Federal Reserve Bank of Dallas

A Robust Test for Weak Instruments with Multiple Endogenous Regressors

Daniel J. Lewis and Karel Mertens

Working Paper 2208 **June 2022**

Research Department

https://doi.org/10.24149/wp2208

A Robust Test for Weak Instruments with Multiple Endogenous Regressors*

Daniel J. Lewis† and Karel Mertens‡

June 12, 2022

Abstract

We extend the popular bias-based test of Stock and Yogo (2005) for instrument strength in linear instrumental variables regressions with multiple endogenous regressors to be robust to heteroskedasticity and autocorrelation. Equivalently, we extend the robust test of Montiel Olea and Pflueger (2013) for one endogenous regressor to the general case with multiple endogenous regressors. We describe a simple procedure for applied researchers to conduct our generalized first-stage test of instrument strength and provide efficient and easy-to-use Matlab code for its implementation. We demonstrate our testing procedures by considering the estimation of the state-dependent effects of fiscal policy as in Ramey and Zubairy (2018).

JEL Classification: C26, C36.

Keywords: Instrumental Variables, Weak Instruments Test, Multiple Endogenous Regressors, Heteroskedasticity, Serial Correlation.

* The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Dallas or New York or the Federal Reserve System. We are grateful to José Luis Montiel Olea and Carolin Pflueger for discussion and sharing computer code, and to Sarah Zubairy for helpful comments. All errors are ours.

† Daniel J. Lewis, Federal Reserve Bank of New York and University of Pennsylvania.

‡ Contact: Karel Mertens, mertens.karel@gmail.com, tel: +(214) 922-6000.

1 Introduction

Empirical researchers using instrumental variables (IV) estimation frequently report first-stage F-statistics – or in the multiple endogenous regressors case, Cragg and Donald (1993) statistics – to assess instrument relevance. These statistics are typically compared with critical values based on the bias of two-stage least-squares (2SLS) relative to the bias of OLS, which are tabulated by Stock and Yogo (2005). However, these test statistics and critical values require the assumption of homoskedastic and serially uncorrelated errors, and are generally invalid if this assumption does not hold. As recently discussed in a survey by Andrews et al. (2019), in practice this often means that researchers assume homoskedastic and serially uncorrelated errors for the purpose of a first-stage test, but make different assumptions when conducting inference for the second-stage parameters of interest.

In an important paper, Montiel Olea and Pflueger (2013) introduce a new statistic – the "effective F-statistic" – that accounts for heteroskedasticity and autocorrelation in the model errors. Their test is based on a Nagar approximation of the 2SLS bias relative to a benchmark, and the limiting distribution and associated critical values depend on the application-specific structure of the robust covariance matrix of the reduced form parameters. A practical limitation, however, is that the test based on the effective F-statistic only applies to IV models with a single endogenous regressor. Andrews et al. (2019) point to the lack of a heteroskedasticity and autocorrelation robust (HAR) test for models with multiple endogenous regressors as an important remaining gap in the IV pre-testing literature.[1] The contribution of this paper is to fill that gap by generalizing the Montiel Olea and Pflueger (2013) test to allow for an arbitrary number of endogenous regressors, and in so doing extend the Stock and Yogo (2005) bias-based test to be heteroskedasticity-autocorrelation robust.

Our generalized test statistic is an extension of the Cragg-Donald statistic proposed by Stock and Yogo (2005), and nests the effective F-statistic as a special case. As in Montiel Olea and Pflueger (2013), we consider the Nagar approximation of the 2SLS bias relative to a benchmark. As in Stock and Yogo (2005), we use a weighted ℓ_2-norm as our bias criterion. Our test therefore also nests the Stock and Yogo (2005) test under conditional homoskedasticity, with numerical differences in the critical values that are due only to the use of the Nagar approximation instead of numerical integration to compute the bias. We show that the worst-case Nagar bias has a sharp upper bound that is inversely proportional to the minimum eigenvalue of the concentration matrix, but depends otherwise only on a covariance

[1] In practice, researchers sometimes report the Kleibergen and Paap (2006) (KP) HAR version of the Cragg and Donald (1993) test statistic. However, the KP statistic and its associated limiting distribution are for tests of non-identification, not weak identification. As discussed in Sanderson and Windmeijer (2016) and Andrews et al. (2019), the KP statistic should also not be compared to the Stock and Yogo (2005) critical values, as it does not share the limiting distribution of their Cragg-Donald statistic except in the special case of homoskedasticity.

matrix that can be consistently estimated using HAR methods. As in the Montiel Olea and Pflueger (2013) test, obtaining this upper bound requires a numerical optimization step. With many endogenous regressors, this step could be very costly. However, our formulation of the optimization problem allows the use of efficient specialized numerical techniques that lead to trivial computation times in typical applications. For very large-dimensional models, we describe a simplified – but more conservative – procedure that avoids numerical optimization altogether.

Our generalized test statistic has a limiting distribution related to a matrix consisting of traces of partitions of a non-central Wishart random matrix. We provide analytical expressions for the cumulants of this matrix of traces, and show that the cumulants of our test statistic are bounded by expressions that depend only on an estimable covariance matrix and the minimum eigenvalue of the concentration matrix. We then approximate the bounding limiting distribution of our test statistic by matching the first three bounding cumulants as in Imhof (1961). We construct critical values testing the null hypothesis that the minimum eigenvalue does not exceed the largest value compatible with a given bias threshold (the instruments are weak), against the alternative that the bias is below that threshold (the instruments are strong).

To demonstrate the practical usefulness of our test, we consider the Ramey and Zubairy (2018) IV estimates of state-dependent government spending multipliers. Starting from existing specifications for estimating the dynamic effects of government spending shocks, the authors introduce a second endogenous regressor by interacting government spending with an indicator for the state of the business cycle or for the monetary policy regime. This application is a good example of how multiple endogenous regressors quickly arise in practice, as it is common for researchers to explore specifications with interactions between an endogenous regressor and other variables. Ramey and Zubairy (2018) use local projections to estimate the effects over different forecast horizons, which means that autocorrelation in the model errors is almost certainly a concern. Indeed, we find that our robust test regularly leads to different conclusions regarding instrument strength than the Stock and Yogo (2005) test which assumes no autocorrelation, or the Montiel Olea and Pflueger (2013) test applied within each of the regime subsamples. Multiple endogenous regressors of course arise in many other contexts as well, including in time series, cross-sectional, as well as panel data models. Our generalized test should therefore be useful to economic practitioners across a broad range of empirical applications.

Like that of Montiel Olea and Pflueger (2013), and the most widely used test of Stock and Yogo (2005), our testing procedure is based on a bias criterion. Stock and Yogo (2005) additionally consider a criterion based on size distortions of second-stage Wald tests. Sanderson

and Windmeijer (2016) also derive a bias-based test, but study a different class of weakly identified models. In particular, they consider cases in which the matrix of first-stage coefficients has a rank deficiency local to unity, as opposed to a rank local to zero as in the asymptotic framework of Stock and Yogo (2005) and Montiel Olea and Pflueger (2013). To date, the Sanderson and Windmeijer (2016) approach has also only been developed under homoskedasticity. Andrews (2018) takes a distinctly different approach, which is broadly applicable to GMM problems, whereby both robust and non-robust confidence sets are computed for the parameters of interest. The relationship between the sets, linked to the size distortion of standard inference, determines whether identification is strong enough to proceed with non-robust methods. Finally, a recent paper by Carrasco and Doukali (2021) introduces a first-stage test that is robust to heteroskedasticity and many instruments. However, as the Montiel Olea and Pflueger (2013) test, their test only applies to models with a single endogenous regressor.

When conducted in conjunction with hypothesis tests for identified parameters, pretests for weak instruments form part of a multiple hypothesis testing problem, which can contribute to size distortions in inference on estimated second-stage parameters; see Lee et al. (2021). In applying our proposed test, we encourage empirical researchers to be mindful of this issue, and also consider robust inference procedures for second-stage parameters, particularly in cases of marginal rejections. However, given the popularity of first-stage tests in applied work, we view a test that allows researchers to make consistent assumptions in both estimation stages as an important improvement over the common practice of assuming away heteroskedastcity and autocorrelation in the first stage, but not in the second.

2 Model and Summary of Testing Procedure

In this section, we provide a non-technical summary of the model assumptions and testing procedures. We also discuss how our test compares to those of Stock and Yogo (2005) and Montiel Olea and Pflueger (2013).

Model Our proposed test applies to linear instrumental variables models with N endogenous regressors and $K \geq N$ instruments,

$$y = Y\beta + u, \tag{1}$$

$$Y = Z\Pi + v, \tag{2}$$

where $\beta \in \mathbb{R}^N$ contains the main parameters of interest, and $\Pi \in \mathbb{R}^{K\times N}$ contains the first-stage parameters. The econometrician observes $y \in \mathbb{R}^T$, $Y \in \mathbb{R}^{T\times N}$, and $Z \in \mathbb{R}^{T\times K}$, where T is the sample size. Without loss of generality, we assume that there are no additional

exogenous regressors, and that $Z'Z/T = I_K$. In the presence of additional exogenous regressors, it suffices to first project y, Y, and Z on these regressors and replace all variables with the resulting projection errors. In addition, the formulas in this paper assume the user has subsequently normalized Z such that its sample average is zero and its sample covariance is the identity matrix.

Consider the reduced form of (1), $y = Z\Pi\beta + w$, where $w = v\beta + u$, and assume that $T^{-\frac{1}{2}}[Z'w \;\; \text{vec}(Z'v)']' \xrightarrow{d} \mathcal{N}(0, \mathbf{W})$, where $\mathbf{W}$ is the asymptotic covariance of the reduced form and first-stage OLS coefficients, and vec is the vectorization operator. The weak instruments test of Stock and Yogo (2005) requires $\mathbf{W}$ to be of the Kronecker form $\Sigma_{wv} \otimes I_K$, where $[w \; v]'[w \; v]/T \xrightarrow{p} \Sigma_{wv}$. The Kronecker structure arises generally only in conditionally homoskedastic and serially uncorrelated models. The purpose of this paper is to let $\mathbf{W}$ be any positive definite matrix, thereby allowing for arbitrary distributional assumptions about the model errors. Montiel Olea and Pflueger (2013) also relax the Kronecker form assumption, but only consider models with $N = 1$. We defer a detailed discussion of all our assumptions to Section 3, but they are otherwise entirely analogous to Stock and Yogo (2005) and Montiel Olea and Pflueger (2013). They include the assumption that the first-stage relationship in (2) is local to zero, and that the instruments are therefore weak under the null hypothesis.

Generalized Test In practice, our test for weak instruments consists of the following steps:

1. Replace all variables with the residuals after projections on any additional exogenous regressors. Normalize the instruments such that the sample average is zero and $Z'Z/T = I_K$.
2. Obtain the residuals from the reduced-form regressions of y on Z and Y on Z, denoted by $\hat{w}_t$ and $\hat{v}_t$, respectively, where $t = 1, ..., T$ indexes individual observations. Compute a heteroskedasticity-robust, heteroskedasticity-and-autocorrelation robust (HAR), or clustered covariance matrix, $\hat{\mathbf{W}}$, of the vector $(\hat{w}_t, \hat{v}_t')' \otimes Z_t$, or, equivalently, the reduced-form and first-stage parameters.
3. Compute the test statistic, $g_{\min}$,

$$g_{\min} \;\; = \;\; T^{-1} \, \text{mineval}\{\hat{\Phi}^{-\frac{1}{2}} Y'ZZ'Y\hat{\Phi}^{-\frac{1}{2}}\},$$

   where $\hat{\Phi} = (I_N \otimes \text{vec}(I_K))'(\hat{\mathbf{W}}_2 \otimes I_K)(I_N \otimes \text{vec}(I_K))$, and $\hat{\mathbf{W}}_2$ is the lower $NK \times NK$ diagonal block of $\hat{\mathbf{W}}$, corresponding to the estimated robust covariance matrix of the first-stage coefficients, and mineval$\{\cdot\}$ is the smallest eigenvalue.
4. For a user-supplied relative bias threshold, τ, (e.g., 0.10), obtain $\lambda_{\min} = \mathcal{B}(\hat{\mathbf{W}})/\tau$ where $\mathcal{B}(\hat{\mathbf{W}})$ is calculated numerically using the Matlab code accompanying this paper. The

function $\mathcal{B}(\hat{\mathbf{W}})$ is closely related to the largest possible Nagar bias relative to the benchmark, and is defined in Theorem 1 with special cases described in the discussion following the statement of the theorem. The scalar $\lambda_{\min}$ is the threshold value for the minimum eigenvalue of the concentration matrix, which is defined in (10).

5. Calculate the first three cumulants of the bounding limiting distribution of the test statistic $g_{\min}$ using the formulas

$$\begin{aligned}
\kappa_1 &= K(1+\lambda_{\min}),\\
\kappa_2 &= 2\big(\operatorname{maxeval}\{(I_N\otimes \operatorname{vec}(I_K))'((\hat{\Sigma}^2\otimes I_K)(I_N\otimes \operatorname{vec}(I_K))\} + 2\lambda_{\min}K\operatorname{maxeval}\{\hat{\Sigma}\}\big),\\
\kappa_3 &= 8\big(\operatorname{maxeval}\{(I_N\otimes \operatorname{vec}(I_K))'((\hat{\Sigma}^3\otimes I_K)(I_N\otimes \operatorname{vec}(I_K))\} + 3\lambda_{\min}K\operatorname{maxeval}\{\hat{\Sigma}\}^2\big),
\end{aligned}$$

   where $\hat{\Sigma} = K(\hat{\Phi}^{-\frac{1}{2}}\otimes I_K)\hat{\mathbf{W}}_2(\hat{\Phi}^{-\frac{1}{2}}\otimes I_K)$ and $\operatorname{maxeval}\{\cdot\}$ is the largest eigenvalue.

