Identification-robust inference under many instruments, heteroskedasticity and invariant moment conditions

Tom Boot* Johannes W. Ligtenberg†

August 29, 2022

Abstract

Identification-robust test statistics are commonly obtained via the continuous updating objective function or its score. When the number of moment conditions grows proportionally with the sample size, the asymptotic distribution of neither of these two objects is known. The main obstacle in establishing the asymptotic distribution is the appearance of a large-dimensional and asymptotically random weighting matrix. We show that when the moment conditions evaluated at the true parameter vector are reflection invariant, this obstacle can be circumvented. In a linear instrumental variables model with many instruments and heteroskedasticity, we show joint asymptotic normality of the objective function and the score statistic under the null. We find a number of additional variance terms that we can consistently estimate to restore the validity of conventional inference procedures in the presence of many instruments.

Keywords: robust inference, many instruments, heteroskedasticity, invariance.

*University of Groningen, t.boot@rug.nl
†University of Groningen, j.w.ligtenberg@rug.nl

1 Introduction

This paper considers a situation where a researcher conducts inference based on a linear instrumental variables (IV) model with (a) many instruments to increase estimation efficiency, (b) no assumptions on the validity of these instruments, and (c) heteroskedasticity. The development of inference procedures under the combination of (a), (b) and (c) has been obstructed by the large-dimensional weighting matrix that appears in identification-robust test statistics that are based on the continuous updating (CU) objective function. Examples include the Anderson-Rubin statistic, Kleibergen's (2005) K statistic, and combinations of these such as the CLR statistic by Moreira (2003).

To circumvent the obstruction posed by the weighting matrix, we show how invariance arguments can be used to obtain the (limiting) distribution of the CU objective function and its score. As a motivating example, we observe that if the moment conditions are orthogonally invariant, the finite sample distribution of the CU objective function is known 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.

We show that when the moment conditions evaluated at the true parameter vector satisfy a 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 suitable for heteroskedastic models as it allows the distribution of the moment conditions to change 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 et al. (2012).

A downside of the Anderson-Rubin 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 distribution of the score function. By deriving a new central limit theorem (CLT) for cubic forms, we show joint asymptotic normality of the score and the AR statistic under many instrument sequences. The variance of the score function is shown to contain several terms that do not appear when the number of instruments grows slower than proportionally with the sample size. We also find that under many instruments *and* heteroskedasticity, the AR statistic and the score are asymptotically dependent.

Obtaining inference that is uniformly valid over the strength of the instruments as well as the number of instruments requires careful implementation. For the AR statistic, we adjust the critical values so that the procedure automatically reduces to the standard procedure under a fixed number of instruments if the number of instruments is indeed small. For the score, we propose an estimator of the variance that reduces to that of Kleibergen (2005) when the number of instruments is small. To enhance power against irrelevant alternatives, we follow the suggestion by Kleibergen (2005) to reject the null hypothesis when the AR statistic and/or the absolute value of the score statistic are large, and control size by choosing the size of each test appropriately.

We assess the finite sample performance of the tests in a simulation that is based on the design in Hausman et al. (2012). The simulation shows that unlike conventional asymptotic approximations, the many instrument identification robust tests have excellent size control regardless of the instrument strength and regardless of the number of instruments. This contrasts with the procedures developed for a fixed number of instruments that 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.

We use our robust tests to revisit the study on the return of education by Angrist and Krueger (1991) using the 1530 instruments suggested in Mikusheva and Sun (2021). We find that the AR test provides rather wide confidence intervals that slightly increase in width when the number of instruments increases. Combining the AR test with the score test leads to confidence intervals for the return on education that are remarkably robust to the number of instruments (k) used: with $k = \{30, 180, 1530\}$ we find that the 95% confidence intervals for the return on education are $[0.05, 0.13]$, $[0.07, 0.13]$, $[0.04, 0.16]$ respectively.

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 et al. (2008) extend the scope of these results by removing the normality assumption. Anatolyev (2019) provides an extensive 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 that remain consistent under many instruments and heteroskedasticity were developed by Hausman et al. (2012), Chao et al. (2012), Chao et al. (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, consistent estimation of the parameters of interest cannot be achieved, 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 et al., 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 and Mikusheva, 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. Bekker and Kleibergen (2003) study the homoskedastic Gaussian IV model and find that under many instrument sequences the score-based statistic by Kleibergen (2002) statistic needs to be rescaled to obtain the familiar χ^2 limiting distribution.

Finally, the combination of robust inference in linear IV models with many instruments and heteroskedasticity that is considered in this paper has been studied recently by Crudu et al. (2021), Mikusheva and Sun (2021), Matsushita and Otsu (2020) and Lim et al. (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 (2020) propose a jackknife LM 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. Critical values that yield good power are then derived using the framework of Andrews (2016).

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.

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 of the CU objective function. Section 5 provides the estimators required to implement the tests and discusses the consistency and asymptotic normality of the continuous updating estimator. Section 6 contains the Monte Carlo results. The empirical application is 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. $A \overset{(\mathrm{E})}{=} B$ means that $\mathrm{E}[A] = \mathrm{E}[B]$. $\mathrm{E}_A[\cdot]$ is the expectation over the distribution of the random variable A. $\rightarrow_d$ denotes convergence in distribution, $\rightarrow_p$ convergence in probability and $\rightarrow_{a.s.}$ 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 some of our results can be applied to any model with invariant moment conditions, our main focus is on the heteroskedastic linear IV model. The model has p endogenous regressors and no other control variables, because we assume these have been partialled out before. Both the first and second stage are 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)* $(\varepsilon_i, \boldsymbol{\eta}_i')$ *is independent over* $i = 1, \ldots, n$, *has mean zero and second moment matrix* $\mathrm{E}[(\varepsilon_i, \boldsymbol{\eta}_i')'(\varepsilon_i, \boldsymbol{\eta}_i')] = \boldsymbol{\Sigma}_i = (\sigma_i^2 \quad \boldsymbol{\sigma}_{12i}'; \quad \boldsymbol{\sigma}_{12i} \quad \boldsymbol{\Sigma}_{22i})$, *(b)* $0 < C^{-1} \leq \lambda_{\min}(\boldsymbol{\Sigma}_i) \leq \lambda_{\max}(\boldsymbol{\Sigma}_i) \leq C < \infty$, *(c) For all* i, $\mathrm{E}[\varepsilon_i^4] \leq C < \infty$ *and* $\mathrm{E}[||\boldsymbol{\eta}_i||^4] \leq C < \infty$, *(d)* $\boldsymbol{z}_i$ *is independent of* $(\varepsilon_j, \boldsymbol{\eta}_j')$ *for all* (i, j).

Parts (a)–(c) are relatively mild assumptions on the first and second stage errors in (1). An alternative to part (d) would be to phrase some of our assumptions conditional on the instruments as for example in Chao et al. (2012).

To estimate $\boldsymbol{\beta}_0$, we have k moment conditions $\boldsymbol{g}_i(\boldsymbol{\beta})$ that 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 $\boldsymbol{G}(\boldsymbol{\beta}) = [\boldsymbol{g}_1(\boldsymbol{\beta}), \ldots, \boldsymbol{g}_n(\boldsymbol{\beta})]'$, such that $\mathrm{rank}(\boldsymbol{G}(\boldsymbol{\beta}_0)) = k$. Define the orthogonal projection matrix $\boldsymbol{P}(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta})(\boldsymbol{G}(\boldsymbol{\beta})'\boldsymbol{G}(\boldsymbol{\beta}))^{-1}\boldsymbol{G}(\boldsymbol{\beta})'$. The continuous updating (CU) objective function introduced by Hansen et al. (1996) can be written as

$$Q(\boldsymbol{\beta}) = \frac{1}{2n}\boldsymbol{\iota}'\boldsymbol{P}(\boldsymbol{\beta})\boldsymbol{\iota}. \tag{2}$$

Note that we can write this function as $Q(\boldsymbol{\beta}) = \frac{k}{2n} + \frac{1}{2n}\sum_{i \neq j}[\boldsymbol{P}(\boldsymbol{\beta})]_{ij}$. 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 rate at which the number of instruments grows is limited to $k^3/n \to 0$.

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

$$\text{AR}(\boldsymbol{\beta}) = 2nQ(\boldsymbol{\beta}). \tag{3}$$

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 reduces to

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

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 because the randomness does not vanish asymptotically. To circumvent this issue, Crudu et al. (2021) and Mikusheva and Sun (2021) replace the heteroskedastic weight matrix by the matrix $\boldsymbol{Z}'\boldsymbol{Z}$. In this paper, we aim to confront the AR statistic from (3) directly.

3 Invariant moment conditions

Invariance conditions are powerful tools to obtain the distribution of test statistics. In our case, the exact finite sample distribution of $Q(\boldsymbol{\beta}_0)$ can be obtained when the moment conditions are orthogonally invariant, i.e. $\boldsymbol{G}(\boldsymbol{\beta}_0) \stackrel{(d)}{=} \boldsymbol{G}(\boldsymbol{\beta}_0)\boldsymbol{Q}$ for any orthogonal matrix $\boldsymbol{Q}$. In this case, $\boldsymbol{\iota}'\boldsymbol{P}(\boldsymbol{\beta}_0)\boldsymbol{\iota}/n \stackrel{(d)}{=} \boldsymbol{z}'\boldsymbol{P}\boldsymbol{z}$ where $\boldsymbol{z}$ is uniformly distributed over the $(n-1)$-dimensional unit sphere and $\boldsymbol{P}$ can be regarded as fixed, see e.g. Vershynin (2018, Chapter 5). As a result, we have that $2Q(\boldsymbol{\beta}_0) \stackrel{(d)}{=} \frac{Z_1}{Z_1+Z_2}$ where $Z_1 \sim \chi^2(k)$ independently of $Z_2 \sim \chi^2(n-k)$. It follows that $2Q(\boldsymbol{\beta}_0) \sim \text{Beta}(k/2, (n-k)/2)$. Importantly, rotational invariance allows us to bypass the fact that the dimensions of the weighting matrix $\boldsymbol{G}(\boldsymbol{\beta}_0)'\boldsymbol{G}(\boldsymbol{\beta}_0)$ can be nonnegligible relative to the sample size.

Unfortunately, rotational invariance is restrictive. In particular, combined with independence of the moment conditions, it implies that the moment conditions are normally distributed. 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 linear IV model in (1), we impose the invariance on the second stage regression errors ε_i.

Assumption A2. *Let $\{r_i\}$ be a sequence of independent Rademacher random variables gathered in the vector $\boldsymbol{r} = (r_1, \ldots, r_n)'$. Then, $\boldsymbol{\varepsilon} \stackrel{(d)}{=} \boldsymbol{D}_r\boldsymbol{\varepsilon}$.*

Note that this assumption implies reflection invariance in the moment conditions, i.e. $\boldsymbol{G}(\boldsymbol{\beta}_0) \stackrel{(d)}{=} \boldsymbol{D}_r\boldsymbol{G}(\boldsymbol{\beta}_0)$. The results in this section in fact apply to any model with reflection invariant moment conditions. The key observation is that Assumption A2 allows the distribution of the moment conditions to differ across i. 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. Define $\boldsymbol{r} = (r_1, \ldots, r_n)'$ as a vector of independent Rademacher random variables, we then have that

$$Q(\boldsymbol{\beta}_0) \stackrel{(d)}{=} Q_r(\boldsymbol{\beta}_0) = \frac{1}{2n}\boldsymbol{r}'\boldsymbol{P}(\boldsymbol{\beta}_0)\boldsymbol{r}. \tag{5}$$

While the exact finite sample distribution of $Q(\boldsymbol{\beta}_0)$ is no longer known, the asymptotic distribution under many instrument sequences can be derived. Since $Q(\boldsymbol{\beta}_0)$ and $Q_r(\boldsymbol{\beta}_0)$ are distributionally equivalent, it suffices to analyze the asymptotic distribution $Q_r(\boldsymbol{\beta}_0)$. Likewise, the AR statistic is defined as $\text{AR}(\boldsymbol{\beta}_0) = 2nQ(\boldsymbol{\beta}_0)$, and we can analyze $\text{AR}_r(\boldsymbol{\beta}_0)$ to find its asymptotic distribution.

Conditional on the moment conditions, the only randomness in $\text{AR}_r(\boldsymbol{\beta}_0)$ comes from the Rademacher random variables, which form a quadratic form with the projection matrix $\boldsymbol{P}(\boldsymbol{\beta}_0)$. 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. *(a)* $\text{rank}[\boldsymbol{P}(\boldsymbol{\beta}_0)] = k$*, (b) for $i = 1, \ldots, n$, $P_{ii}(\boldsymbol{\beta}_0) \leq C < 1$ with probability 1 for all sufficiently large n, (c) Define*

$$\sigma_n^2 = \frac{2}{k}\sum_{i\neq j} P_{ij}(\boldsymbol{\beta}_0)^2, \tag{6}$$

and assume $\sigma_n^2 > 1/C$ with probability 1 for all sufficiently large n.

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). It is required to apply the central limit theorem provided in Lemma A2 by Chao et al. (2012) which leads to the following result.

Corollary 1. *Under Assumptions A2 and A3, when $k \to \infty$ as $n \to \infty$,*

$$\frac{\frac{1}{\sqrt{k}}(\mathrm{AR}_r(\boldsymbol{\beta}_0) - k)}{|\sigma_n|} \to_d N(0,1). \tag{7}$$

Since $\mathrm{AR}_r(\boldsymbol{\beta})$ is distributionally equivalent to $\mathrm{AR}(\boldsymbol{\beta})$, this implies that

$$\frac{\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}_0) - k)}{|\sigma_n|} \to_d N(0,1). \tag{8}$$

This corollary shows that the AR statistic, once shifted and rescaled, no longer has a $\chi^2(k)$ distribution. A similar result is obtained for the AR statistic in a homoskedastic IV model by Anatolyev and Gospodinov (2011). We note that Corollary 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 Corollary 1 requires $k \to \infty$, we can achieve uniform inference across k by testing based on the quantiles from $(\chi^2(k) - k)/\sqrt{2k}$. When k is fixed, $\sigma_n^2 \to_p 2$, and hence, we compare $\mathrm{AR}(\boldsymbol{\beta}_0)$ against a $\chi^2(k)$ distribution. When k increases, the quantiles of $(\chi^2(k)-k)/\sqrt{2k}$ approach that of the standard normal distribution and Corollary 1 applies.

4 Inference based on the score

The AR statistic is not efficient in overidentified models, and may lack power when the number of moments grows proportionally with the sample size. We therefore 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).

By using the same conditioning argument as for the AR statistic, we can obtain the limiting distribution of the first order conditions of the CU objective function. For this, we make the following additional assumption on the IV model in (1).

Assumption A4. *Consider $\boldsymbol{\eta}_i$ and ε_i as in (1). Then,*

$$\boldsymbol{\eta}_i = \varepsilon_i \boldsymbol{a}_i + \boldsymbol{u}_i, \tag{9}$$

where $\boldsymbol{a}_i = \boldsymbol{\sigma}_{21i}/\sigma_i^2$, and $\{\boldsymbol{u}_i, \varepsilon_i, \boldsymbol{z}_i\}$ are mutually independent.

This assumption parametrizes the relation between the first and second stage regression errors to a linear one. This assumption also appears in Bekker and Kleibergen (2003), and it is for example satisfied if $(\varepsilon_i, \boldsymbol{\eta}_i')$ is multivariate normal. It allows us to write $\boldsymbol{x}_i = \bar{\boldsymbol{x}}_i + \varepsilon_i \boldsymbol{a}_i$, where $\bar{\boldsymbol{x}}_i = \bar{\boldsymbol{z}}_i + \boldsymbol{u}_i$. Here $\bar{\boldsymbol{x}}_i$ does not depend on ε_i, which is useful when we apply our invariance condition later on. To increase the flexibility of the model, one could potentially allow for higher-order polynomials in ε_i in (9) at the cost of more elaborate notation.

We note that the assumption on the eigenvalues of the second moment matrix of $(\varepsilon_i, \boldsymbol{\eta}_i')$ in Assumption A1 also has implications for the second moment matrix of the errors $\boldsymbol{u}_i$ defined in Assumption A4. We see that $\boldsymbol{\Sigma}_i^U = \mathrm{E}[\boldsymbol{u}_i\boldsymbol{u}_i'] = \boldsymbol{\Sigma}_{22i} - \sigma_i^{-2}\boldsymbol{\sigma}_{12i}\boldsymbol{\sigma}_{12i}'$, i.e. the Schur complement of $\boldsymbol{\Sigma}_{22i}$. In particular, the bounds on the eigenvalues of $\boldsymbol{\Sigma}_i$ then imply that $0 < C^{-1} \leq \lambda_{\min}(\boldsymbol{\Sigma}_i^U) \leq \lambda_{\max}(\boldsymbol{\Sigma}_i^U) \leq C < \infty$. Moreover, we have $\mathrm{E}[\|u_i\|^4] \leq \mathrm{E}[\|\boldsymbol{\eta}_i\|^4] + \mathrm{E}[\varepsilon_i^4]\|\boldsymbol{a}_i\|^4 \leq C < \infty$.

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 $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{\beta}_0)$. As in Kleibergen (2005) and Newey and Windmeijer (2009), the score of the objective function is given by

$$S_{(i)}(\boldsymbol{\beta}) = \frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i} = -\frac{1}{n}\boldsymbol{x}_{(i)}'(\boldsymbol{I} - \boldsymbol{D}_{P(\boldsymbol{\beta})_L})\boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{\varepsilon}(\boldsymbol{\beta}). \tag{10}$$

Under Assumption A2, and using the decomposition of $\boldsymbol{\eta}_i$ in (9) we find that $S_{(i)}(\boldsymbol{\beta}_0) \overset{(d)}{=} S_{(i),r}(\boldsymbol{\beta}_0)$, where

$$S_{(i),r}(\boldsymbol{\beta}_0) = -\frac{1}{n}\bar{\boldsymbol{x}}_{(i)}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} - \frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}. \tag{11}$$

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 $S_{(i)}(\boldsymbol{\beta}_0)$. What is important to note here is that, 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 need to take into account the randomness that enters via $\boldsymbol{u}_i$.

The conditional expectation and variance of the score are given by the following theorem that requires only the invariance condition Assumption A2.

