck}{n^2} \to_{a.s.} 0.\end{aligned}$$

For the second term on the final line of (B.9), we have

$$\begin{aligned}\frac{1}{n^2}\sum_{i=3}^{n}\sum_{\substack{(k,m)<i\\ k\neq m}}\left(\Psi_{[ik]}^{(h)}\right)^2\left(\Psi_{[im]}^{(h)}\right)^2 &\leq \frac{C}{n^2}\sum_{i,k,m=1}^{n}\left((\Psi_{ik}^{(h)})^2+(\Psi_{ki}^{(h)})^2\right)\left((\Psi_{im}^{(h)})^2+(\Psi_{mi}^{(h)})^2\right) \\ &\leq \frac{C}{n^2}\sum_{i,k,m=1}^{n}\Big[(\Psi_{ik}^{(h)})^2(\Psi_{im}^{(h)})^2+(\Psi_{ik}^{(h)})^2(\Psi_{mi}^{(h)})^2 \\ &\qquad +(\Psi_{ki}^{(h)})^2(\Psi_{im}^{(h)})^2+(\Psi_{ki}^{(h)})^2(\Psi_{mi}^{(h)})^2\Big].\end{aligned}$$

We now show almost sure convergence to zero of the sums over the four terms within the brackets. First,

$$\begin{aligned}\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ik}^{(h)})^2(\Psi_{im}^{(h)})^2 &\leq \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}(\Psi_{ik}^{(h)})^2\right)^2 \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\left(\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i\right)^2 \\ &\leq \frac{1}{n^2}\operatorname{tr}(\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}) \leq \frac{Ck}{n^2} \to_{a.s.} 0.\end{aligned}$$

For the second, and likewise for the third term,

$$\begin{aligned}\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ik}^{(h)})^2(\Psi_{mi}^{(h)})^2 &\leq \frac{1}{n^2}\sum_{i=1}^{n}\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{\Psi}^{(h)}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}|\boldsymbol{e}_i'\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{e}_i||\boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{e}_i| \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}|\max_{j=1,\ldots,n}a_{(h),j}^2|^2 P_{ii} \leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0.\end{aligned}$$

For the fourth and final term we have,

$$\begin{aligned}\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ki}^{(h)})^2(\Psi_{mi}^{(h)})^2 \leq \frac{1}{n^2}\sum_{i,k=1}^{n}(\Psi_{ki}^{(h)})^2\sum_{m=1}^{n}(\Psi_{mi}^{(h)})^2 &\leq \frac{1}{n^2}\sum_{i=1}^{n}\max_{j=1,\ldots,n}a_{(h),j}^4 P_{ii}^2 \\ &\leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0.\end{aligned}$$

Consequently, the quadratic term in (B.8) converges to zero almost surely.

Cubic term From (B.2), the cubic term can be written as,

$$\sum_{i=3}^{n}\mathrm{E}\left[\left(\frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r}r_i\right)^4\middle|\mathcal{J}\right] = \sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r})^4|\mathcal{J},\boldsymbol{U}]|\mathcal{J}].$$

As $\boldsymbol{A}_{-i}$ is symmetric with zeroes on its diagonal, we have by Item 4 of Theorem A.1

$$\begin{aligned}&\sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r})^4|\mathcal{J},\boldsymbol{U}]|\mathcal{J}] \\ &\leq \sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[]12\,\mathrm{tr}(\boldsymbol{A}_{-i}^2)^2 + 48\,\mathrm{tr}(\boldsymbol{A}_{-i}^4) + 32\boldsymbol{\iota}'(\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i})\boldsymbol{\iota}|\mathcal{J}] \\ &\leq \sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[92\,\mathrm{tr}(\boldsymbol{A}_{-i}^2)^2|\mathcal{J}].\end{aligned} \tag{B.10}$$

The second inequality follows since $\boldsymbol{A}_{-i}^2$ is p.s.d., hence $\mathrm{tr}(\boldsymbol{A}_{-i}^4) \leq \mathrm{tr}(\boldsymbol{A}_{-i}^2)^2$, and

$$\begin{aligned}\boldsymbol{\iota}'(\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i})\boldsymbol{\iota} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}\boldsymbol{e}_j)^4 &\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i(\boldsymbol{e}_i'\boldsymbol{A}_{-i}\boldsymbol{e}_j)^2 \\ &= \sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i)^2 \leq \left(\sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i)\right)^2 = \mathrm{tr}(\boldsymbol{A}_{-i}^2)^2.\end{aligned}$$