6. For a given significance level, α, (e.g., 0.05) compute the critical values of the limiting distribution using the Imhof (1961) approximation,

$$\Pr(\chi^2_\nu < (x-\kappa_1)4\omega+\nu), \quad \nu = 8\kappa_2\omega^2, \quad \omega = \kappa_2/\kappa_3,$$

   where χ^2_ν is the (central) chi-squared distribution with (fractional) ν degrees of freedom.

7. Compare the value of the test statistic $g_{\min}$ to the critical value from the previous step divided by K.

We provide a Matlab file, gweakivtest.m, to implement all of these steps.

Simplified Test As in Montiel Olea and Pflueger (2013), we also provide a simplified conservative version of the test that avoids the numerical optimization step. This version of the test may be useful for applications where the number of instruments and/or endogenous regressors is very large. The simplified version of the test follows the same steps as the generalized test, but replaces the sharp bound $\mathcal{B}(\hat{\mathbf{W}})$ with the non-sharp bound $\mathcal{B}^s(\hat{\mathbf{W}})$ defined in Theorem 1. The conservative bound $\mathcal{B}^s(\hat{\mathbf{W}})$ only requires a one-time computation of the largest singular value of two matrices, and is therefore computationally much faster. For the single regressor case, Montiel Olea and Pflueger (2013) prove that $\mathcal{B}(\hat{\mathbf{W}}) \leq 1$, and in that case the simplified procedure can use $\lambda_{\min} = \min\{\mathcal{B}^s(\hat{\mathbf{W}}), 1\}/\tau$. In models with multiple endogenous regressors, however, it is not necessarily the case that $\mathcal{B}(\hat{\mathbf{W}}) \leq 1$. The bound $\mathcal{B}^s(\hat{\mathbf{W}})$ can be smaller or larger than one depending on the application.

Comparison with Existing Critical Values Figure 1 illustrates the critical values that arise from our testing procedure as a function of the number of instrumental variables K. The panels in the first column of Figure 1 show critical values for models with $N = 1$, $N = 2$, and $N = 3$ endogenous regressors for the homoskedastic and serially uncorrelated

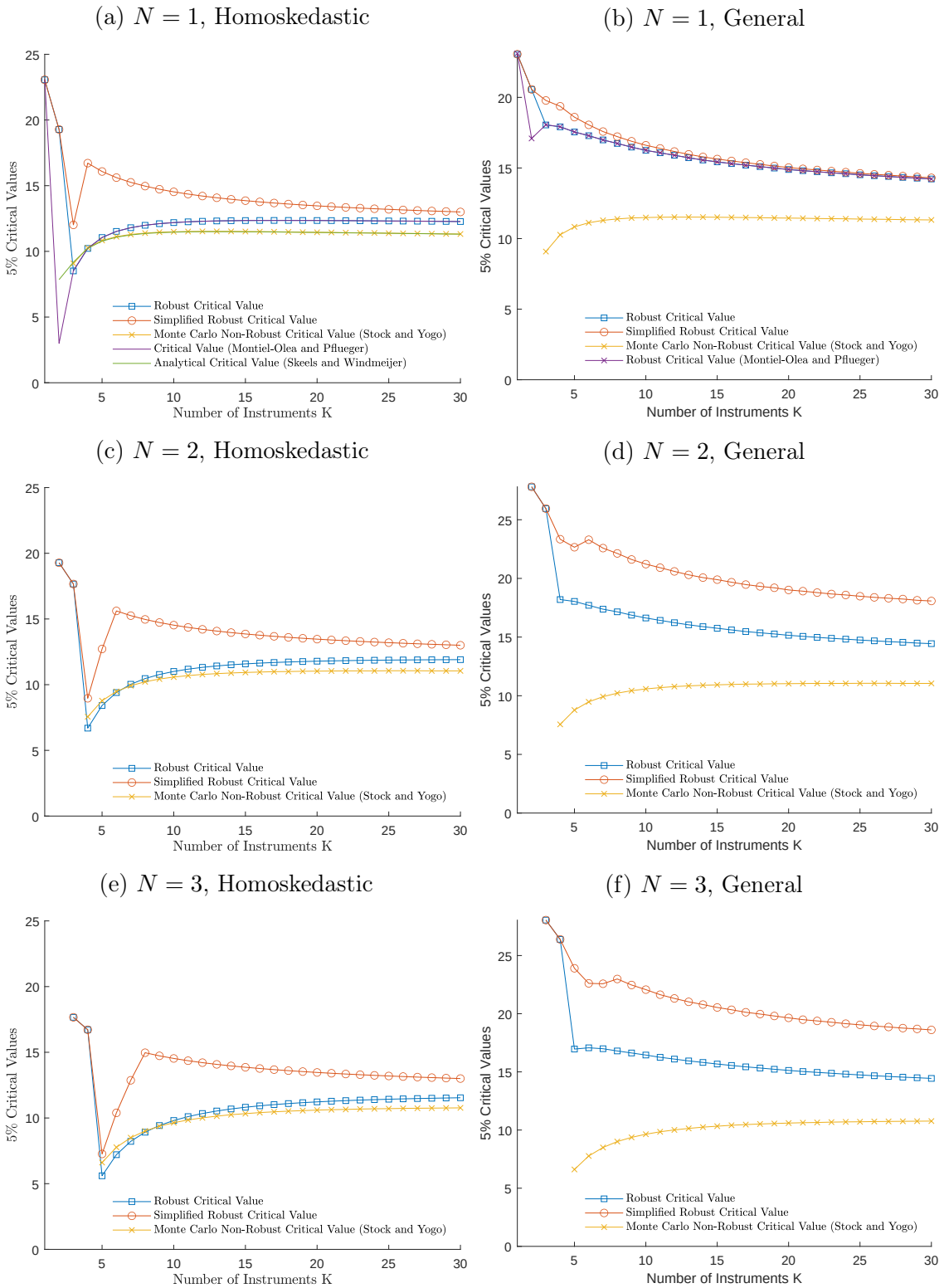
model considered in Stock and Yogo (2005). We show the critical values based on the sharp bound on the Nagar bias, as well as those from the more conservative simplified procedure without the numerical optimization step. Each panel also plots the critical values from the Stock and Yogo (2005) tables for comparison, which are available for $K > N + 1$. For the $N = 1$ case (top left panel), we additionally show the critical values from the Montiel Olea and Pflueger (2013) test for $K \geq 1$, as well as the analytical critical values derived recently for $K > 1$ by Skeels and Windmeijer (2018).

In the homoskedastic and serially uncorrelated model, the critical values only depend on the number of endogenous regressors N and the number of instruments K. In the more general model, the critical values depend additionally on a HAR covariance matrix of the reduced form parameter estimates. The critical values are therefore different for each application. To nevertheless give a sense of the critical values that may arise in practice, the right column in Figure 1 shows an average of the robust critical values for 500 different general covariance matrices drawn randomly from a central Wishart distribution with an identity covariance matrix. Each panel also shows the – in these cases incorrect – Stock and Yogo (2005) non-robust critical values for comparison, and the upper panel ($N = 1$) additionally shows the robust Montiel Olea and Pflueger (2013) critical values.

The main takeaways from Figure 1 are as follows. For models with a single endogenous regressor ($N = 1$), either homoskedastic or heteroskedastic, there are only negligible numerical differences between our critical values and those of Montiel Olea and Pflueger (2013).[2] The only exception is the homoskedastic model with $K = 2$. As we explain below, the reason is that we recommend switching to a more conservative upper bound on the bias when the degree of overidentification is less than two, as the Nagar approximation is potentially poor in those cases. This can be seen in the homoskedastic model for $N = 1$ in the top left panel, as the Montiel Olea and Pflueger (2013) critical value is well below the analytical critical value obtained by Skeels and Windmeijer (2018) when $K = 2$. Whenever $K > N + 1$, the three panels for the homoskedastic model (left column) show that the critical values based on the Nagar bias are relatively close to the Stock and Yogo (2005) values. The robust critical values under a general covariance in the right column of Figure 1, on the other hand, are generally well above the Stock and Yogo (2005) values, which in those cases incorrectly assume homoskedasticity. We emphasize, however, that our robust critical values should be compared against the generalized test statistic in this paper, and not the Cragg-Donald statistic used in the Stock and Yogo (2005) test. Finally, for illustrative purposes Figure 1 also plots the more conservative critical values from our simplified

[2]These differences arise only because we use the Imhof (1961) approximation, matching the first three cumulants of the target distribution, instead of the Patnaik (1949) approximation, matching the first two cumulants. We consistently found the differences in the resulting critical values to be very small.

Figure 1: Comparison of Critical Values



Notes: The left column reports critical values assuming homoskedasticity and serial uncorrelatedness for various numbers of endogenous regressors (N) and instruments (K). The right column repeats the exercise under arbitrary heteroskedasticity and/or autocorrelation. Critical values depend on $\mathbf{W}$ and therefore vary for each application. The figures show averages over 500 draws of $\mathbf{W}$ from a central Wishart distribution with identity covariance matrix for illustrative purposes only. For comparison, we plot applicable critical values from Montiel Olea and Pflueger (2013), Stock and Yogo (2005), and Skeels and Windmeijer (2018).

procedure. We generally do not recommend using these values, as the tighter values will typically be sufficiently easy to compute. The results in Figure 1 suggest, however, that they may be of use in applications where K or N are much larger. One final note concerns the $N = K = 1$ setting, a just-identified model with a single endogenous regressor. In this special case, another alternative is to compare a robust $F-$statistic, or, equivalently, our test statistic, to the size-based critical values of Stock and Yogo (2005) (16.38 for 5% distortion, 8.96 for a 10% distortion), which remain valid in this special case even when homoskedasticity fails to hold, provided a robust test statistic is used, as noted in Andrews et al. (2019). However, our main interest is in models with $N > 1$, for which this option is not available.

3 Testing the Null Hypothesis of Weak Instruments

This section provides the full derivation of our testing procedures. We start by establishing some specific notation: $||U||_2$ is the spectral norm of U (the positive square root of the maximum eigenvalue of UU', also the ℓ_2-norm if U is a vector), $\mathbb{P}^n$ is the set of positive definite $n \times n$ matrices, $\mathbb{O}^{n\times m}$ is the set of $n \times m$ orthogonal real matrices U such that $UU' = I_n$, $\mathcal{K}_{n,m}$ denotes the $n \times m$ commutation matrix such that $\mathcal{K}_{n,m}\,\text{vec}(U) = \text{vec}(U')$ where $U \in \mathbb{R}^{n\times m}$. We also define the special matrix $R_{n,m} = I_n \otimes \text{vec}(I_m)$. The dimension of $R_{n,m}$ is $nm^2 \times n$. For $U \in \mathbb{R}^{nm\times nm}$, the (i,j)-th element of $V = R'_{n,m}(U \otimes I_m)R_{n,m} \in \mathbb{R}^{n\times n}$ is $\text{Tr}(U_{ij})$ where $U_{ij} \in \mathbb{R}^{m\times m}$ is (i,j)-th block of U and $\text{Tr}(\cdot)$ is the trace. For $U \in \mathbb{R}^{nm\times m}$, the i-th element of $V = R'_{n,m}\,\text{vec}(U') \in \mathbb{R}^n$ is equal to $\text{Tr}(U_i)$ where $U_i \in \mathbb{R}^{m\times m}$ is the i-th row block of U. Note that $R'_{n,m}R_{n,m} = mI_N$.

3.1 Weak Instrument Asymptotic Representation of the 2SLS Estimator

The 2SLS estimator for the model in (1)-(2) is

$$\hat{\beta}_{2SLS} = (Y'P_ZY)^{-1}Y'P_Zy, \tag{3}$$

where $P_Z = ZZ'/T$ given the normalization of the instruments. Following Staiger and Stock (1997) and the subsequent literature, we model weak instruments by assuming the first-stage relationship is local-to-zero,

Assumption 1. $\text{vec}(\Pi') = \text{vec}(C')/\sqrt{T} = c/\sqrt{T}$ *where* $C \in \mathbb{R}^{N\times K}$ *is a fixed full rank matrix.*

The next set of assumptions enables us to characterize the weak instrument asymptotic distributions of the 2SLS estimator (and later our test statistic),

Assumption 2. *The following limits hold as $T \to \infty$:*

$$uu'/T \xrightarrow{p} \sigma_u^2 \in \mathbb{R}_+ \ , \ vu'/T \xrightarrow{p} \sigma_{vu} \in \mathbb{R}^N \ , \ vv'/T \xrightarrow{p} \Sigma_v \in \mathbb{P}^N, \tag{2.a}$$

$$T^{-\frac{1}{2}} \begin{bmatrix} Z'w \\ \text{vec}(Z'v) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (\beta' \otimes I_K)(\gamma_2 - c) + \gamma_1 \\ \gamma_2 - c \end{bmatrix} \sim \mathcal{N}(0, \mathbf{W}) \ , \tag{2.b}$$

$$\text{where } \gamma_1 \in \mathbb{R}^K, \ \gamma_2 \in \mathbb{R}^{NK}, \ \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_{12} \\ \mathbf{W}_{12}' & \mathbf{W}_2 \end{bmatrix} \in \mathbb{P}^{(N+1)K}.$$

Defining $\hat{\gamma}_1 = T^{-\frac{1}{2}} Z'y$, $\hat{\gamma}_2 = T^{-\frac{1}{2}} \text{vec}(Z'Y)$, the 2SLS estimator in (3) can be written as

$$\hat{\beta}_{2SLS} = \left(R_{N,K}'(\hat{\gamma}_2\hat{\gamma}_2' \otimes I_K)R_{N,K}\right)^{-1} R_{N,K}' \,\text{vec}(\hat{\gamma}_1\hat{\gamma}_2'). \tag{4}$$

Under Assumptions 1 and 2, $\hat{\gamma}_2 \xrightarrow{d} \gamma_2$ and $\hat{\gamma}_1 \xrightarrow{d} \gamma_1$. By the continuous mapping theorem,

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{d} \beta^*_{2SLS} = \left(R_{N,K}'(\gamma_2\gamma_2' \otimes I_K)R_{N,K}\right)^{-1} R_{N,K}' \,\text{vec}(\gamma_1\gamma_2'). \tag{5}$$

Equation (5) states that $\hat{\beta}^*_{2SLS}$ converges to a quotient of quadratic forms in normal random variables, and $\hat{\beta}_{2SLS}$ is therefore not consistent.