Theorem 1. *Define the information set $\mathcal{J} = \{\varepsilon_i, \boldsymbol{Z}_i'\}_{i=1}^n$. 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*

$$\Omega_{ij}(\boldsymbol{\beta}_0) = \mathrm{E}_r\left[n \cdot S_{(i),r}(\boldsymbol{\beta}_0)S_{(j),r}(\boldsymbol{\beta}_0)\middle|\mathcal{J}\right] = \Omega_{ij}^L(\boldsymbol{\beta}_0) + \Omega_{ij}^H(\boldsymbol{\beta}_0), \tag{12}$$

where

$$\Omega_{ij}^{L}(\boldsymbol{\beta}_0) = \frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'[\boldsymbol{V} - \boldsymbol{D}_P\boldsymbol{V} - \boldsymbol{V}\boldsymbol{D}_P + \boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P + \boldsymbol{D}_P\boldsymbol{D}_V + \dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}]\bar{\boldsymbol{z}}_{(j)}$$
$$+ \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_V - \boldsymbol{D}_V\boldsymbol{D}_P)) + \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})$$
$$- \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}), \tag{13}$$

and

$$\Omega_{ij}^{H}(\boldsymbol{\beta}_0) = \frac{1}{n}\bar{\boldsymbol{z}}_{(i)}'\Big[2\boldsymbol{D}_P\boldsymbol{D}_V + 7\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2$$
$$- 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P} \odot \boldsymbol{P}) \odot \boldsymbol{I} + 3\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}$$
$$- 4\boldsymbol{D}_P(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}) - 4(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{D}_P$$
$$- 2\boldsymbol{D}_P\boldsymbol{V} - 2\boldsymbol{V}\boldsymbol{D}_P + 2\boldsymbol{D}_P^2\boldsymbol{V} + 2\boldsymbol{V}\boldsymbol{D}_P^2\Big]\bar{\boldsymbol{z}}_{(j)} \tag{14}$$
$$- \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_P\boldsymbol{D}_V - 2\boldsymbol{D}_P^2\boldsymbol{D}_V + (\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P} \odot \boldsymbol{P})))$$
$$+ \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) + \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})$$
$$- \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P^2\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}) - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}(\boldsymbol{P} \odot \boldsymbol{P})^2\boldsymbol{D}_{a_{(j)}}).$$

Here, $\boldsymbol{\Sigma}^U(i,j)$ *is an* $n \times n$ *diagonal matrix with the* k*-th diagonal element equal to* $\operatorname{cov}(u_{ki}, u_{kj})$*.*

Proof. See Appendix A.2. □

The variance is decomposed into a component Ω_{ij}^L, which is the conditional expectation of the estimator of the variance of the score when the number of instruments does not increase with the sample size, and a component Ω_{ij}^H. The terms involving $\boldsymbol{a}_{(i)}$ and $\boldsymbol{a}_{(j)}$ cancel when $\boldsymbol{D}_{a_{(i)}} = a_{(i)} \cdot \boldsymbol{I}_n$ for all $i = 1, \ldots, p$. This is in particular true under homoskedasticity.

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)* $\frac{1}{n}\max_{i=1,\ldots,n}\|\bar{\boldsymbol{z}}_i\|^2 \to_{a.s.} 0$*, (c)* $\frac{1}{n}\max_{i=1,\ldots,n}\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\|^2 \to_{a.s.} 0$*, (d)* $\lambda_{\max}(\boldsymbol{V}) \leq C < \infty$ *a.s.n.,* $\min_{i=1,\ldots,n} V_{ii} > 0$ *a.s.n., (e)* $P_{ii} < 1/2$*.*

Part (a) and (b) are relatively standard assumptions under many instruments. (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 part (a) and (b). In particular, part (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. Under homoskedasticity, the first item in part (d) does not appear as in that case $\boldsymbol{V}$ reduces to a projection matrix and it automatically holds. Under homoskedasticity, the second item will only hold when $k/n \to \lambda > 0$ and we assume that the same is true for $\boldsymbol{V}$ as we specify it here. Hence, this limits the results to many instrument sequences. Since the elements P_{ii} are typically of order k/n, part (e) restricts the proportion of the number of instruments relative to the sample size. The condition ensures that the variance matrix of the score has its minimum eigenvalue bounded away from zero.

We can now provide the joint limiting distribution of the Anderson-Rubin statistic and the score evaluated at the true parameter $\boldsymbol{\beta}_0$.

Theorem 2. *Under Assumptions A2 to A5, when $n \to \infty$ and $k/n \to \lambda \in (0,1)$,*

$$\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)^{-1/2}\begin{pmatrix}\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}_0)-k)\\ \sqrt{n}\cdot\boldsymbol{S}(\boldsymbol{\beta}_0)\end{pmatrix} \to_d N(\boldsymbol{0},\boldsymbol{I}_{p+1}). \tag{15}$$

Here $[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{1,1} = \sigma_n^2$ from (6), $[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{2:p+1,2:p+1}$ is given by $\boldsymbol{\Omega}(\boldsymbol{\beta}_0)$ in Theorem 1, and the covariance between the rescaled AR statistic and the score is

$$[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{1,j+1} = [\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)]_{j+1,1} = \frac{2}{\sqrt{n\cdot k}}\,\mathrm{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}), \quad j=1,\dots p, \tag{16}$$

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

Proof. See Appendix B. □

We 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 a homoskedastic setting, we have $\boldsymbol{D}_{a_{(j)}} = \boldsymbol{a}_{(j)}\boldsymbol{I}_n$ and hence $\boldsymbol{\Psi}^{(j)} = \boldsymbol{O}$.

5 Implementation

5.1 An unbiased and consistent variance estimator

To use Corollary 1 and Theorem 2 for testing we require a consistent estimator for $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$, which is given by the following expression.

$$\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0) = \begin{pmatrix} \hat{\sigma}_n^2(\boldsymbol{\beta}_0) & \left[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)\right]'_{2:p,1} \\ \left[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)\right]_{2:p,1} & \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}_0) \end{pmatrix}, \tag{17}$$

with for the variance of the AR statistic $\hat{\sigma}_n^2(\boldsymbol{\beta}_0) = \frac{2}{k}(k - \boldsymbol{\iota}'\boldsymbol{D}_P^2\boldsymbol{\iota})$. We estimate the variance of the score following the decomposition in Theorem 1 as

$$\begin{aligned} \hat{\Omega}_{ij}^L(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{I} - \boldsymbol{D}_{P_\iota})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{D}_{P_\iota})\boldsymbol{x}_{(j)}, \\ \hat{\Omega}_{ij}^H(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}(2\boldsymbol{D}_P\boldsymbol{D}_V + 7\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P + 3\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P} \\ &\quad - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P} \odot \boldsymbol{P}) \odot \boldsymbol{I} - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2 \\ &\quad - 4\boldsymbol{D}_P(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P}) - 4(\dot{\boldsymbol{V}} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{D}_P \\ &\quad - 2\boldsymbol{D}_P\boldsymbol{V} - 2\boldsymbol{V}\boldsymbol{D}_P + 2\boldsymbol{D}_P\boldsymbol{D}_{P_\iota}\boldsymbol{V} + 2\boldsymbol{V}\boldsymbol{D}_{P_\iota}\boldsymbol{D}_P)\boldsymbol{x}_{(j)}. \end{aligned} \tag{18}$$

and for the covariance between the AR statistic and the score, we define for $j = 1, \ldots, p$.

$$[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)]_{1,j+1} = [\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)]_{j+1,1} = \frac{2}{\sqrt{n \cdot k}}\boldsymbol{x}'_{(j)}(\boldsymbol{D}_V - (\boldsymbol{V} \odot \boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon}. \tag{19}$$

Theorem 3. *$\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)$ is conditionally unbiased, i.e.* $\mathrm{E}[\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0)|\mathcal{J}] = \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$. *Also, under Assumptions A2 to A5,* $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_0) \to_p \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$.

Proof. See Appendix A.4. □

While the estimator is conditionally unbiased and consistent, it is not guaranteed to be positive definite. This is a general property of unbiased variance estimators. In different contexts, this is observed in the leave-one-out variance estimators by Kline et al. (2020) and the variance estimator derived by Cattaneo et al. (2018). In the one-dimensional setting, a crude way of dealing with negative variances is to set the variance equal to ∞ when this occurs, such that tests using this variance are never rejected. This is the solution we employ here.

5.2 Robust inference

With our estimator for the variance, we can use Theorem 2 to perform identification robust inference. Using the AR statistic, we obtain a confidence region for $\boldsymbol{\beta}_0$ with asymptotic coverage rate $1-\alpha$ by including all values for $\boldsymbol{\beta}$ for which $(k\hat{\sigma}_n^2)^{-1/2}(\text{AR}(\boldsymbol{\beta})-k) \leq (2k)^{-1/2}(\chi^2(k)_{1-\alpha}-k)$, where $\chi^2(k)_{1-\alpha}$ is the $1-\alpha$ quantile of a $\chi^2(k)$ distribution. With these critical values we compare the AR statistic with $\chi^2(k)$ critical values when k is small, and the recentered and rescaled AR statistic with standard normal critical values when k is large. This yields size correct inference uniformly over the number of instruments.

Similarly, we obtain a confidence region based on the score test by including all values of $\boldsymbol{\beta}$ for which $n\boldsymbol{S}(\boldsymbol{\beta})'\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\boldsymbol{S}(\boldsymbol{\beta})$ is within the $1-\alpha$ quantile of the $\chi^2(p)$ distribution. Given that the score-based test lacks power in regions away from the true value if the objective function is flat, we also combine the AR and score test using the suggestion by Kleibergen (2005) to conduct a test based on the AR statistic at significance level $\alpha_{\text{AR}} = 0.01$ and a test based on the score statistc at significance level $\alpha_S = 0.04$. We then reject the null when either or both these tests reject. Given the possible asymptotic dependence between the AR statistic and the score statistic, this may lead to a slightly conservative test.

6 Simulation results

We test the finite sample performance of the proposed tests in the Monte Carlo setup of Hausman et al. (2012) and Bekker and Crudu (2015). In particular we generate data according to the model in (1), with $n = 800$ and one endogenous regressor, that is $p = 1$. We vary the number of instruments over the grid $k = \{2, 10, 30, 100\}$. When $k = 2$ we take as instruments $\boldsymbol{z}_i = (1, z_{1i})$. Otherwise we use

$$\boldsymbol{z}_i = (1, z_{1i}, z_{1i}^2, z_{1i}^3, z_{1i}^4, z_{1i}D_{i1}, \ldots, z_{1i}D_{i,k-5})', \tag{20}$$

where $z_{1i} \sim N(0,1)$ independent across i and $D_{ij} \in \{0,1\}$ with $\text{P}(D_{ij} = 1) = \frac{1}{2}$ independent across i and j. We vary the relevance of the instruments by setting the first-stage coefficient $\Pi_2^2 = F_0\frac{\sqrt{k}}{n}$ where $F_0 = \{0, 2, 4, \ldots, 20\}$ and $\Pi_i = 0$ for $i = 1, 3, 4, \ldots, k$.

We draw the regression error η_i independently from a standard normal distribution and $\boldsymbol{\varepsilon}$ is generated as $\boldsymbol{\varepsilon} = \rho\boldsymbol{\eta} + \sqrt{\frac{1-\rho^2}{\phi^2+\psi^4}}(\phi\boldsymbol{w}_1 + \psi\boldsymbol{w}_2)$ where $\rho = 0.3$, $\phi = 1.38072$, $\psi = 0.86$, $\boldsymbol{w}_1 \sim N(\boldsymbol{0}, \boldsymbol{D}_{z_1}^2)$ and $\boldsymbol{w}_2 \sim N(\boldsymbol{0}, \psi^2\boldsymbol{I}_n)$. We generate 10,000 data sets using this data generating process.

6.1 Benchmarks

6.1.1 Fixed-k Anderson-Rubin

Under the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ and assuming that k is fixed we have that $\text{AR}(\boldsymbol{\beta}_0) \to_d \chi^2(k)$. We can therefore determine the size by comparing the statistics with $\chi^2(k)$ critical values.

6.1.2 Fixed-k score as in Kleibergen (2005)

The results by Kleibergen (2005) can be used to construct an identification-robust test that is robust against heteroskedasticity. The K-statistic is given by

$$K(\boldsymbol{\beta}_0) = n\boldsymbol{S}(\boldsymbol{\beta}_0)'[\boldsymbol{D}(\boldsymbol{\beta}_0)'\boldsymbol{V}_K^{-1}(\boldsymbol{\beta}_0)\boldsymbol{D}(\boldsymbol{\beta}_0)]^{-1}\boldsymbol{S}(\boldsymbol{\beta}_0) \to_d \chi^2(p), \tag{21}$$

with

$$\begin{aligned} \boldsymbol{V}_K(\boldsymbol{\beta}_0) &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{Z}\varepsilon_i^2 = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2\boldsymbol{Z}, \\ \boldsymbol{D}(\boldsymbol{\beta}_0)\boldsymbol{e}_j &= -\frac{1}{n}\boldsymbol{Z}'\boldsymbol{x}_{(j)} + \frac{1}{n}\boldsymbol{Z}'\boldsymbol{D}_{x_{(j)}}\boldsymbol{P}\boldsymbol{\iota}. \end{aligned} \tag{22}$$

Newey and Windmeijer (2009) show that this statistic is also robust against a slowly increasing number of instruments.

6.1.3 Mikusheva and Sun (2021)

Mikusheva and Sun (2021) develop an AR and a Wald test, which they combine using a pretest for weak identification. Under certain assumptions, that allow for many and weak instruments, their AR test converges as

$$\text{AR}_{MS} = \frac{1}{\sqrt{k\Phi(\boldsymbol{\beta}_0)}}\sum_{i=1}^{n}\sum_{j\neq i} P_{Z,ij}\varepsilon(\boldsymbol{\beta}_0)_i\varepsilon(\boldsymbol{\beta}_0)_j \to_d N(0,1), \tag{23}$$

with

$$\Phi(\boldsymbol{\beta}_0) = \frac{2}{k}\sum_{i=1}^{n}\sum_{j\neq i}\frac{P_{Z,ij}^2}{M_{Z,ii}M_{Z,jj} + M_{Z,ij}^2}[\varepsilon_i(\boldsymbol{\beta}_0)\boldsymbol{e}_i'\boldsymbol{M}_Z\boldsymbol{\varepsilon}(\boldsymbol{\beta}_0)][\varepsilon_j(\boldsymbol{\beta}_0)\boldsymbol{e}_j'\boldsymbol{M}_Z\boldsymbol{\varepsilon}(\boldsymbol{\beta}_0)], \tag{24}$$

for $\boldsymbol{M}_Z = \boldsymbol{I}_n - \boldsymbol{P}_Z$. Due to the quadratic form of AR_{MS}, the test rejects only for large values of the statistic.

When identification is strong enough, in the sense that instruments are not too many nor too weak, Mikusheva and Sun (2021) argue in favor of a jackknife Wald statistic, which in the one-dimensional setting becomes

$$Wald(\beta_0) = \frac{(\hat{\beta}_{JIVE} - \beta_0)^2}{\hat{V}_{MS}} \to_d \chi^2(1), \tag{25}$$

where for $\tilde{P}_{Z,ij} = \frac{P_{Z,ij}^2}{M_{Z,ii}M_{Z,jj}+M_{Z,ij}^2}$ and $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{JIVE}$

$$\begin{aligned}\hat{\beta}_{JIVE} &= \frac{\sum_{i=1}^n \sum_{j\neq i} P_{Z,ij} y_i x_j}{\sum_{i=1}^n \sum_{j\neq i} P_{Z,ij} x_i x_j}, \\ \hat{V}_{MS} &= \frac{\sum_{i=1}^n (\sum_{j\neq i} P_{Z,ij} x_j)^2 \frac{\hat{\varepsilon}_i \boldsymbol{e}_i' \boldsymbol{M}_Z \hat{\boldsymbol{\varepsilon}}}{M_{Z,ii}} + \sum_{i=1}^n \sum_{j\neq i} \tilde{P}_{Z,ij}^2 \boldsymbol{e}_i' \boldsymbol{M}_Z \boldsymbol{x} \hat{\varepsilon}_i \boldsymbol{e}_j' \boldsymbol{M}_Z \boldsymbol{x} \hat{\varepsilon}_j}{(\sum_{i=1}^n \sum_{j\neq i} P_{Z,ij} x_i x_j)^2}.\end{aligned} \tag{26}$$

The AR_{MS} and the jackknife Wald test can be combined using a pretest for weak identification, based on

$$\tilde{F} = \frac{1}{\sqrt{k\hat{\Upsilon}}} \sum_{i=1}^n \sum_{j\neq i} P_{Z,ij} x_i x_j, \tag{27}$$

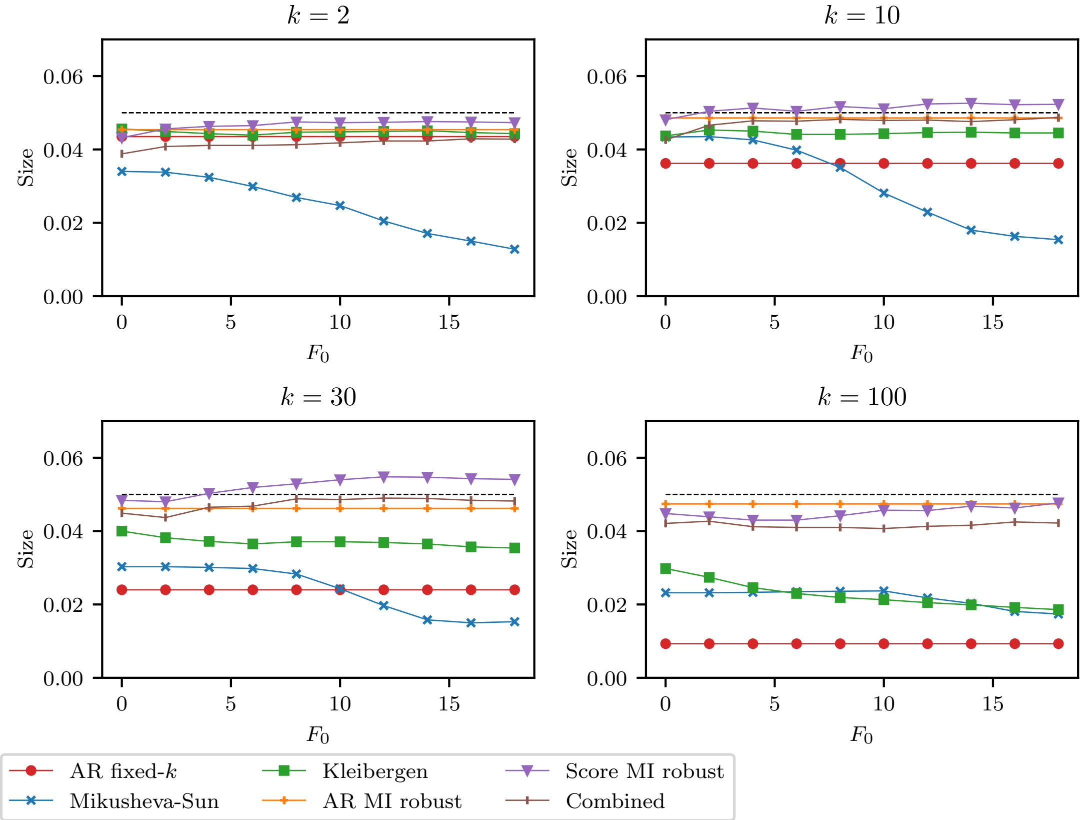
with $\hat{\Upsilon} = \frac{2}{k} \sum_{i=1}^n \sum_{j\neq i} \frac{P_{Z,ij}^2}{M_{Z,ii}M_{Z,jj}+M_{Z,ij}^2} x_i \boldsymbol{e}_i' \boldsymbol{M}_Z \boldsymbol{x} x_j \boldsymbol{e}_j' \boldsymbol{M}_Z \boldsymbol{x}$. Size distortions can be limited by using the AR_{MS} test when $\tilde{F}$ is below a certain cutoff value and the jackknife Wald when it is above. Mikusheva and Sun (2021) show that a cutoff value of 7.15 and 2% and 1% critical values for the two tests yields a combined test with overall asymptotic size smaller than 5%.

6.2 Results

Figure 1 shows the size of the tests for different number of instruments and instrument strength when testing at a $\alpha = 0.05$ confidence level. We see that our many instrument robust AR test, the score test and the combined test show excellent size control regardless of instrument numerosity and strength.

For the size of fixed-k AR statistic, we see that it has close to nominal size for a small number of instruments, regardless of instrument strength. When k increases, its rejection rate quickly falls below 0.05 and the test becomes conservative. The same happens to the fixed-k score statistic. For the test by Mikusheva and Sun (2021) we see in this experiment that the test underrejects for all values of k. Deviations from the nominal size can be as large as for the fixed-k AR statistic.

