UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

A Test for Kronecker Product Structure Covariance Matrix

Guggenberger, P.; Kleibergen, F.; Mavroeidis, S.

DOI
10.1016/j.jeconom.2022.01.005
Publication date
2023
Document Version
Final published version
Published in
Journal of Econometrics

License
CC BY

Link to publication

Citation for published version (APA):
Guggenberger, P., Kleibergen, F., & Mavroeidis, S. (2023). A Test for Kronecker Product Structure Covariance Matrix. *Journal of Econometrics*, *223*(1), 88-112. https://doi.org/10.1016/j.jeconom.2022.01.005

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Nov 2026

Journal of Econometrics 233 (2023) 88–112

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A test for Kronecker Product Structure covariance matrix☆

Patrik Guggenberger [a], Frank Kleibergen [b,*], Sophocles Mavroeidis [c]

[a] Department of Economics, Pennsylvania State University, United States of America
[b] Amsterdam School of Economics, University of Amsterdam, Netherlands
[c] Department of Economics, University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 22 October 2020
Received in revised form 28 January 2022
Accepted 31 January 2022
Available online 4 March 2022

JEL classification:
C12
C26

Keywords:
Covariance matrix
Heteroskedasticity
Invariance
Kronecker product structure
Linear instrumental variables regression model
Reduced rank
Weak identification

ABSTRACT

We propose a test for a covariance matrix to have Kronecker Product Structure (KPS). KPS implies a reduced rank restriction on a certain transformation of the covariance matrix and the new procedure is an adaptation of the Kleibergen and Paap (2006) reduced rank test. To derive the limiting distribution of the Wald type test statistic proves challenging partly because of the singularity of the covariance matrix estimator that appears in the weighting matrix. We show that the test statistic has a χ^2 limiting null distribution with degrees of freedom equal to the number of restrictions tested. Local asymptotic power results are derived. Monte Carlo simulations reveal good size and power properties of the test. Re-examining fifteen highly cited papers conducting instrumental variable regressions, we find that KPS is not rejected in 56 out of 118 specifications at the 5% nominal size.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The robustness properties of nonparametric covariance matrix estimators, like those proposed by White (1980) against heteroskedasticity and by, for example, Newey and West (1987) and Andrews (1991) against heteroskedasticity and autocorrelation, have led to the current default of conducting semi-parametric inference in econometrics. It is well understood that compared to parametrically specified covariance matrix estimators, these robustness properties come at the cost of a large number of additional estimated components. The latter affects the precision of semi-parametric estimators of the structural parameters compared to parametric ones.

For some structural models estimated by the generalized method of moments (GMM), see Hansen (1982), use of nonparametric covariance matrix estimators may also lead to computational challenges for estimation of the structural parameters when using the continuous updating estimator (CUE) of Hansen et al. (1996). Prominent examples of such models are the linear instrumental variables (IV) regression model and the linear factor model in asset pricing. When using a nonparametric covariance matrix estimator, the CUE objective function is often ill behaved, that is, it is flat and/or

☆ Guggenberger gratefully acknowledges the hospitality of the EUI in Florence while parts of the paper were drafted. Mavroeidis gratefully acknowledges the research support of the European Research Council via Consolidator grant number 647152. We would like to thank the Editor Serena Ng, an Associate Editor and two anonymous referees for helpful comments and Lewis McLean for research assistance.

* Corresponding author.
E-mail address: f.r.kleibergen@uva.nl (F. Kleibergen).

https://doi.org/10.1016/j.jeconom.2022.01.005
0304-4076/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

has many local extrema, making the CUE difficult to compute. Part of the appeal of the CUE stems from a number of weak-identification-robust tests based on statistics centered around the CUE for hypotheses involving the structural parameters, see e.g. Kleibergen (2005).

When one uses a Kronecker Product Structure (KPS) covariance matrix estimator instead of a nonparametric one in the CUE objective function in linear IV and factor asset pricing models, the CUE (which is then typically referred to as the limited information maximum likelihood (LIML) estimator) is straightforward to compute. Furthermore, weak-identification-robust tests specified on a subvector of the structural parameter vector with uniformly better power than projected robust full-vector tests are available, see e.g. Guggenberger et al. (2019), Guggenberger et al. (2021), and Kleibergen (2021). The KPS structure of the covariance matrix also allows for an analytical computation of the confidence sets of the structural parameters using the algorithm from Dufour and Taamouti (2005).[1]

The above illustrates the trade-off between, on the one hand, the robustness provided by a nonparametric covariance matrix estimator and, on the other hand, the computational ease and accurate statistical inference provided by a KPS covariance matrix estimator. To help empirical researchers decide when the use of a KPS covariance matrix estimator is justified, we develop a test for the null hypothesis that the covariance matrix $R = E\left(\frac{1}{n}\sum_{i=1}^{n} f_i f_i'\right)$ has KPS, where $f_i = V_i \otimes Z_i$ and $V_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^k$ are uncorrelated random vectors. Here V_i are unobserved error variables (for which consistent estimators are available) and Z_i are observed regressors. This setup encompasses, for example, the linear IV and factor asset pricing models.

The test is based on the insight that KPS implies that a certain invertible transformation $\mathcal{R}(R)$ of R has rank one, see Van Loan and Pitsianis (1993) and (6), and our procedure adapts the Kleibergen and Paap (2006) reduced rank statistic to test for KPS.[2] More precisely, the new test statistic is given as a quadratic form in $vec(\hat{\Lambda})$ with weighting matrix that depends on $vec(\mathcal{R}(\hat{R}))$, where $\hat{R}$ is a sample analogue of R and $\hat{\Lambda}$ is an estimator for a certain matrix that is known to be rank restricted under the null hypothesis (see (16)–(18)). The adaptation of Kleibergen and Paap (2006) is nontrivial partly because the covariance matrix of $vec(\mathcal{R}(\hat{R}))$, that appears in the modified test statistic, is singular. As a consequence, it is a priori not obvious whether the use of the Moore–Penrose generalized inverse in the expression of the Kleibergen and Paap (2006) reduced rank statistic still leads to a χ^2 limiting distribution. To answer that question, we first derive the limiting distribution of $\hat{\Lambda}$ and show the limit to be degenerate Normal. We next establish that the probability limit of the Moore–Penrose inverse of the covariance matrix involved in the Kleibergen and Paap (2006) rank statistic is such that it offsets this degeneracy. As the final result, we conclude that the new KPS test statistic has a χ^2 limiting null distribution with degrees of freedom equal to the number of tested restrictions. We also consider an asymptotic setup where p, k, and n jointly go to infinity and show that the asymptotic null rejection probability of the test is controlled as long as $(pk)^{16} = o(n^3)$. For power considerations, we establish that under sequences of covariance matrices local to KPS the test statistic has a limiting noncentral chi square distribution.

As an important property we show that the proposed test is invariant to orthonormal transformations of the data. In contrast, we show that this is not true for certain alternative tests for KPS that are based on an application of the Kleibergen and Paap (2006) test statistic to a different transformation of the sample covariance matrix estimator that does not lead to a singular covariance matrix (and may therefore a priori seem the more natural choice).

We provide comprehensive Monte Carlo simulations that document good size and power properties of the suggested test. Finally, we apply the new KPS test to various specifications of linear IV models employed in fifteen highly cited empirical studies recently published in top ranked economic journals. We find that for the specifications with independent data and moderate numbers of observations, KPS is not rejected in 24 out of 30 cases at the 5% significance level, while for smaller numbers of observations, it is rejected in 14 out of 28 cases. In specifications with clustered data, KPS is not rejected in 7 out of 17 cases with moderate sample sizes, and 11 out of 35 cases with smaller samples. Overall, KPS is not rejected in 56 out of the 118 specifications that we tested. The relatively high number of non-rejections illustrates the potential importance of the KPS test for applied work.

In a companion paper, Guggenberger et al. (2021), we show how the new KPS test can be used as a key ingredient in a testing procedure with correct asymptotic size for a null hypothesis that restricts the values of a subvector of the structural parameter vector in the linear IV model with a general covariance matrix. The first step of the algorithm uses the KPS test to test the null of a KPS of the covariance matrix of the unrestricted reduced-form sample moment vector. In the second step of the algorithm, the null hypothesis involving the structural parameter is tested using the improved subvector Anderson–Rubin test from Guggenberger et al. (2019) when the test in the first step does not reject and using the AR\AR test from Andrews (2017) otherwise. The AR\AR procedure from Andrews (2017) is an asymptotically size correct inference procedure for testing hypotheses on a subvector of the structural parameters for general covariance matrices but is less powerful than the improved subvector Anderson–Rubin test from Guggenberger et al. (2019) in the linear IV regression model. However, the latter test is asymptotically size correct only when the covariance matrix has KPS. Guggenberger et al. (2021) establish that the resulting two-step procedure has correct asymptotic size and conduct

[1] Dufour and Taamouti (2005) actually assume homoskedasticity, which is a special case of KPS covariance, but their algorithm can be modified to cover the more general case of KPS covariance.

[2] Another adaptation of the Kleibergen and Paap (2006) reduced-rank statistic is by Donald et al. (2007), who develop a test for singularity of a symmetric matrix.

Monte-Carlo experiments which show that it leads to more powerful subvector inference than the AR\AR test in Andrews (2017).

As in the linear IV regression model, KPS of the covariance matrix of the sample moment vector of the linear regression model encompassing linear asset pricing models also leads to improvements in terms of the power of identification robust tests on individual elements of the vector of risk premia and computational ease of obtaining the estimator of the risk premia. There is increasing awareness that risk premia of many risk factors are only weakly identified, see e.g. Kan and Zhang (1999), Kleibergen (2009), and Kleibergen and Zhan (2020). Therefore, it is important to analyze them using inference methods that are robust to weak identification. The current state of the art for conducting weak-factor-robust inference on risk premia is to assume homoskedasticity. Extending homoskedasticity to KPS or even further by extending the switching test procedure from Guggenberger et al. (2021) would extend the scope of the weak-factor-robust inference methods for analyzing the individual risk premia in linear asset pricing models. The KPS test would be an integral part of such extensions.

KPS or separability, which is how other fields sometimes refer to KPS, of the covariance matrix is also studied in the statistics and signal processing literature. The distance to a covariance matrix with KPS is considered in Genton (2007) and Velu and Herman (2017), while Lu and Zimmermann (2005) and Mitchell et al. (2006) analyze the likelihood ratio test of KPS of the covariance matrix of Normally distributed data. They estimate the elements of the KPS covariance matrix using a switching algorithm. Exploiting the reduced rank restriction imposed on the reordered covariance matrix by KPS is also done in Werner et al. (2008). Their results are, however, based on a complex Gaussian distribution for the data, which leads to a degrees of freedom parameter of the χ^2 limiting distribution of their test that is different from the one derived here.

KPS is an example of dimension reduction of a covariance matrix. Other examples of dimension reduction result from shrinking the covariance matrix to a matrix with (much) fewer unrestricted elements to estimate, for example, a scalar multiple of the identity matrix, see e.g. Ledoit and Wolf (2012), or by shrinking the population eigenvalues, see e.g. Ledoit and Wolf (2015) and Ledoit and Wolf (2018).

The paper is organized as follows. In the second section, we introduce the new test for a KPS covariance matrix, derive the asymptotic null distribution of the test statistic, which we denote as KPST, and discuss invariance properties. The third section contains the limiting distribution of the KPST statistic under local alternatives while the fourth section conducts a simulation study to analyze the size and power of the new KPS test. The fifth section summarizes the extensive analysis of testing for a KPS reduced-form covariance matrix in a considerable number of prominent articles. The final sixth section concludes. Proofs and detailed empirical results are given in the Appendix.

We use the vec operator of the matrix A, $vec(A) := (a_1' \dots a_k')' \in \mathbb{R}^{mk}$ for an $m \times k$ dimensional matrix $A = (a_1, \dots, a_k)$. For a symmetric $m \times m$ dimensional matrix A, we also use the $m^2 \times \frac{1}{2}m(m+1)$ dimensional, so-called, duplication matrix D_m which selects the $\frac{1}{2}m(m+1)$ unique elements of A in the $\frac{1}{2}m(m+1)$ dimensional vector $vech(A)$ that vectorizes only the lower triangular part of A:

$$vech(A) = (D_m' D_m)^{-1} D_m' vec(A) \quad \text{and} \quad vec(A) = D_m vech(A).$$

2. A test for Kronecker product structure covariance matrix

We propose a test for a covariance matrix $R \in \mathbb{R}^{kp \times kp}$ to have KPS, where

$$R := E\left(\frac{1}{n}\sum_{i=1}^{n} f_i f_i'\right), \tag{1}$$

for mean zero, independently distributed random vectors $f_i \in \mathbb{R}^{kp}$, $i = 1, \dots, n$, which satisfy

$$f_i := V_i \otimes Z_i \tag{2}$$

with $V_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^k$ uncorrelated random vectors.[3] The specification of f_i fits, for example, a setting where V_i contains the errors of a number of regression equations and Z_i contains the regressors, so that R is then the population counterpart of the sample covariance matrix of the vectors given by the product of these errors and the regressors.

It follows that the covariance matrix has a block structure

$$R := \begin{pmatrix} R_{11} & \cdots & R_{1p} \\ \vdots & \ddots & \vdots \\ R_{p1} & \cdots & R_{pp} \end{pmatrix}, \tag{3}$$

where $R_{jl} \in \mathbb{R}^{k \times k}$, $j, l = 1, \dots, p$. Because $R_{jl} = E\left(\frac{1}{n}\sum_{i=1}^{n} V_{ij} V_{il} Z_i Z_i'\right) = R_{jl}'$, for $V_i = (V_{i1} \dots V_{ip})'$, it follows that R_{jl} is symmetric. We are interested in testing if the covariance matrix R has KPS:

$$H_0 : R = G_1 \otimes G_2 \tag{4}$$

with $G_1 \in \mathbb{R}^{p \times p}$ and $G_2 \in \mathbb{R}^{k \times k}$ symmetric positive definite matrices, against the alternative hypothesis of not having KPS. For normalization purposes, we set one diagonal element equal to one (say the upper left element of G_1).[4] When

[3] The matrix R can depend on the sample size n but for simplicity of notation we do not index R by n.

[4] Normalizing $G_{1,11}$, the upper left element of G_1, to one is an obvious normalization because G_1 is a positive definite matrix (because $G_1 \otimes G_2$ is positive definite) so its diagonal elements are all strictly larger than zero. The normalization does therefore not imply a restriction.