The final line of (B.10) can be further bounded as

$$\begin{aligned}\sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_{-i}^2)|\mathcal{J}] &= \sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}[\mathrm{tr}(\boldsymbol{S}_{i-1}\boldsymbol{A}_i\boldsymbol{S}_{i-1}\boldsymbol{A}_i\boldsymbol{S}_{i-1})|\mathcal{J}] \\ &\leq \sum_{i=3}^{n}\frac{C}{n^2}\mathrm{E}\left[\mathrm{tr}([\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i}]^2 + [\boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}}]^2\right. \\ &\qquad \left. + [\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}]^2 - [\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}]^2)\big|\mathcal{J}\right].\end{aligned} \tag{B.11}$$

To bound these four terms we use the following result

$$\begin{aligned}\frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i)^2\,|\mathcal{J}] &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i|\mathcal{J}] \\ &= \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}^2_{\sum_{h=1}^{p}c_{2n,h}\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}^2_{\sum_{h=1}^{p}c_{2n,h}\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i|\mathcal{J}] \\ &\leq \frac{C}{n^2}\sum_{h=1}^{p}\sum_{i=1}^{n}\mathrm{E}[\bar{x}^4_{(h),i}|\mathcal{J}] \\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{e}_i\|^4 + \mathrm{E}[\|\boldsymbol{U}'\boldsymbol{e}_i\|^4|\mathcal{J}] \to_{a.s.} 0,\end{aligned} \tag{B.12}$$

by Assumption A5 and the finite fourth moment of the elements of $\boldsymbol{U}$. For the first and second term of (B.11), we have by (B.12)

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i})|\mathcal{J}] &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}\left[\left(\sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{P}\boldsymbol{D}^2_{\Phi e_i}\boldsymbol{P}\boldsymbol{e}_j\right)^2\middle|\mathcal{J}\right] \\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}\left[\left(\sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{D}^2_{\Phi e_i}\boldsymbol{e}_j\right)^2\middle|\mathcal{J}\right] \to_{a.s.} 0.\end{aligned}$$

For the third term of (B.11), also by (B.12),

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi})|\mathcal{J}] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{P})|\mathcal{J}] \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[(P_{ii}\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i)^2\,|\mathcal{J}] \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i)^2\,|\mathcal{J}] \to_{a.s.} 0.\end{aligned}$$

And finally, for the fourth term of (B.11), (B.12) implies that

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}|\mathcal{J}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left[\sum_{k=1}^{n}P_{ki}^2(\Phi_{ki})^2\right]^2|\mathcal{J}]$$
$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left[\sum_{k=1}^{n}P_{ii}^2(\Phi_{ki})^2\right]^2|\mathcal{J}]$$
$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[[\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i]^2|\mathcal{J}] \rightarrow_{a.s.} 0.$$

Hence the cubic term converges to zero almost surely. Therefore, the Lyapunov condition is satisfied.

B.2.5 Converging conditional variance

This part of the proofs shows the following convergence result: for any $\epsilon > 0$,

$$\mathrm{P}\left(\left|\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\boldsymbol{r}_{-i},\mathcal{J}] - s_n^2\right| \geq \epsilon \middle| \mathcal{J}\right) \rightarrow_{a.s.} 0.$$

We start by noting that,

$$s_n^2 = \mathrm{E}\left[\left(\sum_{i=3}^{n}y_{in}\right)^2\middle|\mathcal{J}\right] = \mathrm{E}[((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n + o_{a.s.}(1))^2|\mathcal{J}] = 1 + o_{a.s.}(1),$$

where the vanishing part is due to w_{1n}. We can conclude that s_n^2 is bounded and bounded away from zero in probability. Now define $\boldsymbol{r}_{-i} = r_1, \ldots, r_{i-1}$ and write y_{in} in (B.2) as $y_{in} = \Xi_n^{-1/2}(y_{in}^{(1)} + y_{in}^{(2)} + y_{in}^{(1)})$ with

$$y_{in}^{(1)} = \frac{-1}{\sqrt{n}}\boldsymbol{c}_{2n}'\bar{\boldsymbol{X}}'(\boldsymbol{I}_n - 2\boldsymbol{D}_P)\dot{\boldsymbol{V}}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i r_i,$$
$$y_{in}^{(2)} = \frac{-1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{S}_{i-1}\left[(\boldsymbol{\Psi} + \boldsymbol{\Psi}' - 2\boldsymbol{D}_{\Psi}) - 2c_{1n}\gamma_n\dot{\boldsymbol{P}}\right]\boldsymbol{e}_i r_i,$$
$$y_{in}^{(3)} = \frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r}r_i.$$