Finally, we assume that an estimator $\hat{\mathbf{W}}$ is available that is consistent for $\mathbf{W}$.

3.2 Weak Instrument Set and Null Hypothesis

As in Montiel Olea and Pflueger (2013) and Stock and Yogo (2005), our null hypothesis of weak instruments is based on a bias criterion. We consider instruments weak when a weighted ℓ_2 norm of the asymptotic bias $E[\beta^*_{2SLS}]$ is large relative to a benchmark. Define $\mathbf{S} \in \mathbb{P}^{(N+1)K}$ partitioned as $\mathbf{W}$ with

$$\begin{aligned} \mathbf{S}_1 &= \mathbf{W}_1 + (\beta' \otimes I_K)\mathbf{W}_2(\beta \otimes I_K) - (\beta' \otimes I_K)\mathbf{W}_{12}' - \mathbf{W}_{12}(\beta \otimes I_K), \\ \mathbf{S}_{12} &= \mathbf{W}_{12} - (\beta' \otimes I_K)\mathbf{W}_2 \\ \mathbf{S}_2 &= \mathbf{W}_2, \end{aligned} \tag{6}$$

such that $\mathbf{S}$ is the covariance of $T^{-\frac{1}{2}}[Z'u \ \ \text{vec}(Z'v)']'$ for $T \to \infty$.

Our bias criterion is

$$B = \text{Tr}(\mathbf{S}_1)^{-\frac{1}{2}} ||E\left[\beta^*_{2SLS}\right]' R_{N,K}'(\mathbf{S}_2^{\frac{1}{2}} \otimes I_K)||_2, \tag{7}$$

where the weighting matrix is $R_{N,K}'(\mathbf{S}_2 \otimes I_K)R_{N,K}$ and the benchmark for the bias is $\text{Tr}(\mathbf{S}_1)^{\frac{1}{2}}$. When there is just a single endogenous regressor, $N = 1$, the weighting matrix $R_{N,K}'(\mathbf{S}_2 \otimes$

$I_K)R_{N,K}$ becomes the scalar $\text{Tr}(\mathbf{S}_2)$. In that case, $B = E\left[\beta^*_{2SLS}\right]\sqrt{\text{Tr}(\mathbf{S}_2)}/\sqrt{\text{Tr}(\mathbf{S}_1)}$, which is identical to the criterion in Montiel Olea and Pflueger (2013) after replacing $E\left[\beta^*_{2SLS}\right]$ with a Nagar (1959) approximation. Our extension of the benchmark to the $N > 1$ case is analogous to Montiel Olea and Pflueger (2013) and retains the interpretation as the worst-case 2SLS bias as $K \to \infty$, see Appendix A.

In Appendix B, we show that

$$
\begin{aligned}
(8)\ B &= ||\mathbf{h}\rho||_2, \text{ where} \\
\mathbf{h} &= KE\left[\left(R'_{N,K}(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}'\otimes I_K)R_{N,K}\right)^{-1}R'_{N,K}\left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1}\otimes I_K\right)\right], \\
\rho &= \left((R'_{N,K}(\mathbf{S}_2\otimes I_K)R_{N,K}\right)^{-\frac{1}{2}}\otimes I_{K^2}\big)\,\text{vec}\left(\mathbf{S}_{12}\right)/\sqrt{\text{Tr}(\mathbf{S}_1)}\,,
\end{aligned}
$$

and $l = \mathbf{S}_2^{-\frac{1}{2}}c$, $\psi = \mathbf{S}_2^{-\frac{1}{2}}(\gamma_2 - c) \sim \mathcal{N}(0, I_{NK})$, and $\mathcal{S} = ((R'_{N,K}(\mathbf{S}_2\otimes I_K)R_{N,K}/K)^{-\frac{1}{2}}\otimes I_K)\mathbf{S}_2^{\frac{1}{2}}$.[3] In the conditionally homoskedastic and serially uncorrelated model, $\mathcal{S} = I_{NK}$, $\text{Tr}(\mathbf{S}_1)/K = \sigma_u^2$, $R'_{N,K}(\mathbf{S}_2\otimes I_K)R_{N,K}/K = \Sigma_v$, $R'_{N,K}\,\text{vec}(\mathbf{S}_{12})/K = \sigma_{vu}$. In that case B is identical to the absolute bias criterion in Stock and Yogo (2005).

The weak instrument set is defined as $\mathbb{B}_\tau(\mathbf{W}) = \{l \in \mathbb{R}^{NK}, \beta \in \mathbb{R}^N : B > \tau\}$, i.e. as the set associated with bias relative to the benchmark that exceeds a tolerance level τ. This set depends on $\mathbf{W}$, which can be consistently estimated, but also on the NK unknown parameters in l, and on the N unknown parameters in β.

In general, there is no tractable analytical expression for the integral underlying the expectation in $\mathbf{h}$. Whereas Stock and Yogo (2005) evaluate this integral using Monte Carlo methods, we follow Montiel Olea and Pflueger (2013) and adopt a Nagar approximation to $\mathbf{h}$, which we denote by $\mathbf{h}_n$. For reasons that will be clear momentarily, we also reparametrize the functional dependence of the bias on l through

$$
(9) \qquad l = \mathcal{S}^{-1}\sqrt{K}\,\text{vec}(L_0\mathcal{D}_\Lambda^{\frac{1}{2}}Q'_\Lambda),
$$

where $Q_\Lambda \in \mathbb{O}^{N\times N}, \mathcal{D}_\Lambda \in \mathbb{R}^{N\times N}$ contain the eigenstructure of the concentration matrix

$$
(10) \qquad \Lambda = K^{-1}\left(R'_{N,K}(\mathcal{S}ll'\mathcal{S}'\otimes I_K)R_{N,K}\right) \in \mathbb{P}^N,
$$

and $L_0 \in \mathbb{O}^{N\times K}$ is an orthogonal matrix. By definition $\Lambda = Q_\Lambda\mathcal{D}_\Lambda Q'_\Lambda$, where $\mathcal{D}_\Lambda$ is a diagonal matrix containing the eigenvalues $\lambda_i > 0$, $i = 1, ..., N$, and $Q_\Lambda Q'_\Lambda = I_N$. The reparametrization in (9) reformulates the choice of the NK parameters of l as an equivalent choice of the N free parameters in $\mathcal{D}_\Lambda$, the $N^2 - (N+1)N/2$ free parameters of Q_Λ, and

[3] Note that $\mathcal{S}$ is in general non-symmetric, and that $\text{Tr}(\mathcal{S}) = \text{Tr}(\mathcal{S}\mathcal{S}') = NK$.

the $NK - (N+1)N/2$ free parameters of L_0.

Let $B_n(\beta, Q_\Lambda, \mathcal{D}_\Lambda, L_0, \mathbf{W})$ denote the bias after the reparametrization and the Nagar approximation. Let $B_n^*(\mathbf{W}, \lambda_{\min}) = \sup_{\beta, Q_\Lambda, \mathcal{D}_\Lambda, L_0} \{B_n(\beta, Q_\Lambda, \mathcal{D}_\Lambda, L_0, \mathbf{W})\}$ denote the sharp upper bound on the Nagar bias over $\beta \in \mathbb{R}^N$, $Q_\Lambda \in \mathbb{O}^{N\times N}$, $L_0 \in \mathbb{O}^{N\times K}$ and $\mathcal{D}_\Lambda$ in the set of all diagonal matrices with no diagonal element smaller than $\lambda_{\min}$, the smallest eigenvalue of the concentration matrix. The following theorem provides the generalization of the Nagar approximation theorem in Montiel Olea and Pflueger (2013) to the general case with $N \geq 1$ endogenous variables:

Theorem 1 (**Nagar approximation**).

i) The Nagar approximation of the bias in (1) is given by $B_n(\beta, Q_\Lambda, \mathcal{D}_\Lambda, L_0, \mathbf{W}) = ||\mathbf{h}_n \rho||_2$ *where* $\mathbf{h}_n = Q_\Lambda \mathcal{D}_\Lambda^{-\frac{1}{2}} M_1 (\mathcal{D}_\Lambda^{-\frac{1}{2}} Q_\Lambda' \otimes L_0 \otimes L_0) M_2$, *with* $M_1 = R_{N,N}' \left(I_{N^3} + (\mathcal{K}_{N,N} \otimes I_N)\right)$ *and* $M_2 = R_{N,K} R_{N,K}'/(1+N) - I_{NK^2}$.

ii) The Nagar bias has the following bounds:

a) $B_n^*(\mathbf{W}, \lambda_{\min}) = \lambda_{\min}^{-1} \mathcal{B}(\mathbf{W})$, $\mathcal{B}(\mathbf{W}) = K^{-\frac{1}{2}} \sup_{L_0 \in \mathbb{O}^{N\times K}} \{||M_1 (I_N \otimes L_0 \otimes L_0) M_2 \Psi||_2\}$,

b) $B_n^*(\mathbf{W}, \lambda_{\min}) \leq \lambda_{\min}^{-1} \mathcal{B}^s(\mathbf{W})$, $\mathcal{B}^s(\mathbf{W}) = \min\{(2(N+1)/K)^{\frac{1}{2}} ||M_2 \Psi||_2, ||\Psi||_2\}$,

where $\Psi = \left(\mathcal{S}\mathbf{W}_2^{-\frac{1}{2}} [\mathbf{W}_{12} : \mathbf{W}_2]' \otimes I_K\right) R_{N+1,K} (R_{N+1,K}' (\mathbf{W} \otimes I_K) R_{N+1,K})^{-\frac{1}{2}}$.

Proof. See Appendix. □

Part (i) of the theorem provides the generalized analytical expression for the Nagar bias. Part (ii.a) characterizes the sharp upper bound for the Nagar bias, which only depends on the minimum eigenvalue of the concentration matrix and the largest possible singular value of a matrix that only depends on the $NK - (N+1)/2$ nuisance parameters in L_0. In general, numerical optimization of B_n over β, Q_Λ and L_0 is problematic because of the large dimension and the presence of many local maxima. The proof in the Appendix shows that (1) optimizing over β amounts to a straightforward maximum eigenvalue problem, (2) the bias is decreasing in all eigenvalues of the concentration matrix such that the worst case bias occurs when all eigenvalues of Λ equal $\lambda_{\min}$, and (3) in that case the bias no longer depends on Q_Λ. The remaining problem over L_0 in (*ii.a*) has smaller dimension and, importantly, can exploit numerical techniques specialized for optimization over orthogonal matrices. In the code accompanying this paper, we use the curvilinear search algorithm of Wen and Yin (2013) which leads to trivial computation times even for relatively large N and K.[4] Finally, part (*ii.b*) of the theorem provides an alternative – but generally non-sharp – upper bound

[4]The algorithm does not always find the global optimum for any starting value for L_0. Unless the user specifies otherwise, our code takes the maximum over the optima found for 1000 starting values generated by N columns of $K \times K$ matrices drawn from the Haar distribution, i.e uniformly sampled from the space of all orthogonal matrices.

that requires no numerical optimization for cases where N and K are very large.

The practical implementation of our generalized and simplified test is in general based on the bounds in part (ii) of Theorem 1. However, there are two special cases where we choose to deviate from the Nagar bounds stated in the theorem when applying our test in practice. First, when the degree of overidentification is zero or one ($K \leq N+1$), we set $\mathcal{B}(\mathbf{W}) = ||\Psi||_2$, which is more conservative. This choice is motivated by the sharp bound on the Nagar bias in the conditionally homoskedastic and non-serially correlated model, which simplifies in that model to $B_n^*(\mathbf{W}, \lambda_{\min}) = \lambda_{\min}^{-1}|K-(N+1)|/K$. For $K = N+1$, this means that the Nagar approximation implausibly suggests that the 2SLS bias is always zero in the homoskedastic model.[5] The reason is that the threshold value for $\lambda_{\min}$ at the conventional $\tau = 0.10$ level is in a range that is too small for the Nagar bias to continue to provide a reasonable approximation when $K = N+1$. This can be seen in Figure 1a (the $N = 1$ case) from the sharply lower Montiel Olea and Pflueger (2013) critical value at $K = 2$ compared to the analytical value of Skeels and Windmeijer (2018).[6] In the just-identified case, $K = N$, the integral underlying the expectation in $\mathbf{h}$ in (1) does not converge. While the Nagar approximation is not necessarily poor for $K \leq N+1$ when $\mathbf{W}$ does not have the Kronecker form, out of an abundance of caution we prefer to set $\mathcal{B}(\mathcal{W}) = ||\Psi||_2$. For $N = K = 1$, this more conservative bound always coincides with the sharp bound, such that the only case in which our test will in practice provide critical values that are meaningfully different from Montiel Olea and Pflueger (2013) is when $K = 2$.