$p = 1$ or $k = 1$ the null is always true and from now on we assume that $\min\{p, k\} \geq 2$. To measure the distance of the sample covariance matrix estimator from a KPS covariance matrix, we use a convenient (invertible) transformation proposed by Van Loan and Pitsianis (1993).

For a matrix $A \in \mathbb{R}^{kp\times kp}$ with block structure as in (3) define

$$\mathcal{R}(A) := \begin{pmatrix} A_1 \\ \vdots \\ A_p \end{pmatrix} \in \mathbb{R}^{p^2\times k^2} \quad \text{with } A_j := \begin{pmatrix} vec\left(A_{1j}\right)' \\ \vdots \\ vec\left(A_{pj}\right)' \end{pmatrix} \in \mathbb{R}^{p\times k^2}, \tag{5}$$

for $j = 1, \ldots, p$. One can easily show that

$$\mathcal{R}(G_1 \otimes G_2) = vec(G_1)vec(G_2)' \tag{6}$$

and by Theorem 2.1 in Van Loan and Pitsianis (1993), we have

$$\|R - G_1 \otimes G_2\|_F = \left\|\mathcal{R}(R) - vec(G_1)vec(G_2)'\right\|_F,$$

with $\|.\|_F$ the Frobenius or trace norm of a matrix, $\|A\|_F^2 := tr(A'A) = vec(A)'vec(A)$, for any rectangular matrix A. Because $\mathcal{R}(G_1 \otimes G_2)$ is a matrix of rank one, when testing for a KPS, it is more convenient to test for the rank of $\mathcal{R}(R)$ to be one instead of directly testing for KPS of R.

Consider the covariance matrix estimator

$$\hat{R} := \frac{1}{n}\sum_{i=1}^{n} \hat{f}_i\hat{f}_i' \in \mathbb{R}^{kp\times kp} \tag{7}$$

which uses sample values $\hat{f}_i := \hat{V}_i \otimes Z_i$ of the random vectors f_i for some estimated residuals $\hat{V}_i$. We assume that $\hat{f}_i = f_i + o_p(1)$, uniformly over $i = 1, \ldots, n$, as $n \to \infty$. Define the distance from a KPS covariance matrix by the Frobenius norm

$$DS := \min_{G_1>0, G_2>0, G_{1,11}=1} \left\|\mathcal{R}(\hat{R}) - vec(G_1)vec(G_2)'\right\|_F, \tag{8}$$

where $G_1, G_2 > 0$ indicates that $G_1 \in \mathbb{R}^{p\times p}$ and $G_2 \in \mathbb{R}^{k\times k}$ are positive definite symmetric matrices, and $G_{1,11} = 1$ states that the upper left element of G_1 is normalized to 1.

We test for $\mathcal{R}(\hat{R})$ being a rank one matrix using the Kleibergen and Paap (2006) rank statistic. To describe the Kleibergen and Paap (2006) rank statistic consider first a singular value decomposition (SVD) of $\mathcal{R}(\hat{R})$:

$$\mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}', \tag{9}$$

where $\hat{\Sigma} := diag(\hat{\sigma}_1, \ldots, \hat{\sigma}_{\min(p^2,k^2)})$ denotes a $p^2 \times k^2$ dimensional diagonal matrix with the singular values $\hat{\sigma}_j$ ($j = 1, \ldots, \min(p^2, k^2)$) on the main diagonal ordered non-increasingly, and with $\hat{L} \in \mathbb{R}^{p^2\times p^2}$ and $\hat{N} \in \mathbb{R}^{k^2\times k^2}$ orthonormal matrices. Decompose

$$\hat{L} := \begin{pmatrix} \hat{L}_{11} & \hat{L}_{12} \\ \hat{L}_{21} & \hat{L}_{22} \end{pmatrix} = \left(\hat{L}_1 \vdots \hat{L}_2\right), \ \hat{\Sigma} := \begin{pmatrix} \hat{\sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{pmatrix}, \ \hat{N} := \begin{pmatrix} \hat{N}_{11} & \hat{N}_{12} \\ \hat{N}_{21} & \hat{N}_{22} \end{pmatrix} = (\hat{N}_1 \vdots \hat{N}_2), \tag{10}$$

with $\hat{L}_{11} : 1\times 1$, $\hat{L}_{12} : 1\times(p^2-1)$, $\hat{L}_{21} : (p^2-1)\times 1$, $\hat{L}_{22} : (p^2-1)\times(p^2-1)$, $\hat{\sigma}_1 : 1\times 1$, $\hat{\Sigma}_2 : (p^2-1)\times(k^2-1)$, $\hat{N}_{11} : 1\times 1$, $\hat{N}_{12} : 1\times(k^2-1)$, $\hat{N}_{21} : (k^2-1)\times 1$, $\hat{N}_{22} : (k^2-1)\times(k^2-1)$ dimensional matrices.

Theorem 1. *Suppose $\hat{R}$ is positive definite. The distance measure DS in (8) equals the square root of the sum of squares of all but the largest singular value of $\mathcal{R}(\hat{R}) \in \mathbb{R}^{p^2\times k^2}$, i.e. $DS^2 = \sum_{i=2}^{\min(p^2,k^2)} \hat{\sigma}_i^2$, where $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_{\min(p^2,k^2)}$ are the ordered singular values of $\mathcal{R}(\hat{R})$. Furthermore, if $\hat{\sigma}_1 > \hat{\sigma}_2$, then the positive definite symmetric minimizers $\hat{G}_1, \hat{G}_2$ of (8) have the following unique expression:*

$$\begin{aligned} vec(\hat{G}_1) &:= \hat{L}_1/\hat{L}_{11} && \in \mathbb{R}^{p^2\times 1}, \\ vec(\hat{G}_2)' &:= \hat{L}_{11}\hat{\sigma}_1\hat{N}_1' && \in \mathbb{R}^{1\times k^2}. \end{aligned} \tag{11}$$

Proof. See the Appendix. ■

If R is positive definite, and $\hat{R} \xrightarrow{p} R$, then $\hat{R}$ will be positive definite with probability approaching one (w.p.a.1). The choice of normalization in (11) conforms with the normalization of G_1, G_2 in (4) discussed in Footnote above.

The KPST statistic. We use the distance between $\mathcal{R}(\hat{R})$ and a matrix of rank one to test for a KPS of R. The test is based on the limiting distribution of the unique elements of $\hat{R}$ or equivalently $\mathcal{R}(\hat{R})$. These elements result from using the $k^2 \times \frac{1}{2}k(k+1)$

and $p^2 \times \frac{1}{2}p(p+1)$ dimensional duplication matrices D_k and D_p:

$$\begin{aligned} \mathcal{R}(\hat{R}) = &\quad \mathcal{R}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(\hat{V}_i\hat{V}_i' \otimes Z_iZ_i')\right) \\ = &\quad \tfrac{1}{n}\textstyle\sum_{i=1}^{n} vec(\hat{V}_i\hat{V}_i')vec(Z_iZ_i')' \\ = &\quad D_p\hat{R}^*D_k', \end{aligned} \tag{12}$$

with

$$\hat{R}^* := \tfrac{1}{n}\textstyle\sum_{i=1}^{n} vech(\hat{V}_i\hat{V}_i')vech(Z_iZ_i')'. \tag{13}$$

The $\frac{1}{2}p(p+1)\times\frac{1}{2}k(k+1)$ dimensional matrix $\hat{R}^*$ contains the unique elements of $\hat{R}$ and $\mathcal{R}(\hat{R})$. We assume $vec(\hat{R}^*)$ satisfies a central limit theorem:

$$\sqrt{n}(vec(\hat{R}^*) - vec(R^*)) \underset{d}{\to} \psi = vec(\Psi), \tag{14}$$

with $\psi \sim N(0, V_{R^*})$, Ψ a $\frac{1}{2}p(p+1)\times\frac{1}{2}k(k+1)$ dimensional normally distributed random matrix and

$$\begin{aligned} R^* := &\quad E\left(\tfrac{1}{n}\sum_{i=1}^{n} vech(V_iV_i')vech(Z_iZ_i')'\right), \\ V_{R^*} := &\quad \lim_{n\to\infty}\Bigg[E\left(\frac{1}{n}\sum_{i=1}^{n}\left(vech(Z_iZ_i')vech(Z_iZ_i')' \otimes vech(V_iV_i')vech(V_iV_i')'\right)\right) \\ &-E\left(vec(\tfrac{1}{n}\sum_{i=1}^{n} vech(V_iV_i')vech(Z_iZ_i')')\right)E\left(vec(\tfrac{1}{n}\sum_{i=1}^{n} vech(V_iV_i')vech(Z_iZ_i')')\right)'\Bigg]. \end{aligned} \tag{15}$$

In fact, we assume a slightly stronger result, namely, that $\hat{R}^* = R^* + \frac{1}{\sqrt{n}}\Psi + o_p(n^{-\frac{1}{2}})$, holds. A central limit theorem (14) for (possibly) non-identically distributed independent random variables holds under mild conditions, e.g. under the Liapunov's or Lindeberg's condition, see White (1984).

Define

$$\hat{\Lambda} := \left(\hat{L}_{22}\hat{L}_{22}'\right)^{-1/2}\hat{L}_{22}\hat{\Sigma}_2\hat{N}_{22}'\left(\hat{N}_{22}\hat{N}_{22}'\right)^{-1/2} \quad : (p^2-1)\times(k^2-1). \tag{16}$$

It can be shown that $\hat{\Lambda} = vec(\hat{G}_1)_\perp'\mathcal{R}(\hat{R})vec(\hat{G}_2)_\perp$, where

$$\begin{aligned} vec(\hat{G}_1)_\perp &:= \hat{L}_2\hat{L}_{22}^{-1}(\hat{L}_{22}\hat{L}_{22}')^{1/2} &&: p^2\times(p^2-1), \\ vec(\hat{G}_2)_\perp' &:= \left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\hat{N}_{22}'^{-1}\hat{N}_2' &&: (k^2-1)\times k^2, \end{aligned}$$

see Kleibergen and Paap (2006, page 102). We then have

$$\mathcal{R}(\hat{R}) = vec(\hat{G}_1)vec(\hat{G}_2)' + vec(\hat{G}_1)_\perp\hat{\Lambda}vec(\hat{G}_2)_\perp'. \tag{17}$$

Using $\mathcal{R}(R)$, our hypothesis of interest H_0 (4) is transformed into

$$H_0 : \mathcal{R}(R) = vec(G_1)vec(G_2)' \text{ or } H_0 : vec(G_1)_\perp'\mathcal{R}(R)vec(G_2)_\perp = 0, \tag{18}$$

where $vec(G_1)_\perp$ and $vec(G_2)_\perp$ are $p^2\times(p^2-1)$ and $k^2\times(k^2-1)$ dimensional matrices that contain the orthogonal complements of $vec(G_1)$ and $vec(G_2)$, $vec(G_1)_\perp'vec(G_1) \equiv 0$, $vec(G_1)_\perp'vec(G_1)_\perp \equiv I_{p^2-1}$, $vec(G_2)_\perp'vec(G_2) \equiv 0$, $vec(G_2)_\perp'vec(G_2)_\perp \equiv I_{k^2-1}$. The KPST test uses the sample analog of the last component in (18) to test H_0. It further results from identifying $vec(G_1)$ and $vec(G_2)$ using the eigenvectors associated with the first singular value of $\mathcal{R}(R)$.

The Kleibergen and Paap (2006) rank test statistic is a quadratic form of the vectorization of $\hat{\Lambda}$ in (16). Its specification directly extends to the new KPS test but because the covariance matrix of $vec\left(\mathcal{R}(\hat{R})\right)$ is singular, the (degenerate) asymptotic normal distribution of $vec(\hat{\Lambda})$ and the resulting degrees of freedom parameter of the χ^2 limiting distribution of the Kleibergen and Paap (2006) rank test statistic are not obvious.

We define the statistic KPST for testing H_0 in (4) as

$$KPST := n \times vec\left(\hat{\Lambda}\right)'\left(\hat{J}'\hat{V}\hat{J}\right)^{-}vec\left(\hat{\Lambda}\right), \tag{19}$$

where

$$\hat{J} := \left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right), \quad \hat{V} := \widehat{\text{cov}}\left(vec\left(\mathcal{R}(\hat{R})\right)\right) \in \mathbb{R}^{p^2k^2\times p^2k^2}, \tag{20}$$

and

$$\begin{aligned}\widehat{\text{cov}}\left(vec\left(\mathcal{R}(\hat{R})\right)\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(vec(Z_iZ_i')vec(Z_iZ_i')'\otimes vec(\hat{V}_i\hat{V}_i')vec(\hat{V}_i\hat{V}_i')'\right)\\ &\quad - vec\left(\mathcal{R}(\hat{R})\right)vec\left(\mathcal{R}(\hat{R})\right)'\\ &= (D_k\otimes D_p)\widehat{\text{cov}}\left(vec\left(\hat{R}^*\right)\right)(D_k\otimes D_p)',\\ \widehat{\text{cov}}\left(vec\left(\hat{R}^*\right)\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(vech(Z_iZ_i')vech(Z_iZ_i')'\otimes vech(\hat{V}_i\hat{V}_i')vech(\hat{V}_i\hat{V}_i')'\right)\\ &\quad - vec\left(\hat{R}^*\right)vec\left(\hat{R}^*\right)'.\end{aligned} \tag{21}$$

In the Appendix it is shown that the KPST statistic in (19) can be simplified as follows:

$$KPST = n\times\left(vec\left(\hat{\Sigma}_2\right)\right)'\left[(\hat{N}_2\otimes\hat{L}_2)'\hat{V}\left(\hat{N}_2\otimes\hat{L}_2\right)\right]^{-}\left(vec\left(\hat{\Sigma}_2\right)\right). \tag{22}$$

This provides an expression for KPST which is easier to compute. On the other hand, it cannot be directly used to obtain the χ^2 limiting distribution because $\hat{\Sigma}_2$ does not have an asymptotic normal distribution while $vec(\hat{\Lambda})$ does.