Using a conditional version of Chebyshev's inequality, we have

$$
\begin{aligned}
&\mathrm{P}\left(\left|\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\boldsymbol{r}_{-i},\mathcal{J}]-s_n^2(\mathcal{J})\right|\geq\epsilon\middle|\mathcal{J}\right)\\
&=\mathrm{P}\left(\left|\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\mathcal{J}]\right|\geq\epsilon\middle|\mathcal{J}\right)\\
&\leq\frac{1}{\epsilon^2}\,\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[y_{in}^2|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\leq\frac{C}{\epsilon^2}\,\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)}+y_{in}^{(2)}+y_{in}^{(3)})^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)}+y_{in}^{(2)}+y_{in}^{(3)})^2|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\leq\frac{C}{\epsilon^2}\left\{\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})^2|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\right.\\
&\quad+\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(2)})^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(2)})^2|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\quad+\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(3)})^2|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(3)})^2|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\quad+\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})(y_{in}^{(2)})|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})(y_{in}^{(2)})|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\quad+\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})(y_{in}^{(3)})|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(1)})(y_{in}^{(3)})|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\\
&\quad\left.+\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(2)})(y_{in}^{(3)})|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(2)})(y_{in}^{(3)})|\mathcal{J}]\right)^2\middle|\mathcal{J}\right]\right\}
\end{aligned}
\tag{B.13}
$$

Each of these terms converges to zero almost surely. For illustration we show for the cross product between the quadratic and the cubic how this follows below. For the other terms we do this in a separate document that is available upon request.

Define $\hat{\boldsymbol{\Phi}}=\boldsymbol{D}_{\sum_{h=1}^{p}c_{2n,h}\bar{Z}_{(h)}}\boldsymbol{V}\boldsymbol{D}_{\varepsilon}$ and $\hat{\boldsymbol{A}}_{-i}$ as $\boldsymbol{A}_{-i}$ but with $\hat{\boldsymbol{\Phi}}$ instead of $\boldsymbol{\Phi}$ and similarly for other variables that contain $\boldsymbol{\Phi}$. The product between the quadratic and cubic term in (B.13) then is

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{i=3}^{n}\mathrm{E}[(y_{in}^{(2)})(y_{in}^{(3)})|\boldsymbol{r}_{-i},\mathcal{J}]-\sum_{i=3}^{n}\mathrm{E}((y_{in}^{(2)})(y_{in}^{(3)})|\mathcal{J})\right)^2\middle|\mathcal{J}\right]\\
&\leq\frac{C}{n^2}\mathrm{E}\left[\left(\sum_{i=3}^{n}\boldsymbol{r}'\boldsymbol{S}_{i-1}\left[(\boldsymbol{\Psi}+\boldsymbol{\Psi}'-2\boldsymbol{D}_{\Psi})-2c_{1n}\gamma_n\dot{\boldsymbol{P}}\right]\boldsymbol{e}_i\boldsymbol{r}'\boldsymbol{S}_{i-1}\mathrm{E}(\boldsymbol{A}_i|\mathcal{J})\boldsymbol{S}_{i-1}\boldsymbol{r}\right)^2\middle|\mathcal{J}\right]\\
&=\frac{C}{n^2}\mathrm{E}\left[\left(\sum_{i=3}^{n}\boldsymbol{r}'\boldsymbol{S}_{i-1}\left[(\boldsymbol{\Psi}+\boldsymbol{\Psi}'-2\boldsymbol{D}_{\Psi})-2c_{1n}\gamma_n\dot{\boldsymbol{P}}\right]\boldsymbol{e}_i\boldsymbol{r}'\boldsymbol{S}_{i-1}\hat{\boldsymbol{A}}_i\boldsymbol{S}_{i-1}\boldsymbol{r}\right)^2\middle|\mathcal{J}\right]\\
&\leq\frac{C}{n^2}\left(\mathrm{E}\left[\left(\sum_{i=3}^{n}\boldsymbol{r}'\boldsymbol{S}_{i-1}(\boldsymbol{\Psi}+\boldsymbol{\Psi}'-2\boldsymbol{D}_{\Psi})\boldsymbol{e}_i\boldsymbol{r}'\boldsymbol{S}_{i-1}\hat{\boldsymbol{A}}_i\boldsymbol{S}_{i-1}\boldsymbol{r}\right)^2\middle|\mathcal{J}\right]\right.\\
&\qquad\left.+\mathrm{E}\left[\left(\sum_{i=3}^{n}\boldsymbol{r}'\boldsymbol{S}_{i-1}2c_{1n}\gamma_n\dot{\boldsymbol{P}}\boldsymbol{e}_i\boldsymbol{r}'\boldsymbol{S}_{i-1}\hat{\boldsymbol{A}}_i\boldsymbol{S}_{i-1}\boldsymbol{r}\right)^2\middle|\mathcal{J}\right]\right),
\end{aligned}
$$