The other deviation from the bounds in Theorem 1 is more straightforward, and only applies to the simplified (and weakly more conservative) test: when $N = 1$, Montiel Olea and Pflueger (2013) prove that $\mathcal{B}(\mathbf{W}) \leq 1$, and we therefore incorporate this additional bound in our simplified test (although it is not necessarily always tighter). The same bound does not generally hold when $N > 1$.

We conclude this section by defining the null and alternative hypotheses as

$$\begin{aligned} &H_0 : \lambda_{\min} \in \mathcal{H}(\mathbf{W}) \quad \text{vs.} \quad H_1 : \lambda_{\min} \notin \mathcal{H}(\mathbf{W}), \\ &\text{where} \quad \mathcal{H}(\mathbf{W}) = \{\lambda_{\min} \in \mathbb{R}_+ : B_n^*(\lambda_{\min}, \mathbf{W}) > \tau\}, \end{aligned} \tag{11}$$

which are identical to those underlying the tests of Stock and Yogo (2005) and Montiel Olea and Pflueger (2013). Under the null hypothesis, the worst-case bias $B_n^*(\lambda_{\min}, \mathbf{W})$ exceeds a

[5] The matrix M_2 in the Nagar approximation loses rank when $K = N+1$, which translates to a Nagar bias of zero when $\mathbf{W}$ has the Kronecker form.

[6] See Skeels and Windmeijer (2018) for a detailed discussion of the homoskedastic model with $N = 1$ and $K \leq 2$, and Kinal (1980), Phillips (1980), and Basmann (1961) for earlier results. The Stock and Yogo (2005) tables do not report critical values for $K \leq N+1$.

tolerance level τ for at least some values of β. Under the alternative, the worst-case bias is at most τ for any value of β.

3.3 Test Statistic and Critical Values

Given a bias tolerance level τ, Theorem 1 implies that a test of the null hypothesis of weak instruments can be based on a test of whether the minimum eigenvalue of Λ is smaller or equal to the threshold value $\mathcal{B}(\mathbf{W})/\tau$ (or $\mathcal{B}^s(\mathbf{W})/\tau$ in the case of the simplified test).

Our test statistic extends the Cragg and Donald (1993) statistic in the same way that the effective F-statistic in Montiel Olea and Pflueger (2013) extends the regular F-statistic. Specifically, under weak instrument asymptotics

$$\Gamma = \Phi^{-\frac{1}{2}}(Y'P_ZY)\Phi^{-\frac{1}{2}} \xrightarrow{d} R'_{N,K}(W \otimes I_K)R_{N,K}/K, \tag{12}$$

$$g_{\min} = \text{mineval}\{\Gamma\} \xrightarrow{d} \text{mineval}\{R'_{N,K}(W \otimes I_K)R_{N,K}/K\}, \tag{13}$$

where $\Phi = R'_{N,K}(\mathbf{W}_2 \otimes I_K)R_{N,K}$ and the random matrix $W = \mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}'$ has a noncentral Wishart distribution, $W \sim \mathcal{W}(1, \Sigma, \Omega)$, with 1 degree of freedom, covariance matrix $\Sigma = \mathcal{S}\mathcal{S}' \in \mathbb{P}^{NK}$, and a matrix of noncentrality parameters $\Omega = \Sigma^{-1}\mathcal{S}ll'\mathcal{S}'$, where we have used the notation in Muirhead (1982). While W has a noncentral Wishart distribution, critical values for the test statistic $g_{\min}$ require the distribution of mineval$\{R'_{N,K}(W \otimes I_K)R_{N,K}\}$, or the smallest eigenvalue of the $N \times N$ matrix consisting of the traces of the $K \times K$ partitions of W. The distribution of this function of W is unknown, and depends in general on all parameters in Σ and Ω, not just on the threshold for $\lambda_{\min}$.

In practice, we obtain critical values from a bounding distribution of $g_{\min}$. Specifically, we first derive upper bounds for the cumulants of $g_{\min}$ that only depend on $\lambda_{\min}$ and $\mathbf{W}_2$, and then we show that a distribution proposed by Imhof (1961) that matches the upper bounds for the first three cumulants is a conservative limiting distribution for $g_{\min}$ in the right tail.

To understand our approach, consider first the case with $N = 1$, such that $R'_{N,K}(W \otimes I_K)R_{N,K} = \text{Tr}(W)$ is a scalar. The trace of a noncentral Wishart W is a linear combination of noncentral χ^2 variables. While there is no tractable formula for its probability distribution that we are aware of, Mathai (1980) provides an analytical expression for the n-th order cumulant of $\text{Tr}(W)$,

$$\kappa_n = 2^{n-1}(n-1)!\big(\text{Tr}(\Sigma^n) + n\,\text{Tr}(\Sigma^n\Omega)\big)\ . \tag{14}$$

The mean is $\kappa_1 = K(1+\lambda_{\min})$, since $\text{Tr}(\Sigma) = K$ and $\text{Tr}(\Sigma\Omega) = K\,\text{Tr}(\Lambda) = K\Lambda = K\lambda_{\min}$

when $N = 1$. The cumulants for $n > 1$ are bounded by

$$\kappa_n \leq 2^{n-1}(n-1)!\big(\operatorname{Tr}(\Sigma^n) + nK\lambda_{\min}\operatorname{maxeval}\{\Sigma\}^{n-1}\big)\ , \tag{15}$$

where we have used that for a positive semi-definite matrix V, $|\operatorname{Tr}(UV)| \leq \operatorname{maxeval}\{U\}\operatorname{Tr}(V)$, see Fact 8.12.29 in Bernstein (2009), and that $\operatorname{Tr}(\Sigma^n\Omega) \geq 0$.

Next, consider the general case with $N \geq 1$ such that $g_{\min}$ is distributed as the minimum eigenvalue of a matrix with elements that are traces of the $K \times K$ subpartitions of a noncentral Wishart matrix $W \sim \mathcal{W}(1, \Sigma, \Omega)$. Analogously to Stock and Yogo (2005), we use the distribution of $\gamma' R'_{N,K}(W \otimes I_K)R_{N,K}\gamma \geq \operatorname{mineval}\{R'_{N,K}(W \otimes I_K)R_{N,K}\}$ as a bounding distribution, where γ is the eigenvector associated with the minimum eigenvalue of $\Lambda = R'_{N,K}(\Sigma\Omega\otimes I_K)R_{N,K}/K$ and $\gamma'\gamma = 1$. In Appendix D, we extend the results in Mathai (1980) to obtain an analytical expression for the n-th order cumulant of the distribution of $\gamma' R'_{N,K}(W \otimes I_K)R_{N,K}\gamma$,

$$\kappa_n = 2^{n-1}(n-1)!\Big(\operatorname{Tr}\left(((\gamma\gamma' \otimes I_K)\Sigma)^n\right) + n\operatorname{Tr}\left(((\gamma\gamma' \otimes I_K)\Sigma)^n\Omega\right)\Big). \tag{16}$$

For the mean, we have

$$\kappa_1 = \operatorname{Tr}\left((\gamma\gamma' \otimes I_K)\Sigma\right) + \operatorname{Tr}\left((\gamma\gamma' \otimes I_K)\Sigma\Omega\right) = K(1 + \lambda_{\min}), \tag{17}$$

since $\operatorname{Tr}\left((\gamma\gamma'\otimes I_K)\Sigma\right) = \operatorname{Tr}(\Sigma)/N = K$ and $\operatorname{Tr}\left((\gamma\gamma'\otimes I_K)\Sigma\Omega\right) = \gamma' R'(\Sigma\Omega\otimes I_K)R\gamma = K\lambda_{\min}$. We show in Appendix E that the higher-order cumulants have the following upper bounds,

$$\kappa_n \leq 2^{n-1}(n-1)!\Big(\operatorname{maxeval}\{R'_{N,K}(\Sigma^n \otimes I_K)R_{N,K}\} + nK\lambda_{\min}\operatorname{maxeval}\{\Sigma\}^{n-1}\Big). \tag{18}$$

The bounds in (18) nest those of the model with a single endogenous variable in (15). Importantly, the mean and the upper bounds for the higher cumulants depend only $\lambda_{\min}$ and $\mathbf{W}_2$ through Σ.

Note that in the conditionally homoskedastic and serially uncorrelated model, $\Sigma = \mathcal{S} = I_{NK}$, and the bounds in (18) simplify to

$$2^{n-1}(n-1)!\Big(K + nK\lambda_{\min}\Big). \tag{19}$$

These are the cumulants of a non-central chi-squared distribution with K degrees of freedom and non-centrality $K\lambda_{\min}$, which is the exact bounding distribution in the conditionally homoskedastic and serially uncorrelated model, see Stock and Yogo (2005). We therefore know that in that model, the inequalities in (18) hold with equality.

Using the mean and the upper bounds for the second and third cumulant, we construct the Imhof (1961) distribution,

$$\Pr(\chi^2_\nu < (x - \kappa_1)4\omega + \nu), \quad \nu = 8\kappa_2\omega^2, \quad \omega = \kappa_2/\kappa_3, \tag{20}$$

where χ^2_ν has a central chi-squared distribution with ν degrees of freedom. In Appendix F we show that this approximation is conservative in the right tail when using upper bounds for the second and third cumulants as opposed to the exact cumulants.

The end result is a bounding limiting distribution for $g_{\min}$ that is free of nuisance parameters, and leads to a straightforward calculation of the critical values after replacing $\mathbf{W}_2$ with a consistent estimate. As in Stock and Yogo (2005), the use of a bounding limiting distribution means the critical values are conservative in the sense that the null hypothesis is incorrectly rejected with probability less than or equal to α. Figure 1 showed that our critical values are essentially the same as those of Montiel Olea and Pflueger (2013) when $N = 1$, except when $K = 2$ where we opt to be more conservative as explained above. Our critical values are also close to those in the Stock and Yogo (2005) tables – only available for $K > N + 1$ – in the conditionally homoskedastic and serially uncorrelated model. The numerical differences with the Stock and Yogo (2005) values are almost entirely due to the Nagar approximation, since the differences between the Imhof approximation with the exact cumulants and the noncentral chi-squared bounding distribution are very small.

4 Empirical Application

We illustrate our testing procedure in an application by Ramey and Zubairy (2018), who use instrumental variables in local projections to estimate government spending multipliers, using a measure of news about military spending and estimated government spending shocks from Blanchard and Perotti (2002) as instruments. Their key innovation is to allow the effects of government spending to depend on the state of the business cycle, or alternatively on the monetary policy regime. To this end, the authors interact government spending with an indicator for whether the economy is in a period of slack, or an indicator for whether the policy rate is constrained at the zero lower bound (ZLB). Using their notation, Ramey and Zubairy (2018) estimate cumulative multipliers for $h = 0, 1, \ldots$ based on

$$\begin{aligned}\sum_{j=0}^{h} y_{t+j} =& I_{t-1}\left[\gamma_{A,h} + \phi_{A,h}(L)z_{t-1} + m_{A,h}\sum_{j=0}^{h} g_{t+j}\right] \\ &+ (1 - I_{t-1})\left[\gamma_{B,h} + \phi_{B,h}(L)z_{t-1} + m_{B,h}\sum_{j=0}^{h} g_{t+j}\right] + \omega_{t+h},\end{aligned} \tag{21}$$

where h is the horizon in quarters, y_t is GDP divided by trend GDP, I_{t-1} is the regime indicator, z_{t-1} is a vector of controls, $\phi_{A,h}(L)$, $\phi_{A,h}(L)$ are polynomials in the lag operator, g_t is government spending divided by trend GDP, and $m_{A,h}, m_{B,h}$ are the cumulative spending multipliers over h quarters in the respective states. The use of interaction terms involving endogenous regressors is quite common, and is one example of how multiple endogenous regressors often arise in practical applications.

To assess instrument relevance, Ramey and Zubairy (2018) apply the Montiel Olea and Pflueger (2013) test to the individual subsamples implied by the regime indicators, as for each such subsample there is only a single endogenous regressor. However, to assess whether the multiplier estimates are statistically different across regimes, government spending ultimately has to be interacted with the indicators in a single specification as in (21), in which there are not one but two endogenous regressors (N=2). As autocorrelated errors are an inherent feature of local projections such as (21), Ramey and Zubairy (2018) can unfortunately not rely on the Stock and Yogo (2005) test to assess instrument strength for their regression of interest. Our robust test, in contrast, allows for a direct test of instrument relevance for the specifications in (21). We therefore implement our test as described in Section 2 for $\tau = 0.10$ and $\alpha = 0.05$, and using the same Newey and West (1994) automatic bandwidth HAR estimation procedure as Ramey and Zubairy (2018). The instruments are the military news measure and Blanchard and Perotti (2002) shocks interacted with the regime indicator as instruments, such that $K = 4$.