Figure 1: Size under identification robust inference.



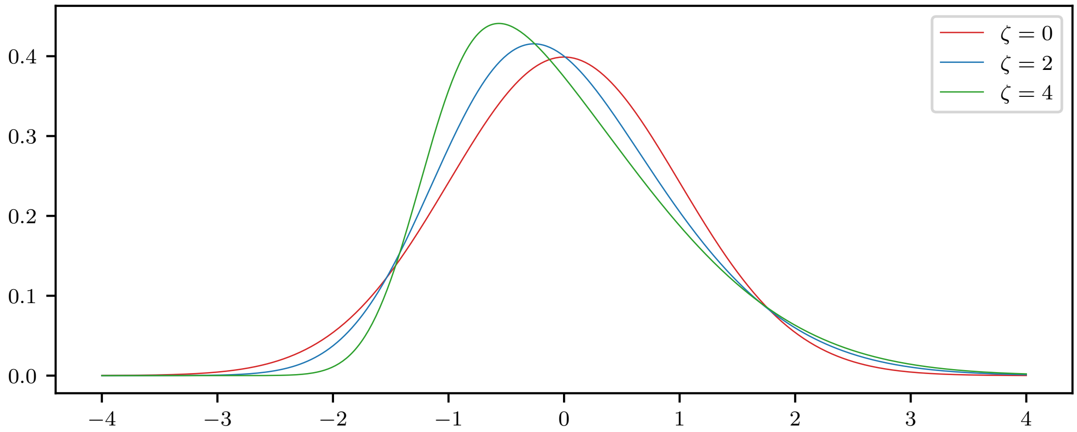
Note: size when testing $H_0\colon \beta = 0$ at $\alpha = 0.05$ based on (i) the fixed-k Anderson-Rubin test, (ii) the combined test by Mikusheva and Sun (2021), (iii) the test by Kleibergen (2005), and (iv-vi) the tests developed here, where we set $\alpha_{\text{AR}} = 0.01$ and $\alpha_S = 0.04$ for the combined test. k denotes the number of instruments and F_0 their strength. The Monte Carlo is the same as in Hausman et al. (2012).

Overall, Figure 1 supports the findings in Theorem 2 as the developed tests attain close to nominal size both under few and many instruments and regardless of whether these instruments are irrelevant or weak.

6.2.1 Moment conditions without invariance

An important issue to consider is what happens to the size of our tests when the invariance assumption on the second stage regression errors is violated. For this we change the distribution of η_i and z_{i1} in (1) to a skewed normal distribution with skewness parameters $\zeta_\eta \in \{0, 2, 4\}$ and $\zeta_z \in \{0, 2, 4\}$. Since in our simulation set up η_i also shows up in the definition of ε_i, the latter will also be skewed. For comparison reasons we rescale and recenter the distribution such that it still has

Figure 2: Probability density functions for the skewed normal distributions.



Note: probability density functions of a rescaled and recentered skewed normal distribution with skewness parameter ζ. The distribution is rescaled and recentered such that it has mean zero and unit variance.

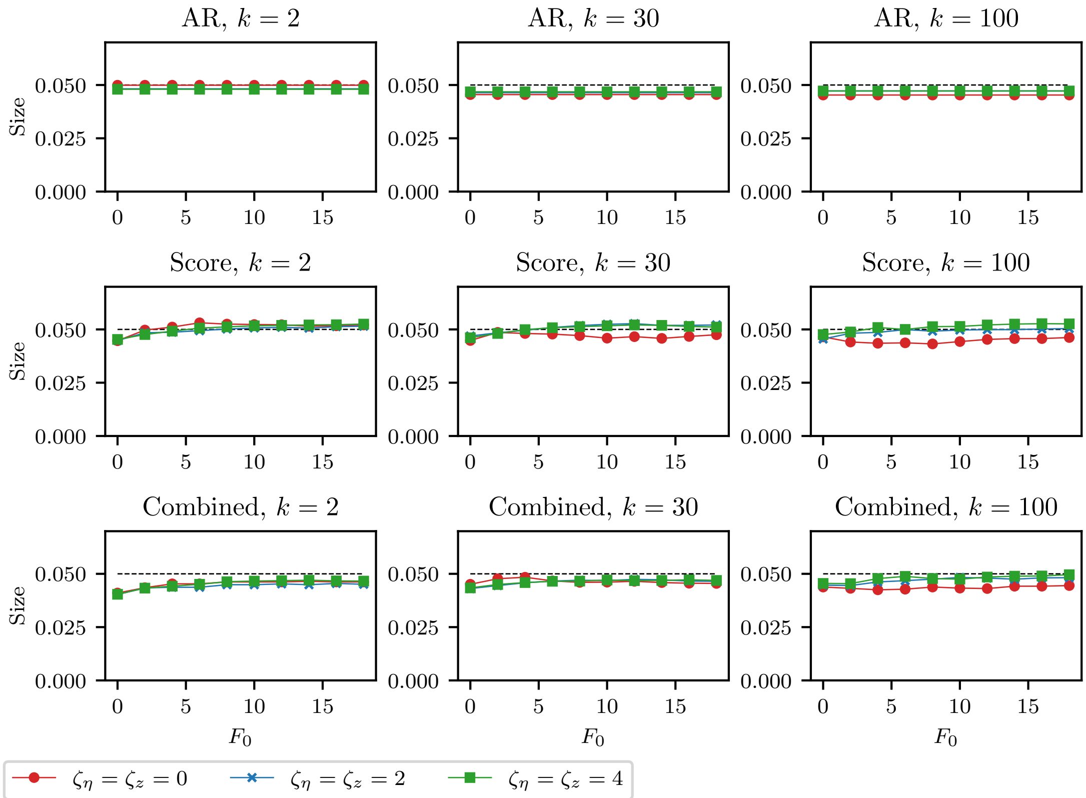
mean zero and unit variance. That is, η_i has probability density function

$$f(x) = \frac{2}{\omega_\eta}\phi\left(\frac{x-\xi_\eta}{\omega_\eta}\right)\Phi\left(\zeta_\eta\frac{x-\xi_\eta}{\omega_\eta}\right), \tag{28}$$

with ϕ and Φ the probability density and cumulative distribution functions of the standard normal distribution and $\omega_\eta = 1/(1-\frac{2\delta_\eta^2}{\pi})$ and $\xi_\eta = -\omega_\eta\delta_\eta\sqrt{2/\pi}$ for $\delta_\eta = \frac{\zeta_\eta}{\sqrt{1+\zeta_\eta^2}}$ and π the mathematical constant. We define the density for z_{i1} similarly. The probability density function of the shifted and rescaled skewed normal distribution as used for η_i and z_{i1} for different values of the skewness parameter is given in Figure 2. If ζ_η and ζ_z are different from zero, both the errors and our instruments are asymmetrically distributed, hence in those cases our invariance assumption will no longer hold.

In Figure 3 we plot the size of our AR, score and combined test for different number of instruments, different instrument strengths and different values for the skewness parameters. We see that in this set up all tests are insensitive to departures from the invariance assumption. The tests remain approximately size correct in all cases considered here and deviations from the 5% line are not consistently larger when there is no invariance, compared to when the invariance condition is satisfied.

Figure 3: Size under identification robust inference with skewed errors.



Note: size when testing $H_0\colon \beta = 0$ at $\alpha = 0.05$ for the tests developed here. For the combined tests we set $\alpha_{\mathrm{AR}} = 0.01$ and $\alpha_S = 0.04$. We consider the Monte Carlo set-up of Hausman et al. (2012), but we use a rescaled and recentred skewed normal distribution for η_i and z_{i1} such that they have mean zero, unit variance and use skewness parameter ζ_η and ζ_z.

7 Application to Angrist and Krueger (1991)

The motivating study for the weak and many instrument literature is Angrist and Krueger (1991), who estimate the effect of an extra year of education on weekly wages. Since time spend in school is a choice variable, it is likely correlated with a person's unobserved characteristics such as its ability. Angrist and Krueger (1991) instrument for education by the quarter in which someone is born. They argue that the quarter of birth is completely random, but correlates with education, as students generally start in the same period of the year, but schooling laws require someone to attend class only until they reach a specific age.

For the period that Angrist and Krueger (1991) consider, schooling laws are not constant over time, nor across states. They therefore interact the quarter of birth dummies with year of birth and state of birth dummies for additional instruments. Depending on the specific combination of interactions the total number of instruments can be 30, 180 or 1530. The authors estimate the return on education for the first two specifications and find a 95% confidence interval of $[0.053; 0.102]$ and $[0.067; 0.096]$ respectively.

Bound et al. (1995) and Staiger and Stock (1997) show that educational attainment and quarter of birth only correlate very weakly. Weak instruments would render the statistical tests used by Angrist and Krueger (1991) invalid. Controlling for weak instruments, Staiger and Stock (1997) find confidence intervals that are up to 13 percentage points wider and can contain zero. Hansen et al. (2008) find that rather than a weak instrument problem, the presence of many instruments disturbs the distribution of the test statistics. Correcting for the large number of instruments ($k = 180$), they find a confidence interval of $[0.078; 0.134]$.

None of these previously mentioned studies considered the case with 1530 instruments and, more importantly, corrected for possible heteroskedasticity. Only the recent study by Mikusheva and Sun (2021) considers these two cases. Using 180 and 1530 instruments they find confidence intervals of $[0.066; 0.13]$ and $[0.024; 0.12]$ respectively using their pre-testing procedure and $[-0.047; 0.20]$ and $[0.008; 0.20]$ using their AR statistic.

Section 6 shows that the test by Mikusheva and Sun (2021) suffers from small size distortions, making the confidence intervals not fully reliable. We therefore revisit the Angrist and Krueger (1991) study and construct confidence intervals for the return on schooling by inverting our AR, score and the combined test. Details of the implementation can be found in Appendix C.

Table 1 shows the confidence intervals for the return to education based on inverting the AR test, the score test, and their combination where we consider $k = \{30, 180, 1530\}$ instruments. If we first focus on the confidence intervals of the AR test at $\alpha = 0.05$ and $k = 30$ we see an interval that is substantially wider than the ones by Angrist and Krueger (1991) and Hansen et al. (2008). The coefficient is also no longer significantly positive. When we use 180 instruments our confidence interval is shifted upward relative to the one from the AR statistic by Mikusheva and Sun (2021), resulting in a significantly positive coefficient. Moreover, the confidence region is less wide, suggesting an improved efficiency by the use of many instruments. However, for $k = 1530$ the confidence interval is wider than for 180 instruments and is no longer significantly positive.

To discuss the results of the score based test, first consider Figure 4 where we plot the score and the AR statistic. As discussed in Section 5.2, we see that there

Table 1: Confidence intervals for the return on education.

		AR		Score		Combined	
		Lower	Upper	Lower	Upper	Lower	Upper
$\alpha = 0.1$	$k = 30$	0.01	0.17	0.05	0.12	0.05	0.12
	$k = 180$	0.02	0.19	0.07	0.13	0.07	0.13
	$k = 1530$	0.01	0.21	0.04	0.15	0.04	0.15
$\alpha = 0.05$	$k = 30$	0.00	0.18	0.05	0.13	0.05	0.13
	$k = 180$	0.01	0.20	0.07	0.13	0.07	0.13
	$k = 1530$	0.00	0.23	0.04	0.16	0.04	0.16
$\alpha = 0.01$	$k = 30$	-0.01	0.20	0.04	0.14	0.03	0.14
	$k = 180$	0.00	0.22	0.06	0.14	0.06	0.14
	$k = 1530$	-0.03	0.26	0.04	0.17	0.04	0.17

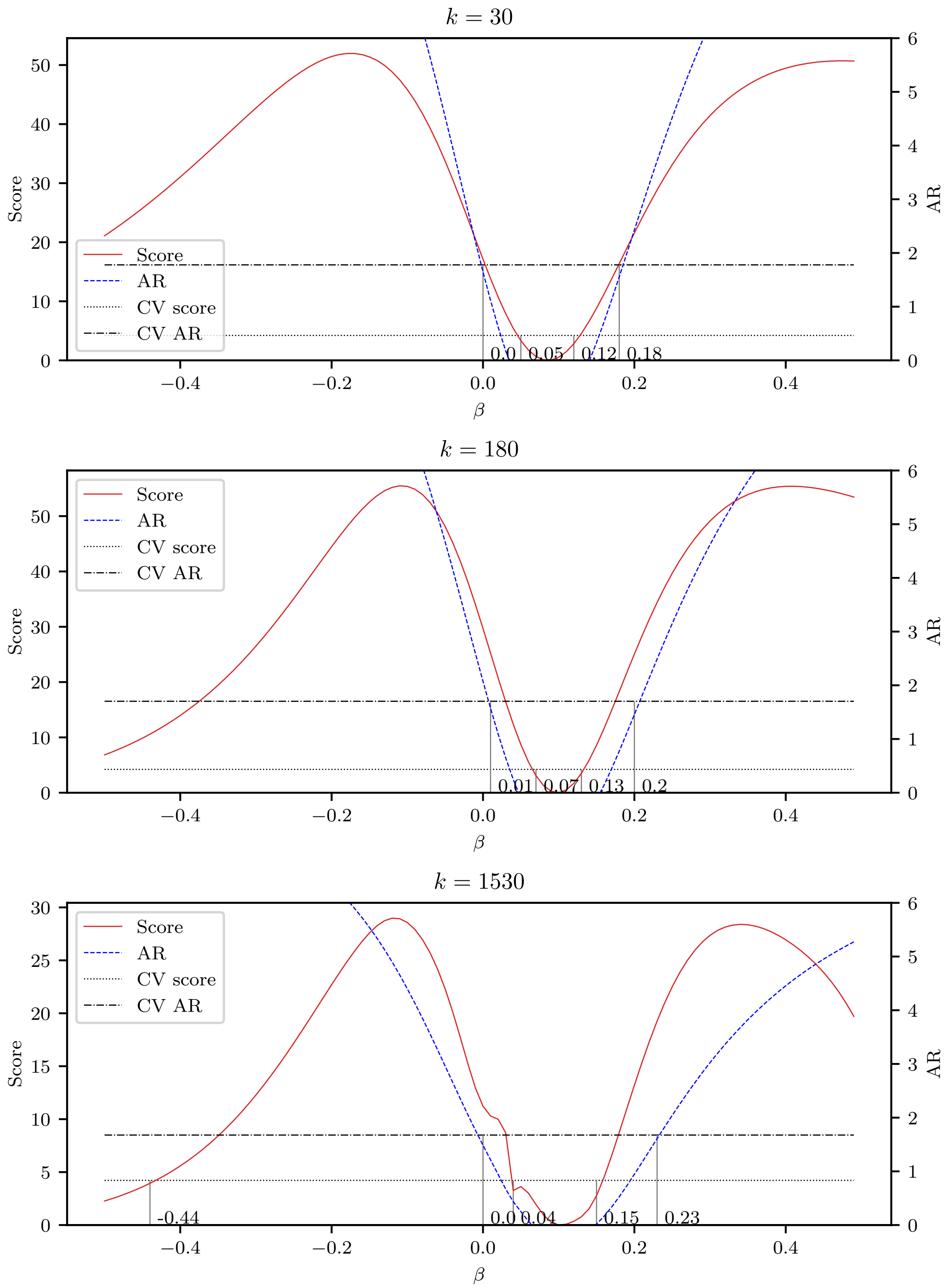
Note: confidence intervals for the return on education based on inverting the AR, the score and their combination. α is the confidence level. For the combination, $\alpha_{\text{AR}} = 0.2\alpha$ and $\alpha_S = 0.8\alpha$. k is the number of instruments. For the score, only the part of the confidence interval around the minimizer of the objective function is reported.

are multiple intervals of β for which the score test cannot reject, since the objective function flattens out for β's further away from the minimum. Intuitively, we only want to consider the region around the minimum of the objective function. These are the numbers reported in Table 1. This approach can be made formal using the both the AR and the score statistic and setting the significance levels such that the overall size is controlled. We report these results in the final two columns of Table 1. We find that the combined test yields substantially narrower confidence intervals than the AR test. The lower bounds are close to the lower bounds found by Angrist and Krueger (1991). The upper bound from the score test on the other hand is higher than the ones found in the original paper. The confidence intervals are remarkably robust to the included number of instruments.

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. Monte Carlo simulations show close to nominal size regardless of the strength of the instruments and the number of instruments. We apply our new tests to the return to education study by Angrist and Krueger (1991) and find that our confidence intervals are robust to the number of included instruments.

Figure 4: AR and score statistic in the Angrist and Krueger (1991) application.



Note: these figures show the value of the AR and score statistic together with their 0.05 critical values (CV) for different β and number of instruments k. The intersections are marked and can be used to construct confidence intervals for β.

References

Anatolyev, S. (2019). Many instruments and/or regressors: a friendly guide. *Journal of Economic Surveys*, 33(2):689–726.

Anatolyev, S. and Gospodinov, N. (2011). Specification testing in models with many instruments. *Econometric Theory*, 27(2):427–441.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46 – 63.

Andrews, D. W., Marmer, V., and Yu, Z. (2019). On optimal inference in the linear IV model. *Quantitative Economics*, 10(2):457–485.

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica*, 84(6):2155–2182.

Andrews, I. and Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84(4):1571–1612.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

Bao, Y. and Ullah, A. (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*, 140(5):1193–1205.

Bekker, P. and Kleibergen, F. (2003). Finite-sample instrumental variables inference using an asymptotically pivotal statistic. *Econometric Theory*, 19(5):744–753.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–681.

Bekker, P. A. and Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics*, 185(2):332–342.

Bekker, P. A. and Lawford, S. (2008). Symmetry-based inference in an instrumental variable setting. *Journal of Econometrics*, 142(1):28–49.

Bekker, P. A. and van der Ploeg, J. (2005). Instrumental variable estimation based on grouped data. *Statistica Neerlandica*, 59(3):239–267.

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

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Canay, I. A., Romano, J. P., and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica*, 85(3):1013–1030.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association*, 113(523):1350–1361.

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

Chao, J. C. and Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5):1673–1692.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K., and Woutersen, T. (2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory*, 28(1):42–86.

Crudu, F., Mellace, G., and Sándor, Z. (2021). Inference in instrumental variable models with heteroskedasticity and many instruments. *Econometric Theory*, 37(2):281–310.

Donald, S. G. and Newey, W. K. (2000). A jackknife interpretation of the continuous updating estimator. *Economics Letters*, 67(3):239–243.

Efron, B. (1969). Student's t-test under symmetry conditions. *Journal of the American Statistical Association*, 64(328):1278–1302.

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

Hansen, C., Hausman, J., and Newey, W. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26(4):398–422.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

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(371):693–700.

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

Lim, D., Wang, W., and Zhang, Y. (2022). A conditional linear combination test with many weak instruments. *arXiv preprint arXiv:2207.11137*.

Magnus, J. R. and Neudecker, H. (1998). *Matrix differential calculus: with applications in statistics and econometrics*. John Wiley & Sons Ltd.

Matsushita, Y. and Otsu, T. (2020). *Jackknife Lagrange multiplier test with many weak instruments*. LSE, STICERD.