due to the odd number of Rademacher random variables in $y_{in}^{(1)}y_{in}^{(2)}$. We only bound the second term. The first term follows by similar arguments and an eigenvalue bound on $\boldsymbol{\Psi}+\boldsymbol{\Psi}'-2\boldsymbol{D}_{\Psi}$. By completing the square we obtain

$$
\begin{aligned}
&\frac{C}{nk}\mathrm{E}[(\sum_{i,j=3}^{n}\boldsymbol{r}'\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{j-1}\boldsymbol{r}\boldsymbol{r}'\hat{\boldsymbol{A}}_{-i}\boldsymbol{r}\boldsymbol{r}'\hat{\boldsymbol{A}}_{-j}\boldsymbol{r}|\mathcal{J}]\\
&=\frac{C}{nk}\sum_{i,j=3}^{n}\mathrm{tr}(\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{j-1}[20(\hat{\boldsymbol{A}}_{-i}\odot\hat{\boldsymbol{A}}_{-j})-6\boldsymbol{I}\odot(\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j}+\hat{\boldsymbol{A}}_{-j}\hat{\boldsymbol{A}}_{-i})\\
&\quad+4\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j}+4\hat{\boldsymbol{A}}_{-j}\hat{\boldsymbol{A}}_{-i}+\mathrm{tr}(\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j})\boldsymbol{I}])\\
&=\frac{C}{nk}\sum_{i,j=3}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{j-1}[20(\hat{\boldsymbol{A}}_{-i}\odot\hat{\boldsymbol{A}}_{-j})-6\boldsymbol{I}\odot(\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j}+\hat{\boldsymbol{A}}_{-j}\hat{\boldsymbol{A}}_{-i})+4\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j}\\
&\quad+4\hat{\boldsymbol{A}}_{-j}\hat{\boldsymbol{A}}_{-i}+\mathrm{tr}(\hat{\boldsymbol{A}}_{-i}\hat{\boldsymbol{A}}_{-j})\boldsymbol{I}]\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i,
\end{aligned}
$$

by Item 7 of Theorem A.1. For sake of space, we focus on the first term, which can be written as

$$\begin{aligned}
&\frac{C}{nk}\sum_{i,j=3}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{j-1}(\hat{\boldsymbol{A}}_{-i}\odot\hat{\boldsymbol{A}}_{-j})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&=\frac{C}{nk}\sum_{i,j=3}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{j-1}\boldsymbol{S}_{i-1}(\hat{\boldsymbol{A}}_{i}\odot\hat{\boldsymbol{A}}_{j})\boldsymbol{S}_{i-1}\boldsymbol{S}_{j-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&=\frac{C}{nk}[2\sum_{i=3}^{n}\sum_{j=i+1}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\hat{\boldsymbol{A}}_{i}\odot\hat{\boldsymbol{A}}_{j})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i+\sum_{i=3}^{n}\boldsymbol{e}_i\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\hat{\boldsymbol{A}}_{i}\odot\hat{\boldsymbol{A}}_{i})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i]\\
&=\frac{C}{nk}[2\sum_{i=3}^{n}\sum_{j=i+1}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\sum_{r=1}^{4}\hat{\boldsymbol{A}}_{i}^{(r)}\odot\sum_{s=1}^{4}\hat{\boldsymbol{A}}_{j}^{(s)})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&\quad+\sum_{i=3}^{n}\boldsymbol{e}_i'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\sum_{r=1}^{4}\hat{\boldsymbol{A}}_{i}^{(r)}\odot\sum_{s=1}^{4}\hat{\boldsymbol{A}}_{i}^{(s)})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i],
\end{aligned}$$