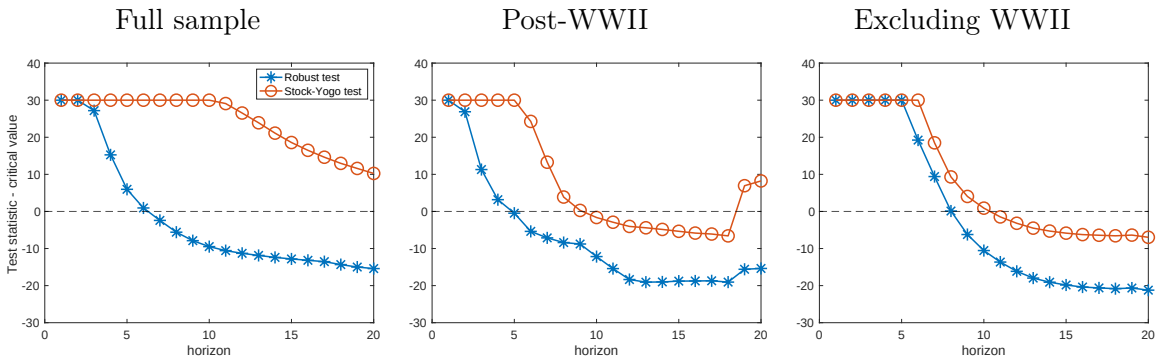
Figure 2 reports the results for the main specifications and sample periods discussed in Ramey and Zubairy (2018). The first row is for specifications that interact government spending with a measure of slack based on the unemployment rate.[7] The starred blue line plots the difference between our robust test statistic and the associated critical value, capped at 30 for visibility across horizons. Although not reported by Ramey and Zubairy (2018), for illustrative purposes the circled red line plots the difference between the Cragg and Donald (1993) test statistic and the Stock and Yogo (2005) critical value. The second row in Figure 2 shows the corresponding results for specifications where government spending is interacted by an indicator of whether monetary policy is constrained by the ZLB.

Similar to the regime-specific results reported by Ramey and Zubairy (2018), our robust test rejects that the instruments are weak at relatively short horizons across all samples. According to our test, issues with instrument relevance start to arise in the specifications with the slack indicator after horizons between 5 to 8 quarters depending on the sample pe-

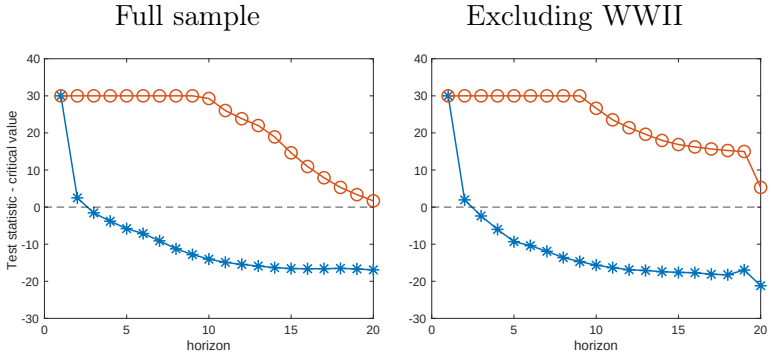
[7]The results in the first row of Figure 2 can be compared to the Montiel Olea and Pflueger (2013) test results with the combined instruments in the second and third column of Figure 4 in Ramey and Zubairy (2018), which report the individual test results for the regressions in the regime-determined subsamples. The results in the second row can similarly be compared to Figure 10 in Ramey and Zubairy (2018).

Figure 2: Test Results for Ramey and Zubairy (2018) Specifications Across Horizons

a. *Government Spending Interacted with Indicator of Slack*



b. *Government Spending Interacted with ZLB Indicator*



Notes: The top row reports the difference between test statistics and critical values for $\tau = 0.10$ and $\alpha = 0.05$ across horizons for specifications with government spending interacted with an indicator for whether the economy was in a state of slack for different sample periods: 1890-2015, 1947-2015 (post-WWII), and 1890-2015 excluding WWII. The second row reports the same difference for specifications with government spending interacted with an indicator for whether monetary policy is constrained by the zero lower bound for different sample periods (1890-2015 and 1890-2015 excluding WWII.). The starred blue line shows the difference between our robust test statistic and the associated critical values across horizons. The circled red line shows the difference between the Cragg and Donald (1993) statistic and critical values from Stock and Yogo (2005). As in Ramey and Zubairy (2018), we cap the results at 30 for visibility.

riod. In the specifications with the ZLB indicator, instrument relevance becomes a concern for horizons as short as 2 quarters. Comparison with the results from the Stock and Yogo (2005) test demonstrates the importance of allowing for heteroskedasticity and autocorrelation in the first-stage testing procedures. For three of the five specifications considered in Figure 2, the Stock and Yogo (2005) test leads to a rejection of weak instruments at all horizons considered. For the remaining two specifications, the Stock and Yogo (2005) test rejects weak instruments for an additional 2-4 quarters compared to the robust test. Our robust test results also vary meaningfully from the regime-specific robust tests reported by Ramey and Zubairy (2018) from subsample regressions with $N = 1$ and $K = 2$, which either reject weak instruments across all forecast horizons considered, or else fail to reject only for substantially longer forecast windows. The main implication is that researchers interested in estimating longer horizon state-dependent multipliers as in Ramey and Zubairy (2018) should consider robust inference procedures.

5 Concluding Remarks

First-stage tests as those proposed in Stock and Yogo (2005) or, more recently, Montiel Olea and Pflueger (2013), are a widely-used diagnostic tool to assess instrument relevance in empirical applications that involve instrumental variables. When researchers are not comfortable imposing homoskedasticity assumptions for second stage inference, they should also avoid imposing such assumptions in first-stage testing procedures. In this paper, we generalize the testing approach of Montiel Olea and Pflueger (2013) to provide a first-stage test that is valid under heteroskedasticity and autocorrelation regardless of the number of endogenous regressors. The computer code accompanying this paper provides empirical researchers with an easy-to-use bias-based first-stage test under assumptions that match those imposed for second-stage inference. Our generalization of the Nagar aproximation to the 2SLS bias should also permit extensions to the methods in Ganics et al. (2021) to construct confidence intervals for the 2SLS bias, which we leave for future work.

References

Alzer, Horst (1997). "On Some Inequalities for the Gamma and Psi Functions". In: *Mathematics of Computation* 66.217, pp. 373–389. DOI: 10.1090/S0025-5718-97-00807-7.

Andrews, Isaiah (2018). "Valid Two-Step Identification-Robust Confidence Sets for GMM". In: *The Review of Economics and Statistics* 100.2, pp. 337–348. DOI: 10.1162/REST_a_00682.

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019). "Weak Instruments in Instrumental Variables Regression: Theory and Practice". In: *Annual Review of Economics* 11.1, pp. 727–753. DOI: 10.1146/annurev-economics-080218-025643.

Basmann, R. L. (1961). "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Over-Identified Cases". In: *Journal of the American Statistical Association* 56.295, pp. 619–636. DOI: 10.1080/01621459.1961.10480649.

Bernstein, Dennis S. (2009). *Matrix Mathematics, Theory, Facts and Formulas*. Princeton: Princeton University Press. DOI: 10.1515/9781400833344.

Blanchard, Olivier and Roberto Perotti (2002). "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output*". In: *The Quarterly Journal of Economics* 117.4, pp. 1329–1368. DOI: 10.1162/003355302320935043.

Carrasco, Marine and Mohamed Doukali (2021). "Testing overidentifying restrictions with many instruments and heteroscedasticity using regularised jackknife IV". In: *The Econometrics Journal* 25.1, pp. 71–97. DOI: 10.1093/ectj/utab020.

Cragg, John G. and Stephen G. Donald (1993). "Testing Identifiability and Specification in Instrumental Variable Models". In: *Econometric Theory* 9.2, pp. 222–240. DOI: 10.1017/S0266466600007519.

Ganics, Gergely, Atsushi Inoue, and Barbara Rossi (2021). "Confidence Intervals for Bias and Size Distortion in IV and Local Projections-IV Models". In: *Journal of Business & Economic Statistics* 39.1, pp. 307–324. DOI: 10.1080/07350015.2019.1660175.

Horn, Roger A. and Charles R. Johnson (2013). *Matrix Analysis*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139020411.

Imhof, J. P. (Dec. 1961). "Computing the distribution of quadratic forms in normal variables". In: *Biometrika* 48.3-4, pp. 419–426. DOI: 10.1093/biomet/48.3-4.419.

Kinal, Terrence W. (1980). "The Existence of Moments of k-Class Estimators". In: *Econometrica* 48.1, pp. 241–249. DOI: 10.2307/1912027.

Kleibergen, Frank and Richard Paap (2006). "Generalized reduced rank tests using the singular value decomposition". In: *Journal of Econometrics* 133.1, pp. 97–126. DOI: 10.1016/j.jeconom.2005.02.011.

Kollo, Tõnu and Dietrich von Rosen (Dec. 1995). "Approximating by the Wishart distribution". In: *Annals of the Institute of Statistical Mathematics* 47, pp. 767–783. DOI: 10.1007/BF01856546.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack R. Porter (2021). *Valid t-ratio Inference for IV*. Working Paper 29124. NBER. DOI: 10.3386/w29124.

Magnus, Jan R. and Heinz Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Oxford: John Wiley and Sons Ltd. DOI: 10.1002/9781119541219.

Mathai, Arak (Jan. 1980). "Moments of the trace of a noncentral Wishart matrix". In: *Communications in Statistics - Theory and Methods* 9, pp. 795–801. DOI: 10.1080/03610928008827921.

Montiel Olea, José Luis and Carolin Pflueger (2013). "A Robust Test for Weak Instruments". In: *Journal of Business & Economic Statistics* 31.3, pp. 358–369. DOI: 10.1080/00401706.2013.806694.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory*. Hoboken, New Jersey: John Wiley and Sons Ltd. DOI: 10.1002/9780470316559.

Nagar, A. L. (1959). "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations". In: *Econometrica* 27.4, pp. 575–595. DOI: 10.2307/1909352.

Newey, Whitney K. and Kenneth D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *The Review of Economic Studies* 61.4, pp. 631–653. DOI: 10.2307/2297912.

Patnaik, P.B. (1949). "The Non-Central Chi-Square and F-Distributions and Their Aplications". In: *Biometrika* 36, pp. 202–232.

Phillips, P. C. B. (1980). "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing n + 1 Endogenous Variables". In: *Econometrica* 48.4, pp. 861–878.

Ramey, Valerie A. and Sarah Zubairy (2018). "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data". In: *Journal of Political Economy* 126.2, pp. 850–901. DOI: 10.1086/696277.

Sanderson, Eleanor and Frank Windmeijer (2016). "A weak instrument F-test in linear IV models with multiple endogenous variables". In: *Journal of Econometrics* 190.2. Endogeneity Problems in Econometrics, pp. 212–221. DOI: https://doi.org/10.1016/j.jeconom.2015.06.004.

Skeels, Christopher L. and Frank Windmeijer (2018). "On the Stock–Yogo Tables". In: *Econometrics* 6.4. DOI: 10.3390/econometrics6040044.

Staiger, Douglas and James H. Stock (1997). "Instrumental Variables Regression with Weak Instruments". In: *Econometrica* 65.3, pp. 557–586.

Stock, James and Motohiro Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models*. Ed. by Donald W.K. Andrews. New York: Cambridge University Press, pp. 80–108. DOI: 10.1017/CBO9780511614491.

Wen, Zaiwen and Wotao Yin (2013). "A Feasible Method for Optimization with Orthogonality Constraints". In: *Mathematical Programming* 142, pp. 397–434. DOI: 10.1007/s10107-012-0584-1.

A Many-Weak-Instruments Limit

Define $\sigma_1^2 = \lim_{K\to\infty} \text{Tr}(\mathbf{S}_1)/K \in \mathbb{R}$, $\Sigma_2 = \lim_{K\to\infty} R'_{N,K}(\mathbf{S}_2 \otimes I_K)R_{N,K}/K \in \mathbb{R}^{N\times N}$, $\sigma_{21} = \lim_{K\to\infty} R'_{N,K}\,\text{vec}(\mathbf{S}_{12})/K \in \mathbb{R}^N$, and the concentration matrix $\Lambda = K^{-1}\left(R'_{N,K}(\mathcal{S}ll'\mathcal{S}' \otimes I_K)R_{N,K}\right) \in \mathbb{R}^{N\times N}$.

Consider the limits

$$\begin{aligned}
L_1 &= \operatorname*{plim}_{K\to\infty} K^{-1}\left(R'_{N,K}(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_K)R_{N,K}\right) \\
&= \lim_{K\to\infty} K^{-1}\left(R'_{N,K}(\mathcal{S}\mathcal{S}' \otimes I_K)R_{N,K}\right) + \lim_{K\to\infty}\Lambda \\
&= \lim_{K\to\infty} K^{-1}(KI_N) + \lim_{K\to\infty}\Lambda \\
&= I_N + \lim_{K\to\infty}\Lambda \\
L_2 &= \operatorname*{plim}_{K\to\infty} K^{-1}\left(R'_{N,K}(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_K)\right)(\Phi^{-\frac{1}{2}} \otimes I_{K^2})\,\text{vec}\,(\mathbf{S}_{12}) \\
&= \lim_{K\to\infty} K^{-1}R'_{N,K}((K\Sigma_2)^{-\frac{1}{2}} \otimes I_{K^2})\,\text{vec}\,(\mathbf{S}_{12}) \\
&= \lim_{K\to\infty} K^{-\frac{3}{2}}\Sigma_2^{-\frac{1}{2}}R'_{N,K}\,\text{vec}\,(\mathbf{S}_{21}) \\
&= K^{-\frac{1}{2}}\Sigma_2^{-\frac{1}{2}}\sigma_{21}
\end{aligned}$$

Given the definitions of ρ and $\mathbf{h}$ in Appendix B, the bias relative to the benchmark is

$$\begin{aligned}
\operatorname*{plim}_{K\to\infty}\rho'\mathbf{h}'\mathbf{h}\rho &= K^{-1}\left(L_1^{-1}L_2/\sqrt{\sigma_1^2}\right)'\left(L_1^{-1}L_2/\sqrt{\sigma_1^2}\right) \\
&= K^{-1}\left(\Sigma_2^{-\frac{1}{2}}\sigma_{21}\sigma_1^{-1}\right)'\left(I_N + \lim_{K\to\infty}\Lambda\right)^{-2}\left(\Sigma_2^{-\frac{1}{2}}\sigma_{21}\sigma_1^{-1}\right).
\end{aligned}$$

As in Montiel Olea and Pflueger (2013), the worst-case of this bias occurs when $\lim_{K\to\infty}\Lambda = 0$, with the remaining expression maximized when first-stage and reduced-form errors are perfectly correlated, in which case it is equal to 1. This shows that the benchmark is equal to the worst-case many instruments 2SLS bias.