Mikusheva, A. and Sun, L. (2021). Inference with many weak instruments. *The Review of Economic Studies (forthcoming)*.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):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(3):821–841.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

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

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

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

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability & Its Applications*, 5(3):302–305.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):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 Expectation over higher order forms in Rademacher random variables

Theorem A.1. *Consider a random $n \times 1$ 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. $\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{r}] = -2\,\mathrm{tr}(\boldsymbol{D}_{A_1}\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1)\,\mathrm{tr}(\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1\boldsymbol{A}_2) + \mathrm{tr}(\boldsymbol{A}_1'\boldsymbol{A}_2)$.

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} \tag{A.1}
$$

5.

$$
\begin{aligned}
\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}] = \mathrm{tr}(\boldsymbol{D}_{A_4A_1}\boldsymbol{D}_{A_2\boldsymbol{A}_3}) + \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}\boldsymbol{\iota} \\
- 2\,\mathrm{tr}(\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_{A_4}) + \boldsymbol{\iota}'(\boldsymbol{A}_4 \odot \boldsymbol{A}_1 \odot (\boldsymbol{A}_3'\boldsymbol{A}_2'))\boldsymbol{\iota}.
\end{aligned} \tag{A.2}
$$

If in addition $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are symmetric matrices with zero diagonal, we have

6.

$$\begin{aligned}\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{r})^2(\boldsymbol{r}'\boldsymbol{A}_2\boldsymbol{r})^2] = {} & 4\operatorname{tr}(\boldsymbol{A}_1^2)\operatorname{tr}(\boldsymbol{A}_2^2) + 8\operatorname{tr}^2(\boldsymbol{A}_1\boldsymbol{A}_2) + 32\operatorname{tr}(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_1\boldsymbol{A}_2) \\ & + 16\operatorname{tr}(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_2\boldsymbol{A}_1) - 32\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_1^2)(\boldsymbol{I}\odot\boldsymbol{A}_2^2)\boldsymbol{\iota} \\ & - 64\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_1\boldsymbol{A}_2)(\boldsymbol{I}\odot\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{\iota} \\ & + 32\boldsymbol{\iota}'(\boldsymbol{A}_1\odot\boldsymbol{A}_1\odot\boldsymbol{A}_2\odot\boldsymbol{A}_2)\boldsymbol{\iota}.\end{aligned} \tag{A.3}$$

7.

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

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

2. See Ullah (2004), Appendix A5.

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

$$\mathrm{E}[\boldsymbol{v}'\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}] = \underbrace{\mathrm{E}[\boldsymbol{v}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}]}_{(I)} + \underbrace{\mathrm{E}[\boldsymbol{v}'\boldsymbol{\Delta}\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}]}_{(II)}. \tag{A.5}$$

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

$$(I) = \mathrm{E}\left[\sum_{i,j,k=1}^{n} v_i a_{1,ij} a_{2,jk} r_j r_k\right] = \sum_{i,j=1}^{n} v_i a_{1,ij} a_{2,jj} = \boldsymbol{v}'\boldsymbol{A}_1\boldsymbol{D}_{A_2}\boldsymbol{\iota}. \tag{A.6}$$

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

$$(II.a) = \sum_{i\neq j} v_i a_{1,ji} a_{2,ij} = \boldsymbol{\iota}\boldsymbol{A}_1 \odot \boldsymbol{A}_2'\boldsymbol{v} - \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{D}_{A_2}\boldsymbol{v}. \tag{A.7}$$

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

$$(II.b) = \sum_{i\neq j} v_i a_{1,jj} a_{2,ji} = \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{A}_2\boldsymbol{v} - \boldsymbol{\iota}'\boldsymbol{D}_{A_1}\boldsymbol{D}_{A_2}\boldsymbol{v}. \tag{A.8}$$

4. We decompose the expectation as

$$
\begin{aligned}
\mathrm{E}[\boldsymbol{r}'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{r}\boldsymbol{r}'\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{r}\boldsymbol{r}')] \\
&= \mathrm{E}[\mathrm{tr}\,(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{I}+(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I}))\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4(\boldsymbol{I}+(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})))] \\
&= \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4)]}_{(I)} \\
&\quad + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4)]}_{(II)} \\
&\quad + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I}))]}_{(II')} \\
&\quad + \underbrace{\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I}))]}_{(III)}.
\end{aligned} \tag{A.9}
$$

Starting with (I), we have that

$$
\begin{aligned}
(I) &= \mathrm{E}\left[\sum_{i=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4 e_i\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_r\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{D}_r\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{A}_4\boldsymbol{e}_i\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_4\boldsymbol{e}_i \\
&= \sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{A}_4\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{e}_j \\
&= \mathrm{tr}(\boldsymbol{D}_{A_4A_1}\boldsymbol{D}_{A_2A_3}).
\end{aligned} \tag{A.10}
$$

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 $\mathrm{E}[\delta_{kl}^2] = \mathrm{E}[r_k^2 r_l^2] = 1$.

$$
\begin{aligned}
(II) &= \mathrm{E}\left[\sum_{i=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{D}_r\boldsymbol{A}_2(\boldsymbol{r}\boldsymbol{r}'-\boldsymbol{I})\boldsymbol{A}_3\boldsymbol{D}_r\boldsymbol{A}_4\boldsymbol{e}_i\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j,m,k,l=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_r\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{e}_m\delta_{mk}\boldsymbol{e}_k'\boldsymbol{A}_3\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{D}_r\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{A}_4\boldsymbol{e}_i\right].
\end{aligned} \tag{A.11}
$$

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),

$$
(II.a) = \sum_{i=1}^{n}\sum_{j\neq l}\boldsymbol{e}_i'\boldsymbol{A}_1\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{A}_2\boldsymbol{e}_j\boldsymbol{e}_l'\boldsymbol{A}_3\boldsymbol{e}_l\boldsymbol{e}_l'\boldsymbol{A}_4\boldsymbol{e}_i
$$

$$
\begin{aligned}
&= \sum_{j \neq l} \boldsymbol{e}_l' \boldsymbol{D}_{A_3} \boldsymbol{A}_4 \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{e}_j \\
&= \boldsymbol{\iota}' \boldsymbol{D}_{A_3} \boldsymbol{A}_4 \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{\iota} \underbrace{- \operatorname{tr}\left(\boldsymbol{D}_{A_3} \boldsymbol{A}_4 \boldsymbol{A}_1 \boldsymbol{D}_{A_2}\right)}_{(II.a.2)}.
\end{aligned}
$$

For case (b), we have

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

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.9) is

$$
(III) = \mathrm{E}\left[\sum_{i=1}^{n} \sum_{j,k,m,l,s} a_{1,ij} a_{2,jk} a_{3,ml} a_{4,ls} r_j r_l \delta_{km} \delta_{si}\right]. \tag{A.12}
$$

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.

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

$$= \sum_{i=1}^{n} \sum_{k \neq i} \boldsymbol{e}_i' (\boldsymbol{A}_1 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_4) \boldsymbol{e}_k$$

$$= \boldsymbol{\iota}' (\boldsymbol{A}_1 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_4) \boldsymbol{\iota} \underbrace{- \operatorname{tr}((\boldsymbol{A}_1 \odot \boldsymbol{A}_3)(\boldsymbol{A}_2 \odot \boldsymbol{A}_4))}_{(III.a.2)}.$$

$$(III.b) = \sum_{i=1}^{n} \sum_{j,m \neq i} a_{1,ij} a_{2,ji} a_{3,mj} a_{4,jm}$$

$$= \operatorname{tr}(\boldsymbol{A}_1' \boldsymbol{A}_2' \odot \boldsymbol{A}_3' \boldsymbol{A}_4') + (III.a.2).$$

$$(III.c) = \sum_{i=1}^{n} \sum_{j \neq i, j \neq m, m \neq i} a_{1,ij} a_{2,jj} a_{4,im} a_{3,mi}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} a_{1,ij} a_{2,jj} \boldsymbol{e}_i' \boldsymbol{A}_4 \boldsymbol{A}_3 \boldsymbol{e}_i - a_{1,ij} a_{2,jj} a_{4,ij} a_{3ji} - a_{1,ij} a_{2,jj} a_{4,ii} a_{3,ii}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \boldsymbol{e}_i \boldsymbol{D}_{A_4 A_3} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{e}_j - \boldsymbol{e}_i' \boldsymbol{D}_{A_3} \boldsymbol{D}_{A_4} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{e}_j - \boldsymbol{e}_i' (\boldsymbol{A}_1 \boldsymbol{D}_{A_2}) \odot \boldsymbol{A}_4 \odot \boldsymbol{A}_1' \boldsymbol{e}_j$$

$$= \boldsymbol{\iota}' \boldsymbol{D}_{A_4 A_3} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{\iota} - \operatorname{tr}(\boldsymbol{D}_{A_4 A_3} \boldsymbol{A}_1 \boldsymbol{D}_{A_2}) - \boldsymbol{\iota}' \boldsymbol{D}_{A_3} \boldsymbol{D}_{A_4} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{\iota}$$
$$+ \operatorname{tr}(\boldsymbol{D}_{A_3} \boldsymbol{D}_{A_4} \boldsymbol{A}_1 \boldsymbol{D}_{A_2}) - \boldsymbol{\iota}' (\boldsymbol{A}_1 \boldsymbol{D}_{A_2}) \odot \boldsymbol{A}_4 \odot \boldsymbol{A}_3' \boldsymbol{\iota} + \operatorname{tr}((\boldsymbol{A}_1 \boldsymbol{D}_{A_2}) \odot \boldsymbol{A}_4 \odot \boldsymbol{A}_3').$$

where the first and last term on the last line are equal.

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

$$\begin{array}{llll}
(c.1) & = \boldsymbol{\iota}' \boldsymbol{D}_{A_4 A_3} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{\iota}, & (c.2) & = -\operatorname{tr}(\boldsymbol{D}_{A_4 A_3} \boldsymbol{A}_1 \boldsymbol{D}_{A_2}), \\
(c.3) & = -\boldsymbol{\iota}' \boldsymbol{D}_{A_3} \boldsymbol{D}_{A_4} \boldsymbol{A}_1 \boldsymbol{D}_{A_2} \boldsymbol{\iota}, & (c.4) & = \operatorname{tr}(\boldsymbol{D}_{A_3} \boldsymbol{D}_{A_4} \boldsymbol{A}_1 \boldsymbol{D}_{A_2}), \\
(c.5) & = -\boldsymbol{\iota}' (\boldsymbol{A}_1 \boldsymbol{D}_{A_2}) \odot \boldsymbol{A}_4 \odot \boldsymbol{A}_3' \boldsymbol{\iota}, & (d.1) & = \boldsymbol{\iota}' \boldsymbol{D}_{A_4} \boldsymbol{A}_3' \boldsymbol{A}_1 (\boldsymbol{I} \odot \boldsymbol{A}_2) \boldsymbol{\iota}, \\
(d.2) & = -\operatorname{tr}(\boldsymbol{D}_{A_4} \boldsymbol{A}_3' \boldsymbol{A}_1 \boldsymbol{D}_{A_2}), & (e.1) & = \boldsymbol{\iota}' ((\boldsymbol{A}_3 \odot \boldsymbol{A}_1') \boldsymbol{A}_4) \odot \boldsymbol{A}_2 \boldsymbol{\iota}, \\
(e.3) & = -\boldsymbol{\iota}' ((\boldsymbol{A}_1 \odot \boldsymbol{A}_3' \odot \boldsymbol{I}) \boldsymbol{A}_2) \odot \boldsymbol{A}_4 \boldsymbol{\iota}, & (g.1) & = \boldsymbol{\iota}' (\boldsymbol{A}_1 \odot (\boldsymbol{A}_3 (\boldsymbol{A}_2' \odot \boldsymbol{A}_4))) \boldsymbol{\iota}, \\
(g.2) & = -\operatorname{tr}((\boldsymbol{A}_1 \odot (\boldsymbol{A}_3 (\boldsymbol{A}_2' \odot \boldsymbol{A}_4)))), & (h.1) & = \boldsymbol{\iota}' (\boldsymbol{A}_1 \odot \boldsymbol{A}_2') \boldsymbol{A}_4' \boldsymbol{D}_{A_3} \boldsymbol{\iota}, \\
(h.5) & = -\boldsymbol{\iota}' ((\boldsymbol{A}_1 \odot \boldsymbol{A}_2') \odot \boldsymbol{I})' A_4 (\boldsymbol{A}_3' \odot \boldsymbol{I}) \boldsymbol{\iota}, & (i.1) & = \boldsymbol{\iota}' \boldsymbol{D}_{A_1} \boldsymbol{A}_2 \boldsymbol{A}_4' \boldsymbol{D}_{A_3} \boldsymbol{\iota}.
\end{array} \tag{A.13}$$

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

$$(III.c) = (c.1) + (c.2) + (c.3) + (c.4) + (c.5) + (c.4),$$

$$(III.d) = (d.1) + (d.2) + (c.3)^* + (c.4) + (c.3) + (c.4),$$

$$(III.e) = (e.1) + (c.5)^* + (e.3) + (c.4) + (c.5)^* + (c.4),$$

$$(III.f) = (c.1)^* + (c.5)^* + (c.3)^* + (c.4) + (c.2)^* + (c.4),$$

$$(III.g) = (g.1) + (g.2)^* + (g.2) + (c.4) + (d.2)^* + (c.4),$$

$$(III.h) = (h.1) + (g.2)^* + (c.2)^* + (c.4) + (h.5) + (c.4),$$

$$(III.i) = (i.1) + (e.3) + (h.5)^* + (c.4) + (h.5) + (c.4),$$
$$(III.j) = (h.1)^* + (g.2) + (h.5)^* + (c.4) + (c.2) + (c.4).$$

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*

$$\lambda_{\max}(\boldsymbol{A} \odot \boldsymbol{B}) \leq \lambda_{\max}(\boldsymbol{A} \otimes \boldsymbol{B}) \leq \max\{\lambda_{\max}(\boldsymbol{A})\,\lambda_{\max}(\boldsymbol{B}), \lambda_{\min}(\boldsymbol{A})\,\lambda_{\min}(\boldsymbol{B})\}, \tag{A.14}$$

and

$$\begin{aligned}\lambda_{\min}(\boldsymbol{A} \odot \boldsymbol{B}) &\geq \lambda_{\min}(\boldsymbol{A} \otimes \boldsymbol{B}) \\ &\geq \min\{\lambda_{\min}(\boldsymbol{A})\,\lambda_{\min}(\boldsymbol{B}), \lambda_{\min}(\boldsymbol{A})\,\lambda_{\max}(\boldsymbol{B}), \lambda_{\max}(\boldsymbol{A})\,\lambda_{\min}(\boldsymbol{B})\}.\end{aligned} \tag{A.15}$$

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

$$\begin{aligned}\lambda_{\max}(\boldsymbol{A} \odot \boldsymbol{B}) &= \max_{\boldsymbol{v}:\boldsymbol{v}'\boldsymbol{v}=1} \boldsymbol{v}'(\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{v} = \max_{\boldsymbol{u}:\boldsymbol{u}'\boldsymbol{u}=1} \boldsymbol{u}'(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{u} \\ &\leq \max_{\boldsymbol{w}:\boldsymbol{w}'\boldsymbol{w}=1} \boldsymbol{w}'(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{w} = \lambda_{\max}(\boldsymbol{A} \otimes \boldsymbol{B}),\end{aligned} \tag{A.16}$$

and

$$\begin{aligned}\lambda_{\min}(\boldsymbol{A} \odot \boldsymbol{B}) &= \min_{\boldsymbol{v}:\boldsymbol{v}'\boldsymbol{v}=1} \boldsymbol{v}'(\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{v} = \min_{\boldsymbol{u}:\boldsymbol{u}'\boldsymbol{u}=1} \boldsymbol{u} \\ &\geq \min_{\boldsymbol{w}:\boldsymbol{w}'\boldsymbol{w}=1} \boldsymbol{w}'(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{w} = \lambda_{\min}(\boldsymbol{A} \otimes \boldsymbol{B}),\end{aligned} \tag{A.17}$$

where $\boldsymbol{u}$ follows the structure above and $\boldsymbol{w}$ is any vector in $\mathbb{R}^{n^2}$.

The last set of inequalities 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 $i, j = 1, \ldots, n$. □

Corollary 2. *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 1

Proof. The first order conditions for the CUE are given by

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i} = -\frac{1}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{I} - \boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon}. \tag{A.18}$$

Under Assumption A2, the first order conditions satisfy

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i} &\overset{(d)}{=} -\frac{1}{n}(\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)})'(\boldsymbol{I} - \boldsymbol{D}_r\boldsymbol{D}_{Pr})\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{\varepsilon} \\ &= -\frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} \\ &\quad -\frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r} \overset{(\mathrm{E}_r)}{=} 0.\end{aligned} \tag{A.19}$$

This proves the first statement.

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

$$\begin{aligned}\mathrm{E}\left[n\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_i}\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_j}|\mathcal{J}\right] = \mathrm{E}\Bigg[&\underbrace{\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{x}_{(j)}}_{(I)} + \underbrace{\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{D}_{P\iota}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{D}_{P\iota}\boldsymbol{x}_{(j)}}_{(II)} \\ &\underbrace{-\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{D}_{P\iota}\boldsymbol{x}_{(j)}}_{(III)}\underbrace{-\frac{1}{n}\boldsymbol{x}'_{(i)}\boldsymbol{D}_{P\iota}\boldsymbol{V}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{V}\boldsymbol{x}_{(j)}}_{(IV)}|\mathcal{J}\Bigg].\end{aligned} \tag{A.20}$$

Using that $\boldsymbol{x}_{(i)} = \bar{\boldsymbol{x}}_{(i)} + \boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)}$ we get that (I) is distributed equivalently to

$$\begin{aligned}(I) &\overset{(d)}{=} \frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} + \boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} \\ &\quad + \frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} \\ &\overset{\mathrm{E}_r[\cdot]}{=} \frac{1}{n}\Big[\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} - 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{a_{(j)}}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\,\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) \\ &\quad + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})\Big] \\ &\overset{\mathrm{E}_U[\cdot]}{=} \frac{1}{n}\Big[\underbrace{\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{V}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})}_{\text{Kleibergen's } K} \\ &\quad - 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{a_{(j)}}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\,\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})\Big].\end{aligned} \tag{A.21}$$

For (II), we have

$$\begin{aligned}(II) &\overset{(d)}{=} \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} \\ &\quad + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r} + \frac{1}{n}\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_{\bar{x}_{(i)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{r}\end{aligned}$$

$$
\begin{aligned}
&+\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}\\
&\overset{\mathrm{E}_r[\cdot]}{=}\frac{1}{n}\bar{\boldsymbol{x}}'_{(i)}\{3\boldsymbol{D}_P\boldsymbol{D}_V+8\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P-12\boldsymbol{D}_P\boldsymbol{D}_V\boldsymbol{D}_P+4(\boldsymbol{V}\odot\boldsymbol{P}\odot\boldsymbol{P})\\
&-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\\
&-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{x}}_{(j)}\\
&-\frac{2}{n}\operatorname{tr}((\boldsymbol{I}\odot\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})+\frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\\
&+\frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})\\
&\overset{\mathrm{E}_U[\cdot]}{=}\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{Kleibergen's } K \text{ (I)}}\\
&+\underbrace{\frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})+\frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{Kleibergen's } K \text{ (II)}}\\
&+\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})\\
&-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\\
&-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)}\\
&+\frac{2}{n}\operatorname{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})))\\
&-\frac{2}{n}\operatorname{tr}((\boldsymbol{I}\odot\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})+\frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})+\frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}).
\end{aligned}
$$