where $\hat{\boldsymbol{A}}_i^{(r)}$ for $r=1,\dots,4$ are the four terms between the $\boldsymbol{S}_{i-1}$ in $\boldsymbol{A}_{-i}$ from (B.2), but with $\boldsymbol{\Phi}$ substituted by $\hat{\boldsymbol{\Phi}}$. Again, only consider the first term, which consist of 16 cross products for the different r and s. Let $\sum_{j=i+1}^{n}\boldsymbol{e}_j'\boldsymbol{e}_j=\boldsymbol{I}_n-\boldsymbol{S}_{i-1}-\boldsymbol{e}_i'\boldsymbol{e}_i=\tilde{\boldsymbol{I}}_{in}$. Then for $r=1,s=1$

$$\begin{aligned}
&\frac{C}{nk}\sum_{i=3}^{n}\sum_{j=i+1}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\hat{\boldsymbol{A}}_{i}^{(1)}\odot\hat{\boldsymbol{A}}_{j}^{(1)})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&=\frac{C}{nk}\sum_{i=3}^{n}\sum_{j=i+1}^{n}\boldsymbol{e}_j'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\boldsymbol{D}_{\hat{\Phi}e_i}\dot{\boldsymbol{P}}\odot\boldsymbol{D}_{\hat{\Phi}e_j}\dot{\boldsymbol{P}})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&=\frac{C}{nk}\sum_{i=3}^{n}\sum_{j=i+1}^{n}\sum_{k,l<i}\dot{P}_{jk}\hat{\Phi}_{ki}\hat{\Phi}_{kj}\dot{P}_{kl}^{2}\dot{P}_{li}\\
&=\frac{C}{nk}\sum_{i=3}^{n}\sum_{k<i}\boldsymbol{e}_k'\dot{\boldsymbol{P}}\tilde{\boldsymbol{I}}_{in}\hat{\boldsymbol{\Phi}}'\boldsymbol{e}_k\boldsymbol{e}_k'\hat{\boldsymbol{\Phi}}\boldsymbol{e}_i\boldsymbol{e}_k'(\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&=\frac{C}{nk}\sum_{i=3}^{n}\boldsymbol{e}_i'\hat{\boldsymbol{\Phi}}'\boldsymbol{D}_{\dot{P}\tilde{I}_{in}\hat{\Phi}'}\boldsymbol{S}_{i-1}(\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i\\
&\le\frac{C}{nk}\sum_{i=3}^{n}[\boldsymbol{e}_i'\hat{\boldsymbol{\Phi}}'\boldsymbol{D}^2_{\dot{P}\tilde{I}_{in}\hat{\Phi}'}\hat{\boldsymbol{\Phi}}\boldsymbol{e}_i\boldsymbol{e}_i'\dot{\boldsymbol{P}}\boldsymbol{S}_{i-1}(\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})\boldsymbol{S}_{i-1}(\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})\boldsymbol{S}_{i-1}\dot{\boldsymbol{P}}\boldsymbol{e}_i]^{\frac{1}{2}}\\
&\le\frac{C}{nk}\sum_{i=3}^{n}[\boldsymbol{e}_i'\boldsymbol{P}(\sum_{h=1}^{p}\boldsymbol{D}_{\bar{z}_{(h)}})\boldsymbol{D}^2_{\dot{V}\tilde{I}_{in}P(\sum_{h=1}^{p}D_{\bar{z}_{(h)}})}(\sum_{h=1}^{p}\boldsymbol{D}_{\bar{z}_{(h)}})\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\dot{\boldsymbol{P}}^2\boldsymbol{e}_i]^{\frac{1}{2}},
\end{aligned}$$

by the Cauchy-Schwarz inequality twice and because $\lambda_{\max}((\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}})(\dot{\boldsymbol{P}}\odot\dot{\boldsymbol{P}}))\le C$. Now since for any $n\times n$ matrix $\boldsymbol{A}$ we have $\lambda_{\max}(\boldsymbol{D}_A^2)=\max_{j=1,\dots,n}(\boldsymbol{e}_j'\boldsymbol{A}\boldsymbol{e}_j)^2=\max_{j=1,\dots,n}\boldsymbol{e}_j'\boldsymbol{A}'\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}\boldsymbol{e}_j\le\max_{j=1,\dots,n}\boldsymbol{e}_j'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{e}_j\le\lambda_{\max}(\boldsymbol{A}'\boldsymbol{A})$ and $\lambda_{\max}(\dot{\boldsymbol{V}})\le C$