The KPST statistic.* For comparison, we now introduce an alternative test statistic KPST* that fits more naturally into the Kleibergen and Paap (2006) framework. However, unlike KPST, KPST* turns out not to be invariant to orthonormal transformations of the data. Because $vec(G_1) = D_p vech(G_1)$, $vec(G_2) = D_k vech(G_2)$, the hypothesis of interest (18) can also be specified as:

$$\text{H}_0 : R^* = vech(G_1)vech(G_2)' \text{ or } \text{H}_0 : vech(G_1)_\perp' R^* vech(G_2)_\perp = 0, \tag{23}$$

where $vech(G_1)_\perp$ and $vech(G_2)_\perp$ are $\frac{1}{2}p(p+1)\times(\frac{1}{2}p(p+1)-1)$ and $\frac{1}{2}k(k+1)\times(\frac{1}{2}k(k+1)-1)$ dimensional matrices that contain the orthogonal complements of $vech(G_1)$ and $vech(G_2)$, $vech(G_1)_\perp' vech(G_1) = 0$, $vech(G_1)_\perp' vech(G_1)_\perp = I_{\frac{1}{2}p(p+1)-1}$, $vech(G_2)_\perp' vech(G_2) = 0$, $vech(G_2)_\perp' vech(G_2)_\perp = I_{\frac{1}{2}k(k+1)-1}$. This specification of the hypothesis fits directly in the setup of the Kleibergen and Paap (2006) rank test because the covariance matrix of $\hat{R}^*$ is non-singular. The corresponding specification of $vec(\hat{\Lambda})$ converges to a Normally distributed random vector. The specification of the null hypothesis in (23) allows us to easily infer the number of restrictions tested, which equals $\left(\frac{1}{2}k(k+1)-1\right)\left(\frac{1}{2}p(p+1)-1\right)$. However, the resulting rank statistic does not equal KPST in (19). Specifically, define the SVD of $\hat{R}^* = \hat{L}^*\hat{\Sigma}^*\hat{N}^{*\prime}$, where

$$\hat{L}^* := \begin{pmatrix}\hat{L}_{11}^* & \hat{L}_{12}^*\\ \hat{L}_{21}^* & \hat{L}_{22}^*\end{pmatrix},\ \hat{\Sigma}^* := \begin{pmatrix}\hat{\sigma}_1^* & 0\\ 0 & \hat{\Sigma}_2^*\end{pmatrix},\ \hat{N}^* := \begin{pmatrix}\hat{N}_{11}^* & \hat{N}_{12}^*\\ \hat{N}_{21}^* & \hat{N}_{22}^*\end{pmatrix},$$

with $\hat{L}_{11}^* : 1\times 1$, $\hat{L}_{12}^* : 1\times\left(\frac{1}{2}p(p+1)-1\right)$, $\hat{L}_{21}^* : (\frac{1}{2}p(p+1)-1)\times 1$, $\hat{L}_{22}^* : (\frac{1}{2}p(p+1)-1)\times(\frac{1}{2}p(p+1)-1)$, $\hat{\sigma}_1^* : 1\times 1$, $\hat{\Sigma}_2^* : (\frac{1}{2}p(p+1)-1)\times(\frac{1}{2}k(k+1)-1)$, $\hat{N}_{11}^* : 1\times 1$, $\hat{N}_{12}^* : 1\times(\frac{1}{2}k(k+1)-1)$, $\hat{N}_{21}^* : (\frac{1}{2}k(k+1)^2-1)\times 1$, $\hat{N}_{22}^* : (\frac{1}{2}k(k+1)-1)\times(\frac{1}{2}k(k+1)-1)$ dimensional matrices and $\hat{L}_2^* = (\hat{L}_{12}^{*\prime} \vdots \hat{L}_{22}^{*\prime})'$, $\hat{N}_2^* = (\hat{N}_{12}^{*\prime} \vdots \hat{N}_{22}^{*\prime})'$. The Kleibergen and Paap (2006) statistic for testing (23) using $\hat{R}^*$ is

$$\begin{aligned}&KPST^* := n\times vec\left(\hat{\Lambda}^*\right)'\left(\hat{J}^{*\prime}\hat{V}^*\hat{J}^*\right)^{-}vec\left(\hat{\Lambda}^*\right),\ \text{where}\\ &\hat{\Lambda}^* := \left(\hat{L}_{22}^*\hat{L}_{22}^{*\prime}\right)^{-1/2}\hat{L}_{22}^*\hat{\Sigma}_2^*\hat{N}_{22}^{*\prime}\left(\hat{N}_{22}^*\hat{N}_{22}^{*\prime}\right)^{-1/2},\\ &\hat{J}^* := \left(\left(N_{22}^*N_{22}^{*\prime}\right)^{1/2}N_{22}^{*\prime-1}\left[N_{12}^{*\prime} \vdots N_{22}^{*\prime}\right]\otimes\left(L_{22}^*L_{22}^{*\prime}\right)^{1/2}L_{22}^{*\prime-1}\left[L_{12}^{*\prime} \vdots L_{22}^{*\prime}\right]\right),\\ &\hat{V}^* := \widehat{\text{cov}}\left(vec\left(\hat{R}^*\right)\right)\in\mathbb{R}^{\left(\frac{1}{4}p(p+1)k(k+1)\right)\times\left(\frac{1}{4}p(p+1)k(k+1)\right)},\end{aligned} \tag{24}$$

see Corollary 1 in Kleibergen and Paap (2006).

Asymptotic theory and invariance to orthonormal transformations. The statistics KPST in (19) and KPST* in (24) are not identical, and, unlike the proposed KPST statistic, tests of the KPS hypothesis based on the KPST* statistic are not invariant to orthonormal transformations of the data, as stated in the following Theorem.

Theorem 2. *Assume $E\left(\|f_i\|^8\right) < \kappa$ for some $\kappa < \infty$, $\hat{f}_i = f_i + o_p(1)$, uniformly for $i = 1,\ldots,n$, as $n\to\infty$, and the central limit theorem in (14) holds in the slightly stronger version $\hat{R}^* = R^* + \frac{1}{\sqrt{n}}\Psi + o_p(n^{-\frac{1}{2}})$. Then, under H_0, for KPST and KPST* defined in (19) and (24), respectively, the following hold:*

a.

$$KPST \underset{d}{\to} \chi^2_{df}$$

as $n \to \infty$ (for fixed p and k) with degrees of freedom

$$df := \left(\tfrac{1}{2}k(k+1)-1\right)\left(\tfrac{1}{2}p(p+1)-1\right). \tag{25}$$

b.

$$KPST^* \underset{d}{\to} \chi^2_{df}$$

as $n \to \infty$ (for fixed p and k) with df as given in (25).
c. *The statistics KPST and KPST* are in general not numerically identical. While KPST is invariant to orthonormal transformations of the data in $\hat{V}_i$ and Z_i, KPST* is not invariant to such transformations.*
d. *For sequences p, k, n that satisfy*

$$\tfrac{(pk)^{16}}{n^3} \to 0, \tag{26}$$

we have

$$\lim_{n,p,k\to\infty} \Pr\left[KPST < \chi^2_{df,1-\alpha}\right] \le \alpha, \tag{27}$$

where $\chi^2_{df,1-\alpha}$ denotes the $1-\alpha$ quantile of a χ^2_{df} distribution.

Proof. See the Appendix.[5] ■

We define the new KPST test as follows: it rejects H_0 in (4) at nominal size α if

$$KPST > \chi^2_{df,1-\alpha}. \tag{28}$$

Based on Theorem 2a and d, the resulting test has limiting null rejection probability bounded by α.

Theorem 2c shows that the rank-one tests KPST and KPST* that are based on $\mathcal{R}(\hat{R})$ and $\hat{R}^*$ respectively are not identical even though they are testing the same underlying hypothesis that the covariance R matrix has KPS. This difference occurs because these are Wald tests, and Wald statistics are in general not invariant to non-linear transformations.

Theorem 2d provides a sufficient condition for uniform convergence of $\hat{\Lambda}$ and its covariance matrix estimator for settings where p, k, and n jointly go to infinity so the main results for the limiting distribution of KPST remain unaltered. It is needed to assess the validity of the asymptotic approximation for settings where p and k are relatively large compared to the number of observations n.

The conditions in Theorem 2d are weaker than those in Newey and Windmeijer (2009). Newey and Windmeijer (2009) prove the validity of the asymptotic approximation of test statistics where the number of observations grows faster than the cube of the number of moment restrictions. The number of moment restrictions here is proportional to $(pk)^2$ so their rate would be $(pk)^6/n \to 0$ which is more restrictive than the rate in (26).

Invariance to nonsingular transformations. Theorem 2c shows the KPST is invariant to orthonormal transformations, but it is still not invariant to general nonsingular transformations of the data. To ensure invariance to nonsingular transformations, we need to normalize the data as in Kleibergen and Paap (2006). Specifically, the KPST statistic is computed using the moment vector

$$\hat{f}_i = C_1'\widehat{V}_i \otimes C_2' Z_i, \tag{29}$$

where C_1 and C_2 are the Cholesky factors of the inverse of the second moments of $\hat{V}_i$ and Z_i, i.e., $C_1C_1' = \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_i\widehat{V}_i'\right)^{-1}$ and $C_2C_2' = \left(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\right)^{-1}$. To see why this normalization yields invariance, let A be a nonsingular $k\times k$ matrix, and define the transformed instruments $Z_{A,i} := AZ_i$. Let C_{2A} denote the Cholesky factor of $\left(\frac{1}{n}\sum_{i=1}^n AZ_iZ_i'A'\right)^{-1}$. The KPST statistic with the original instruments Z_i is computed using $C_2'Z_i$ in the moment vector (29), while the KPST with the transformed instruments AZ_i uses $C_{2A}'AZ_i$ in the same formula (29). Therefore, the transformation from $C_2'Z_i$ to $C_{2A}'Z_{A,i}$ is given by $T_A := C_{2A}'AC_2^{-1\prime}$, i.e., $C_{2A}'Z_{A,i} = C_{2A}'AZ_i = T_A\left(C_2'Z_i\right)$. Now, observe that T_A is an orthonormal matrix, because $T_A'T_A = C_2^{-1}A'C_{2A}C_{2A}'AC_2^{-1\prime} = C_2^{-1}A'\left(\frac{1}{n}\sum_{i=1}^n AZ_iZ_i'A'\right)^{-1}AC_2^{-1\prime} = C_2^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\right)^{-1}C_2^{-1\prime} = I_k$. Hence, invariance follows from Theorem 2c. The exact same argument can be made about rotations of the reduced form errors $\hat{V}$.

Clustered data. In case of clustered data, we assume there are n clusters of N_i observations each, so the total number of data points is $\sum_{i=1}^n N_i$:

$$f_i = \sum_{j=1}^{N_i} f_{ij}, \tag{30}$$

[5] We thank an anonymous associate editor for pointing at the vech operator and duplication matrix to simplify the proof and exposition.

for mean zero kp dimensional random vectors f_{ij}, $j = 1, \ldots, N_i$, $i = 1, \ldots, n$. Observations f_{ij} within cluster i can be arbitrarily dependent, i.e., $E\left(f_{ij}f_{is}\right)$ is unrestricted for all $j, s = 1, \ldots, N_i$, while observations across clusters are independent. The $kp \times kp$ dimensional (positive semi-definite) covariance matrix of the sample moments then results as:

$$R = \frac{1}{n}\sum_{i=1}^{n} E(f_i f_i'). \tag{31}$$

3. Limiting distribution of KPST under local alternatives

To analyze the power of KPST under local alternatives, we construct the limiting distribution of the KPST statistic under alternatives where the covariance matrix of the moments $R \in \mathbb{R}^{kp\times kp}$ is local to KPS:

$$\mathrm{H}_1 : R = (G_1 \otimes G_2) + \frac{1}{\sqrt{n}}A_0, \tag{32}$$

where $G_1 \in \mathbb{R}^{p\times p}$ and $G_2 \in \mathbb{R}^{k\times k}$ are symmetric positive definite matrices, and $A_0 \in \mathbb{R}^{kp\times kp}$ is a fixed symmetric matrix. The best-fitting KPS approximation of R under H_1 w.r.t. Frobenius norm, defined as $\bar{G}_{1,n} \otimes \bar{G}_{2,n}$, where $\bar{G}_{1,n}, \bar{G}_{2,n}$ solve $\min_{\bar{G}_1>0,\bar{G}_2>0} \left\|(G_1 \otimes G_2) + \frac{1}{\sqrt{n}}A_0 - \bar{G}_1 \otimes \bar{G}_2\right\|_F$, will in general differ from $G_1 \otimes G_2$. That is, $\bar{G}_{1,n} \neq G_1$ and $\bar{G}_{2,n} \neq G_2$, unless A_0 lies in the span of the orthogonal complement of $G_1 \otimes G_2$. However, under the local alternatives (32), $\bar{G}_{1,n} \to G_1$ and $\bar{G}_{2,n} \to G_2$. This needs to be taken into account when we characterize the asymptotic distribution of the KPST statistic under the local alternatives in (32).

The re-arranged matrix $\mathcal{R}(R)$ under H_1 is:

$$\begin{aligned} \mathcal{R}(R) = &\quad vec(G_1)vec(G_2)' + \frac{1}{\sqrt{n}}\mathcal{R}(A_0) \\ = &\quad vec(\bar{G}_{1,n})vec(\bar{G}_{2,n})' + vec(\bar{G}_{1,n})_\perp \Lambda_n vec(\bar{G}_{2,n})'_\perp, \end{aligned} \tag{33}$$

with

$$\Lambda_n = vec(\bar{G}_{1,n})'_\perp \mathcal{R}(R) vec(\bar{G}_{2,n})_\perp. \tag{34}$$

The decomposition in the last line of (33) is identical to the one in (17).