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

For (III),

$$
\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[\cdot]}{=}-\frac{3}{n}\bar{\boldsymbol{x}}'_{(j)}\boldsymbol{D}_P\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)}+\frac{2}{n}\bar{\boldsymbol{x}}'_{(j)}\boldsymbol{D}_P^2\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)}\\
&\quad+\frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})-\frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})\\
&\quad-\frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\\
&\overset{\mathrm{E}_U[\cdot]}{=}\underbrace{-\frac{1}{n}\bar{\boldsymbol{z}}_{(j)}\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(i)}-\frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})-\frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{Kleibergen's } K}\\
&\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}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{D}_V)\\
&\quad+\frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})-\frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})-\frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}).
\end{aligned}
$$

Lastly by symmetry,

$$\begin{aligned}
(IV) &\overset{\mathrm{E}_r[\cdot]}{=} -\frac{3}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{D}_P\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} + \frac{2}{n}\bar{\boldsymbol{x}}'_{(i)}\boldsymbol{D}_P^2\boldsymbol{V}\bar{\boldsymbol{x}}_{(j)} \\
&\quad + \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) \\
&\quad - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) \\
&\overset{\mathrm{E}_U[\cdot]}{=} \underbrace{-\frac{1}{n}\bar{\boldsymbol{z}}_{(i)}\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)} - \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) - \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P\boldsymbol{D}_V)}_{\text{Kleibergen's } K} \\
&\quad - \frac{2}{n}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\boldsymbol{D}_V) \\
&\quad + \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}})\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) - \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}).
\end{aligned} \tag{A.22}$$

Together, we find

$$\Omega_{ij}(\boldsymbol{\beta}_0) = \mathrm{E}_r\left[n\cdot S_{(i),r}(\boldsymbol{\beta}_0)S_{(j),r}(\boldsymbol{\beta}_0)\middle|\mathcal{J}\right] = \Omega^L_{ij}(\boldsymbol{\beta}_0) + \Omega^H_{ij}(\boldsymbol{\beta}_0), \tag{A.23}$$

where

$$\begin{aligned}
\Omega^L_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\bar{\boldsymbol{z}}'_{(i)}[\boldsymbol{V} - \boldsymbol{D}_P\boldsymbol{V} - \boldsymbol{V}\boldsymbol{D}_P + \boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P + \boldsymbol{D}_P\boldsymbol{D}_V + \dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P}]\bar{\boldsymbol{z}}_{(j)} \\
&\quad + \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_V - \boldsymbol{D}_V\boldsymbol{D}_P)) + \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) - \frac{1}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}), \\
\Omega^H_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\bar{\boldsymbol{z}}'_{(i)}\Big[2\boldsymbol{D}_P\boldsymbol{D}_V + 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 - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot\boldsymbol{I} + 3\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P} \\
&\qquad - 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\boldsymbol{V} - 2\boldsymbol{V}\boldsymbol{D}_P + 2\boldsymbol{D}_P^2\boldsymbol{V} + 2\boldsymbol{V}\boldsymbol{D}_P^2\Big]\bar{\boldsymbol{z}}_{(j)} \\
&\quad - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{D}_P\boldsymbol{D}_V - 2\boldsymbol{D}_P^2\boldsymbol{D}_V + (\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}))) \\
&\quad + \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) + \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) \\
&\quad - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_P^2\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}) - \frac{2}{n}\operatorname{tr}(\boldsymbol{D}_{a_{(i)}}(\boldsymbol{P}\odot\boldsymbol{P})^2\boldsymbol{D}_{a_{(j)}}).
\end{aligned} \tag{A.24}$$

□

A.3 Proof of Theorem 2

We defer the proof of Theorem 2 to Appendix B, due to its length.

A.4 Proof of Theorem 3

A.4.1 Unbiasedness

Proof. To show that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_0)$ is (conditionally) unbiased, we start by analyzing the variance of the score. The variance estimator was decomposed into

$$\begin{aligned}
\hat{\Omega}^L_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}(\boldsymbol{I}-\boldsymbol{D}_{P_\iota})\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{D}_{P_\iota})\boldsymbol{x}_{(j)},\\
\hat{\Omega}^H_{ij}(\boldsymbol{\beta}_0) &= \frac{1}{n}\boldsymbol{x}'_{(i)}(2\boldsymbol{D}_P\boldsymbol{D}_V + 7\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P + 3\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P}\\
&\quad - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot\boldsymbol{I} - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2\\
&\quad - 4\boldsymbol{D}_P(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P}) - 4(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P\\
&\quad - 2\boldsymbol{D}_P\boldsymbol{V} - 2\boldsymbol{V}\boldsymbol{D}_P + 2\boldsymbol{D}_P\boldsymbol{D}_{P_\iota}\boldsymbol{V} + 2\boldsymbol{V}\boldsymbol{D}_{P_\iota}\boldsymbol{D}_P)\boldsymbol{x}_{(j)}.
\end{aligned} \tag{A.25}$$

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

$$\begin{aligned}
\boldsymbol{x}'_{(j)}\boldsymbol{V}\boldsymbol{x}_{(i)} &= \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{a}'_{(j)}\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)} + \boldsymbol{a}'_{(j)}\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)}\\
&\overset{(d)}{=} \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \boldsymbol{r}'\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_\varepsilon\boldsymbol{V}\bar{\boldsymbol{x}}_{(i)} + \bar{\boldsymbol{x}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(i)}}\boldsymbol{r}.
\end{aligned} \tag{A.26}$$

This implies that

$$\mathrm{E}[\boldsymbol{x}'_{(j)}\boldsymbol{V}\boldsymbol{x}_{(i)}|\mathcal{J}] = \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{V}) + \mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}). \tag{A.27}$$

Similarly, we obtain

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

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}_P\boldsymbol{D}_V\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \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}_P^2\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}), \\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(i)}, \\
\boldsymbol{x}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \bar{\boldsymbol{z}}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\bar{\boldsymbol{z}}_{(i)}, \\
\boldsymbol{x}'_{(j)}((\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot\boldsymbol{I})\boldsymbol{x}_{(i)} &= \boldsymbol{z}'_{(j)}((\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot\boldsymbol{I})\boldsymbol{z}_{(i)} \\
&\quad + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})) \\
&\quad + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}(\boldsymbol{P}\odot\boldsymbol{P})^2\boldsymbol{D}_{a_{(j)}}), \qquad \text{(A.28)} \\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(i)}, \\
\boldsymbol{x}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \bar{\boldsymbol{z}}'_{(j)}(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(i)}, \\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \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{V}) \\
&\quad + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_P^2), \\
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P^2\boldsymbol{V}\boldsymbol{x}_{(i)} &\overset{(\mathrm{E})}{=} \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_P^2\boldsymbol{V}\bar{\boldsymbol{z}}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(j,i)}\boldsymbol{D}_P^2\boldsymbol{V}) \\
&\quad + \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_P^3).
\end{aligned}
$$

Finally, note that

$$
\boldsymbol{x}'_{(j)}\boldsymbol{D}_P\boldsymbol{D}_{P\iota}\boldsymbol{V}\boldsymbol{x}_{(i)} = \bar{\boldsymbol{z}}_{(j)}\boldsymbol{D}_P^2\boldsymbol{V}\boldsymbol{z}_{(i)} + \mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_P^2\boldsymbol{D}_V) + \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_P\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}). \qquad \text{(A.29)}
$$

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 $\mathrm{E}[\hat{\sigma}_n^2(\boldsymbol{\beta}_0)] = \mathrm{E}[\frac{2}{k}(k - \boldsymbol{\iota}'\boldsymbol{D}_P^2\boldsymbol{\iota})] = \frac{2}{k}(\sum_{i=1}^n P_{ii} - \sum_{i=1}^n P_{ii}^2) =$

$\frac{2}{k}(\sum_{i,j=1}^n P_{ij}^2 - \sum_{i=1}^n P_{ii}^2) = \frac{2}{k}(\sum_{i\neq j} P_{ij}^2) = \sigma_n^2$ and

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{1,j+1}(\boldsymbol{\beta}_0) &= \frac{2}{\sqrt{n\cdot k}}\boldsymbol{x}_j'(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon} \\
&= \frac{2}{\sqrt{n\cdot k}}\,\mathrm{E}[\bar{\boldsymbol{x}}_j'(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{D}_{a_{(j)}}(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon}] \\
&\overset{(d)}{=} \frac{2}{\sqrt{n\cdot k}}\,\mathrm{E}[\bar{\boldsymbol{x}}_j'(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r))\boldsymbol{D}_r^2\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{r} \\
&\quad + \boldsymbol{r}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}(\boldsymbol{D}_V - (\boldsymbol{V}\odot\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r))\boldsymbol{D}_r^2\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{r}] \\
&\overset{(E)}{=} \frac{2}{\sqrt{n\cdot k}}[\mathrm{tr}(\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_\varepsilon) - \boldsymbol{\iota}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{\iota}] \\
&= \frac{2}{\sqrt{n\cdot k}}[\mathrm{tr}(\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P) - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P)] \\
&= \frac{2}{\sqrt{n\cdot k}}\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_P) \\
&= \frac{2}{\sqrt{n\cdot k}}\,\mathrm{tr}(\boldsymbol{\Psi}^{(h)}\odot\boldsymbol{P}),
\end{aligned} \tag{A.30}
$$

where the part with $\hat{\boldsymbol{x}}_{(j)}$ has expectation zero, due to the odd number of Rademacher random variables. For the expectation of the first term of the part that remains we used Theorem A.1.

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

A.4.2 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 $\boldsymbol{x}_{(i),r} = \bar{\boldsymbol{z}}_{(i)} + \boldsymbol{u}_{(i)} + \boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(i)}$. To show consistency, we need to show for some matrix $\boldsymbol{A}_r$ that possibly depends on the vector of Rademacher random variables $\boldsymbol{r}$, that

$$
n^{-2}\,\mathrm{E}[(\boldsymbol{x}_{(i),r}'\boldsymbol{A}_r\boldsymbol{x}_{(j),r} - \mathrm{E}[\boldsymbol{x}_{(i),r}'\boldsymbol{A}_r\boldsymbol{x}_{(j),r}|\mathcal{J}])^2|\mathcal{J}] \to_p 0. \tag{A.31}
$$

For $\boldsymbol{A}_r$ we have to 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) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}$, (d) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r$ and (e) $\boldsymbol{A}_r = \boldsymbol{D}_r\boldsymbol{D}_P\boldsymbol{D}_{Pr}\boldsymbol{V}$. Cases (a) and (b) will cover the consistency of the terms listed in (A.27) and (A.28) that are all of the form $\boldsymbol{x}_{(i)}'\boldsymbol{A}_r\boldsymbol{x}_{(j)}$. For all these terms $\lambda_{\max}(\boldsymbol{A}\odot\boldsymbol{A}) \leq C$ *a.s.n.* and $\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}') \leq C$ *a.s.n.*, which we will use repeatedly below. We will also often invoke the bound that for a random vector $\boldsymbol{w}$ with independent elements that have bounded fourth moment, we have $\mathrm{E}[(\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w} - \mathrm{E}[\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w}])^2|\boldsymbol{A}] \leq C\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}')$, see for instance Whittle (1960).

For each of the cases (a)-(e), we decompose (A.31) into three parts that will be treated separately,

$$
\begin{aligned}
&n^{-2}\,\mathrm{E}[(\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r}-\mathrm{E}[\boldsymbol{x}'_{(i),r}\boldsymbol{A}_r\boldsymbol{x}_{(j),r}|\mathcal{J}])^2|\mathcal{J}]\\
&\quad\le 4\,\underbrace{n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}-\mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(I)}\\
&\qquad+4\,\underbrace{n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j),r}-\mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(II)}\\
&\qquad+4\,\underbrace{n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}-\mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}_r\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}]}_{(III)}.
\end{aligned}
\tag{A.32}
$$

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

$$
\begin{aligned}
(a.I)&=n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}-\mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{A}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\
&=n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}r)^2|\mathcal{J}]\\
&=n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}})+n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(j)}}\dot{\boldsymbol{A}}\boldsymbol{D}_{\bar{z}_{(i)}})\\
&\le 2\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}})\left(\frac{1}{n^2}\sum_{k=1}^n\bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^n\bar{z}^4_{(j),k}\right)^{1/2}\to_{a.s.}0,
\end{aligned}
\tag{A.33}
$$

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

$$
\begin{aligned}
(a.II)&=n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)}-\mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}_r\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\
&\le 2n^{-1}\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}})\frac{1}{n}\sum_{k=1}^n\mathrm{E}[u^4_{(i),k}]^{1/2}\,\mathrm{E}[u^4_{(j),k}]^{1/2}\to_{a.s.}0,
\end{aligned}
\tag{A.34}
$$

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

We now continue with $(b.I)-(b.III)$. 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

$$
\begin{aligned}
(b.II)&=n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}\boldsymbol{u}_{(j)}-\mathrm{E}[\boldsymbol{u}'_{(i)}\boldsymbol{A}\boldsymbol{u}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\
&=n^{-2}\,\mathrm{E}[(\boldsymbol{u}'_{(i)}\boldsymbol{A}\boldsymbol{u}_{(j)})^2|\mathcal{J}]-n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_A)^2\\
&=n^{-2}\,\mathrm{E}[\sum_{k,k',l,l'}u_{(i),k}u_{(i),k'}u_{(j),l}u_{(j),l'}A_{kl}A_{k'l'}]-n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\Sigma^U(i,j)}\boldsymbol{D}_A)^2\\
&\le n^{-2}\boldsymbol{u}'_{(i)}\boldsymbol{D}_{u_{(i)}}(\boldsymbol{A}\odot\boldsymbol{A})\boldsymbol{D}_{u_{(j)}}\boldsymbol{u}_{(j)}+n^{-2}\boldsymbol{u}'_{(j)}\boldsymbol{D}_{u_{(i)}}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{D}_{u_{(i)}}\boldsymbol{u}_{(j)}\\
&\le n^{-1}(\lambda_{\max}(\dot{\boldsymbol{A}}\odot\dot{\boldsymbol{A}})+\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}'))\frac{1}{n}\sum_{k=1}^n\mathrm{E}[u^4_{(i),k}]^{1/2}\,\mathrm{E}[u^4_{(j),k}]^{1/2}\to_{a.s.}0.
\end{aligned}
\tag{A.35}
$$

Finally, $(b.III)$ satisfies

$$\begin{aligned}(b.III) &= n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_r\boldsymbol{A}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\ &= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{r})^2|\mathcal{J}] - \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}})^2\\ &\leq Cn^{-2}\,\mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{D}_\varepsilon\boldsymbol{A}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_{a_{(i)}})\\ &\leq Cn^{-2}\,\mathrm{tr}(\boldsymbol{D}_\varepsilon\boldsymbol{A}\boldsymbol{D}_\varepsilon^2\boldsymbol{A}'\boldsymbol{D}_\varepsilon).\end{aligned} \tag{A.36}$$

Using the specific expressions for $\boldsymbol{A}$ as in (A.27) and (A.28), we see that $(b.III) \to_{a.s.} 0$.

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

$$\begin{aligned}(c.I) &= n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\ &= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\ &= n^{-2}\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}^2_{V\bar{z}_{(j)}}\boldsymbol{D}_{\bar{z}_{(i)}}) - 2n^{-2}\,\mathrm{tr}(\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P)\\ &\quad + n^{-2}\boldsymbol{\iota}'(\boldsymbol{P}\odot\boldsymbol{P}\odot(\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}))\boldsymbol{\iota}\\ &= n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)} - 2n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\\ &\quad + n^{-2}\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(i)}}(\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\bar{\boldsymbol{z}}_{(j)}\\ &\leq \left(\frac{1}{n^2}\sum_{k=1}^n \bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^n(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{V}\boldsymbol{e}_k)^4\right)^{1/2} \to_{a.s.} 0,\end{aligned} \tag{A.37}$$

with the convergence implied by Assumption A5.

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

$$\begin{aligned}(c.III) &= n^{-2}\,\mathrm{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \mathrm{E}[\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)}|\mathcal{J}])^2|\mathcal{J}]\\ &= n^{-2}\,\mathrm{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{r} - \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}]\\ &\leq n^{-2}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}^2_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}) \to_{a.s.} 0.\end{aligned} \tag{A.38}$$

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

$$(d.I) = n^{-2}\,\mathrm{E}[(\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)} - \mathrm{E}[\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}])^2|\mathcal{J}]. \tag{A.39}$$

Notice that

$$\begin{aligned}n^{-1}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_r\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_r\bar{\boldsymbol{z}}_{(j)} &= n^{-1}\boldsymbol{\iota}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\iota}\\ &= n^{-1}\boldsymbol{\iota}'(\boldsymbol{D}_P + \boldsymbol{D}_r\dot{\boldsymbol{P}})\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}(\boldsymbol{D}_P + \boldsymbol{D}_r\dot{\boldsymbol{P}})\boldsymbol{\iota}\\ &= n^{-1}\bar{\boldsymbol{z}}'_{(i)}\boldsymbol{D}_P\boldsymbol{V}\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)} + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_P\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{\iota}\\ &\quad + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{V}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_P\boldsymbol{\iota}\\ &\quad + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_V\boldsymbol{D}_{\bar{z}_{(j)}}\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{\iota}\\ &\quad + n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{\iota}.\end{aligned} \tag{A.40}$$

The second and third term have expectation equal to zero. The fourth term has expectation $n^{-1}\bar{\boldsymbol{z}}'_{(i)}(\boldsymbol{D}_V\boldsymbol{D}_P(\boldsymbol{I}-\boldsymbol{D}_P)\bar{\boldsymbol{z}}_{(j)}$. 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

$$\operatorname{tr}(\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}}) + n^{-1}\sum_{k=1}^{n}\sum_{l\neq k} r_k r_l \boldsymbol{r}'_{-kl}\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{r}_{-kl}. \tag{A.41}$$

Squaring and taking the expectation, we get the bound

$$\begin{aligned}
&\frac{2}{n^2}\operatorname{E}[\operatorname{tr}(\dot{\boldsymbol{P}}\boldsymbol{D}_r\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_r\dot{\boldsymbol{P}})^2|\mathcal{J}] + \frac{4}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\operatorname{E}[(\boldsymbol{r}'_{-kl}\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{r}_{-kl})^2|\mathcal{J}]\\
&\leq \frac{2}{n^2}\operatorname{E}[(\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(i)}}(\dot{\boldsymbol{V}}\odot\dot{\boldsymbol{P}}^2)\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{r})^2] + \frac{4}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\operatorname{tr}(\boldsymbol{D}_{Pe_k}\boldsymbol{D}_{\bar{z}_{(i)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{D}_{Pe_l}\boldsymbol{D}_{Pe_l}\boldsymbol{D}_{\bar{z}_{(j)}}\dot{\boldsymbol{V}}\boldsymbol{D}_{\bar{z}_{(i)}}\boldsymbol{D}_{Pe_k})\\
&\leq C\left(\frac{1}{n^2}\sum_{k=1}^{n}\bar{z}^4_{(i),k}\frac{1}{n^2}\sum_{k=1}^{n}\bar{z}^4_{(j),k}\right)^{1/2} \to_{a.s.} 0.
\end{aligned} \tag{A.42}$$

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

$$\begin{aligned}
(d.III) &= n^{-2}\operatorname{E}[(\boldsymbol{a}'_{(i)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_{Pr}\boldsymbol{V}\boldsymbol{D}_{Pr}\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} - \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}]\\
&= n^{-2}\operatorname{E}[(\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{r})^2|\mathcal{J}] - \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}))^2|\mathcal{J}]^2\\
&\leq n^{-2}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}^2_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}) \to_{a.s.} 0.
\end{aligned} \tag{A.43}$$

Parts $(e.I)-(e.III)$ follow using the same techniques used to establish $(a)-(d)$.