we have that $\lambda_{\max}(\boldsymbol{D}^2_{\dot{V}\tilde{I}_{in}P(\sum_{h=1}^p D_{\bar{z}_{(h)}})}) \leq C\,\lambda_{\max}([\sum_{h=1}^p \boldsymbol{D}_{\bar{z}_{(h)}}]^2) = C\max_{j=1,\ldots,n}\|\bar{\boldsymbol{z}}_i\|^2$. Therefore the equation above becomes

$$\frac{C}{nk}\sum_{i=3}^{n}[\boldsymbol{e}_i'\boldsymbol{P}(\sum_{h=1}^{p}\boldsymbol{D}_{\bar{z}_{(h)}})\boldsymbol{D}^2_{\dot{V}\tilde{I}_{in}P(\sum_{h=1}^p D_{\bar{z}_{(h)}})}(\sum_{h=1}^{p}\boldsymbol{D}_{\bar{z}_{(h)}})\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\dot{\boldsymbol{P}}^2\boldsymbol{e}_i]^{\frac{1}{2}}$$
$$\leq \frac{C}{nk}\sum_{i=3}^{n}[(\max_{j=1,\ldots,n}\|\bar{\boldsymbol{z}}_i\|^2)^2\boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\dot{\boldsymbol{P}}^2\boldsymbol{e}_i]^{\frac{1}{2}} \leq \frac{C}{nk}\sum_{i=3}^{n}\max_{j=1,\ldots,n}\|\bar{\boldsymbol{z}}_i\|^2\boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{e}_i \to_{a.s.} 0,$$

by Assumption A5. The other combinations of r and s can be shown to converge to zero using similar arguments. Continuing like this we can show that (B.13) converges to zero almost surely.

B.3 Unconditional distribution of $\boldsymbol{t}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n$ by Lebesgue's dominated convergence theorem

To obtain the unconditional distribution, note that for some $\epsilon > 0$, say $\epsilon = 1$, we have

$$\sup_n \mathrm{E}([|\,\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} < y|\mathcal{J})|^{1+\epsilon}]$$
$$= \sup_n \mathrm{E}[(\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} < y|\mathcal{J}))^2] \leq \sup_n \mathrm{E}[1^2] \leq \infty.$$

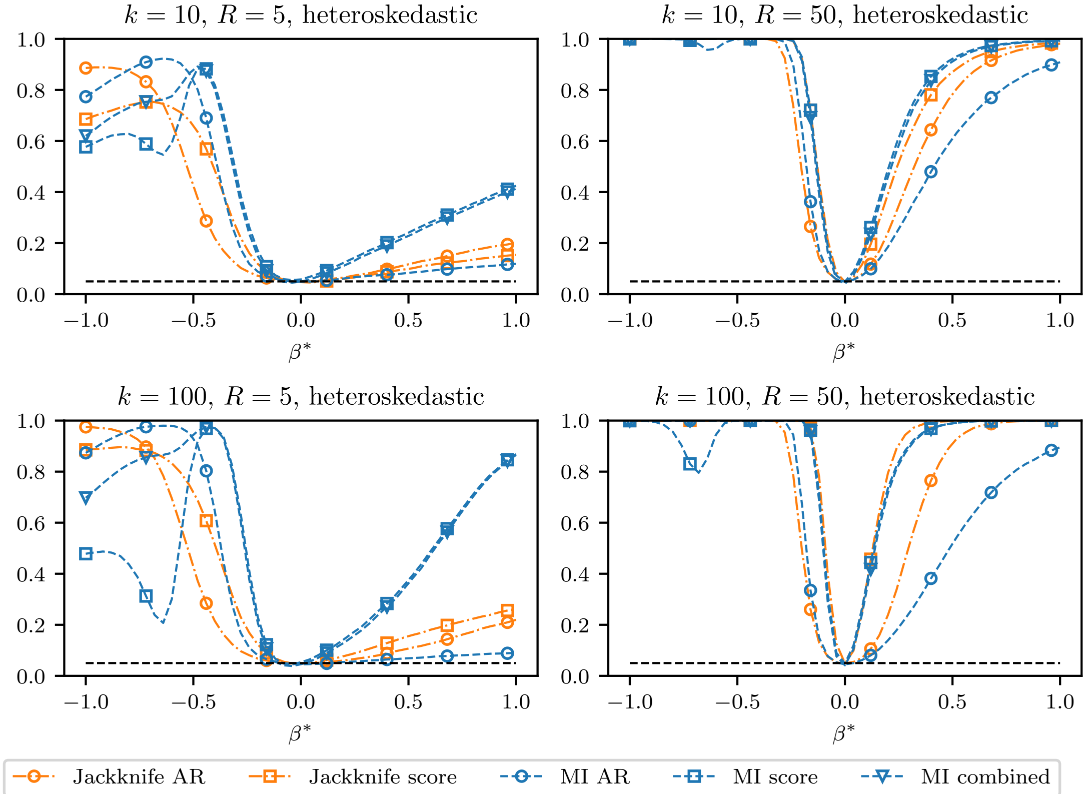
Therefore, $\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} < y|\mathcal{J})$ is uniformly integrable (Billingsley, 1995, p. 338) and we can apply a version of Lebesgue's dominated convergence theorem (Billingsley, 1995, Theorem 25.12)