Theorem 3. *Under local to KPS sequences of covariance matrices as in (32) and for mean zero, independently distributed random vectors $f_i \in \mathbb{R}^{kp}$ with finite eighth moments,*

$$KPST \underset{d}{\to} \chi^2_{df}(\delta)$$

as $n \to \infty$ (with k, p fixed), where

$$\begin{aligned} \delta := &\quad vec(a_0)' \left[\left(vec(G_2)'_\perp \otimes vec(G_1)'_\perp\right)(D_k \otimes D_p)V_{R^*}\right. \\ &\quad \left.(D_k \otimes D_p)'\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right)\right]^- vec(a_0), \end{aligned} \tag{35}$$

V_{R^*} has been defined in (15), and

$$a_0 := vec(G_1)'_\perp \mathcal{R}(A_0) vec(G_2)_\perp \in \mathbb{R}^{\left(\frac{1}{2}k(k+1)-1\right)\times\left(\frac{1}{2}p(p+1)-1\right)}.$$

Proof. See the Appendix. ■

4. Simulation study on size and power

Size. We evaluate the accuracy of the limiting distribution in Theorem 2 to approximate the finite sample distribution of the KPST statistic. We do so in a small simulation experiment using the linear regression model:

$$Y_i = Z_i'\Pi + V_i, \qquad i = 1, \ldots, n, \tag{36}$$

where Y_i is a p dimensional vector of dependent variables, Z_i is a k dimensional vector of explanatory (exogenous) variables and V_i is a p dimensional vector of errors. We further set Π to zero (which is without loss of generality because KPST uses the residual vectors) and generate the Z_i's independently from $N(0, I_k)$ distributions and V_i given Z_i independently from a $N\left(0, h\left(Z_i\right) I_p\right)$ distribution. We consider two different specifications of $h(Z_i)$. The first leads to homoskedasticity and has $h(Z_i) = 1$ while the second leads to (scalar) heteroskedasticity and has $h\left(Z_i\right) = \|Z_i\|^2/k$. For each case, we compute null rejection probabilities (NRPs) using the three conventional nominal significance levels of 10%, 5% and 1%. The NRPs are computed using 40,000 Monte Carlo replications for the KPST test that uses chi-square critical values based on the results from Theorem 2. Table 1 reports the NRPs when the sample size depends on the dimensions p and k, specifically $n = (kp)^{16/3}$, in accordance with Theorem 2. We notice only a slight underrejection in some cases, but in the remaining cases the NRPs are not significantly different from the test's nominal levels.

Table 1
Rejection frequencies (in percentages) of KPST test at various significance levels. χ^2_{df} critical values. $n = (pk)^{16/3}$, df: number of restrictions given in Eq. (25), m: number of estimated parameters. Computed using 40,000 MC replications.

Data Generating Process:					Homoskedastic			Scalar heteroskedastic		
p	k	n	df	m	10%	5%	1%	10%	5%	1%
2	2	1626	4	9	10.0	5.1	1.0	9.7	4.4	0.7
2	3	14130	10	18	10.0	5.0	0.8	9.3	4.2	0.7
2	4	65536	18	30	9.4	5.0	0.9	9.7	4.9	0.9
2	5	215444	28	45	9.8	4.7	0.9	9.8	5.1	1.0
3	2	14130	10	18	10.2	5.0	0.9	10.0	4.7	0.9
3	3	122827	25	36	9.7	4.9	1.0	9.8	5.0	0.9

Table 2
Rejection frequencies (in percentages) of KPST test at various significance levels. χ^2_{df} critical values. $n = (pk)^4$, df: number of restrictions given in Eq. (25), m: number of estimated parameters. Computed using 40,000 MC replications.

Data Generating Process:					Homoskedastic			Scalar heteroskedastic		
p	k	n	df	m	10%	5%	1%	10%	5%	1%
2	2	256	4	9	11.2	5.3	0.9	11.4	4.8	0.5
2	3	1296	10	18	10.2	4.9	0.9	9.3	4.0	0.5
2	4	4096	18	30	9.9	5.1	1.0	9.1	4.2	0.8
2	5	10000	28	45	9.7	4.6	0.8	8.8	4.0	0.6
2	6	20736	40	63	10.0	5.1	1.0	9.5	4.5	0.7
2	7	38416	54	84	9.8	4.8	0.9	9.5	4.5	0.8
3	2	1296	10	18	9.9	4.8	0.7	9.0	3.7	0.5
3	3	6561	25	36	9.8	5.0	0.9	9.6	4.4	0.7
3	4	20736	45	60	10.7	5.6	1.2	10.2	5.1	0.9
3	5	50625	70	90	10.4	5.2	1.0	10.2	5.0	0.7
3	6	104976	100	126	10.2	5.0	1.1	10.1	5.0	1.0
3	7	194481	135	168	10.2	5.0	1.0	10.0	5.0	1.0

Table 2 reports NRPs with a smaller sample size $n = (pk)^4$. In this case, we find some modest deviations from the nominal size but these are generally quite small.

To investigate NRPs in smaller samples, Figs. 1 to 3 show the NRPs as a function of the sample size n for smaller sample sizes than in Tables 1, 2 for different settings of p and k. Depending on the value of the latter, the NRPs are close to the nominal level for values of n much smaller than $(pk)^4$. For larger values of pk, we therefore do not (like for the smaller values of pk) show the rejection frequencies all the way up to $n = (pk)^{\frac{16}{3}}$, i.e. the value indicated by Theorem 2d, but just to $(pk)^4$, which is for $p = 2$, $k = 7$ at the bottom right hand side of Fig. 1, equal to approximately 40,000, and for $p = 5$, $k = 4$ at the bottom right hand side of Fig. 2 equal to 160,000 (note that the horizontal axis is in log-scale). In many cases, the NRPs are still much closer to their nominal significance levels than indicated by this rate. For example, when $p = k = 2$ and testing at the 5% significance level, the NRP is close to the nominal level for sample size of around 100. More striking is that when $p = 2$ and $k = 5$ the KPST test at 5% nominal size has NRPs close to the nominal size for values of n around 200. Figs. 1–3 also show that the KPST test generally over-rejects for small n. Moreover, the over-rejection is increasing in the dimensions k and p and can be very substantial for very small n, see Fig. 3, as is the case for any Wald test when the number of restrictions is large relative to the sample size. Therefore, it is of interest to investigate the possibility of small-sample corrections, e.g., following the bootstrap approach of Chen and Fang (2019). From a practical perspective, this over-rejection means that rejection of KPS with small sample sizes, which happens only a few times in the applications reported in Section 5, could be due to a significantly higher type 1 error probability than the nominal size of the test.[6]

Power. We simulate the power of the KPST test using the asymptotic χ^2 critical values stated in Theorem 2. The Data Generating Process (DGP) is given by a model with $p = k = 2$, where $Y_i = Z_i\Pi + V_i$ and $\Pi = 0$, see (36). The two dimensional vectors containing the regressors Z_i and errors V_i are simulated according to:

$$V_i \sim iid \begin{cases} N(0, \Omega_1), \\ N(0, \Omega_2), \end{cases} \quad Z_i \sim iid \begin{cases} N\left(0, Q_{zz,1}\right), & i = 1, \ldots, [n/2] \\ N\left(0, Q_{zz,2}\right), & i = [n/2]+1, \ldots, n, \end{cases} \tag{37}$$

[6] When KPST is used as a pre-test in a two-step procedure, such as the subvector Anderson–Rubin test of Guggenberger et al. (2021), that involves choosing a second-step test that is robust to violation of KPS when the KPST rejects in the first step, over-rejection will only affect the power but not the overall size of the two-step procedure.

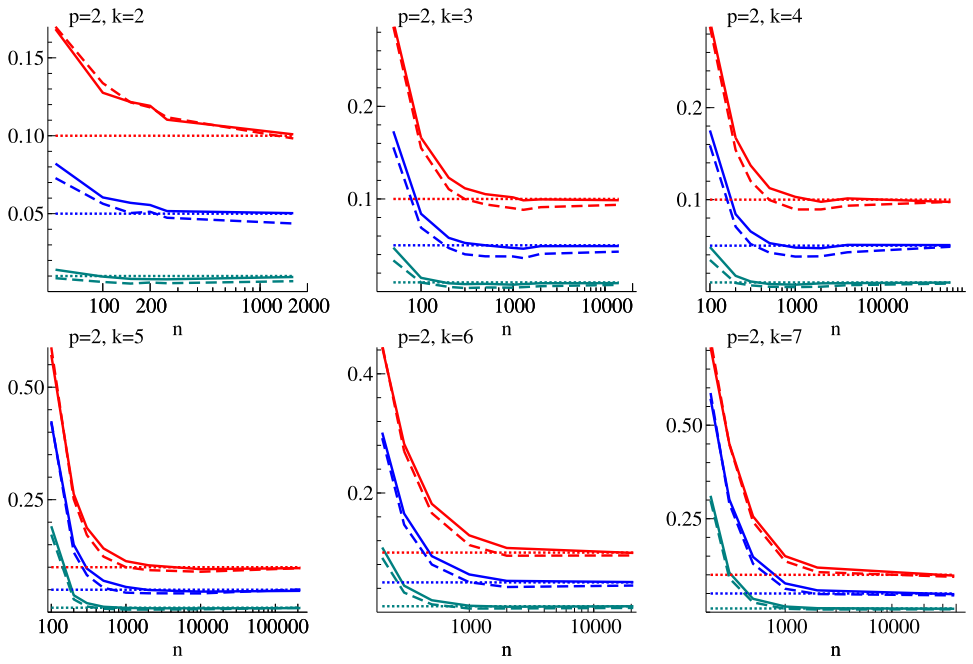

Fig. 1. Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

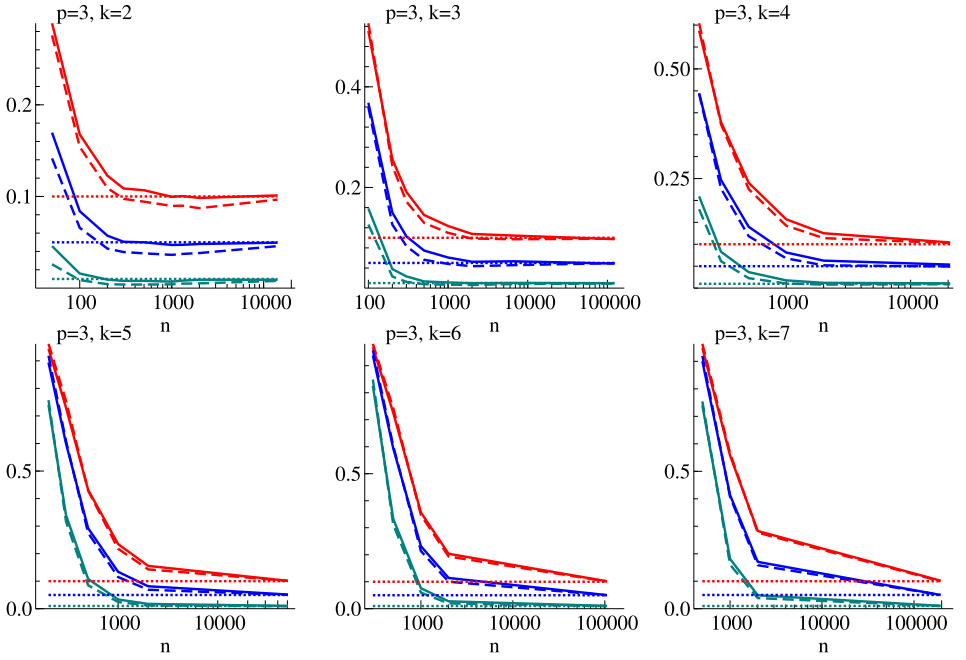

Fig. 2. Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with $\Omega_1 = diag\,(b, 1)$, $\Omega_2 = diag\,(1, b)$, $Q_{zz,1} = diag\,(1, c)$, $Q_{zz,2} = diag\,(c, 1)$, and

$$b := \frac{1}{2}\frac{\sigma}{\sqrt{n}} - \frac{1}{2}\sqrt{\frac{\sigma}{\sqrt{n}}\left(\frac{\sigma}{\sqrt{n}} + 8\right)} + 1, \quad c := \frac{1}{2}\frac{\sigma}{\sqrt{n}} + \frac{1}{2}\sqrt{\frac{\sigma}{\sqrt{n}}\left(\frac{\sigma}{\sqrt{n}} + 8\right)} + 1, \tag{38}$$

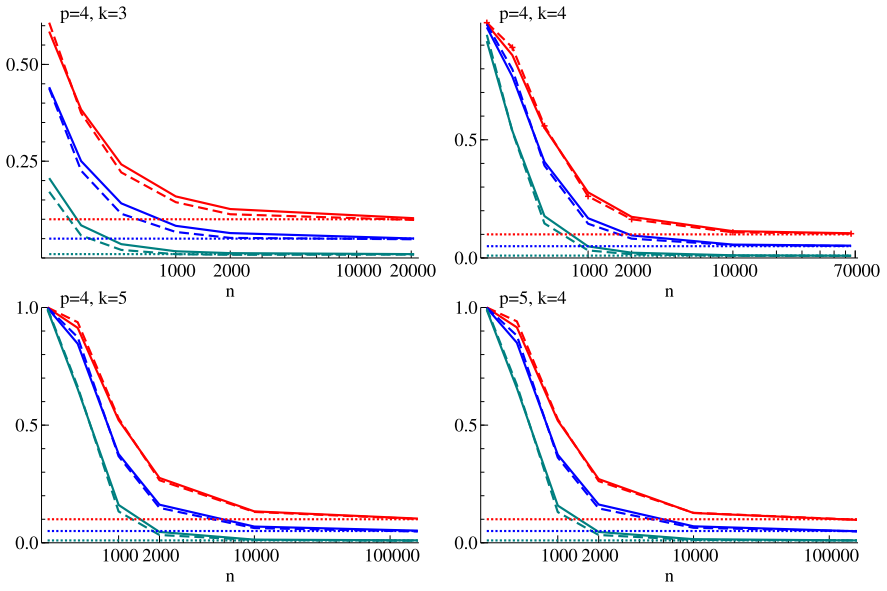

Fig. 3. Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for $\sigma \in [0, \sqrt{n})$. The covariance matrix R is then such that:

$$\begin{aligned} R = & \ \tfrac{1}{n} var\left(\textstyle\sum_{i=1}^{n}(V_i \otimes Z_i)\right) = \tfrac{1}{2} diag\,(b+c, 1+bc, 1+bc, b+c) \\ = & \ \underbrace{I_4}_{G_1 \otimes G_2} + \tfrac{\sigma}{\sqrt{n}} \times diag\,(1,-1,-1,1)\,, \end{aligned} \tag{39}$$

and $G_1 = G_2 = I_2$. Because

$$\mathcal{R}(diag\,(1,-1,-1,1)) = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \tag{40}$$

$vec(G_1)'\mathcal{R}(diag\,(1,-1,-1,1))vec(G_2) = 0$, the re-arranged specification of R in (33) equals:

$$\begin{aligned} \mathcal{R}(R) = & \ vec(G_1)vec(G_2)' + \tfrac{\sigma}{\sqrt{n}} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \\ = & \ vec(G_1)vec(G_2)' + \tfrac{1}{\sqrt{n}} vec(G_1)_\perp a_0 vec(G_2)'_\perp, \end{aligned} \tag{41}$$

where

$$vec(G_1)_\perp = vec(G_2)_\perp = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ -1 & 0 & 0 \end{pmatrix}, \ a_0 = \sigma \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sigma e_1 e_1', \ e_1 = (1,0,0)', \tag{42}$$

is such that the local deviation from KPS lies in the orthogonal complement of $vec(G_1)$ and $vec(G_2)$. The non-centrality parameter of the non-central χ^2 limiting distribution follows from (35). Note that

$$(e_1 \otimes e_1)' \left[\left([vec(G_2)]'_\perp \otimes [vec(G_1)]'_\perp \right) \text{cov}\left(vec\left(\mathcal{R}(\hat{R}) \right) \right) \left([vec(G_2)]_\perp \otimes [vec(G_1)]_\perp \right) \right]^{-} (e_1 \otimes e_1) = \frac{1}{4}, \tag{43}$$

where $G_i = I_2$ for $i = 1, 2$. Note also that $vec\,(a_0) = 2\sigma(e_1 \otimes e_1)$. Thus, the non-centrality parameter is

$$\delta = \tfrac{1}{4}\sigma^2. \tag{44}$$

For $\sigma = 0$, R has KPS, so the null hypothesis in (4) holds. For the limiting case of $\sigma = \sqrt{n} : b = 0$, so Ω_1 and Ω_2 are singular.