B Derivation of $B = ||\mathbf{h}\rho||_2$

Recall the expression for β^*_{2SLS} in equation (5) in the main text,

$$\beta^*_{2SLS} = \left(R'_{N,K}(\gamma_2\gamma_2' \otimes I_K)R_{N,K}\right)^{-1}R'_{N,K}\,\text{vec}(\gamma_1\gamma_2').$$

From Assumption 2.b it follows that $\gamma_1 = \mathbf{S}_{12}\mathbf{S}_2^{-1}(\gamma_2 - c) + \epsilon$, where ϵ is mean zero and independent of γ_2. Substituting into (5) and taking expectations yields

$$E\left[\beta^*_{2SLS}\right] = E\left[\left(R'_{N,K}(\gamma_2\gamma_2' \otimes I_K)R_{N,K}\right)^{-1} R'_{N,K}\,\mathrm{vec}(\mathbf{S}_{12}\mathbf{S}_2^{-1}(\gamma_2 - c)\gamma_2')\right]. \tag{B.1}$$

Defining $l = \mathbf{S}_2^{-\frac{1}{2}}c$, $\psi = \mathbf{S}_2^{-\frac{1}{2}}(\gamma_2 - c) \sim \mathcal{N}(0, I_{NK})$ and using $\mathrm{vec}(UV) = (V' \otimes I)\,\mathrm{vec}(U)$, the bias in (B.1) can be rewritten as

$$\begin{aligned} E\left[\beta^*_{2SLS}\right] = E\Bigg[&\left(R'_{N,K}(\mathbf{S}_2^{\frac{1}{2}}(l+\psi)(l+\psi)'(\mathbf{S}_2^{\frac{1}{2}})' \otimes I_K)R_{N,K}\right)^{-1} \\ &\times R'_{N,K}\left(\mathbf{S}_2^{\frac{1}{2}}(l+\psi)\psi'\mathbf{S}_2^{-\frac{1}{2}} \otimes I_K\right)\Bigg]\,\mathrm{vec}\left(\mathbf{S}_{12}\right). \end{aligned} \tag{B.2}$$

Define $\Phi = R'_{N,K}(\mathbf{S}_2 \otimes I_K)R_{N,K}$, the matrix $\mathcal{S} = ((\Phi/K)^{-\frac{1}{2}} \otimes I_K)\mathbf{S}_2^{\frac{1}{2}}$, and $\rho = \left((R'_{N,K}(\mathbf{S}_2 \otimes I_K)R_{N,K})^{-\frac{1}{2}} \otimes I_{K^2}\right)\mathrm{vec}\left(\mathbf{S}_{12}\right)/\sqrt{\mathrm{Tr}(\mathbf{S}_1)}$. The unweighted bias becomes

$$\begin{aligned} E\left[\beta^*_{2SLS}\right] = K\Phi^{-\frac{1}{2}}E\big[&\left(R'_{N,K}(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_K)R_{N,K}\right)^{-1} \\ &\times R'_{N,K}\left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_K\right)\big]\rho\sqrt{\mathrm{Tr}(\mathbf{S}_1)}. \end{aligned} \tag{B.3}$$

Using the definition of the bias criterion in (1),

$$B^2 \;=\; \frac{E\left[\beta^*_{2SLS}\right]' R'_{N,K}(\mathbf{S}_2 \otimes I_K)R_{N,K}E\left[\beta^*_{2SLS}\right]}{\mathrm{Tr}(\mathbf{S}_1)} = \rho'\mathbf{h}'\mathbf{h}\rho, \tag{B.4}$$

where $\mathbf{h} = KE\left[\left(R'_{N,K}(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_K)R_{N,K}\right)^{-1} R'_{N,K}\left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_K\right)\right]$. Therefore, $B = ||\mathbf{h}\rho||_2$.

C Proof of Theorem 1

Define the function $h : \mathbb{R}^{NK} \mapsto \mathbb{R}^{N \times NK^2}$

$$h(\psi) \;=\; \left(R'(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_K)R\right)^{-1} R'\left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_K\right). \tag{C.1}$$

The Nagar approximation $\mathbf{h}_n$ of $\mathbf{h}$ as defined in (1) is the expectation of the second-order Taylor expansion of $Kh(\psi)$ evaluated at $\psi = 0$, which is given in vectorized form by

$$\mathrm{vec}(\mathbf{h}_n) = \frac{K}{2}(\mathbf{1}'_{NK} \otimes I_{(NK)^2})\left(\sum_{j=1}^{KN}\left(e_j^{NK}(e_j^{NK})' \otimes I_{(NK)^2}\right)(\nabla_2 h(0))_j\right), \tag{C.2}$$

where $\mathbf{1}_m$ is the $m \times 1$ vector of ones, e_j^m the $m \times 1$ vector with j-th element equal to one and zeros in all other rows, and $\nabla_2 h(\psi)$ is $(NK)^3 \times NK$ second matrix derivative of h. Using

the matrix differentiation rules in Magnus and Neudecker (2019), $\nabla_2 h(\psi)$ is given by

$$\begin{aligned}\nabla_2 h(\psi) &= (I_{NK} \otimes A_3(\psi))\nabla A_1(\psi) + (A_1(\psi)' \otimes I_{(NK)^2})\nabla A_3(\psi) \\ &\quad + (I_{NK} \otimes A_4(\psi))\nabla A_2(\psi) + (A_2(\psi)' \otimes I_{(NK)^2})\nabla A_4(\psi),\end{aligned}$$

where

$$\begin{aligned}
A_1(\psi) &= -(C_1(\psi) \otimes C_1(\psi))B_1 \left((\mathcal{S} \otimes \mathcal{S}(l+\psi)) + (\mathcal{S}(l+\psi) \otimes \mathcal{S})\right) \\
A_2(\psi) &= B_2 \left(\left((\mathcal{S}^{-1})' \otimes \mathcal{S}(l+\psi)\right) + \left((\mathcal{S}^{-1})'\psi \otimes \mathcal{S}\right)\right) \\
A_3(\psi) &= \left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_K\right)' R_{N,K} \otimes I_N \\
A_4(\psi) &= I_{NK^2} \otimes C_1(\psi) \\
\nabla A_1(\psi) &= \left(((\mathcal{S} \otimes \mathcal{S}(l+\psi)) + (\mathcal{S}(l+\psi) \otimes \mathcal{S}))' B_1' \otimes I_{N^2}\right) C_2(\psi) - (I_{NK} \otimes (C_1(\psi) \otimes C_1(\psi)))B_3 \\
\nabla A_2(\psi) &= (I_{NK} \otimes B_2)\left((\mathrm{vec}((\mathcal{S}^{-1})') \otimes I_{NK})\mathcal{S} + (\mathcal{K}_{NK,NK} \otimes I_{NK})(I_{NK} \otimes \mathrm{vec}(\mathcal{S}))(\mathcal{S}^{-1})'\right) \\
\nabla A_3(\psi) &= (I_N \otimes \mathcal{K}_{N,NK^2} \otimes I_N)(I_{(NK)^2} \otimes \mathrm{vec}(I_N))\mathcal{K}_{N,NK^2} A_2(\psi) \\
\nabla A_4(\psi) &= (I_{NK^2} \otimes \mathcal{K}_{N,NK^2} \otimes I_N)(\mathrm{vec}(I_{NK^2}) \otimes I_{N^2})A_1(\psi) \\
B_1 &= (R_{N,K}' \otimes R_{N,K}')(I_{NK} \otimes \mathcal{K}_{K,NK} \otimes I_K)\left(I_{(NK)^2} \otimes \mathrm{vec}(I_K)\right) \\
B_2 &= (I_{NK^2} \otimes R_{N,K}')(I_{NK} \otimes \mathcal{K}_{K,NK} \otimes I_K)\left(I_{(NK)^2} \otimes \mathrm{vec}(I_K)\right) \\
B_3 &= (I_{NK} \otimes B_1)\left(\mathrm{vec}(\mathcal{S} \otimes I_{NK})\mathcal{S} + (\mathcal{K}_{NK,NK} \otimes I_{NK})(I_{NK} \otimes \mathrm{vec}(\mathcal{S}))\mathcal{S}\right) \\
C_1(\psi) &= \left(R_{N,K}'(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_K)R_{N,K}\right)^{-1} \\
C_2(\psi) &= (I_N \otimes \mathcal{K}_{N,N} \otimes I_N)\left[\mathrm{vec}(A_0(\psi)) \otimes I_{N^2} \ : \ I_{N^2} \otimes \mathrm{vec}(A_0(\psi))\right]\left[A_1(\psi)' \ : \ A_1(\psi)'\right]'.
\end{aligned}$$

Writing (C.2) in matrix form and simplifying yields

$$\mathbf{h}_n = \Lambda^{-1}\Big(R_{N,K}' - (\mathrm{vec}(\Lambda^{-1}) \otimes I_N)'(I_N \otimes \mathcal{K}_{N,N})\big((I_{N^2} + \mathcal{K}_{N,N})(I_N \otimes L) \otimes L\big)\Big), \tag{C.3}$$

where $L = \sqrt{K}R_{N,K}'(\mathcal{S}l \otimes I_K)$ and Λ is the concentration matrix defined in (10). Using the eigenvalue decomposition $\Lambda = Q_\Lambda \mathcal{D}_\Lambda Q_\Lambda'$, and the fact that (9) implies $L_0 = Q_\Lambda' \Lambda^{-\frac{1}{2}} L$, (C.3) can be rewritten as

$$\mathbf{h}_n = Q_\Lambda \mathcal{D}_\Lambda^{-\frac{1}{2}} M_1 (\mathcal{D}_\Lambda^{-\frac{1}{2}} Q_\Lambda \otimes L_0 \otimes L_0) M_2, \tag{C.4}$$

where $M_1 = R_{N,N}'\left(I_{N^3} + (\mathcal{K}_{N,N} \otimes I_N)\right)$ and $M_2 = R_{N,K}R_{N,K}'/(1+N) - I_{NK^2}$, which proves Part (i).

Recall the definition of ρ in (1) in the main text

$$\rho = \left((R'(\mathbf{S}_2 \otimes I_K)R)^{-\frac{1}{2}} \otimes I_{K^2}\right)\mathrm{vec}\left(\mathbf{S}_{12}\right) / \sqrt{\mathrm{Tr}(\mathbf{S}_1)}.$$

Using the definitions of $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_{12}$ in (6),

$$\text{(C.5)} \quad \text{Tr}(\mathbf{S}_1) = \text{Tr}((\tilde{\beta}' \otimes I_K)\mathbf{W}(\tilde{\beta} \otimes I_K)) = \tilde{\beta}' R'_{N+1,K}(\mathbf{W} \otimes I_K)R_{N+1,K}\tilde{\beta},$$

$$\text{(C.6)} \quad \text{vec}(\mathbf{S}_{12})' = \text{vec}((\tilde{\beta}' \otimes I_K)[\mathbf{W}_{12} : \mathbf{W}_2])' = \text{vec}(\tilde{\beta}' R'_{N+1,K}([\mathbf{W}_{12} : \mathbf{W}_2] \otimes I_K))',$$

where $\tilde{\beta} = [1 : -\beta']'$. Substituting into the definition of ρ yields

$$\text{(C.7)} \quad \rho = K^{-\frac{1}{2}}\Psi x/\sqrt{x'x},$$

where $x = (R'_{N+1,K}(\mathbf{W}\otimes I_K)R_{N+1,K})^{\frac{1}{2}}\tilde{\beta}$ and $\Psi = \left(\mathcal{S}\mathbf{W}_2^{-\frac{1}{2}}[\mathbf{W}_{12} : \mathbf{W}_2]'\otimes I_K\right)R_{N+1,K}(R'_{N+1,K}(\mathbf{W}\otimes I_K)R_{N+1,K})^{-\frac{1}{2}}$. Since,

$$\text{(C.8)} \quad \sup_{\beta\in\mathbb{R}^N} ||\mathbf{h}_n\rho||_2 = K^{-\frac{1}{2}} \sup_{x\in\mathbb{R}^{N+1}} \frac{||\mathbf{h}_n\Psi x||_2}{||x||_2} = K^{-\frac{1}{2}}||\mathbf{h}_n\Psi||_2,$$

the optimization of the Nagar bias over β amounts to the taking the largest singular value of the matrix $\mathbf{h}_n\Psi$.