$$\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y) = \mathrm{E}[\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y|\mathcal{J})]$$
$$= \mathrm{E}[\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_{nr} < y|\mathcal{J})] \to_{a.s.} \mathrm{E}[\Phi(y)] = \Phi(y).$$

B.4 Distribution of $\boldsymbol{Y}_n$ by the Cramér-Wold theorem

We have shown that for any $\boldsymbol{\alpha}$ we have $(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{Z}$, with $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_{p+1})$. Then also $C(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d C(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{Z}$ and by the Cramér-Wold theorem (Billingsley, 1995, T29.4) $\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d \boldsymbol{Z}$. □

Figure 10: Power under identification robust inference for $a_i = |Z_{i1}|$.



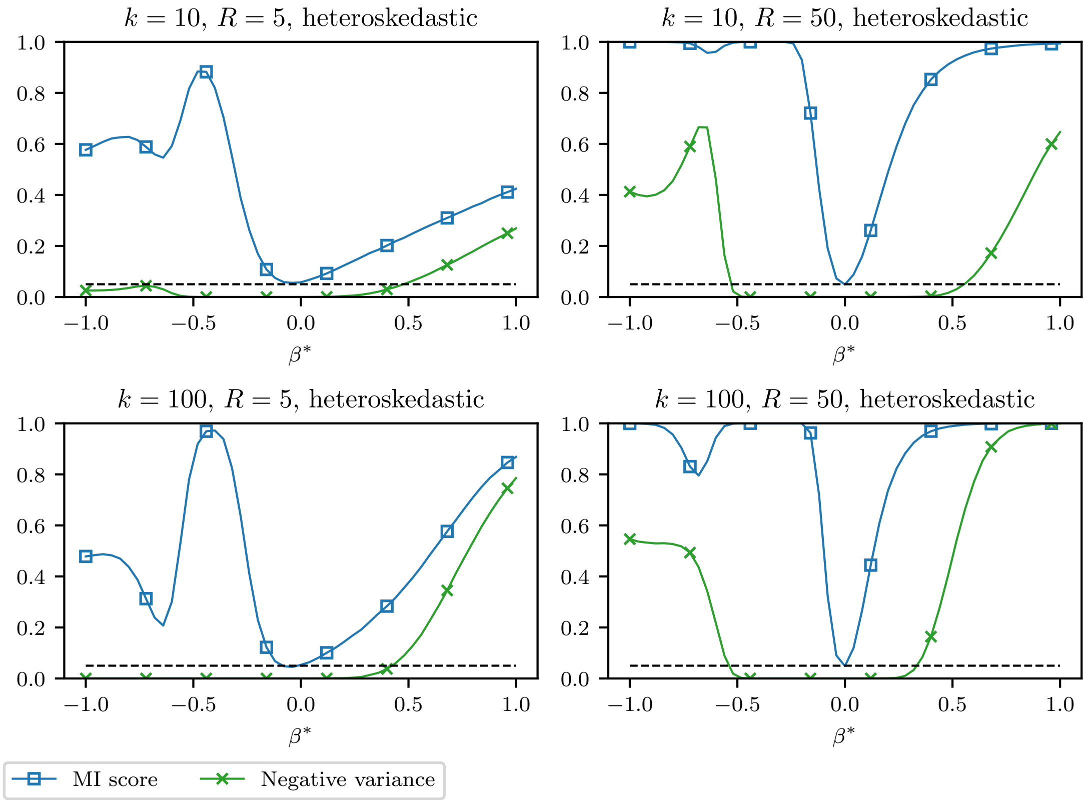
Note: power when testing $H_0 : \beta = 0$ when the true $\beta = \beta^*$ at $\alpha = 0.05$ based on the jackknife Anderson-Rubin test without crossfit variance (Crudu et al., 2021; Mikusheva and Sun, 2022), the jackknife score test without crossfit variance (Matsushita and Otsu, 2022) and the tests developed here. k denotes the number of instruments, R their strength and the invariance assumption is satisfied. The combined test uses $\alpha_{\text{AR}} = 0.01$. The Monte Carlo is described in Section 6.