Lastly, we consider the estimator of the covariance between the AR and the score statistic. From (16) and (19) we can bound the variance of $[\hat{\boldsymbol{\Sigma}}_{n,r}(\boldsymbol{\beta}_0)]_{1,j}$ as

$$\begin{aligned}
&\operatorname{E}[([\hat{\boldsymbol{\Sigma}}_{n,r}(\boldsymbol{\beta}_0)]_{1,j})^2|\mathcal{J}]\\
&\leq \frac{4}{nk}\operatorname{E}[(\operatorname{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}) - (\bar{\boldsymbol{z}}_{(j)} + \boldsymbol{D}_r\boldsymbol{D}_\varepsilon\boldsymbol{a}_{(j)} + \boldsymbol{u}_{(j)})'(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]\\
&\leq \frac{C}{nk}(\operatorname{E}[(\operatorname{tr}(\boldsymbol{\Psi}^{(j)}\odot\boldsymbol{P}) - \boldsymbol{a}'_{(j)}\boldsymbol{D}_r\boldsymbol{D}_\varepsilon(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]\\
&\quad + \operatorname{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \operatorname{E}[(\boldsymbol{u}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}])\\
&= \frac{C}{nk}(\operatorname{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \operatorname{E}[(\boldsymbol{u}'_{(j)}(\boldsymbol{D}_V - \boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]).
\end{aligned} \tag{A.44}$$

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

$$
\begin{aligned}
&\frac{C}{nk}\mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}(\boldsymbol{D}_V - |\mathcal{J}(\boldsymbol{V}\odot\boldsymbol{P}))\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}]\\
&\leq \frac{C}{nk}(\mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_r\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}] + \mathrm{E}[(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_r(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon})^2|\mathcal{J}])\\
&= \frac{C}{nk}(\mathrm{E}[\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{r}\boldsymbol{r}'\boldsymbol{D}_V\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\bar{\boldsymbol{z}}_{(j)}|\mathcal{J}]\\
&\quad + \mathrm{E}[\boldsymbol{r}'\boldsymbol{D}_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_{\bar{z}_{(j)}}\boldsymbol{r}|\mathcal{J}]) \qquad \text{(A.45)}\\
&= \frac{C}{nk}([\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P^3\bar{\boldsymbol{z}}_{(j)})] + [\boldsymbol{\varepsilon}'\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{\varepsilon}])\\
&= \frac{C}{nk}([\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P^3\bar{\boldsymbol{z}}_{(j)})] + [\boldsymbol{\iota}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{\iota}])\\
&\leq \frac{C}{nk}([\mathrm{tr}(\bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P^3\bar{\boldsymbol{z}}_{(j)})] + \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)}) \to_{a.s.} 0,
\end{aligned}
$$

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

$$
\boldsymbol{\iota}'\boldsymbol{D}_\varepsilon\boldsymbol{D}_P(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}^2_{\bar{z}_{(j)}}(\boldsymbol{V}\odot\boldsymbol{P})\boldsymbol{D}_P\boldsymbol{D}_\varepsilon\boldsymbol{\iota} \leq \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_\varepsilon\boldsymbol{D}_V^2\boldsymbol{D}_\varepsilon\bar{\boldsymbol{z}}_{(j)} = \bar{\boldsymbol{z}}'_{(j)}\boldsymbol{D}_V\boldsymbol{D}_P\bar{\boldsymbol{z}}_{(j)}. \qquad \text{(A.46)}
$$

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

Appendix B Central limit theorem

The proof of Theorem 2 is very long. We therefore only give an outline of it and explain the most important steps. We defer the details to a separate document that is available upon request.

The proof of the CLT 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

$$
\begin{pmatrix} \frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta}) - k) \\ \sqrt{n}\cdot\boldsymbol{S} \end{pmatrix} = \boldsymbol{Y}_n, \qquad \text{(B.1)}
$$

such that it is a martingale difference array.

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

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 $\mathrm{AR}(\boldsymbol{\beta})$ has the same distribution as $\mathrm{AR}_r(\boldsymbol{\beta})$. Therefore $\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta})-k) \overset{(d)}{=} \frac{1}{\sqrt{k}}(\mathrm{AR}_r(\boldsymbol{\beta})-k)$. Then defining

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

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

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

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_h} &= -\frac{1}{n}\boldsymbol{x}'_{(h)}(\boldsymbol{I}-\boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon} \\
&= -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}(\boldsymbol{I}-\boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{D}_{a_{(h)}}(\boldsymbol{I}-\boldsymbol{D}_{P\iota})\boldsymbol{V}\boldsymbol{\varepsilon}\right] \\
&\overset{(d)}{=} -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{r}\right] \\
&= -\frac{1}{n}\left[\bar{\boldsymbol{x}}'_{(h)}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r} + \boldsymbol{r}'\boldsymbol{\Psi}^{(h)}\boldsymbol{r} - \boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{r}\right],
\end{aligned} \tag{B.3}$$

where $\boldsymbol{\Psi}^{(h)} \equiv \boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}$. Define $\boldsymbol{\Phi}^{(h)} \equiv \boldsymbol{D}_{\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon$. We can rewrite the final term as

$$\begin{aligned}
\boldsymbol{r}'\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{r} &= \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}) + \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{\Delta}) \qquad \boldsymbol{\Delta} \equiv \boldsymbol{r}\boldsymbol{r}' - \boldsymbol{I}_n \\
&= \mathrm{tr}(\boldsymbol{\Phi}^{(h)}\boldsymbol{D}_r) + \mathrm{tr}(\boldsymbol{P}\boldsymbol{D}_r\boldsymbol{\Phi}^{(h)}\boldsymbol{\Delta}) \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{\substack{i,j,k \\ i\neq k}} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j, k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} P_{ij}\Phi jj^{(h)}r_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} \varepsilon_iV_{ij}\varepsilon_j\bar{x}_{h,j}V_{jj}\varepsilon_jr_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k + \sum_{i\neq k} P_{ii}\Phi^{(h)}_{ik}r_k + \sum_{i\neq j} P_{ii}\Phi^{(h)}_{ij}r_i \\
&= \bar{\boldsymbol{x}}'_{(h)}\boldsymbol{D}_V\boldsymbol{D}_\varepsilon\boldsymbol{r} + 2\sum_{j\neq i}\Phi^{(h)}_{ji}P_{jj}r_i + \sum_{i\neq j\neq k\neq i} P_{ij}\Phi^{(h)}_{jk}r_ir_jr_k.
\end{aligned} \tag{B.4}$$

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, $\mathrm{tr}(\boldsymbol{\Psi}^{(h)}) = 0$.

We conclude that

$$\sqrt{n}\frac{\partial Q(\boldsymbol{\beta})}{\partial\beta_h} \overset{(d)}{=} -\frac{1}{\sqrt{n}}\sum_{j\neq i}\Phi_{ji}^{(h)}(1-2P_{jj})r_i - \frac{1}{\sqrt{n}}\sum_{j\neq i}\Psi_{ji}^{(h)}r_jr_i + \frac{1}{\sqrt{n}}\sum_{i\neq j\neq k\neq i}P_{ij}\Phi_{jk}^{(h)}r_ir_jr_k$$
$$= w_{1n,S}^{(h)} + \sum_{i=3}^{n} y_{in,S}^{(h)}, \tag{B.5}$$

where

$$w_{1n,S}^{(h)} = -\frac{1}{\sqrt{n}}\sum_{j\neq 1}\Phi_{j1}^{(h)}(1-2P_{jj})r_1 - \frac{1}{\sqrt{n}}\sum_{j\neq 2}\Phi_{j2}^{(h)}(1-2P_{jj})r_2 - \frac{1}{\sqrt{n}}\Psi_{[21]}^{(h)}r_2r_1,$$
$$y_{in,S}^{(h)} = \left[-\frac{1}{\sqrt{n}}\sum_{j\neq i}\Phi_{ji}^{(h)}(1-2P_{jj}) - \frac{1}{\sqrt{n}}\sum_{j<i}\Psi_{[ij]}^{(h)}r_j + \frac{1}{\sqrt{n}}\sum_{l<j<i}A_{[ijl]}^{(h)}r_jr_l\right]\cdot r_i, \tag{B.6}$$

and $\Psi_{[ij]}^{(h)} = \Psi_{ij}^{(h)} + \Psi_{ji}^{(h)}$, $A_{[ijk]}^{(h)} = A_{ijk}^{(h)} + A_{ikj}^{(h)} + A_{jik}^{(h)} + A_{jki}^{(h)} + A_{kij}^{(h)} + A_{kji}^{(h)}$, $A_{ijk}^{(h)} = P_{ij}\Phi_{jk}^{(h)}$.

Then the full statistic becomes

$$\boldsymbol{Y}_n = \begin{pmatrix}\frac{1}{\sqrt{k}}(\mathrm{AR}(\boldsymbol{\beta})-k)\\ \sqrt{n}\cdot\boldsymbol{S}\end{pmatrix} = \begin{pmatrix}w_{1n,\mathrm{AR}}\\ \boldsymbol{w}_{1n,S}\end{pmatrix} + \sum_{i=3}^{n}\begin{pmatrix}y_{in,\mathrm{AR}}\\ \boldsymbol{y}_{in,S}\end{pmatrix}. \tag{B.7}$$

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

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

$$(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n = w_{1n} + \sum_{i=3}^{n} y_{in}, \tag{B.8}$$

where we define

$$w_{1n} = \Xi_n^{-1/2}\left[c_{1n}w_{1n,\mathrm{AR}} + \boldsymbol{c}_{2n}'\boldsymbol{w}_{1n,S}\right],$$
$$y_{in} = \Xi_n^{-1/2}\left[-\frac{1}{\sqrt{n}}\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii}) - \frac{1}{\sqrt{n}}\sum_{j<i}(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ij]} - 2c_{1n}\gamma_nP_{ij})r_j + \frac{1}{\sqrt{n}}\sum_{l<j<i}\boldsymbol{c}_{2n}'\boldsymbol{a}_{[ijl]}r_lr_j\right]\cdot r_i, \tag{B.9}$$

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 then also $\boldsymbol{c}_n'\boldsymbol{c}_n\leq C$, 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. Let $\boldsymbol{\Psi} = \boldsymbol{M}\boldsymbol{D}_{\sum_{h=1}^{p} c_{2n,h} a_{(h)}} \boldsymbol{P}$ and $\boldsymbol{\Phi} = \boldsymbol{D}_{\sum_{h=1}^{p} c_{2n,h} \bar{x}_{(h)}} \boldsymbol{V}\boldsymbol{D}_{\varepsilon}$ then we can write

$$\begin{aligned} y_{in} = \Xi_n^{-1/2} \Big\{ &- \frac{1}{\sqrt{n}} \boldsymbol{c}_{2n}' \bar{\boldsymbol{X}}' (\boldsymbol{I}_n - 2\boldsymbol{D}_P) \dot{\boldsymbol{V}} \boldsymbol{D}_{\varepsilon} \boldsymbol{e}_i \\ &- \frac{1}{\sqrt{n}} \boldsymbol{r}' \boldsymbol{S}_{i-1} \left[(\boldsymbol{\Psi} + \boldsymbol{\Psi}' - 2\boldsymbol{D}_{\Psi}) - 2c_{1n}\gamma_n \dot{\boldsymbol{P}} \right] \boldsymbol{e}_i + \frac{1}{\sqrt{n}} \boldsymbol{r}' \boldsymbol{A}_{-i} \boldsymbol{r} \Big\} \cdot r_i, \\ \boldsymbol{A}_{-i} = \boldsymbol{S}_{i-1} \boldsymbol{A}_i \boldsymbol{S}_{i-1} &= \boldsymbol{S}_{i-1} [\dot{\boldsymbol{P}} \boldsymbol{D}_{\Phi e_i} + \boldsymbol{D}_{\Phi e_i} \dot{\boldsymbol{P}} + \boldsymbol{P} \boldsymbol{e}_i \boldsymbol{e}_i' \boldsymbol{\Phi} - \boldsymbol{D}_{P e_i} \boldsymbol{D}_{e_i' \Phi}] \boldsymbol{S}_{i-1}. \end{aligned} \tag{B.10}$$

To see that the last term in (B.10) equals the last term of y_{in} in (B.9) 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

$$\begin{aligned} \boldsymbol{r}' \boldsymbol{S}_{i-1} \dot{\boldsymbol{P}} \boldsymbol{D}_{\Phi e_i} \boldsymbol{S}_{i-1} \boldsymbol{r} &= \sum_{h=1}^{p} c_{2n,h} [\sum_{l<j<i} P_{jl} \bar{x}_{(h),l} V_{li} \varepsilon_i r_j r_l + \sum_{l<j<i} P_{lj} \bar{x}_{(h),l} V_{ji} \varepsilon_i r_j r_l] \\ &= \sum_{h=1}^{p} c_{2n,h} [\sum_{l<j<i} A_{jli}^{(h)} r_j r_l + \sum_{l<j<i} A_{lji}^{(h)} r_j r_l], \end{aligned} \tag{B.11}$$

the second term

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

and third term

$$\begin{aligned} &\boldsymbol{r}' \boldsymbol{S}_{i-1} (\boldsymbol{P} \boldsymbol{e}_i \boldsymbol{e}_i' \boldsymbol{\Phi}^{(h)} - \boldsymbol{D}_{P e_i} \boldsymbol{D}_{e_i' \Phi^{(h)}}) \boldsymbol{S}_{i-1} \boldsymbol{r} \\ &= \sum_{h=1}^{p} c_{2n,h} [\sum_{l<j<i} P_{ji} \bar{x}_{(h),i} V_{il} \varepsilon_l r_j r_l + \sum_{l<j<i} P_{li} \bar{x}_{(h),i} V_{ij} \varepsilon_j r_j r_l] \\ &= \sum_{h=1}^{p} c_{2n,h} [\sum_{l<j<i} A_{jil}^{(h)} r_j r_l + \sum_{l<j<i} A_{lij}^{(h)} r_j r_l]. \end{aligned} \tag{B.12}$$

Furthermore note that $\boldsymbol{A}_{-i}$ is a symmetric matrix with all diagonal elements equal to zero.

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

We start by noting the following result.

Result 1. For $\boldsymbol{A}$ $m \times n$ and $\boldsymbol{B}$ $n \times m$ with $n \geq m$, the eigenvalues of $\boldsymbol{AB}$ equal those of $\boldsymbol{BA}$ plus $n - m$ zeroes (Magnus and Neudecker, 1998, Ch. 1 T9). Hence if $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices $\lambda_{\max}(\boldsymbol{AB}) = \lambda_{\max}(\boldsymbol{BA})$.

Next, using Assumption A5, we have that

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{j=1}^{n} \|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_j\varepsilon_j\|^4 \\
&\leq \frac{1}{n^2}\max_{l=1,\ldots,n} \|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_l\varepsilon_l\|^2 \sum_{j=1}^{n} \|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_j\varepsilon_j\|^2 \\
&\leq \frac{o_{a.s.}(1)}{n}\sum_{h=1}^{p}\sum_{j=1}^{n}(\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_j\varepsilon_j)^2 \\
&= \frac{o_{a.s.}(1)}{n}\sum_{h=1}^{p}\sum_{j=1}^{n}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_{\varepsilon}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\bar{\boldsymbol{Z}}\boldsymbol{e}_h \qquad \text{(B.13)} \\
&= \frac{o_{a.s.}(1)}{n}\sum_{h=1}^{p}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\bar{\boldsymbol{Z}}\boldsymbol{e}_h \\
&= \frac{b_n}{n}\sum_{h=1}^{p}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\bar{\boldsymbol{Z}}\boldsymbol{e}_h \\
&\leq \frac{o_{a.s.}(1)C}{n}\lambda_{\max}(\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{Z}')\sum_{h=1}^{p}\boldsymbol{e}_h'\bar{\boldsymbol{Z}}'\bar{\boldsymbol{Z}}\boldsymbol{e}_h \to_{a.s.} 0,
\end{aligned}
$$

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

Also under Assumption A5 and by the finite kurtosis of $\boldsymbol{U}$

$$\begin{aligned}
&\mathrm{E}[\frac{1}{n^2}\sum_{i,j=1}^{n}\|\phi_{ij}\|^4|\mathcal{J}] \\
&= \mathrm{E}[\frac{1}{n^2}\sum_{i,j=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i V_{ij}\varepsilon_j\|^4|\mathcal{J}] \\
&= \mathrm{E}[\frac{1}{n^2}\sum_{i,j=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4 (V_{ij}\varepsilon_j)^4|\mathcal{J}] \\
&\leq \mathrm{E}[\frac{1}{n^2}\sum_{i,j,k=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4 (\boldsymbol{e}_i'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_i)^2(\boldsymbol{e}_i'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_i)^2|\mathcal{J}] \qquad \text{(B.14)}\\
&= \mathrm{E}[\frac{1}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4 V_{ii}^2|\mathcal{J}] \\
&\leq \mathrm{E}[\frac{C}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{X}}'\boldsymbol{e}_i\|^4|\mathcal{J}] \\
&\leq \frac{C}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{e}_i\|^4 + \mathrm{E}[\|\boldsymbol{U}'\boldsymbol{e}_i\|^4|\mathcal{J}] \to_{a.s.} 0.
\end{aligned}$$

Furthermore we have

$$\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}] \\
&= \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\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{u}_{(h)}\boldsymbol{u}_{(h)}'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{u}_{(h)}|\mathcal{J}] \\
&= \frac{C}{n^2}\sum_{h=1}^{p}\Bigg(\sum_{i,j=1}^{n}\mathrm{E}[u_{(h),j}^4|\mathcal{J}]](\boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_j)^2 \\
&\quad + 2\sum_{i,j,k=1}^{n}\mathrm{E}[u_{(h),j}^2u_{(h),k}^2|\mathcal{J}]](\boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_k)^2 \\
&\quad + \sum_{i,j,k=1}^{n}\mathrm{E}[u_{(h),j}^2u_{(h),k}^2|\mathcal{J}]]\boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_j\boldsymbol{e}_k'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_k\Bigg) \\
&\leq \frac{C}{n^2}\sum_{h=1}^{p}\Bigg(C\sum_{i,j=1}^{n}\mathrm{E}[u_{(h),j}^4|\mathcal{J}]]\boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{e}_j + 3\sum_{i=1}^{n}\mathrm{E}[u_{(h),j}^2u_{(h),k}^2|\mathcal{J}]](\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i)^2\Bigg) \\
&\leq \frac{C}{n^2}\sum_{h=1}^{p}\Bigg(C\sum_{j=1}^{n}\mathrm{E}[u_{(h),j}^4|\mathcal{J}]]\boldsymbol{e}_j'\boldsymbol{V}\boldsymbol{e}_j + 3C\sum_{i=1}^{n}\mathrm{E}[u_{(h),j}^2u_{(h),k}^2|\mathcal{J}]]\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i\Bigg) \to_{a.s.} 0,
\end{aligned}$$

(B.15)

because by independence of the $u_{(h),i}$ there exists a similar result as Item 2 of Theorem A.1 for $\boldsymbol{u}_{(h)}$, and by Assumption A5 and Result 1 and the finite kurtosis of $\boldsymbol{U}$.