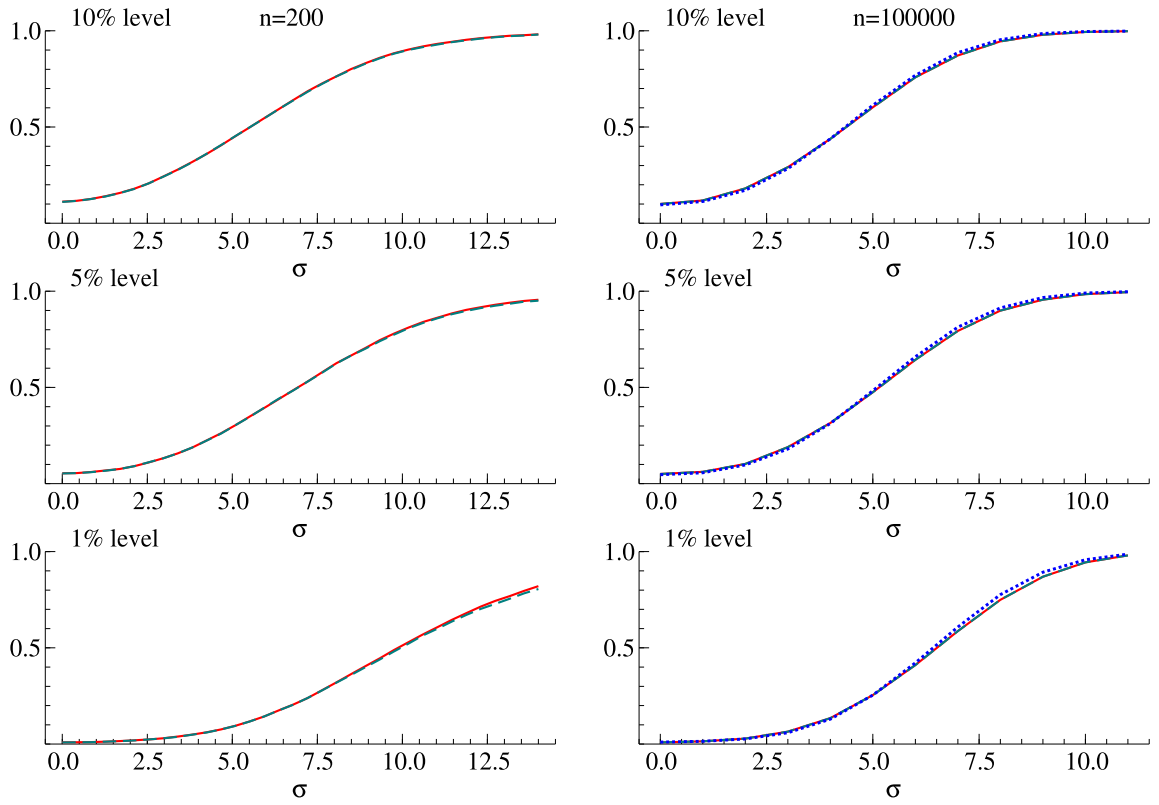

Fig. 4. Power of KPST (solid red) and KPST* (dashed green) tests with sample size $n = 200$ (left) and $n = 10^5$ (right). Asymptotic approximation from Theorem 3 (dotted blue) superimposed on the right. σ measures deviation from KPS in Frobenius norm. Computed using 10,000 Monte Carlo replications.

We compute the power function of the KPST test at three significance levels 10%, 5% and 1% using 10,000 Monte Carlo replications. For comparison, we also compute the power of the non-invariant KPST* test that rejects H_0 if the statistic KPST* in (24) exceeds the corresponding $1-\alpha$ quantile of χ^2_{df} with degrees of freedom df given in Theorem 2b, which are the same critical values as for the KPST test in (28). The results are reported graphically in Fig. 4. The left-hand-side graphs in Fig. 4 show that for a moderate sample of size $n = 200$ both tests have good and essentially identical power. Moreover, as the sample size increases, the power function of both tests approaches the noncentral χ^2 asymptotic approximation in Theorem 3 indicated in blue on the right-hand-side graphs of Fig. 4 for $n = 100,000$. Results for other sample sizes are qualitatively similar and are omitted in the interest of brevity. In particular, KPST has nontrivial power even for small samples.

5. Empirical applications

We investigate whether KPS covariance matrices are potentially relevant for applied work. To do so, we apply the KPST test to the covariance matrices of estimators in published empirical studies. We consider fifteen highly cited papers conducting linear IV regressions from top journals in economics and test for KPS of the joint covariance matrix of the (unrestricted reduced form) least squares estimators which result from regressing all endogenous variables on the instruments.[7] Tables 6 and 7 in the Supplementary Appendix report the results of the KPST test for the 118 different specifications we analyzed. Table 6 does so for the studies using independent data (sixty specifications) while Table 7 lists the results for studies with clustered data (fifty eight specifications). Because these tables are rather extensive, Tables 3 and 4 report a summary of our findings on the KPST tests.

Table 3, summarizing our results on KPS tests for the papers using independent data, shows considerable support for KPS covariance matrices especially when the number of observations is not too large. For the 60 different specifications using independent data reported in Table 3, KPS is rejected at the 5% nominal size for only about one third of them, namely for 22.

[7] Both the endogenous variables and the instruments are first regressed on the control, or included exogenous, variables and only the residuals from these regressions are used.

Table 3
Summary of results of 5% significance level KPST tests for specifications in papers using independent observations.

Paper	#specifications	KPS rejection	# observations
Tanaka et al. (2010)	2	none	moderate
Nunn (2008)	4	4	small
Acemoglu and Johnson (2005)	24	10	small
Hansford and Gomez (2010)	2	2	huge
Alesina et al. (2013)	6	1	moderate
Yogo (2004)	22	5	moderate

Table 4
Summary of results of 5% significance level KPST tests for specifications in papers using clustered observations.

Paper	#specific.	KPS rej.	# obs.	Clustered KPS rej.	# clusters
Duranton and Turner (2011)	8	6	large	5	moderate
Acemoglu et al. (2008)	9	7	large	5	moderate
Johnson et al. (2006)	4	4	huge	4	huge
Parker et al. (2013)	2	2	huge	2	huge
Autor et al. (2013)	18	18	large	13	small
Autor and Dorn (2013)	7	7	huge	7	small
Acemoglu et al. (2011)	1	1	small	1	very small
Miguel et al. (2004)	3	0	large	3	small
Voors et al. (2012)	6	1	moderate	0	small

Table 4, summarizing the test results for papers using clustered data, shows that for the 58 different specifications with clustered data, KPS is rejected at the 5% nominal size for 46 specifications when using the unrestricted covariance matrix estimator (7) and for 40 when using the clustered covariance matrix estimator (31). The number of observations in the involved papers using clustered data is typically much larger than for the papers using independent observations which largely explains our different findings for independent compared to clustered observations.

Summarizing, our analysis of the KPS of covariance matrices of moment condition vectors in a considerable number of prominent empirical studies shows that KPS is often not rejected especially for moderate sample sizes.

6. Conclusion

We propose a test for the null of a covariance matrix of a vector of moment equations to have a KPS. The test is an extension of the Kleibergen and Paap (2006) rank test and is easy to use. We apply it to data used in a considerable number of prominent applied studies conducting IV regressions and find that KPS of the covariance matrix of the least squares estimator of the unrestricted reduced form is often not rejected for moderate sample sizes. In linear IV regression, a KPS covariance matrix brings considerable advantages for both computation and inference in weakly identified settings. Given the common occurrence of weak identification in applications, our empirical findings underscore the contribution that the use of KPS covariance matrices can make in applied work.

In a companion paper, Guggenberger et al. (2021), we develop a two-step test procedure that in the first step uses the new KPS covariance matrix test and, depending on its outcome, in the second step conducts a weak-identification-robust test on a subset of the structural parameters based either on an improved powerful subvector AR test or based on the AR/AR test that is robust to arbitrary forms of conditional heteroskedasticity. The two-step procedure is constructed such that its asymptotic size is bounded by the nominal size. A promising area for application of testing for KPS is in linear factor models for establishing risk premia. The default setting in this area is to assume homoskedasticity and weak identification is often present.

To further improve the approximation of the finite sample distribution of the KPST statistic, it would also be of interest to investigate whether the bootstrap can deliver refinements as Chen and Fang (2019) show for rank tests on general matrices. We leave this important extension for future work.

Appendix. Proofs

Proof of Theorem 1. For a given nonzero matrix $A \in \mathbb{R}^{m\times n}$ with SVD $A = Udiag(\sigma_1, \ldots, \sigma_p)V'$ for singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ with $p = \min\{m, n\}$, rectangular $diag(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}$, orthogonal matrices $U = [u_1, \ldots, u_m] \in \mathbb{R}^{m\times m}$, and $V = [v_1, \ldots, v_n] \in \mathbb{R}^{n\times n}$, Lemma 2 in Guggenberger et al. (2021) states that a minimizing argument in the minimization problem

$$\min_{B\in\mathbb{R}^{m\times n},\ rk(B)=1} \|A - B\|_F^2 \tag{45}$$

is given by $\widehat{B} = \sigma_1 u_1 v_1'$ and the minimum equals $\sum_{i=2}^{p} \sigma_i^2$. Furthermore, it is shown that if $\sigma_1 > \sigma_2$ then $\widehat{B} = \sigma_1 u_1 v_1'$ is the unique minimizer.

Theorem 5.8 in Van Loan and Pitsianis (1993) states that if $A \in \mathbb{R}^{pk\times pk}$ is symmetric and positive definite then minimizers $\tilde{G}_1$ and $\tilde{G}_2$ for the problem

$$\min_{G_1\in\mathbb{R}^{p\times p},G_2\in\mathbb{R}^{k\times k}} \|A - G_1\otimes G_2\|_F^2 \tag{46}$$

exist that are also symmetric and positive definite. Because $\hat{R}$ is symmetric by construction and positive definite by assumption, (46) and $\|\hat{R} - G_1\otimes G_2\|_F = \|\mathcal{R}(\hat{R}) - vec(G_1)vec(G_2)'\|_F$ (which holds by Theorem 2.1 in Van Loan and Pitsianis, 1993) imply that symmetric positive definite matrices $\hat{G}_1$ and $\hat{G}_2$ exist that minimize $\|\mathcal{R}(\hat{R}) - vec(G_1)vec(G_2)'\|_F^2$ over $G_1 \in \mathbb{R}^{p\times p}$, $G_2 \in \mathbb{R}^{k\times k}$. Therefore, because the rank of $vec(\hat{G}_1)vec(\hat{G}_2)'$ is one, $\widehat{B} := vec(\hat{G}_1)vec(\hat{G}_2)'$ is a minimizer in the problem $\min_{B\in\mathbb{R}^{p^2\times k^2},\ rk(B)=1} \|\mathcal{R}(\hat{R}) - B\|_F^2$. But by (45) with A playing the role of $\mathcal{R}(\hat{R})$ we know that the minimum equals $\sum_{i=2}^{\min\{p^2,k^2\}} \hat{\sigma}_i^2$, where $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \cdots \geq \hat{\sigma}_{\min\{p^2,k^2\}} \geq 0$ denote the singular values of $\mathcal{R}(\hat{R}) \in \mathbb{R}^{p^2\times k^2}$. This establishes the claimed formula for DS^2.

Next, if $\hat{\sigma}_1 > \hat{\sigma}_2$, then, by the uniqueness part of Lemma 2 in Guggenberger et al. (2021), the minimizing $\hat{G}_1$ and $\hat{G}_2$ in (46) satisfy $vec(\hat{G}_1)vec(\hat{G}_2)' = \hat{\sigma}_1\hat{L}_1\hat{N}_1'$. There are many different ways we can define the argmins such that $vec\left(\hat{G}_1\right)$ and $vec\left(\hat{G}_2\right)$ are proportional to $\hat{L}_1$ and $\hat{N}_1$, respectively. Because sign-definiteness and symmetry are not affected by normalization by a positive constant, any normalization will produce estimates $\hat{G}_1, \hat{G}_2$ that are symmetric and positive definite. The normalization we use is $vec(\hat{G}_1) = \hat{L}_1/\hat{L}_{11}$, so that $vec(\hat{G}_2) = \hat{L}_{11}\hat{\sigma}_1\hat{N}_1$, i.e., the upper left element of $\hat{G}_1$ is normalized to 1. We know that $\hat{L}_{11} \neq 0$ for otherwise, one diagonal block of the covariance matrix $\hat{R}$ would be zero, which would contradict the assumption that $\hat{R}$ is positive definite. This establishes (11).