Appendix C Additional simulation results

C.1 Power comparison with jackknife tests

Figure 10 shows the power of the jackknife AR by Crudu et al. (2021) and Mikusheva and Sun (2022) and the jackknife score by Matsushita and Otsu (2022) both without crossfit variance for the DGPs described in Section 6. The panels show that there is no clear ordering in power of the tests. For the AR test we find that the jackknife approach delivers higher power when β^* is positive, while the continuous updating based AR delivers higher power when β^* is negative. For the score, we generally find higher power for the continuous updating based score, although this has a power dip for negative values of β^*.

Figure 11: Power of the many instrument robust score test and fraction of negative variance estimates for heteroskedastic data.



Note: power when testing $H_0 : \beta = 0$ when the true $\beta = \beta^*$ at $\alpha = 0.05$ based on the many instrument robust score test together with the fraction of negative variance estimates. k denotes the number of instruments, R their strength and the invariance assumption is satisfied. The Monte Carlo is described in Section 6.

C.2 Power derived from negative variances

In this subsection we show the power that the many instrument robust score test obtains due to rejecting the null when finding a negative variance. We plot the rejection rate and the fraction of negative variances in Figure 11 for the heteroskedastic case where $a_i = |Z_{i1}|$. We see that we only observe negative variance estimates under the alternatives, thus increasing the power without affecting the size. Moreover, the number negative variance estimates increases with the distance between the true and hypothesized value of β and thus counteracts the decrease in power against distant alternatives from which score tests generally suffer.

Unreported simulation results show that for the case of a single instrument and weak or irrelevant instruments negative variances also occur when the null hypothesis is satisfied. Consequently, the many instrument robust score test becomes oversized in these cases. This is also what we observe in the upper right panel of Figure 5. There the many instrument robust score test has rejection rates slightly above the desired five percent line. This is not surprising as we only show consistency of the variance estimator under many-instrument sequences.

Appendix D Details of the applications

Table 2 shows the exact values of the confidence intervals and the point estimates for the Card (2009) application that are shown graphically in Figure 1.

Table 2: 95% confidence interval and point estimates for the negative inverse elasticity of substitution between immigrants and natives.

	High school equivalent workers			College equivalent workers		
	Lower	Point	Upper	Lower	Point	Upper
OLS	-0.0406	-0.0297	-0.0188	-0.0728	-0.0576	-0.0424
Bartik	-0.0569	-0.0367	-0.0164	-0.0999	-0.0734	-0.0420
2SLS	-0.0482	-0.0363	-0.0244	-0.0794	-0.0621	-0.0448
Fixed-k AR	-0.1141		0.0072	-0.1593		0.0289
MI AR	-0.0899		0.0008	-0.1342		0.0000
Fixed-k score	-0.0905		0.0001	-0.1275		-0.0227
MI score	-0.0754		-0.0021	-0.1172		-0.0312
MI combined	-0.0759		-0.0011	-0.1181		-0.0307

Note: 95% confidence intervals for the negative inverse elasticity of substitution between natives and immigrants. Confidence intervals and point estimates are constructed using (i) OLS (ii) 2SLS with the Bartik instrument (iii) 2SLS (iv) the fixed-k AR statistic (v) the fixed-k score statistic by (vi-viii) the tests developed here.