Next, note that

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

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

$$\begin{aligned}
&\mathrm{E}\left[\|\boldsymbol{c}_{2n}'\boldsymbol{w}_{1n,S}\|^4\middle|\mathcal{J}\right] \leq C\cdot\mathrm{E}\left[\|\boldsymbol{w}_{1n,S}\|^4\middle|\mathcal{J}\right]\\
&= C\,\mathrm{E}\left[\|\frac{-1}{\sqrt{n}}\sum_{j\neq 1}\boldsymbol{\phi}_{j1}(1-2P_{jj})r_1 - \frac{1}{\sqrt{n}}\sum_{j\neq 2}\boldsymbol{\phi}_{j2}(1-2P_{jj})r_2 - \frac{1}{\sqrt{n}}\boldsymbol{\psi}_{[21]}r_2r_1\|^4\middle|\mathcal{J}\right]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\left[\|\sum_{j\neq 1}\boldsymbol{\phi}_{j1}(1-2P_{jj})\|^4 + \|\sum_{j\neq 2}\boldsymbol{\phi}_{j2}(1-2P_{jj})\|^4 + \|\boldsymbol{\psi}_{[21]}\|^4\middle|\mathcal{J}\right]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\Big[\|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_1\varepsilon_1\|^4 + \|\boldsymbol{U}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_1\varepsilon_1\|^4 + \|\boldsymbol{\phi}_{11}(1-2P_{11})\|^4\\
&\quad + \|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_2\varepsilon_2\|^4 + \|\boldsymbol{U}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_2\varepsilon_2\|^4 + \|\boldsymbol{\phi}_{22}(1-2P_{22})\|^4\\
&\quad + p\max_{h=1,\dots,p}(\Psi_{[21]}^{(h)})^4\Big|\mathcal{J}\Big]\\
&\leq \frac{C}{n^2}\,\mathrm{E}\left[\sum_{j=1}^n\|\bar{\boldsymbol{Z}}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_j\varepsilon_j\|^4 + \|\boldsymbol{U}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_{\varepsilon^2}\boldsymbol{Z})^{-1}\boldsymbol{z}_j\varepsilon_j\|^4 + \|\boldsymbol{\phi}_{11}\|^4 + \|\boldsymbol{\phi}_{22}\|^4 + C\middle|\mathcal{J}\right]\\
&\rightarrow_{a.s.} 0,
\end{aligned} \tag{B.17}$$

where for the final line we use (B.13), (B.14), (B.15), Assumption A5 and that

$$\Psi_{jk}^{(h)} = \boldsymbol{e}_j'\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{e}_k \leq \boldsymbol{e}_j\boldsymbol{M}\boldsymbol{e}_j\boldsymbol{e}_k'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}^2}\boldsymbol{P}\boldsymbol{e}_k \leq \max_{i=1,\dots,n} a_{(h),i}^2 \leq C \quad a.s.n. \tag{B.18}$$

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

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

$$(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n = \sum_{i=3}^n y_{in} + o_p(1). \tag{B.19}$$

B.2.2 Martingale difference sequence

Define the σ-fields $\mathcal{F}_{i,n} = \sigma(r_1,\dots,r_i)$ 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 $\boldsymbol{\Omega}$. Let $\boldsymbol{v}$ be any p dimensional vector. Then

$$\boldsymbol{v}'\boldsymbol{\Omega}\boldsymbol{v} = \frac{1}{n}\boldsymbol{v}[\bar{\boldsymbol{Z}}'\tilde{\boldsymbol{\Omega}}\bar{\boldsymbol{Z}} + \mathrm{E}[\boldsymbol{U}'\tilde{\boldsymbol{\Omega}}\boldsymbol{U}|\mathcal{J}] + \boldsymbol{T}]\boldsymbol{v}, \tag{B.20}$$

where from (12) we have $\tilde{\boldsymbol{\Omega}} = \boldsymbol{V} - 3\boldsymbol{D}_P\boldsymbol{D}_V + 2\boldsymbol{D}_P^2\boldsymbol{V} + 2\boldsymbol{V}\boldsymbol{D}_P^2 - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P})\odot \boldsymbol{I} + 8\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P + 4\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P} - 4\boldsymbol{D}_P^2\dot{\boldsymbol{V}}\boldsymbol{D}_P - 4\boldsymbol{D}_P\dot{\boldsymbol{V}}\boldsymbol{D}_P^2 - 4(\dot{\boldsymbol{V}}\odot\boldsymbol{P}\odot\boldsymbol{P})\boldsymbol{D}_P - 4\boldsymbol{D}_P(\dot{\boldsymbol{V}}\odot \boldsymbol{P}\odot\boldsymbol{P}) - 3\boldsymbol{D}_P\dot{\boldsymbol{V}} - 3\dot{\boldsymbol{V}}\boldsymbol{D}_P$ and $T_{ij} = \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) - 2\,\mathrm{tr}(\boldsymbol{D}_P^2\boldsymbol{D}_{a_{(i)}}\boldsymbol{D}_{a_{(j)}}) - 2\,\mathrm{tr}((\boldsymbol{I}_n\odot \boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P})\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}) + 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}) + 2\,\mathrm{tr}(\boldsymbol{D}_P\boldsymbol{P}\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}) - \mathrm{tr}(\boldsymbol{D}_{a_{(i)}}\boldsymbol{P}\boldsymbol{D}_{a_{(j)}}\boldsymbol{P})$.

Each of the three terms is bigger or equal to zero and for the second we have

$$\begin{aligned}\mathrm{E}[\boldsymbol{v}'\boldsymbol{U}'\tilde{\boldsymbol{\Omega}}\boldsymbol{U}\boldsymbol{v}|\mathcal{J}] &= \mathrm{E}[\sum_{i,j,k,l=1}^{n} v_iU_{ji}\tilde{\Omega}_{jk}U_{kl}v_l|\mathcal{J}] = \mathrm{E}[\sum_{i,j,l=1}^{n} v_iU_{ji}\tilde{\Omega}_{jj}U_{jl}v_l|\mathcal{J}] \\ &= \mathrm{E}[\boldsymbol{v}'\boldsymbol{U}'\boldsymbol{D}_{\tilde{\Omega}}\boldsymbol{U}\boldsymbol{v}|\mathcal{J}] \geq \lambda_{\min}(\boldsymbol{D}_{\tilde{\Omega}})\,\mathrm{E}[\boldsymbol{v}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{v}|\mathcal{J}],\end{aligned} \tag{B.21}$$

with the diagonal elements of $\boldsymbol{D}_{\tilde{\Omega}}$ bounded away from zero because

$$\begin{aligned}\tilde{\Omega}_{ii} &= \boldsymbol{e}_i'(\boldsymbol{V} - 3\boldsymbol{D}_P\boldsymbol{D}_V - 2(\boldsymbol{V}\boldsymbol{D}_\varepsilon\odot\boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P}\odot\boldsymbol{P}) + 4\boldsymbol{D}_P^2\boldsymbol{D}_V)\boldsymbol{e}_i \\ &= V_{ii}(1 - 3P_{ii} + 4P_{ii}^2) - 2\sum_{j=1}^{n} V_{ij}^2\varepsilon_j^2P_{ji}^2 \\ &\geq V_{ii}(1 - 3P_{ii} + 4P_{ii}^2) - 2\sum_{j=1}^{n} V_{ij}^2\varepsilon_j^2P_{ii}P_{jj} \\ &\geq V_{ii}(1 - 3P_{ii} + 4P_{ii}^2) - 2\max_{k=1,\dots,n} P_{kk}\sum_{j=1}^{n} V_{ij}^2\varepsilon_j^2P_{ii} \\ &\geq V_{ii}(1 - 3P_{ii} + 4P_{ii}^2) - V_{ii}P_{ii} \\ &= V_{ii}(1 - 4P_{ii} + 4P_{ii}^2) > 0,\end{aligned} \tag{B.22}$$

for $P_{ii} < \frac{1}{2}$ and by Assumption A5. Then because $\mathrm{E}[\boldsymbol{U}'\boldsymbol{U}|\mathcal{J}]$ is positive definite, we conclude that (B.21) is bounded away from zero and hence so is (B.20).

Now let $\boldsymbol{b} = [\boldsymbol{\Sigma}]_{2:p+1,1}$ the covariance between the AR statistic and the score. Then note that $\det(\boldsymbol{\Sigma}_n) = \det(\boldsymbol{\Omega})\det(\boldsymbol{\Omega} - \boldsymbol{b}\boldsymbol{b}'\sigma_n^{-2})$ 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. Denote $\boldsymbol{\rho}$ 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

$$\begin{aligned}\det(\boldsymbol{\Sigma}_n) &= \det(\boldsymbol{\Omega})\det(\boldsymbol{\Omega} - \boldsymbol{b}\boldsymbol{b}'\sigma_n^{-2}) \\ &= \det(\boldsymbol{\Omega})\det(\boldsymbol{\Omega} - \boldsymbol{D}_\rho\boldsymbol{\Omega}\boldsymbol{D}_\rho) \\ &= \det(\boldsymbol{\Omega})\det((\boldsymbol{I} + \boldsymbol{D}_\rho)\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{D}_\rho)) \\ &= \det(\boldsymbol{\Omega})\det(\boldsymbol{I} + \boldsymbol{D}_\rho)\det(\boldsymbol{\Omega})\det(\boldsymbol{I} - \boldsymbol{D}_\rho) > 0,\end{aligned} \tag{B.23}$$

if $\rho_i \neq \pm 1$ for all i.

Therefore $\boldsymbol{\Sigma}_n^{-1}$ exists.

B.2.4 Lyapunov condition

We will now show that

$$\begin{aligned}\sum_{i=3}^{n} \text{E}[y_{in}^4|\mathcal{J}] = \sum_{i=3}^{n} \text{E}[(\Xi_n^{-1/2} \Bigg[&-\frac{1}{\sqrt{n}}\sum_{j\neq i} \boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii}) \\ &- \frac{1}{\sqrt{n}}\sum_{j<i}(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ij]} - 2c_{1n}\gamma_n P_{ij})r_j + \frac{1}{\sqrt{n}}\sum_{l<j<i}\boldsymbol{c}_{2n}'\boldsymbol{a}_{[ijk]}r_l r_j\Bigg] r_i)^4|\mathcal{J}] \\ \leq C\Xi_n^{-2}\sum_{i=3}^{n} &\underbrace{\text{E}[\left(-\frac{1}{\sqrt{n}}\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii})r_i\right)^4|\mathcal{J}]}_{\text{linear}} \\ &+ \underbrace{\text{E}[\left(\frac{1}{\sqrt{n}}\sum_{j<i}(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ij]} - 2c_{1n}\gamma_n P_{ij})r_j r_i\right)^4|\mathcal{J}]}_{\text{quadratic}} \\ &+ \underbrace{\text{E}[\left(\frac{1}{\sqrt{n}}\sum_{l<j<i}\boldsymbol{c}_{2n}'\boldsymbol{a}_{[ijk]}r_l r_j r_i\right)^4|\mathcal{J}]}_{\text{cubic}} \to_{a.s.} 0.\end{aligned} \tag{B.24}$$

First of all we require Ξ_n to be bounded away from zero to assert that Ξ_n^{-2} is finite.

$$\Xi_n = \text{var}(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-\frac{1}{2}}\boldsymbol{Y}_n|\mathcal{J}) = (\boldsymbol{\alpha}'\boldsymbol{\alpha})\,\text{var}(w_{1n} + \sum_{i=3}^{n} y_{in}|\mathcal{J}) = (\boldsymbol{\alpha}'\boldsymbol{\alpha})(1 + o_{a.s.}(1)) > 0, \tag{B.25}$$

since $\boldsymbol{\alpha}'\boldsymbol{\alpha} > 0$. Next we will consider the linear, quadratic and cubic terms one by one.

Linear term For the term linear in $\boldsymbol{r}$, it suffices to show that

$$\text{E}[\frac{1}{n^2}\sum_{i=3}^{n}\left(\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii})\right)^4|\mathcal{J}] \to_{a.s.} 0. \tag{B.26}$$

We have that

$$
\begin{aligned}
&\mathrm{E}[\frac{1}{n^2}\sum_{i=3}^{n}\left(\sum_{j\neq i}\boldsymbol{c}_{2n}'\boldsymbol{\phi}_{ji}(1-2P_{ii})\right)^4|\mathcal{J}]\\
&\leq \mathrm{E}[\frac{C}{n^2}\sum_{i=3}^{n}(1-2P_{ii})^4\|\sum_{j=1}^{n}\boldsymbol{\phi}_{ji}-\boldsymbol{\phi}_{ii}\|^4|\mathcal{J}]\\
&\leq \mathrm{E}[\frac{C}{n^2}\sum_{i=3}^{n}\left(\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i\|^4+\|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i\|^4+\|\boldsymbol{\phi}_{ii}\|^4\right)|\mathcal{J}]\\
&\leq \mathrm{E}[\frac{C}{n^2}\sum_{i=1}^{n}\left(\|\bar{\boldsymbol{Z}}'\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i\|^4+\|\boldsymbol{U}'\boldsymbol{V}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i\|^4+\|\boldsymbol{\phi}_{ii}\|^4\right)|\mathcal{J}]\rightarrow_{a.s.}0,
\end{aligned}
\tag{B.27}
$$

since $(1-2P_{ii})^2<1$ and by Assumption A5, (B.13), (B.14) and (B.15).

Quadratic term For the term quadratic in $\boldsymbol{r}$ in (B.24), we first notice that

$$
\frac{1}{n^2}\sum_{i=3}^{n}\mathrm{E}[\|\sum_{j<i}\gamma_n P_{ij}r_i r_j\|^4|\mathcal{J}]\leq\frac{\gamma_n^2}{n^2}\sum_{i=3}^{n}\left(\sum_{j<i}P_{ij}^4+3\sum_{\substack{(j,m)<i\\ j\neq m}}P_{ij}^2P_{im}^2\right)\leq C\frac{k}{nk}\rightarrow 0. \tag{B.28}
$$

Similarly,

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=3}^{n}\mathrm{E}\left[\|\sum_{k<i}\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}r_i r_k\|^4\middle|\mathcal{J}\right]\\
&\leq\frac{C}{n^2}\sum_{i=3}^{n}\sum_{k<i}\sum_{l<i}\sum_{m<i}\sum_{s<i}|\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[il]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[im]}||\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[is]}|\,\mathrm{E}[r_k r_l r_m r_s|\mathcal{J}]\\
&\leq\frac{C}{n^2}\sum_{i=3}^{n}\left(\sum_{k<i}\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}\right)^4+3\sum_{\substack{(k,m)<i\\ k\neq m}}\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[ik]}\right)^2\left(\boldsymbol{c}_{2n}'\boldsymbol{\psi}_{[im]}\right)^2\right)\\
&\leq\frac{C}{n^2}\sum_{i=3}^{n}\left(\sum_{k<i}\|\boldsymbol{\psi}_{[ik]}\|^4+3\sum_{\substack{(k,m)<i\\ k\neq m}}\|\boldsymbol{\psi}_{[ik]}\|^2\|\boldsymbol{\psi}_{[im]}\|^2\right)\\
&\leq\frac{C}{n^2}\sum_{i=3}^{n}\left(\sum_{k<i}\sum_{h=1}^{p}\left(\Psi_{[ik]}^{(h)}\right)^4+3\sum_{\substack{(k,m)<i\\ k\neq m}}\sum_{h=1}^{p}\left(\Psi_{[ik]}^{(h)}\right)^2\sum_{h=1}^{p}\left(\Psi_{[im]}^{(h)}\right)^2\right).
\end{aligned}
\tag{B.29}
$$

To bound this expression, note that by (B.18) $\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{e}_i\leq C$ *a.s.n.* Also, for any vector $\boldsymbol{v}$

$$
\boldsymbol{v}'\boldsymbol{\Psi}^{(h)'}\boldsymbol{\Psi}^{(h)}\boldsymbol{v}=\boldsymbol{v}'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{v}\leq\max_{i=1,\ldots,n}a_{(h),i}^2\cdot\boldsymbol{v}'\boldsymbol{P}\boldsymbol{v}. \tag{B.30}
$$

This implies that

$$\sum_{i=1}^{n}(\Psi_{ij}^{(h)})^2 \leq \max_{i=1,\dots,n} a_{(h),i}^2 \cdot P_{jj} \leq C \quad a.s.n. \tag{B.31}$$

Then,

$$\begin{aligned}
\frac{1}{n^2}\sum_{i,k=1}^{n}(\Psi_{ik}^{(h)})^4 &\leq \frac{1}{n^2}\sum_{i,k=1}^{n}\left(\sum_{j=1}^{n}(\Psi_{jk}^{(h)})^2\right)(\Psi_{ik}^{(h)})^2 \\
&\leq \frac{1}{n^2}\sum_{i,k=1}\max_{j=1,\dots,n} a_{(h),j}P_{kk}(\Psi_{ik}^{(h)})^2 \\
&\leq \frac{1}{n^2}\max_{j=1,\dots,n} a_{(h),j}\sum_{k=1}^{n}P_{kk}\sum_{i=1}^{n}(\Psi_{ik}^{(h)})^2 \\
&\leq \frac{1}{n^2}\max_{j=1,\dots,n} a_{(h),j}^2\sum_{k=1}^{n}P_{kk}^2 \\
&\leq \frac{Ck}{n^2} \to_{a.s.} 0.
\end{aligned} \tag{B.32}$$

Using this result, we have that

$$\frac{1}{n^2}\sum_{i=3}^{n}\sum_{k<i}\left(\Psi_{[ik]}^{(h)}\right)^4 \leq \frac{1}{n^2}\sum_{k,i=1}^{n}\left(\Psi_{ik}^{(h)}+\Psi_{ki}^{(h)}\right)^4 \leq \frac{C}{n^2}\sum_{k,i=1}^{n}(\Psi_{ik}^{(h)})^4+(\Psi_{ki}^{(h)})^4 \leq \frac{Ck}{n^2} \to_{a.s.} 0. \tag{B.33}$$

And also we find that

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=3}^{n}\sum_{\substack{(k,m)<i \\ k\neq m}}\left(\Psi_{[ik]}^{(h)}\right)^2\left(\Psi_{[im]}^{(h)}\right)^2 &\leq \frac{C}{n^2}\sum_{i,k,m=1}^{n}\left((\Psi_{ik}^{(h)})^2+(\Psi_{ki}^{(h)})^2\right)\left((\Psi_{im}^{(h)})^2+(\Psi_{mi}^{(h)})^2\right) \\
&\leq \frac{C}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ik}^{(h)})^2(\Psi_{im}^{(h)})^2+(\Psi_{ik}^{(h)})^2(\Psi_{mi}^{(h)})^2 \\
&\quad+(\Psi_{ki}^{(h)})^2(\Psi_{im}^{(h)})^2+(\Psi_{ki}^{(h)})^2(\Psi_{mi}^{(h)})^2.
\end{aligned} \tag{B.34}$$

For which we use

$$\begin{aligned}
\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ik}^{(h)})^2(\Psi_{im}^{(h)})^2 &\leq \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}(\Psi_{ik}^{(h)})^2\right)^2 \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\left(\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i\right)^2 \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i \\
&\leq \frac{1}{n^2}\,\mathrm{tr}(\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}) \\
&\leq \frac{1}{n^2}\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}) \\
&\leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0,
\end{aligned} \tag{B.35}$$

and

$$\begin{aligned}
\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ik}^{(h)})^2(\Psi_{mi}^{(h)})^2 &\leq \frac{1}{n^2}\sum_{i=1}^{n}\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{\Psi}^{(h)}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}|\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{\Psi}^{(h)}\boldsymbol{e}_i||\boldsymbol{e}_i'\boldsymbol{\Psi}^{(h)}\boldsymbol{\Psi}^{(h)\prime}\boldsymbol{e}_i| \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}|\boldsymbol{e}_i'\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{e}_i||\boldsymbol{e}_i'\boldsymbol{P}\boldsymbol{D}_{a_{(h)}}\boldsymbol{M}\boldsymbol{D}_{a_{(h)}}\boldsymbol{P}\boldsymbol{e}_i| \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}|\max_{j=1,\ldots,n} a_{(h),j}^2|^2 P_{ii} \\
&\leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0.
\end{aligned} \tag{B.36}$$

Similarly we find

$$\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ki}^{(h)})^2(\Psi_{im}^{(h)})^2 \leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0. \tag{B.37}$$

Lastly,

$$\frac{1}{n^2}\sum_{i,k,m=1}^{n}(\Psi_{ki}^{(h)})^2(\Psi_{mi}^{(h)})^2 \leq \frac{1}{n^2}\sum_{i,k=1}^{n}(\Psi_{ki}^{(h)})^2\sum_{m=1}^{n}(\Psi_{mi}^{(h)})^2 \leq \frac{1}{n^2}\sum_{i=1}^{n}\max_{j=1,\ldots,n} a_{(h),j}^4 P_{ii}^2 \leq \frac{Ck}{n^2} \rightarrow_{a.s.} 0. \tag{B.38}$$

Consequently, the quadratic term converges to zero almost surely.