Proof of Eq. (22). Because $\hat{L}_{22}^{-1}(\hat{L}_{22}\hat{L}_{22}')^{1/2}$ and $\hat{N}_{22}^{-1}\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}$ are invertible which follows from expressions stated in the proof of Theorem 2a, KPST can be rewritten as:

$$\begin{aligned}
KPST = \ & n\times\left[vec(vec(\hat{G}_1)_\perp'\mathcal{R}(\hat{R})vec(\hat{G}_2)_\perp)\right]' \\
& \left[\left(vec(\hat{G}_2)_\perp'\otimes vec(\hat{G}_1)_\perp'\right)\hat{V}\left(vec(\hat{G}_2)_\perp\otimes vec(\hat{G}_1)_\perp\right)\right]^- \\
& \left[vec(vec(\hat{G}_1)_\perp'\mathcal{R}(\hat{R})vec(\hat{G}_2)_\perp)\right] \\
= \ & \left(vec\left(\hat{L}_2'\hat{L}\hat{\Sigma}\hat{N}'\hat{N}_2\right)\right)'\left(\hat{N}_{22}^{-1}\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\otimes\hat{L}_{22}^{-1}(\hat{L}_{22}\hat{L}_{22}')^{1/2}\right) \\
& \left[\left(\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\hat{N}_{22}'^{-1}\otimes(\hat{L}_{22}\hat{L}_{22}')^{1/2}\bar{L}_{22}^{-1\prime}\right)\left(\hat{N}_2\otimes\hat{L}_2\right)'(D_k\otimes D_p)\hat{V}_{\hat{R}^*}\right. \\
& \left.(D_k\otimes D_p)'\left(\hat{N}_2\otimes\hat{L}_2\right)\left(\hat{N}_{22}^{-1}\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\otimes\hat{L}_{22}^{-1}(\hat{L}_{22}\hat{L}_{22}')^{1/2}\right)\right]^- \\
& \left(\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\hat{N}_{22}'^{-1}\otimes(\hat{L}_{22}\hat{L}_{22}')^{1/2}\hat{L}_{22}^{-1\prime}\right)\left(vec\left(\hat{L}_2'\hat{L}\hat{\Sigma}\hat{N}'\hat{N}_2\right)\right) \\
= \ & n\times\left(vec\left(\hat{\Sigma}_2\right)\right)'\left[(\hat{N}_2\otimes\hat{L}_2)'\hat{V}\left(\hat{N}_2\otimes\hat{L}_2\right)\right]^-\left(vec\left(\hat{\Sigma}_2\right)\right).
\end{aligned}$$

Proof of Theorem 2a. The hypothesis of interest in (18) is: $H_0 : vec(G_1)_\perp'\mathcal{R}(R)vec(G_2)_\perp = 0$. We test this hypothesis using a SVD of $\mathcal{R}(\hat{R})$:

$$\mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}',$$

whose elements using (12) and the orthonormality of $\hat{L}$ and $\hat{N}$ can be specified as

$$\hat{L} = (D_p\hat{A} \vdots D_{p\perp}), \qquad \hat{N} = (D_k\hat{B} \vdots D_{k\perp}),$$
$$\hat{\Sigma} = \begin{pmatrix}\hat{\sigma}_1 & 0\\ 0 & \hat{\Sigma}_2\end{pmatrix}, \qquad \hat{\Sigma}_2 = \begin{pmatrix}\hat{\Sigma}_{22} & 0\\ 0 & 0\end{pmatrix},$$

where $\hat{A}$ is a $\frac{1}{2}p(p+1)\times\frac{1}{2}p(p+1)$ dimensional matrix, $\hat{A}'D_p'D_p\hat{A} = I_{\frac{1}{2}p(p+1)}$, $\hat{B}$ is a $\frac{1}{2}k(k+1)\times\frac{1}{2}k(k+1)$ dimensional matrix, $\hat{B}'D_k'D_k\hat{B} = I_{\frac{1}{2}k(k+1)}$, $\hat{\Sigma}_{22}$ is a diagonal $(\frac{1}{2}p(p+1)-1)\times(\frac{1}{2}k(k+1)-1)$ dimensional matrix, $D_{p\perp}$ and $D_{k\perp}$ are $p^2\times\frac{1}{2}p(p-1)$ and $k^2\times\frac{1}{2}k(k-1)$ dimensional matrices which are the orthogonal complements of D_p and D_k, $D_p'D_{p\perp}\equiv 0$, $D_{p\perp}'D_{p\perp}\equiv I_{\frac{1}{2}p(p-1)}$, $D_k'D_{k\perp}\equiv 0$ and $D_{k\perp}'D_{k\perp}\equiv I_{\frac{1}{2}k(k-1)}$. We also use an identical SVD of the population counterpart $\mathcal{R}(R)$ of $\mathcal{R}(\hat{R})$:

$$\mathcal{R}(R) = L\Sigma N',$$

with an identical specification of its elements (but without "^") and where under H_0 : $\Sigma_{22} = 0$.

To obtain the limit distribution of the sample analog of the parameter tested under H_0 recall from below (16) that

$$\hat{\Lambda} = \quad vec(\hat{G}_1)'_\perp \mathcal{R}(\hat{R}) vec(\hat{G}_2)_\perp. \tag{47}$$

Next, we use that $vec(\hat{G}_1)_\perp = vec(G_1)_\perp + O_p(n^{-\frac{1}{2}})$, $vec(\hat{G}_2)_\perp = vec(G_2)_\perp + O_p(n^{-\frac{1}{2}})$, which holds under our imposed conditions, see Kleibergen and Paap (2006), the assumption $\hat{R}^* = R^* + \frac{1}{\sqrt{n}}\Psi + o_p(n^{-\frac{1}{2}})$, and $D_p R^* D_k' = vec(G_1) vec(G_2)'$, which holds under H_0. Thus, under H_0

$$\begin{aligned}\hat{\Lambda} = & \left[vec(G_1)_\perp + O_p(n^{-\frac{1}{2}})\right]' \left[vec(G_1)vec(G_2)' + \tfrac{1}{\sqrt{n}} D_p \Psi D_k' + o_p(n^{-\frac{1}{2}})\right] \left[vec(G_2)_\perp + O_p(n^{-\frac{1}{2}})\right] \\ = & \ \tfrac{1}{\sqrt{n}} vec(G_1)'_\perp D_p \Psi D_k' vec(G_2)_\perp + o_p(n^{-\frac{1}{2}}).\end{aligned}$$

To construct the limit distribution of $\hat{\Lambda}$, recall that A and B were defined from $L = (D_p A \vdots D_{p\perp})$ and $N = (D_k B \vdots D_{k\perp})$, and partition them as

$$A = \begin{pmatrix} a_1 & A_2 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & B_2 \end{pmatrix},$$

where $a_1 : \frac{1}{2}p(p+1) \times 1$, $A_2 : \frac{1}{2}p(p+1) \times (\frac{1}{2}p(p+1) - 1)$, $b_1 : \frac{1}{2}k(k+1) \times 1$, $B_2 : \frac{1}{2}k(k+1) \times (\frac{1}{2}k(k+1) - 1)$. Then,

$$\begin{aligned} vec(G_1)_\perp = & \ L_2 L_{22}^{-1} (L_{22} L_{22}')^{1/2}, & vec(G_2)_\perp = & \ N_2 N_{22}^{-1} \left(N_{22} N_{22}'\right)^{1/2}, \\ L_2 = & \begin{pmatrix} e'_{1,\frac{1}{2}p(p+1)} A_2 & 0 \\ D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix}, & N_2 = & \begin{pmatrix} e'_{1,\frac{1}{2}k(k+1)} B_2 & 0 \\ D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix}, \\ L_{22} = & \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix}, & N_{22} = & \begin{pmatrix} D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix}, \end{aligned}$$

where we use that $D_p = (e_{1,\frac{1}{2}p(p+1)} \vdots D_{2,p}')'$, for $D_{2,p} : (p^2-1) \times \frac{1}{2}p(p+1)$ and $D_k = (e_{1,\frac{1}{2}k(k+1)} \vdots D_{2,k}')'$, for $D_{2,k} : (k^2-1) \times \frac{1}{2}k(k+1)$ with $e_{1,i}$ the first i dimensional unity vector (i.e. the first column of I_i). We partition $D_{p\perp} = (0 \vdots D_{2,p\perp}')'$, where $D_{2,p\perp} : (p^2-1) \times \frac{1}{2}p(p-1)$, $D_{2,p\perp}' D_{2,p\perp} = I_{\frac{1}{2}p(p-1)}$, and $D_{k\perp} = (0 \vdots D_{2,k\perp}')'$, where $D_{2,k\perp} : (k^2-1) \times \frac{1}{2}k(k-1)$, $D_{2,k\perp}' D_{2,k\perp} = I_{\frac{1}{2}k(k-1)}$, where the specifications of $D_{p\perp}$ and $D_{k\perp}$ result from those of D_p and D_k.

We next use the spectral decompositions of $A_2' D_{2,p}' D_{2,p} A_2 : (\frac{1}{2}p(p+1) - 1) \times (\frac{1}{2}p(p+1) - 1)$ and $B_2' D_{2,k}' D_{2,k} B_2 : (\frac{1}{2}k(k+1) - 1) \times (\frac{1}{2}k(k+1) - 1)$:

$$\begin{aligned} A_2' D_{2,p}' D_{2,p} A_2 = & \ L_{D_{2p}A_2} \Lambda^2_{D_{2,p}A_2} L'_{D_{2p}A_2} \\ B_2' D_{2,k}' D_{2,k} B_2 = & \ L_{D_{2k}B_2} \Lambda^2_{D_{2,k}B_2} L'_{D_{2k}B_2}, \end{aligned}$$

with $L_{D_{2p}A_2}$ and $L_{D_{2k}B_2}$ orthonormal $(\frac{1}{2}p(p+1)-1) \times (\frac{1}{2}p(p+1)-1)$ and $(\frac{1}{2}k(k+1)-1) \times (\frac{1}{2}k(k+1)-1)$ dimensional matrices and $\Lambda^2_{D_{2,p}A_2}$ and $\Lambda^2_{D_{2,k}B_2}$ diagonal $(\frac{1}{2}p(p+1)-1) \times (\frac{1}{2}p(p+1)-1)$ and $(\frac{1}{2}k(k+1)-1) \times (\frac{1}{2}k(k+1)-1)$ dimensional matrices with the squared singular values in non-increasing order on the diagonal. We note that $A_2' D_{2,p}' D_{2,p} A_2$ is invertible. This results since $A_2' D_p' D_p A_2 = A_2' D_{2,p}' D_{2,p} A_2 + A_2' e_{1,\frac{1}{2}p(p+1)} e'_{1,\frac{1}{2}p(p+1)} A_2 = I_{\frac{1}{2}p(p+1)}$ so $A_2' D_{2,p}' D_{2,p} A_2 = I_{\frac{1}{2}p(p+1)} - A_2' e_{1,\frac{1}{2}p(p+1)} e'_{1,\frac{1}{2}p(p+1)} A_2$. Only when $e'_{1,\frac{1}{2}p(p+1)} A_2 A_2' e_{1,\frac{1}{2}p(p+1)} = 1$ is this of lower rank since the specification then corresponds with a projection matrix. This is, however, not possible given the specification of $L = (D_p A \vdots D_{p\perp})$ which is orthonormal so $L'L = LL' = I_{p^2}$. The quadratic form (inner product) of the top row of L is thus equal to one. Given the specification of D_p, $D_{p\perp}$ has only zeros on the first row. Next, the L_{11} element is unequal to zero because R_{11} is a positive definite covariance matrix. Since the L_{11} element is unequal to zero, the length of the vector of the remaining elements on the first row of L cannot be equal to one. This implies that $e'_{1,\frac{1}{2}p(p+1)} A_2 A_2' e_{1,\frac{1}{2}p(p+1)} \neq 1$ so $A_2' D_{2,p}' D_{2,p} A_2$ is invertible and $B_2' D_{2,k}' D_{2,k} B_2$ as well. A further consequence is that L_{22} and N_{22} are invertible and similarly $\hat{L}_{22}$ and $\hat{N}_{22}$.

The above spectral decompositions feature in the SVDs of L_{22}, and N_{22}, which we can specify as:

$$\begin{aligned} L_{22} = & \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \\ = & \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p}A_2} \Lambda^2_{D_{2,p}A_2} L'_{D_{2p}A_2})^{-\frac{1}{2}} L_{D_{2p}A_2} \Lambda_{D_{2,p}A_2} L'_{D_{2,p}A_2} & D_{2,p\perp} \end{pmatrix} \\ = & \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p}A_2} \Lambda^2_{D_{2,p}A_2} L'_{D_{2p}A_2})^{-\frac{1}{2}} L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \Lambda_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\ & \begin{pmatrix} L'_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix}, \end{aligned}$$

$$
\begin{aligned}
(L_{22}L_{22}')^{\frac{1}{2}} &= \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix}\begin{pmatrix} \Lambda_{D_{2,p}A} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\
&\quad \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix}', \\
L_{22}^{-1}\left(L_{22}L_{22}'\right)^{1/2} &= \begin{pmatrix} L_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix}\begin{pmatrix} L'_{D_{2p}A_2}(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}P'_{D_{2p}A_2})^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix} \\
&= \begin{pmatrix} (L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}P'_{D_{2p}A_2})^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix} = \begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
vec\,(G_1)_\perp = L_2L_{22}^{-1}\left(L_{22}L_{22}'\right)^{1/2} &= \begin{pmatrix} D_pA_2 & D_{p\perp} \end{pmatrix}\begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix}, \\
vec\,(G_2)_\perp = N_2N_{22}^{-1}\left(N_{22}N_{22}'\right)^{1/2} &= \begin{pmatrix} D_kB_2 & D_{k\perp} \end{pmatrix}\begin{pmatrix} (B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}' \end{pmatrix},
\end{aligned} \tag{48}
$$

where in the third line of the decomposition of L_{22}, we have the three components that result from a SVD of L_{22}.