Next, note that $\mathbf{h}_n\mathbf{h}'_n = Q_\Lambda\mathcal{D}_h Q'_\Lambda$ where $\mathcal{D}_h = (K - 2(1+N))\mathcal{D}_\Lambda^{-2} + \mathcal{D}_\Lambda^{-\frac{1}{2}}M_1(\mathcal{D}_\Lambda^{-1} \otimes I_{N^2})M'_1\mathcal{D}_\Lambda^{-\frac{1}{2}}$ is a diagonal matrix, such that Q_Λ are eigenvectors of $\mathbf{h}_n\mathbf{h}'_n$ and $\mathcal{D}_h$ contains the eigenvalues. The i-th diagonal element of $\mathcal{D}_h$ is

$$\text{(C.9)} \quad \frac{1}{\lambda_i^2}\left(K - N + 1 + \sum_{j\neq i}^{N}\frac{\lambda_i}{\lambda_j}\right) > 0,$$

where $\lambda_i > 0$ is the i-th eigenvalue of Λ. Each eigenvalue of $\mathbf{h}_n\mathbf{h}'_n$ is decreasing in all eigenvalues of Λ. Making the dependence on $\mathcal{D}_\Lambda$ explicit by the notation $\mathbf{h}_n(\mathcal{D}_\Lambda)$ and fixing Q_Λ, it is therefore the case that $\lambda_{\min}^{-2}\mathbf{h}_n(I_N)'\mathbf{h}_n(I_N) - \mathbf{h}_n(\mathcal{D}_\Lambda)'\mathbf{h}_n(\mathcal{D}_\Lambda)$ is positive semidefinite for all $\mathcal{D}_\Lambda$ with $\lambda_{\min}$ as the smallest diagonal element. This in turn implies that $\lambda_{\min}^{-2}\Psi'\mathbf{h}_n(I_N)'\mathbf{h}_n(I_N)\Psi - \Psi'\mathbf{h}_n(\mathcal{D}_\Lambda))'\mathbf{h}_n(\mathcal{D}_\Lambda))\Psi$ is positive semidefinite, see Proposition 8.1.2 in Bernstein (2009). It follows from Weyl's inequality that $\lambda_{\min}^{-1}||\mathbf{h}_n(I_N)\Psi||_2 \geq ||\mathbf{h}_n(\mathcal{D}_\Lambda))\Psi||_2$, see for example Theorem 8.4.9 in Bernstein (2009). Therefore,

$$\text{(C.10)} \quad \sup_{\mathcal{D}_\Lambda:\lambda_i\geq\lambda_{\min}} K^{-\frac{1}{2}}||\mathbf{h}_n(\mathcal{D}_\Lambda)\Psi||_2 = K^{-\frac{1}{2}}\lambda_{\min}^{-1}||Q_\Lambda M_1(Q_\Lambda \otimes L_0 \otimes L_0)M_2\Psi||_2,$$

which states that largest bias occurs when all eigenvalues of the concentration matrix are equal to the smallest eigenvalue, and therefore when $\mathcal{D}_\Lambda = \lambda_{\min}I_N$. Finally, $Q_\Lambda M_1(Q_\Lambda \otimes L_0 \otimes L_0)M_2\Psi = M_1(I_N \otimes L_0 \otimes L_0)M_2\Psi$ for any Q_Λ, and therefore the sharp upper bound

for the bias does not depend on Q_Λ. This means the sharp upper bound is

$$B_n^*(\mathbf{W}, \lambda_{\min}) = \lambda_{\min}^{-1} K^{-\frac{1}{2}} \sup_{L_0 \in \mathbb{O}^{N \times K}} \{||M_1(I_N \otimes L_0 \otimes L_0)M_2\Psi||_2\}, \tag{C.11}$$

which concludes the proof of Part $(ii.a)$

Turning to Part $(ii.b)$, the upper bound $B^*(\mathbf{W}, \lambda_{\min})_n \leq \lambda_{\min}^{-1}(2(N+1)/K)^{\frac{1}{2}}||M_2\Psi||_2$ follows from $K^{-\frac{1}{2}}||\mathbf{h}_n(\lambda_{\min}I_N)\Psi||_2 \leq K^{-\frac{1}{2}}\lambda_{\min}^{-1}||M_1||_2||(I_N \otimes L_0 \otimes L_0)||_2||M_2\Psi||_2$ and the fact that $||M_1||_2 = (2(N+1))^{\frac{1}{2}}$ and $||(I_N \otimes L_0 \otimes L_0)||_2 = 1$. The inequality follows from Proposition 9.6.1 in Bernstein (2009). Finally, the upper bound, $B_n^*(\mathbf{W}, \lambda_{\min}) \leq \lambda_{\min}^{-1}||\Psi||_2$, follows from $K^{-\frac{1}{2}}||\mathbf{h}_n(\lambda_{\min}I_N)\Psi||_2 \leq K^{-\frac{1}{2}}||\mathbf{h}_n(\lambda_{\min}I_N)||_2||\Psi||_2 = \lambda_{\min}^{-1}||\Psi||_2$ since $||\mathbf{h}_n(\lambda_{\min}I_N)||_2 = K^{\frac{1}{2}}\lambda_{\min}^{-1}$, see (C.9).

D Cumulants of Traces of Partitions of a Non-Central Wishart Matrix

In this section, we derive expressions for cumulants of the matrix of the traces of $K \times K$ partitions of a non-central Wishart matrix.

The Laplace transform of the trace of a non-central Wishart distribution is given in Mathai (1980), equation (1.6), from which it follows immediately that the cumulant generating function is

$$K_{\mathrm{Tr}(W)}(t) = -\frac{1}{2}\mathrm{Tr}\,(\Omega) - \frac{K}{2}\ln|I_{NK} - 2\Sigma| + \frac{1}{2}\mathrm{Tr}\,((I_{NK} - 2\Sigma)^{-1}\Omega). \tag{D.12}$$

However, we are interested in the traces of each $K \times K$ block of W. We follow e.g., Muirhead (1982) or Kollo and Rosen (1995) in evaluating the cumulant generating function for a submatrix,

$$K_{\mathrm{Tr}(W)}(T_N) = -\frac{1}{2}\mathrm{Tr}\,(\Omega) - \frac{K}{2}\ln|I_{NK} - 2M(T_N)\Sigma| + \frac{1}{2}\mathrm{Tr}\,((I_{NK} - 2M(T_N)\Sigma)^{-1}\Omega), \tag{D.13}$$

where T_N is a $N \times N$ matrix and

$$M(T_N) = \sum_{i,j=1,\ldots,N} t_{ij}M_{ij}, \quad M_{ij} = e_j e_i', \tag{D.14}$$

where e_i is the ith block of K columns of the matrix I_{NK}, so that $M_{ij}W$ is matrix containing the ith block of K rows of W in its jth block of K rows, and zero otherwise. Indexing each selection matrix M_{ij} to a scalar value t_{ij} yields the cumulant generating function of the trace of $M_{ij}W$, analogously to Mathai (1980), when the remainder of T_N is set to zero. Note that $\mathrm{Tr}(M_{ij}W) = \mathrm{Tr}(W_{ij})$, since the jth diagonal block of $M_{ij}W$ is W_{ij}, and all other diagonal blocks are zero.

The ij entry of $R'_{N,K}(W \otimes I_K)R_{N,K}$ corresponds to the trace of the ij $K \times K$ block of W. Thus, the nth cumulants of $R'_{N,K}(W \otimes I_K)R_{N,K}$ are obtained by taking the coefficients on $\frac{T_N^n}{n!}$ in the Taylor expansion of $K_{\mathrm{Tr}(W)}(T_N)$, evaluated at $T_N = 0$. Let $\iota(i)$ denote the index of some $K \times K$ block of a $NK \times NK$ matrix. Then $\kappa_n(W)^{\iota(1),\iota(2),\dots}$ denotes the nth cumulant of $\mathrm{Tr}(W_{\iota(1)})$ with $\mathrm{Tr}(W_{\iota(2)})\dots$ (i.e. the covariance for $n = 2$, etc.). Taking such partial derivatives yields the following expressions:

$$\kappa_1^{\iota(1)}(W) = \frac{K}{2}\,\mathrm{Tr}(2M_{\iota(1)}\Sigma) + \frac{1}{2}\,\mathrm{Tr}(2M_{\iota(1)}\Sigma\Omega) \tag{D.15}$$

$$\kappa_2^{\iota(1),\iota(2)}(W) = \frac{K}{2}\,\mathrm{Tr}(2^2 M_{\iota(2)}\Sigma M_{\iota(1)}\Sigma) + \frac{1}{2}\sum_{p\in\mathcal{P}(\iota(1),\iota(2))}\mathrm{Tr}(2^2 M_{p(1)}\Sigma M_{p(2)}\Sigma\Omega) \tag{D.16}$$

$$\begin{aligned}\kappa_3^{\iota(1),\iota(2),\iota(3)}(W) =& \frac{K}{2}\sum_{p\in\mathcal{P}(\iota(2),\iota(3))}\mathrm{Tr}(2^3 M_{p(1)}\Sigma M_{p(2)}\Sigma M_{\iota(1)}\Sigma) \\ &+ \frac{1}{2}\sum_{p\in\mathcal{P}(\iota(1),\iota(2),\iota(3))}\mathrm{Tr}(2^3 M_{p(1)}\Sigma M_{p(2)}\Sigma M_{p(3)}\Sigma\Omega)\end{aligned} \tag{D.17}$$

$$\begin{aligned}\kappa_n^{\iota(1),\iota(2),\dots}(W) =& 2^{n-1}\left(K\sum_{p\in\mathcal{P}(\iota(2),\iota(3),\dots)}\mathrm{Tr}(M_{p(1)}\Sigma M_{p(2)}\Sigma\dots M_{p(n-1)}\Sigma M_{\iota(1)}\Sigma)\right. \\ &\left.+\sum_{p\in\mathcal{P}(\iota(1),\iota(2),\iota(3),\dots)}\mathrm{Tr}(M_{p(1)}\Sigma M_{p(2)}\Sigma M_{p(3)}\Sigma\dots M_{p(n)}\Sigma\Omega)\right),\end{aligned} \tag{D.18}$$

where $\mathcal{P}()$ denotes the set of all permutations of the indices in the argument and $p(i)$ denotes the ith index in a given permutation. Note that for $N = 1$, the formulas collapse to those for the trace in Mathai (1980).

We next prove a lemma relating $\mathrm{Tr}(M_{\iota(1)}\Sigma M_{\iota(2)}\Sigma M_{\iota(3)}\Sigma\dots M_{\iota(n)}\Sigma)$ to $\mathrm{Tr}(\Sigma_{\iota(1)}\Sigma_{\iota(2)}\Sigma_{\iota(3)}\dots\Sigma_{\iota(n)})$. For this purpose, denote the row block index of $\iota(i)$ as $\iota(i)_1$ and the column block index as $\iota(i)_2$, so $M_{\iota(i)} = M_{\iota(i)_1,\iota(i)_2}$. Additionally, let $B_{i\bullet}$ denote the ith block of K rows of the matrix B, and similarly $B_{\bullet i}$ for the block of columns.

Lemma 1. $\mathrm{Tr}(M_{\iota(1)}\Sigma M_{\iota(2)}\Sigma\dots M_{\iota(n)}\Sigma) = \mathrm{Tr}(\Sigma_{\iota(1)_1,\iota(2)_2}\Sigma_{\iota(2)_1,\iota(3)_2}\dots\Sigma_{\iota(n)_1,\iota(1)_2})$.

Proof. For a general symmetric matrix B, the matrix $M_{\iota(1)}B$ has the $\iota(1)_1$ block of K rows of B transferred to its $\iota(1)_2$ block of rows, with the remainder zeros. The product of a sequence of such matrices, $M_{\iota(1)}B\dots M_{\iota(n-1)}B$ contains zeros except for the $\iota(1)_1$ block of K rows. Suppose that this block of rows can be written as $B_{\iota(1)_1,\iota(2)_2}B_{\iota(2)_1,\iota(3)_2}\dots B_{\iota(n-1)_1\bullet}$. Then the product of $M_{\iota(1)}B\dots M_{\iota(n-1)}BM_{\iota(n)}B$ is also a matrix of zeros except for the $\iota(1)_2$ block of K rows, which are equal to $B_{\iota(1)_1,\iota(2)_2}B_{\iota(2)_1,\iota(3)_2}\dots B_{\iota(n-1)_1\iota(n)_2}B_{\iota(n)_1\bullet}$. Consider as a base case $M_{\iota(1)}BM_{\iota(2)}B$. The product is zeros except for the $\iota(1)_2$ block of K rows, which is equal to $B_{\iota(1)_1,\iota(2)_2}B_{\iota(2)_1\bullet}$. Thus, by induction, it follows that the product of $M_{\iota(1)}B\dots BM_{\iota(n)}B$ is a matrix of zeros, except for the $\iota(1)_2$ block of K rows,

which are equal to $B_{\iota(1)_1,\iota(2)_2}B_{\iota(2)_1,\iota(3)_2}\dots B_{\iota(n)_1,\cdot}$, for all n. As an immediate consequence, $\mathrm{Tr}(M_{\iota(1)}B\dots BM_{\iota(n)}B) = \mathrm{Tr}(B_{\iota(1)_1,\iota(2)_2}B_{\iota(2)_1,\iota(3)_2}\dots B_{\iota(n)_1,\iota(1)_2})$, since the latter argument is the single non-zero diagonal block. Letting $B = \Sigma$ and applying the preceding result yields the stated lemma. □

Applying Lemma 1 to the above cumulants yields,

(D.19)

$$\kappa_1^{\iota(1)}(W) = \frac{K}{2}\mathrm{Tr}(2\Sigma_{\iota(1)}) + \frac{1}{2}\mathrm{Tr}(2\Sigma_{\iota(1)}\Omega)$$

(D.20)

$$\kappa_2^{\iota(1),\iota(2)}(W) = \frac{K}{2}\mathrm{Tr}(2^2\Sigma_{\iota(2)_1,\iota(1)_2}\Sigma_{\iota(1)_1,\iota(2)_2}) + \frac{1}{2}\sum_{p\in\mathcal{P}(\iota(1),\iota(2))}\mathrm{Tr}(2^2\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1,p(1)_2}\Omega)$$

(D.21)

$$\begin{aligned}\kappa_3^{\iota(1),\iota(2),\iota(3)}(W) =& \frac{K}{2}\sum_{\iota\in\mathcal{P}(\iota(2),\iota(3))}\mathrm{Tr}(2^3\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1,\iota(1)_2}\Sigma_{\iota(1)_1,p(1)_2}) \\ &+ \frac{1}{2}\sum_{\iota\in\mathcal{P}(\iota(1),\iota(2),\iota(3))}\mathrm{Tr}(2^3\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1p(3)_2}\Sigma_{p(3)_1,p(1)_2}\Omega)\end{aligned}$$