Cubic term From (B.10), the cubic term can be written as,

$$\sum_{i=3}^{n} \mathrm{E}[((\frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r})r_i)^4|\mathcal{J}] = \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r})^4|\mathcal{J},\boldsymbol{U}]|\mathcal{J}]. \tag{B.39}$$

Since $\boldsymbol{A}_{-i}$ is symmetric and has diagonal elements equal to zero, we have by Item 4 of Theorem A.1

$$\begin{aligned}\sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{E}[(\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r})^4|\mathcal{J},\boldsymbol{U}]|\mathcal{J}] &= \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[]12\,\mathrm{tr}(\boldsymbol{A}_{-i}^2)^2 + 48\,\mathrm{tr}(\boldsymbol{A}_{-i}^4) - 96\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_{-i}^2)^2\boldsymbol{\iota} \\ &\quad + 32\boldsymbol{\iota}'(\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i})\boldsymbol{\iota}|\mathcal{J}] \\ &\le \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[92\,\mathrm{tr}(\boldsymbol{A}_{-i}^2)^2 - 96\boldsymbol{\iota}'(\boldsymbol{I}\odot\boldsymbol{A}_{-i}^2)^2\boldsymbol{\iota}|\mathcal{J}] \\ &\le \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[92\,\mathrm{tr}(\boldsymbol{A}_{-i}^2)^2|\mathcal{J}],\end{aligned} \tag{B.40}$$

where the second inequality follows since $\boldsymbol{A}_{-i}^2$ is positive semidefinite, hence $\mathrm{tr}(\boldsymbol{A}_{-i}^4) \le \mathrm{tr}(\boldsymbol{A}_{-i}^2)^2$, and

$$\begin{aligned}\boldsymbol{\iota}'(\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i}\odot\boldsymbol{A}_{-i})\boldsymbol{\iota} &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}\boldsymbol{e}_j)^4 \le \sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i(\boldsymbol{e}_i'\boldsymbol{A}_{-i}\boldsymbol{e}_j)^2 \\ &= \sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i)^2 \le \left(\sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{A}_{-i}^2\boldsymbol{e}_i)\right)^2 = \mathrm{tr}(\boldsymbol{A}_{-i}^2)^2.\end{aligned} \tag{B.41}$$

Now,

$$\begin{aligned}\sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_{-i}^2)|\mathcal{J}] &= \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{tr}(\boldsymbol{S}_{i-1}\boldsymbol{A}_i\boldsymbol{S}_{i-1}\boldsymbol{A}_i\boldsymbol{S}_{i-1})|\mathcal{J}] \\ &\le \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{tr}(\boldsymbol{A}_i^2)|\mathcal{J}] \\ &\le \sum_{i=3}^{n} \frac{C}{n^2}\mathrm{E}[\mathrm{tr}([\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i}]^2 + [\boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}}]^2 + [\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}]^2 - [\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}]^2)|\mathcal{J}].\end{aligned} \tag{B.42}$$

To bound these four terms we use the following result

Result 2.

$$\begin{aligned}\frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i)^2|\mathcal{J}] &= \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i|\mathcal{J}]\\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i|\mathcal{J}]\\ &= \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\boldsymbol{e}_i'\boldsymbol{D}_\varepsilon\boldsymbol{V}\boldsymbol{D}^2_{\sum_{h=1}^{p}c_{2n,h}\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}^2_{\sum_{h=1}^{p}c_{2n,h}\bar{x}_{(h)}}\boldsymbol{V}\boldsymbol{D}_\varepsilon\boldsymbol{e}_i|\mathcal{J}] \qquad (B.43)\\ &\leq \frac{C}{n^2}\sum_{h=1}^{p}\sum_{i=1}^{n}\mathrm{E}[\bar{x}^4_{(h),i}|\mathcal{J}]\\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\|\bar{\boldsymbol{Z}}'\boldsymbol{e}_i\|^4+\mathrm{E}[\|\boldsymbol{U}'\boldsymbol{e}_i\|^4|\mathcal{J}] \to_{a.s.} 0,\end{aligned}$$

by Assumption A5 and the finite kurtosis of $\boldsymbol{U}$.

We can then bound the four terms as follows. For the first and second term we have by Result 2

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i}\dot{\boldsymbol{P}}\boldsymbol{D}_{\Phi e_i})|\mathcal{J}] &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{D}_{\Phi e_i}\boldsymbol{P}\boldsymbol{D}_{\Phi e_i})|\mathcal{J}]\\ &= \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{D}_{\Phi e_i}\boldsymbol{P}\boldsymbol{D}_{\Phi e_i}\boldsymbol{P})|\mathcal{J}]\\ &= \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left(\sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{P}\boldsymbol{D}_{\Phi e_i}\boldsymbol{P}\boldsymbol{D}_{\Phi e_i}\boldsymbol{P}\boldsymbol{e}_j\right)^2|\mathcal{J}] \qquad (B.44)\\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left(\sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{P}\boldsymbol{D}^2_{\Phi e_i}\boldsymbol{P}\boldsymbol{e}_j\right)^2|\mathcal{J}]\\ &\leq \frac{C}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left(\sum_{j=1}^{n}\boldsymbol{e}_j'\boldsymbol{D}^2_{\Phi e_i}\boldsymbol{e}_j\right)^2|\mathcal{J}] \to_{a.s.} 0.\end{aligned}$$

Third, by Result 2

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}|\mathcal{J}] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{P}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{P})|\mathcal{J}]\\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[(P_{ii}\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i)^2|\mathcal{J}] \qquad (B.45)\\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{e}_i'\boldsymbol{\Phi}\boldsymbol{\Phi}'\boldsymbol{e}_i)^2|\mathcal{J}] \to_{a.s.} 0.\end{aligned}$$

Fourth, by Result 2

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\mathrm{tr}^2(\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}\boldsymbol{D}_{Pe_i}\boldsymbol{D}_{e_i'\Phi}|\mathcal{J}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left[\sum_{k=1}^{n}P_{ki}^2(\Phi_{ki})^2\right]^2|\mathcal{J}]$$
$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[\left[\sum_{k=1}^{n}P_{ii}^2(\Phi_{ki})^2\right]^2|\mathcal{J}] \tag{B.46}$$
$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}[[\boldsymbol{e}_i'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{e}_i]^2|\mathcal{J}] \to_{a.s.} 0.$$

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

B.2.5 Converging conditional variance

First note that

$$s_n^2 = \mathrm{E}[\left(\sum_{i=3}^{n}y_{in}\right)^2|\mathcal{J}] = \mathrm{E}[((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n + o_{a.s.}(1))^2|\mathcal{J}] = 1 + o_{a.s.}(1), \tag{B.47}$$

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.10) as $y_{in} = \Xi_n^{-1/2}(y_{in}^{(1)} + y_{in}^{(2)} + y_{in}^{(1)})$ with

$$\begin{aligned} y_{in}^{(1)} &= \frac{-1}{\sqrt{n}}\boldsymbol{c}_{2n}'\bar{\boldsymbol{X}}'(\boldsymbol{I}_n - 2\boldsymbol{D}_P)\dot{\boldsymbol{V}}\boldsymbol{D}_{\varepsilon}\boldsymbol{e}_i r_i, \\ y_{in}^{(2)} &= \frac{-1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{S}_{i-1}\left[(\boldsymbol{\Psi} + \boldsymbol{\Psi}' - 2\boldsymbol{D}_{\Psi}) - 2c_{1n}\gamma_n\dot{\boldsymbol{P}}\right]\boldsymbol{e}_i r_i, \\ y_{in}^{(3)} &= \frac{1}{\sqrt{n}}\boldsymbol{r}'\boldsymbol{A}_{-i}\boldsymbol{r}r_i. \end{aligned} \tag{B.48}$$

Then we need to show that for any $\epsilon > 0$,

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

$$
\begin{aligned}
&\le \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}]\right.\right. \\
&\quad \left.\left. - \sum_{i=3}^{n} \mathrm{E}[(y_{in}^{(1)} + y_{in}^{(2)} + y_{in}^{(3)})^2 | \mathcal{J}]\right)^2 \right| \mathcal{J}\Bigg] \\
&= \frac{C}{\epsilon^2} \mathrm{E}\left[\left(\sum_{i=3}^{n} \mathrm{E}[(y_{in}^{(1)})^2 + (y_{in}^{(2)})^2 + (y_{in}^{(3)})^2 + 2(y_{in}^{(1)})(y_{in}^{(2)}) + 2(y_{in}^{(1)})(y_{in}^{(3)})\right.\right. \\
&\quad + 2(y_{in}^{(2)})(y_{in}^{(3)}) | \boldsymbol{r}_{-i}, \mathcal{J}] - \sum_{i=3}^{n} \mathrm{E}[(y_{in}^{(1)})^2 + (y_{in}^{(2)})^2 + (y_{in}^{(3)})^2 \\
&\quad \left.\left. +2(y_{in}^{(1)})(y_{in}^{(2)}) + 2(y_{in}^{(1)})(y_{in}^{(3)}) + 2(y_{in}^{(2)})(y_{in}^{(3)}) | \mathcal{J}]\right)^2 \right| \mathcal{J}\Bigg] \\
&\le \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 \right| \mathcal{J}\right] \qquad \text{(B.49)} \\
&\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 \right| \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 \right| \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 \right| \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 \right| \mathcal{J}\right] \\
&\quad + \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 \right| \mathcal{J}\right] \Bigg\} \to_{a.s.} 0.
\end{aligned}
$$

Here the second equality follows since each of the cross products has expectation zero, the first inequality is a conditional Markov inequality and the second inequality uses that $\Xi_n^{-1/2}$ is bounded. Each of these terms can be shown to converge to zero almost surely. The general procedure is

1. Take the expectation over $\boldsymbol{U}$ and $\boldsymbol{r}$ in the second term within the square.

2. Complete the square.

3. If needed split the term in a part that depends on $\boldsymbol{U}$ and a part that does not.

4. Take the expectation over $\boldsymbol{r}$.

5. Bound each of the remaining terms.

This procedure yields a large number of terms that need to be bounded, because when we complete the square we often get a double sum over products of four terms, since each term in (B.49) contains two squares and a sum within one of these squares. By taking the expectation over $\boldsymbol{r}$ we obtain a number of different forms of these products. If furthermore a product contains $\boldsymbol{A}_{-i}$, which itself is a sum of four terms, this yields 16 different cross products that need to be bounded. We do this in a separate document.

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

$$\begin{aligned}\sup_n \mathrm{E}([|\,\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y|\mathcal{J})|^{1+\epsilon}] &= \sup_n \mathrm{E}[(\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y|\mathcal{J}))^2] \\ &\leq \sup_n \mathrm{E}[1^2] \leq \infty.\end{aligned} \tag{B.50}$$

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

$$\begin{aligned}&\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y) = \mathrm{E}[\mathrm{P}((\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n < y|\mathcal{J})] \\ &\to \mathrm{E}[\Phi(y)] = \Phi(y).\end{aligned} \tag{B.51}$$

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 $\boldsymbol{Z} \sim N(0,1)$. 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) $\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \to_d \boldsymbol{Z}$. Since a linear combination of the elements of $\boldsymbol{Z}$ is normally distributed, each of the elements is as well. In particular, $\boldsymbol{Z} \sim N(0, \boldsymbol{I}_{p+1})$ and hence $\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{Y}_n \sim N(0, \boldsymbol{I}_{p+1})$, which concludes the proof.

Appendix C Details of the application to Angrist and Krueger (1991)

We consider the following model from Angrist and Krueger (1991)

$$\begin{aligned}\tilde{\boldsymbol{y}} &= \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{\gamma} + \tilde{\boldsymbol{\varepsilon}} \\ \tilde{\boldsymbol{X}} &= \tilde{\boldsymbol{Z}}\boldsymbol{\Pi} + \boldsymbol{W}\boldsymbol{\Gamma} + \tilde{\boldsymbol{V}},\end{aligned} \tag{C.1}$$

Table 2: Description of the data used in estimation.

Name	Description	Included in $\boldsymbol{W}$	Name dataset
Constant	Constant	Yes	
RACE	1 if black	Yes	v19
MARRIED	1 if married	Yes	v10
SMSA	1 if resides in city center	Yes	v20
NEWENG	1 if New England state	Yes	v13
MIDATL	1 if Middle Atlantic state	Yes	v11
ENOCENT	1 if East North Central state	Yes	v4
WNOCENT	1 if West North Central state	Yes	v24
SOATL	1 if South Atlantic state	Yes	v21
ESOCENT	1 if East South Central state	Yes	v6
WSOCENT	1 if West South Central state	Yes	v25
MT	1 if Mountain state	Yes	v12
Ydummies	9 year dummies	Yes	
Sdummies	49 state dummies	If $k = 180$ or 1530	
SOB	State of birth	No	v17
QOB	Quarter of birth	No	v18
YOB	Year of birth	No	v27
LWKLYWGE	Log-weakly wage	No	v9
EDUC	Years of education	No	v4

where the dependent variable $\boldsymbol{y}$ is $n \times 1$, the endogenous regressor $\boldsymbol{X}$ is $n \times p$, the included instruments $\boldsymbol{W}$ are $n \times q$ and the excluded instruments $\boldsymbol{Z}$ are $n \times k$.

Following Mikusheva and Sun (2021) we focus on the specification for column 6 in Tables V and VII of Angrist and Krueger (1991). That is, $\boldsymbol{y}$ is the 1980 log-weekly wage of men born between 1930 and 1939; $\boldsymbol{X}$ is the years of education; $\boldsymbol{W}$ are control variables detailed in Table 2; and $\boldsymbol{Z}$ are 30, 180 or 1530 instruments. For the 30 instrument setting we use 3 quarter of birth dummies and 27 year of birth and quarter of birth interactions. For the 180 we use in addition 150 quarter of birth and state of birth interactions. For the 1530 we use in addition 1350 quarter of birth, year of birth and state of birth interactions.

To cast above model in the framework of (1) we partial out the included instruments. That is, we premultiply the equation in (C.1) with $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'$ to obtain (1) with $\boldsymbol{y} = \boldsymbol{M}\tilde{\boldsymbol{y}}$ and similarly for the other variables.

We obtain confidence intervals by calculating the AR, the score and the joint statistic for a grid of βs and add the point to the confidence interval whenever the statistic is below the corresponding critical value. The grid consists of 100 evenly spaced points between -0.5 and 0.49.

The data comes from the NEW7080.rar file from the Angrist Data Archive. Table 2 details which variables from the data set we use. We consider the cohort of men born between 1930 and 1939 and thus have 329,509 observations.

Due to the large sample size we cannot use some the formulas in Section 5 directly as they involve $n \times n$ matrices that cannot be handled easily by today's computers. We therefore rewrite some of the terms. Define $\tilde{\boldsymbol{P}} = \boldsymbol{D}_\varepsilon \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2 \boldsymbol{Z})^{-1/2}$ such that $\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}' = \boldsymbol{P}$ and similarly for $\tilde{\boldsymbol{V}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{D}_\varepsilon^2 \boldsymbol{Z})^{-1/2}$, then the most troublesome terms in (A.25) can be written as

$$
\begin{aligned}
\boldsymbol{x}'(\boldsymbol{V} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{x} &= \operatorname{tr}(\boldsymbol{D}_x \boldsymbol{V} \boldsymbol{D}_x (\boldsymbol{P} \odot \boldsymbol{P})) \\
&= \operatorname{tr}(\tilde{\boldsymbol{V}}' \boldsymbol{D}_x (\boldsymbol{P} \odot \boldsymbol{P}) \boldsymbol{D}_x \tilde{\boldsymbol{V}}) \\
&= \sum_{i=1}^{k} \boldsymbol{e}_i' \tilde{\boldsymbol{V}}' \boldsymbol{D}_x (\boldsymbol{P} \odot \boldsymbol{P}) \boldsymbol{D}_x \tilde{\boldsymbol{V}} \boldsymbol{e}_i \\
&= \sum_{i=1}^{k} \boldsymbol{\iota}' \boldsymbol{D}_{\tilde{V} e_i} \boldsymbol{D}_x (\boldsymbol{P} \odot \boldsymbol{P}) \boldsymbol{D}_x \boldsymbol{D}_{\tilde{\boldsymbol{V}} e_i} \boldsymbol{\iota} \\
&= \sum_{i=1}^{k} \operatorname{tr}(\boldsymbol{D}_{\tilde{\boldsymbol{V}} e_i} \boldsymbol{D}_x \boldsymbol{P} \boldsymbol{D}_x \boldsymbol{D}_{\tilde{V} e_i} \boldsymbol{P}) \\
&= \sum_{i=1}^{k} \operatorname{tr}(\tilde{\boldsymbol{P}}' \boldsymbol{D}_{\tilde{V} e_i} \boldsymbol{D}_x \boldsymbol{P} \boldsymbol{D}_x \boldsymbol{D}_{\tilde{\boldsymbol{V}} e_i} \tilde{\boldsymbol{P}}),
\end{aligned} \tag{C.2}
$$

and similarly for $\boldsymbol{x}'\boldsymbol{D}_P(\boldsymbol{V} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{x}$ and $\boldsymbol{x}'(\boldsymbol{V} \odot \boldsymbol{P} \odot \boldsymbol{P})\boldsymbol{D}_P\boldsymbol{x}$.

$$
\begin{aligned}
\boldsymbol{x}'(\boldsymbol{V}\boldsymbol{D}_\varepsilon \odot \boldsymbol{V}\boldsymbol{D}_\varepsilon)(\boldsymbol{P} \odot \boldsymbol{P}) \odot \boldsymbol{I})\boldsymbol{x} &= \sum_{i,j=1}^{n} x_i^2 \varepsilon_i^2 V_{ij} \varepsilon_j^4 \\
&= \sum_{i=1}^{n} x_i^2 \varepsilon_i^2 (\boldsymbol{e}_i' \boldsymbol{V} \boldsymbol{D}_\varepsilon)^{\odot 4} \boldsymbol{\iota},
\end{aligned} \tag{C.3}
$$

where $^{\odot}$ is the element wise power. This expression can efficiently be calculated for smaller subsamples of the total data set.