Then, under H_0:

$$
\begin{aligned}
\sqrt{n}\hat{\Lambda} &= vec(G_1)_\perp'D_p\Psi D_k'vec(G_2)_\perp + o_p(1) \\
&= (L_{22}L_{22}')^{1/2}L_{22}^{-1}L_2'D_p\Psi D_k'N_2N_{22}^{-1}\left(N_{22}N_{22}'\right)^{1/2} + o_p(1) \\
&= \begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix}'\begin{pmatrix} D_pA_2 & D_{p\perp} \end{pmatrix}'D_p\Psi D_k' \\
&\quad \begin{pmatrix} D_kB_2 & D_{k\perp} \end{pmatrix}\begin{pmatrix} (B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}' \end{pmatrix} + o_p(1) \\
&= \begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix}'\begin{pmatrix} A_2'D_p'D_p\Psi D_k'D_kB_2 & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad \begin{pmatrix} (B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}' \end{pmatrix} + o_p(1) \\
&= D_{2,p}A_2\bar{\Lambda}B_2'D_{2,k}' + o_p(1),
\end{aligned} \tag{49}
$$

where

$$
\bar{\Lambda} := (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p\Psi D_k'D_kB_2(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}},
$$

which is a $\left(\frac{1}{2}p(p+1)-1\right)\times\left(\frac{1}{2}k(k+1)-1\right)$ normally distributed random matrix with mean zero. The covariance matrix of $vec(\bar{\Lambda})$ equals

$$
\begin{aligned}
V_{vec(\bar{\Lambda})} = &\left((B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \otimes (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p\right)V_{R^*} \\
&\times\left((B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \otimes (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p\right)'.
\end{aligned}
$$

The above implies that the limiting distribution of $\sqrt{n}\hat{\Lambda}$ is degenerate Normal because $D_{2,p}A_2$ and $D_{2,k}B_2$ are $(p^2-1)\times(\frac{1}{2}p(p+1)-1)$ and $(k^2-1)\times(\frac{1}{2}k(k+1)-1)$ dimensional matrices, respectively, and so the number of rows exceeds the number of columns.

We now apply a weak law of large numbers to the sample average $\hat{V}$ defined in (19). The matrix $\hat{V}$ contains summands of eighth order products of f_i and the weak law of large numbers holds by the assumption that $E\left(\|f_i\|^8\right) < \kappa$. To derive the limit of the covariance matrix estimator in the KPST statistic, the following derivations are important:

$$
\begin{aligned}
&\left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right)'\hat{V}\left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right) \\
&\underset{p}{\to} (vec(G_2)_\perp \otimes vec(G_1)_\perp)'(D_k\otimes D_p)V_{R^*}(D_k\otimes D_p)'(vec(G_2)_\perp \otimes vec(G_1)_\perp) \\
&= \left(\begin{pmatrix} D_kB_2 & D_{k\perp} \end{pmatrix}\begin{pmatrix} (B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}' \end{pmatrix}\otimes \right. \\
&\quad \left.\begin{pmatrix} D_pA_2 & D_{p\perp} \end{pmatrix}\begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}' \end{pmatrix}\right)'(D_k\otimes D_p)V_{R^*}(D_k\otimes D_p)'
\end{aligned}
$$

$$
\begin{aligned}
&\left(\begin{pmatrix} D_kB_2 & D_{k\perp}\end{pmatrix}\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}'\end{pmatrix}\otimes\right.\\
&\left.\begin{pmatrix} D_pA_2 & D_{p\perp}\end{pmatrix}\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}'\end{pmatrix}\right)\\
=&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)\\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \\ 0\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p \\ 0\end{pmatrix}\right)V_{R^*}\\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \\ 0\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p \\ 0\end{pmatrix}\right)'\\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)'\\
=&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)\begin{pmatrix}V_{vec(\bar\Lambda)} & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)',
\end{aligned}
\tag{50}
$$

where the specifications of $vec(G_1)_\perp$ and $vec(G_2)_\perp$ result from (48). The convergence behavior of KPST is then characterized by:

$$
\begin{aligned}
KPST =&n\times\left[vec(vec(G_1)_\perp'\mathcal{R}(\hat R)vec(G_2)_\perp)\right]'\\
&\left[(vec(G_2)_\perp\otimes vec(G_1)_\perp)'(D_k\otimes D_p)V_{R^*}(D_k\otimes D_p)'(vec(G_2)_\perp\otimes vec(G_1)_\perp)\right]^-\\
&\left[vec(vec(G_1)_\perp'\mathcal{R}(\hat R)vec(G_2)_\perp)\right]+o_p(1)\\
=&vec(\bar\Lambda)'(D_{2,k}B_2\otimes D_{2,p}A_2)'\left(\begin{pmatrix} D_{2,k}B_2(B_2'D_{2,k}'D_{2,k}B_2)^{-1} & D_{2,k\perp}\end{pmatrix}\otimes\right.\\
&\left.\begin{pmatrix} D_{2,p}A_2(A_2'D_{2,p}'D_{2,p}A_2)^{-1} & D_{2,p\perp}\end{pmatrix}\right)\begin{pmatrix}V_{vec(\bar\Lambda)} & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}^-\\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-1}B_2'D_{2,k}' \\ D_{2,k\perp}'\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-1}A_2'D_{2,p}' \\ D_{2,p\perp}'\end{pmatrix}\right)\\
&(D_{2,k}B_2\otimes D_{2,p}A_2)vec(\bar\Lambda)+o_p(1)\\
=&vec(\bar\Lambda)'V_{vec(\bar\Lambda)}^{-1}vec(\bar\Lambda)+o_p(1)\xrightarrow[d]{}\chi^2_{df},
\end{aligned}
$$

with $df=\left(\frac{1}{2}p(p+1)-1\right)\left(\frac{1}{2}k(k+1)-1\right)$. The first equality substitutes $vec(\hat G_1)_\perp$ and $vec(\hat G_2)_\perp$ by their limits. The second equality follows from (47), vectorizing (49), and the last line of (50). It also uses that the Moore–Penrose inverse of the expression on the last line of (50) equals

$$
\begin{aligned}
&\left[\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)\begin{pmatrix}V_{vec(\bar\Lambda)} & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\right.\\
&\left.\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)'\right]^-\\
=&\left(\begin{pmatrix} D_{2,k}B_2(B_2'D_{2,k}'D_{2,k}B_2)^{-1} & D_{2,k\perp}\end{pmatrix}\otimes\right.\\
&\left.\begin{pmatrix} D_{2,p}A_2(A_2'D_{2,p}'D_{2,p}A_2)^{-1} & D_{2,p\perp}\end{pmatrix}\right)\begin{pmatrix}V_{vec(\bar\Lambda)} & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}^-\\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-1}B_2'D_{2,k}' \\ D_{2,k\perp}'\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-1}A_2'D_{2,p}' \\ D_{2,p\perp}'\end{pmatrix}\right)
\end{aligned}
$$

which follows because

$$\begin{pmatrix} D_{2,k}B_2(B_2'D_{2,k}'D_{2,k}B_2)^{-1} & D_{2,k\perp} \end{pmatrix}' \begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp} \end{pmatrix} = \begin{pmatrix} I_{\frac{1}{2}k(k+1)-1} & 0 \\ 0 & I_{\frac{1}{2}k(k-1)} \end{pmatrix} = I_{k^2-1}$$

as $D_{2,k\perp} = I_{\frac{1}{2}k(k-1)}$ and the same argument can be applied to the other component. The third equality then follows from

$$\begin{aligned} \begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp} \end{pmatrix}^{-1} &= \begin{pmatrix} (A_2'D_{2,p}'D_{2,p}A_2)^{-1} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp} \end{pmatrix}' \\ \begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp} \end{pmatrix}^{-1} &= \begin{pmatrix} (B_2'D_{2,k}'D_{2,k}B_2)^{-1} & 0 \\ 0 & I_{\frac{1}{2}k(k-1)} \end{pmatrix} \begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp} \end{pmatrix}'. \end{aligned}$$

b. We show that if $\mathcal{R}(\hat{R}) = D_p\hat{R}^*D_k'$ is replaced with

$$\bar{R} := D_p(D_p'D_p)^{-\frac{1}{2}}\hat{R}^*(D_k'D_k)^{-\frac{1}{2}}D_k'$$

in the definition of KPST, one obtains KPST* in (24). To show this, we use SVDs of $\bar{R} = \bar{L}\bar{\Sigma}\bar{N}'$ and $\hat{R}^* = \hat{L}^*\hat{\Sigma}^*\hat{N}^{*\prime}$ which are related through:

$$\begin{aligned} \bar{L} &= \left(D_p(D_p'D_p)^{-\frac{1}{2}}\hat{L}^* \vdots D_{p\perp} \right) \\ \bar{\Sigma} &= \begin{pmatrix} \hat{\Sigma}^* & 0 \\ 0 & 0 \end{pmatrix} \\ \bar{N} &= \left(D_k(D_k'D_k)^{-\frac{1}{2}}\hat{N}^* \vdots D_{k\perp} \right). \end{aligned}$$

To show that KPST using $\bar{R}$, indicated by $\text{KPST}_{\bar{R}}$, equals KPST*, we analyze $\text{KPST}_{\bar{R}}$:

$$\begin{aligned} KPST_{\bar{R}} =& n \times \left[vec(vec(\bar{G}_1)_\perp'\bar{R}vec(\bar{G}_2)_\perp) \right]' \\ & \left[\left(vec(\bar{G}_2)_\perp' \otimes vec(\bar{G}_1)_\perp' \right) \hat{V}_{\bar{R}} \left(vec(\bar{G}_2)_\perp \otimes vec(\bar{G}_1)_\perp \right) \right]^- \\ & \left[vec(vec(\bar{G}_1)_\perp'\bar{R}vec(\bar{G}_2)_\perp) \right], \qquad \hat{V}_{\bar{R}} := \widehat{\text{cov}}\left(vec\left(\bar{R}\right) \right) \\ =& vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix} \right)' \left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{1/2} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right)' \\ & \left\{ \left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{1/2} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right) \right. \\ & \left[\left(\left(D_k(D_k'D_k)^{-\frac{1}{2}}\hat{N}_2^* \vdots D_{k\perp} \right) \otimes \left(D_p(D_p'D_p)^{-\frac{1}{2}}\hat{L}_2^* \vdots D_{p\perp} \right) \right)' \right. \\ & \left(D_k(D_k'D_k)^{-\frac{1}{2}} \otimes D_p(D_p'D_p)^{-\frac{1}{2}} \right) \hat{V}_{\hat{R}^*} \left(D_k(D_k'D_k)^{-\frac{1}{2}} \otimes D_p(D_p'D_p)^{-\frac{1}{2}} \right)' \\ & \left(\left(D_k(D_k'D_k)^{-\frac{1}{2}}\hat{N}_2^* \vdots D_{k\perp} \right) \otimes \left(D_p(D_p'D_p)^{-\frac{1}{2}}\hat{L}_2^* \vdots D_{p\perp} \right) \right) \\ & \left. \left(\bar{N}_{22}^{-1} \left(\bar{N}_{22}\bar{N}_{22}'\right)^{\frac{1}{2}} \otimes \bar{L}_{22}^{-1}(\bar{L}_{22}\bar{L}_{22}')^{1/2} \right) \right\}^{-1} \\ & \left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{\frac{1}{2}'} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right) vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix} \right) \\ =& vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix} \right)' \left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{1/2} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right)' \\ & \left[\left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{1/2} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right) \right. \\ & \left(\begin{pmatrix} \hat{N}_2^* & 0 \end{pmatrix} \otimes \begin{pmatrix} \hat{L}_2^* & 0 \end{pmatrix} \right)' \hat{V}_{\hat{R}^*} \left(\begin{pmatrix} \hat{N}_2^* & 0 \end{pmatrix} \otimes \begin{pmatrix} \hat{L}_2^* & 0 \end{pmatrix} \right) \\ & \left. \left(\bar{N}_{22}^{-1} \left(\bar{N}_{22}\bar{N}_{22}'\right)^{\frac{1}{2}} \otimes \bar{L}_{22}^{-1}(\bar{L}_{22}\bar{L}_{22}')^{1/2} \right) \right]^{-1} \\ & \left(\left(\bar{N}_{22}\bar{N}_{22}'\right)^{\frac{1}{2}'} \bar{N}_{22}'^{-} \otimes (\bar{L}_{22}\bar{L}_{22}')^{1/2}\bar{L}_{22}'^{-} \right) vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix} \right) \end{aligned}$$

$$
\begin{aligned}
&= vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix}\right)' \Big[\left(\begin{pmatrix} \hat{N}_2^* & 0 \end{pmatrix} \otimes \begin{pmatrix} \hat{L}_2^* & 0 \end{pmatrix}\right)' \\
&\quad \hat{V}_{\hat{R}^*} \left(\begin{pmatrix} \hat{N}_2^* & 0 \end{pmatrix} \otimes \begin{pmatrix} \hat{L}_2^* & 0 \end{pmatrix}\right)\Big]^- vec\left(\begin{pmatrix} \hat{\Sigma}_2^* & 0 \\ 0 & 0 \end{pmatrix}\right) \\
&= n \times vec(\hat{\Sigma}_2^*)' \left[\left(\hat{N}_2^{*\prime} \otimes \hat{L}_2^{*\prime}\right) \hat{V}_{\hat{R}^*} \left(\hat{N}_{2^*}^* \otimes \hat{L}_2^*\right)\right]^{-1} vec(\hat{\Sigma}_{2,R^*}) \qquad (51) \\
&= KPST^*,
\end{aligned}
$$

which is the KPST expression using $\hat{R}^*$ so it differs from KPST. Here we used that

$$
\begin{aligned}
\hat{V}_{\bar{R}} &= \left(D_k(D_k'D_k)^{-\frac{1}{2}} \otimes D_p(D_p'D_p)^{-\frac{1}{2}}\right) \hat{V}_{\hat{R}^*} \left(D_k(D_k'D_k)^{-\frac{1}{2}} \otimes D_p(D_p'D_p)^{-\frac{1}{2}}\right)', \\
\bar{L}_2 &= \left(D_p(D_p'D_p)^{-\frac{1}{2}} \hat{L}_{2,R^*} \vdots D_{p\perp}\right), \\
\bar{N}_2 &= \left(D_k(D_k'D_k)^{-\frac{1}{2}} \hat{N}_{2,R^*} \vdots D_{k\perp}\right),
\end{aligned}
$$

and that an expression like (22) can be similarly shown to apply to KPST*.