(D.22)

$$\begin{aligned}\kappa_n^{\iota(1),\iota(2),\dots}(W) =& 2^{n-1}\left(K\sum_{p\in\mathcal{P}(\iota(2),\iota(3),\dots)}\mathrm{Tr}(\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1,p(3)_2}\dots\Sigma_{p(n-1)_1,\iota(1)_2}\Sigma_{\iota(1)_1,p(1)_2})\right. \\ &\left.+\sum_{p\in\mathcal{P}(\iota(1),\iota(2),\dots)}\mathrm{Tr}(\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1,p(3)_2}\dots\Sigma_{p(n)_1,p(1)_2}\Omega)\right).\end{aligned}$$

We ultimately need the cumulants of $\gamma' R'_{N,K}(W\otimes I_K)R_{N,K}\gamma$. Using the preceding expressions for cumulants of $R'_{N,K}(W\otimes I_K)R_{N,K}$, we can compute the cumulants of such quadratic forms. The nth cumulant of the quadratic form $\gamma' A\gamma$ for a random matrix A is given by

$$\kappa_n(\gamma' A'\gamma) = \sum_{\iota(1)_1=1}^{N}\sum_{\iota(1)_2=1}^{N}\dots\sum_{\iota(n)_1=1}^{N}\sum_{\iota(n)_2=1}^{N}\left(\prod_{j=1}^{n}\gamma_{\iota(j)_1}\gamma_{\iota(j)_2}\right)\kappa_n^{\iota(1),\dots,\iota(n)}(A), \tag{D.23}$$

where $\iota(i)$ denote indices of individual elements of A. Given the previously derived expressions for the cumulants of the entries of $R'_{N,K}(W\otimes I_K)R_{N,K}$, we can now compute the cumulants of $\gamma' R'_{N,K}(W\otimes I_K)R_{N,K}\gamma$, noting that the cumulants for the ij entry are equal to those for the trace of the ij $K\times K$ block of W, $\mathrm{Tr}(W_{ij})$.

Plugging in the first summation in $\kappa_n^{\iota(1),\dots\iota(n)}(W)$, equation (D.22),

(D.24)
$$\begin{aligned}
&\sum_{\iota(1)_1=1}^{N}\sum_{\iota(1)_2=1}^{N}\cdots\sum_{\iota(n)_1=1}^{N}\sum_{\iota(n)_2=1}^{N}\left(\prod_{j=1}^{n}\gamma_{\iota(j)_1}\gamma_{\iota(j)_2}\right)\sum_{p\in\mathcal{P}(\iota(2),\iota(3),\dots)}\operatorname{Tr}\left(\Sigma_{p(1)_1,p(2)_2}\Sigma_{p(2)_1,p(3)_2}\cdots\right.\\
&\qquad\left.\Sigma_{p(n-1)_1,\iota(1)_2}\Sigma_{\iota(1)_1,p(1)_2}\right)\\
=&\sum_{p\in\mathcal{P}(\iota(2),\iota(3),\dots)}\sum_{\iota'(1)_1=1}^{N}\sum_{\iota'(1)_2=1}^{N}\cdots\sum_{\iota'(n)_1=1}^{N}\sum_{\iota'(n)_2=1}^{N}\left(\prod_{j=1}^{n}\gamma_{\iota'(j)_1}\gamma_{\iota'(j)_2}\right)\operatorname{Tr}(\Sigma_{\iota'(1)}\dots\Sigma_{\iota'(n)})\\
=&(n-1)!\sum_{\iota'(1)_1=1}^{N}\sum_{\iota'(1)_2=1}^{N}\cdots\sum_{\iota'(n)_1=1}^{N}\sum_{\iota'(n)_2=1}^{N}\left(\prod_{j=1}^{n}\gamma_{\iota'(j)_1}\gamma_{\iota'(j)_2}\right)\operatorname{Tr}(\Sigma_{\iota'(1)}\dots\Sigma_{\iota(n)'}),
\end{aligned}$$

where we used a change of indices to move from the first line to the second (recognizing that each set of permuted indices on the blocks of Σ is just the index for some other block of Σ indexed by $\iota'(i)$) and in moving to the third observed that the summand of the outer summation does not depend on the indices of that summation. Note that, by definition,

$$\operatorname{Tr}(((\gamma\gamma'\otimes I_K)B)^n)=\sum_{\iota(1)_1=1}^{N}\sum_{\iota(1)_2=1}^{N}\cdots\sum_{\iota(n)_1=1}^{N}\sum_{\iota(n)_2=1}^{N}\left(\prod_{j=1}^{n}\gamma_{\iota(j)_1}\gamma_{\iota(j)_2})\right)\operatorname{Tr}(B_{\iota(1)}\dots B_{\iota(n)}).$$

Thus, the expression further simplifies to

$$(n-1)!\operatorname{Tr}(((\gamma\gamma'\otimes I_K)\Sigma)^n). \tag{D.25}$$

Next, we can apply the same steps to the second summation in the cumulants to obtain

$$n!\operatorname{Tr}(((\gamma\gamma'\otimes I_K)\Sigma)^n\Omega). \tag{D.26}$$

Combining both terms yields the expression (16) in the main text.

E Upper Bounds on the Cumulants of $g_{\min}$

Recall the expression for the cumulants in the main text

$$\kappa_n \;=\; 2^{n-1}(n-1)!\Big(\operatorname{Tr}\left(((\gamma\gamma'\otimes I_K)\Sigma)^n\right)+n\operatorname{Tr}\left(((\gamma\gamma'\otimes I_K)\Sigma)^n\Omega\right)\Big),$$

with $\gamma'\gamma=1$, and note that we are proving upper bounds for cumulants of order $n>1$.

Using the fact that for a positive semi-definite matrix V, $|\operatorname{Tr}(UV)|\leq \operatorname{maxeval} U\operatorname{Tr}(V)$,

see Fact 8.12.29 in Bernstein (2009), and the fact that $\mathrm{Tr}\left(((\gamma\gamma' \otimes I_K)\Sigma)^n\Omega\right) \geq 0$, we have

$$\begin{aligned}\mathrm{Tr}\left(((\gamma\gamma' \otimes I_K)\Sigma)^n\Omega\right) &\leq \mathrm{maxeval}\{((\gamma\gamma' \otimes I_K)\Sigma)^{n-1}\}\,\mathrm{Tr}\left((\gamma\gamma' \otimes I_K)\Sigma\Omega\right) \\ &= K\lambda_{\min}\,\mathrm{maxeval}\{((\gamma\gamma' \otimes I_K)\Sigma)^{n-1}\},\end{aligned} \tag{E.27}$$

where the last step follows from $\mathrm{Tr}\left((\gamma\gamma' \otimes I_K)\Sigma\Omega\right) = \gamma' R'(\Sigma\Omega \otimes I_K)R\gamma = K\lambda_{\min}$.

Next note that

$$\begin{aligned}\mathrm{maxeval}\{((\gamma\gamma' \otimes I_K)\Sigma)^{n-1}\} &= (\mathrm{maxeval}\{(\gamma\gamma' \otimes I_K)\Sigma\})^{n-1} \\ &= \left(\mathrm{maxeval}\{\Sigma^{\frac{1}{2}}(\gamma\gamma' \otimes I_K)\Sigma^{\frac{1}{2}}\}\right)^{n-1} \\ &\leq (\mathrm{maxeval}\{\Sigma\}\,\mathrm{maxeval}\{(\gamma\gamma' \otimes I_K)\})^{n-1} \\ &= \mathrm{maxeval}\{\Sigma\}^{n-1},\end{aligned} \tag{E.28}$$

where the inequality follows from Ostrowski's theorem, see for example Theorem 4.5.9 in Horn and Johnson (2013), and the last step is due to the fact that the matrix $\gamma\gamma'$ has only one non-zero eigenvalue that is equal to one. We therefore have the inequality

$$\mathrm{Tr}\left(((\gamma\gamma' \otimes I_K)\Sigma)^n\Omega\right) \leq K\lambda_{\min}\,\mathrm{maxeval}\{\Sigma\}^{n-1}. \tag{E.29}$$

Using the Lieb-Thirring inequality for positive semi-definite matrices, see Bernstein (2009) Fact 8.12.17,

$$\mathrm{Tr}(((\gamma\gamma' \otimes I_K)\Sigma)^n) \leq \mathrm{Tr}((\gamma\gamma' \otimes I_K)^n\Sigma^n) = \mathrm{Tr}(((\gamma\gamma')^n \otimes I_K)\Sigma^n) = \mathrm{Tr}((\gamma\gamma' \otimes I_K)\Sigma^n),$$

where the last equality results from the fact that the matrix $\gamma\gamma'$ has only one non-zero eigenvalue that is equal to one. Since $\mathrm{Tr}((\gamma\gamma' \otimes I_K)\Sigma^n) = \gamma' R'(\Sigma^n \otimes I_K)R\gamma$ with $\gamma'\gamma = 1$, we have

$$\mathrm{Tr}(((\gamma\gamma' \otimes I_K)\Sigma)^n) \leq \mathrm{maxeval}\{R'(\Sigma^n \otimes I_K)R\}. \tag{E.30}$$

Applying the two inequalities (E.29) and (E.30) leads to the upper bounds in (18) of the main text.

F Conservative Imhof Approximation

In the general case with $\Sigma \neq I_{NK}$, we use the approximation in Imhof (1961) for the cdf of quadratic forms in normal variables,

$$\text{(F.31)} \quad \Pr(\mathcal{F} < x) \approx \Pr(\chi^2_\nu < (x-\kappa_1)4\omega + \nu) = \int_{\kappa_1 - \nu(4\omega)^{-1}}^{x} \phi(z)dz \quad \text{, where}$$

$$\nu = 8\kappa_2\omega^2 \quad ; \quad \omega = \kappa_2/\kappa_3 \quad ; \quad \phi(z) = \left(1 + \frac{z-\kappa_1}{2\kappa_2\omega}\right)^{\nu/2-1} e^{-\frac{\nu}{2}\left(1+\frac{z-\kappa_1}{2\kappa_2\omega}\right)} \frac{(\nu/2)^{\nu/2-1}\omega}{2^{\nu/2-2}\Gamma(\nu/2)}.$$

This approximation matches the first three central moments of the true distribution of $\mathcal{F}$.

The pdf $\phi(z)$ has a mode at $z^m = \kappa_1 - (2\omega)^{-1}$ if $\nu \geq 2$, and at zero otherwise. The critical value associated with the upper α-percentile is implicitly defined by $\alpha = \int_{x(\alpha)}^{\infty} \phi(z)dz$. To find the largest possible critical value among all possible distributions, we solve the following optimization problem:

$$\text{(F.32)} \qquad \max_{\kappa_1,\kappa_2,\kappa_3} \; x(\alpha) \;\; \text{s.t.} \;\; \kappa_n \leq \bar{\kappa}_n \text{ for } n = 1,2,3 \; .$$

The Kuhn-Tucker conditions are[1]

$$\text{(F.33)} \qquad \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial \kappa_n} dz = \mu_n,$$

together with $\mu_n \geq 0$, $n = 1,2,3$, the constraints and the complementary slackness conditions, where μ_n are the multipliers times $\phi(x(\alpha)) > 0$. The partial derivatives are

$$\text{(F.34)} \qquad \frac{\partial \phi(z)}{\partial \kappa_1} = \frac{1 + (z-\kappa_1)2\omega}{2\kappa_2\omega}\left(1 + \frac{z-\kappa_1}{2\kappa_2\omega}\right)^{-1} \phi(z),$$

$$\text{(F.35)} \qquad \frac{\partial \phi(z)}{\partial \kappa_2} = \frac{\phi(z)}{\kappa_2} G_1\left((z-\kappa_1)4\omega + \nu\right),$$

$$\text{(F.36)} \qquad \frac{\partial \phi(z)}{\partial \kappa_3} = \frac{\phi(z)}{\kappa_3} G_2\left((z-\kappa_1)4\omega + \nu\right),$$

where

$$\text{(F.37)} \qquad G_1(y) = -\frac{1}{2}\left(y - 2\nu(\nu-2)/y + \nu\right) + 3/2(\ln(y/2) - \psi(\nu/2))\nu,$$

$$\text{(F.38)} \qquad G_2(y) = \frac{1}{2}\left(y - \nu(\nu-2)/y\right) - (\ln(y/2) - \psi(\nu/2))\nu,$$

[1] This follows from the implicit function theorem and Leibniz's rule: $1 = -\phi(x(\alpha))\frac{\partial x(\alpha)}{\partial y} + \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial y} dz \Rightarrow \frac{\partial x(\alpha)}{\partial y} = \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial y} dz / \phi(x(\alpha))$ with $\phi(x(\alpha)) > 0$ for $\alpha \in (0,1)$.

and $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function (the logarithmic derivative of the gamma function $\Gamma(x)$). From Alzer (1997) (equation 2.2), we know that

$$1/\nu < \ln(\nu/2) - \psi(\nu/2) < 2/\nu. \tag{F.39}$$

For $n = 1$, the LHS of (F.33) is always positive to the right of the mode, which means the constraint on the mean ($n = 1$) is always binding. The Alzer bounds imply that in the right tail of any optimal distribution, the LHS of (F.33) is always strictly positive for $n = 2, 3$, which means that the constraints are also binding as long as α is sufficiently small. In other words, the Imhof approximation matching the upper bounds for the cumulants is a conservative approximation for the right tail of the true distribution of $g_{\min}$.