Because $\hat{R}^*$ has the non-degenerate limiting distribution (14), the limiting distribution of KPST using $\hat{R}^*$ directly results from Kleibergen and Paap (2006, Corollary 1) and is also χ^2_{df}.

c. To show (non-) invariance to orthonormal transformations of $\hat{V}_i$ and Z_i, we consider a $p \times p$ dimensional orthonormal matrix Q using which we rotate $\hat{V}_i$ to become $Q\hat{V}_i$ so

$$
\begin{aligned}
vec(Q\hat{V}_i\hat{V}_i'Q') &= (Q \otimes Q)vec(\hat{V}_i\hat{V}_i') \\
&= (Q \otimes Q)D_p vech(\hat{V}_i\hat{V}_i') \\
vech(Q\hat{V}_i\hat{V}_i'Q') &= (D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p vech(\hat{V}_i\hat{V}_i'),
\end{aligned}
$$

which implies that if we also rotate Z_i by the $k \times k$ orthonormal matrix H:

$$
\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_{i=1}^n vec(Q\hat{V}_i\hat{V}_i'Q')vec(HZ_iZ_i'H')' = \\
&(Q \otimes Q)D_p \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n vec(\hat{V}_i\hat{V}_i')vec(Z_iZ_i')'\right] D_k'(H \otimes H)',
\end{aligned}
$$

and

$$
\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_{i=1}^n vech(Q\hat{V}_i\hat{V}_i'Q')vech(HZ_iZ_i'H')' = \\
&(D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n vech(\hat{V}_i\hat{V}_i')vech(Z_iZ_i')'\right] \\
&\times D_k'(H \otimes H)'D_k(D_k'D_k)^{-1}.
\end{aligned}
$$

Hence, because Q and H are orthonormal, this implies that the different components of the SVD decomposition of $\mathcal{R}(\hat{R})$ in (9) with the transformed $\hat{R}$ become $(Q \otimes Q)\hat{L}$, $\hat{\Sigma}$ and $(H \otimes H)\hat{N}$. Because these rotations also transform the covariance matrix $\hat{V}$ to $(Q \otimes Q)\hat{V}(H \otimes H)'$, it immediately follows from the expression in KPST in (22) that KPST is invariant to rotations of V_i and Z_i.

Let $\hat{R}_T^*$ be the $\hat{R}^*$ in (13) obtained using the transformed data $Q\hat{V}_i$ and HZ_i, whose SVD is

$$
\begin{aligned}
\hat{R}_T^* &= \tfrac{1}{n}\textstyle\sum_{i=1}^n vech(Q\hat{V}_i\hat{V}_i'Q')vech(HZ_iZ_i'H')' \\
&= (D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n vech(\hat{V}_i\hat{V}_i')vech(Z_iZ_i')'\right] \\
&\quad D_k'(H \otimes H)'D_k(D_k'D_k)^{-1} \\
&= (D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p\hat{L}^*\hat{\Sigma}^*\hat{N}^{*\prime}D_k'(H \otimes H)'D_k(D_k'D_k)^{-1} \\
&= \hat{L}_T^*\hat{\Sigma}_T^*\hat{N}_T^{*\prime},
\end{aligned}
$$

with $\hat{L}_T^*$ and $\hat{N}_T^*$ orthonormal $(\frac{1}{2}p(p+1)-1) \times (\frac{1}{2}p(p+1)-1)$ and $(\frac{1}{2}k(k+1)-1) \times (\frac{1}{2}k(k+1)-1)$ dimensional matrices and $\hat{\Sigma}_T^*$ a diagonal $(\frac{1}{2}p(p+1)-1) \times (\frac{1}{2}p(p+1)-1)$ dimensional matrix with the singular values in non-increasing order on the main diagonal. Because $(D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p$ and $(D_k'D_k)^{-1}D_k'(H \otimes H)'D_k$ are not orthonormal, it follows that

$$
\hat{L}_T^* \neq (D_p'D_p)^{-1}D_p'(Q \otimes Q)D_p\hat{L}^* \text{ and } \hat{N}_T^* \neq (D_k'D_k)^{-1}D_k'(H \otimes H)'D_k\hat{N}^*
$$

and therefore also

$$
\hat{\Sigma}_T^* \neq \hat{\Sigma}^*.
$$

When substituting these expressions into the expression of KPST* in (51), it then follows that KPST* is not invariant to orthonormal transformations of the data.

d. Under H_0 and joint limit sequences of k, p and n, we have to consider all components of $vec(\hat{\Lambda})$ in (16) and its covariance matrix estimator.

$$
\begin{aligned}
vec(\hat{\Lambda}) &= \left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right)' vec(\mathcal{R}(\hat{R})) \\
&= \left(\left[vec(G_2)_\perp + vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(G_1)_\perp + vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' \\
&\quad vec\left(\mathcal{R}(R) + \mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) \\
&= (vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec(\mathcal{R}(R)) + \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes I_{p^2-1}\right)' \\
&\quad vec(vec(G_1)_\perp' \mathcal{R}(R)) + (vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(I_{k^2-1} \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec(\mathcal{R}(R) vec(G_2)_\perp) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec(\mathcal{R}(R)) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) \\
&= (vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec(\mathcal{R}(R)) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left[\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right] \\
&= a + b + c
\end{aligned}
$$

for

$$
\begin{aligned}
a &:= (vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right), \\
b &:= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec(\mathcal{R}(R)) + \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right), \\
c &:= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right).
\end{aligned}
$$

In the derivation above, we use that under H_0, $\mathcal{R}(R) = vec(G_1) vec(G_2)'$, see (6). Therefore $vec(G_1)_\perp' \mathcal{R}(R) = 0$, $\mathcal{R}(R) vec(G_2)_\perp = 0$. The limit behavior of KPST results from the limit behavior of a. We specify both $vec(G_1)_\perp$ and $vec(G_2)_\perp$, whose dimensions increase as k and p get larger, as orthonormal matrices, $vec(G_1)_\perp' vec(G_1)_\perp \equiv I_{\frac{1}{2}k(k-1)}$ and $vec(G_2)_\perp' vec(G_2)_\perp \equiv I_{\frac{1}{2}p(p-1)}$. Hence the length of each column of $vec(G_1)_\perp$ and $vec(G_2)_\perp$ equals one and does not change when k and/or p increase.

From (14), it follows that $\mathcal{R}(\hat{R}) - \mathcal{R}(R) = O_p(n^{-\frac{1}{2}})$, and so the same holds for the convergences rates of $vec(\hat{G}_1)_\perp - vec(G_1)_\perp$ and $vec(\hat{G}_2)_\perp - vec(G_2)_\perp$, see Kleibergen and Paap (2006). Because $vec(\hat{G}_1)_\perp$ and $vec(\hat{G}_2)_\perp$ are solved from $\mathcal{R}(\hat{R})$, it follows that $\mathcal{R}(\hat{R}) - \mathcal{R}(R)$, $vec(\hat{G}_1)_\perp - vec(G_1)_\perp$ and $vec(\hat{G}_2)_\perp - vec(G_2)_\perp$ are all jointly dependent. In a limiting sequence

where the dimensions p and k jointly increase with the sample size n, we then have the following convergence rates:

$$\begin{aligned}
&1. && (vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) && = O_p\left(n^{-\frac{1}{2}}\right)\\
&2. && \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) && = O_p\left(\tfrac{k^2}{n}\right)\\
&3. && \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) && = O_p\left(\tfrac{p^2}{n}\right)\\
&4. && \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec(\mathcal{R}(R)) && = O_p\left(\tfrac{(pk)^2}{n}\right)\\
&5. && \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right) && = O_p\left(\tfrac{p^2k^2}{n\sqrt{n}}\right).
\end{aligned}$$

The individual elements of each of the above five components result from multiplying the first KPS matrix with the second vectorized matrix. This multiplication implies that the individual elements equal weighted summations where the number of elements where we sum over increases with the sequence of k and p. This affects the convergence rate of the individual elements. The convergence rate of the individual elements is then a function of the sum of the involved weights and the convergence rates of the multiplied components. Along these lines, we next establish the convergence rate for, say, the qth element of each of the five components in the above expression:

$$\begin{aligned}
&1. \left[(vec(G_2)_\perp \otimes vec(G_1)_\perp)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_q\\
&\quad = \sum_{i=1}^{p^2}\sum_{j=1}^{k^2} \quad [vec(G_2)_\perp]_{jm}\,[vec(G_1)_\perp]_{il}\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i},
\end{aligned}$$

for $m = 1+\lfloor (q-1)/(k^2-1)\rfloor$, $l = q-(p^2-1)(m-1)$, with $\lfloor b\rfloor$ the entier function of a scalar b, which is of order $O_p\left(n^{-\frac{1}{2}}\right)$. This convergence rate follows because $vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)$ is $O_p\left(n^{-\frac{1}{2}}\right)$ and $vec(G_1)_\perp$ and $vec(G_2)_\perp$ are both orthonormal matrices. The sum of the weights $[vec(G_2)_\perp]_{jm}$ and $[vec(G_1)_\perp]_{il}$ $i = 1,\ldots,p^2$, $j = 1,\ldots,k^2$ in the above summation is therefore finite and does not grow with the sequence of k and p. Hence, it does not affect the convergence rate which then results from $\mathcal{R}(\hat{R}) - \mathcal{R}(R) = O_p\left(n^{-\frac{1}{2}}\right)$.

$$\begin{aligned}
&2. \left[\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_q\\
&\quad = \sum_{i=1}^{p^2}\sum_{j=1}^{k^2} \quad \left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_{jm}[vec(G_1)_\perp]_{il}\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i},
\end{aligned}$$

for $m = 1 + \lfloor (q-1)/(k^2-1)\rfloor$, $l = q-(p^2-1)(m-1)$, which is of order $O_p\left(\frac{k^2}{n}\right)$. This order results from the k^2 dependent components $\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_{jm}$ and $\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i}$ that we sum over and that the sum of the weights in the summation is proportional to k^2. Each of the (dependent) components in $\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_{jm}$ and $\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i}$ are $O_p(n^{-\frac{1}{2}})$ so summing over k^2 of them and multiplying through results in $O_p\left(\frac{k^2}{n}\right)$. The additional weights $[vec(G_1)_\perp]_{il}$, $i = 1,\ldots,p^2$, are again such that their sum is finite so it does not grow with the sequence of k and p because $vec(G_1)_\perp$ is orthonormal. Hence, they do not affect the convergence rate.

$$\begin{aligned}
&3. \left[\left([vec(G_2)_\perp] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_q\\
&\quad = \sum_{i=1}^{p^2}\sum_{j=1}^{k^2} \quad [vec(G_2)_\perp]_{jm}\left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]_{il}\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i},
\end{aligned}$$

which is of order $O_p\left(\frac{p^2}{n}\right)$. The argument for this convergence rate is identical to the one for 2.

$$\begin{aligned}
&4. \left[\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(R)\right)\right]_q\\
&\quad = \sum_{i=1}^{p^2}\sum_{j=1}^{k^2} \quad \left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_{jm}\left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]_{il}[vec\left(\mathcal{R}(R)\right)]_{(j-1)k^2+i},
\end{aligned}$$

which is of order $O_p\left(\frac{(pk)^2}{n}\right)$. This order results from the double sum over p^2 random variables in $\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]$ and k^2 random variables in $\left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]$ which are dependent. The sum of the weights is then proportional to $(pk)^2$ and because the convergence rates of $\left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]$ and $\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]$ are both $O_p(n^{-\frac{1}{2}})$, this

then leads to the $O_p\left(\frac{(pk)^2}{n}\right)$ convergence rate.

$$5.\left[\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)' vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_q$$
$$= \sum_{i=1}^{p^2}\sum_{j=1}^{k^2} \left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_{jm}\left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]_{il}\left[vec\left(\mathcal{R}(\hat{R}) - \mathcal{R}(R)\right)\right]_{(j-1)k^2+i},$$

is of order $O_p\left(\frac{(pk)^2}{n\sqrt{n}}\right)$ which follows along the lines of the above results.

For the limit behavior of $\sqrt{n}\hat{\Lambda}$ to just result from 1 (and in consequence, the limit distribution of KPST to remain unaffected) it is then sufficient to have joint limit sequences that satisfy:

$$\frac{(pk)^2}{\sqrt{n}} \to 0.$$

For the estimator of the covariance matrix of $\hat{\Lambda}$, we further have

$$\begin{aligned}
&\left(\left[vec(\hat{G}_2)\right]_\perp' \otimes \left[vec(\hat{G}_1)\right]_\perp'\right)\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R})))\left(\left[vec(\hat{G}_2)\right]_\perp' \otimes \left[vec(\hat{G}_1)\right]_\perp'\right) &=\\
&\left(\left[vec(G_2)_\perp + vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes \left[vec(G_1)_\perp + vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right)'\\
&\left(\text{cov}(vec(\mathcal{R}(\hat{R}))) + \widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \text{cov}(vec(\mathcal{R}(\hat{R})))\right)\\
&\left(\left[vec(G_2)_\perp + vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(G_1)_\perp + vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) &=\\
&\left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\text{cov}(vec(\mathcal{R}(\hat{R})))\left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)' + U &=\\
&A_1 + B_1 + B_2 + B_3 + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + D_1 + \ldots
\end{aligned}$$

where below we show that the maximal convergence rates besides the zero-th order component are $O_p(n^{-\frac{1}{2}})$, $O_p\left(\frac{(pk)^2}{n}\right)$, $O_p\left(\frac{k^4}{n}\right)$, $O_p\left(\frac{p^4}{n}\right)$ and $O_p\left(\frac{k^4p^4}{n\sqrt{n}}\right)$. All these rates appear in an identical manner in the inverse of the estimator of the covariance matrix.[8] When taking the resulting inverse and accounting for the summations over the k^2p^2 components in $vec(\hat{\Lambda})$, we obtain a slightly stronger condition than just for $\hat{\Lambda}$:

$$\frac{(pk)^{16}}{n^3} \to 0,$$

which results from the $O_p(n^{-\frac{1}{2}})$ components from the inverse of the covariance matrix estimator paired with the $O_p\left(\frac{(pk)^2}{n}\right)$ components from $\hat{\Lambda}$ corrected for the multiplication by n and the double summation over p^2k^2 components.[9] The rate that would result from $\hat{\Lambda}$ is $\frac{(pk)^{12}}{n^3} \to 0$. The convergence rate is in between the rate implied by Newey and Windmeijer (2009) which would be $\frac{k^4p^4}{n}$ for $\hat{\Lambda}$ and $\frac{k^6p^6}{n}$ for convergence of the test statistic which is slightly stricter than our rate of $\frac{(pk)^{16}}{n^3} \to 0$.

Below, we state the rates of the different A, B, C and D (third order error) components where we only provide the rate for one of the D components because we just showed that they do not lead to the largest error rate because the $O_p\left(\frac{k^4p^4}{n\sqrt{n}}\right)$ is less than the $O_p\left(\frac{p^2k^2}{n}\right)$ that results from some of the C components.

$$\begin{aligned}
A_1 &= \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\text{cov}(vec(\mathcal{R}(\hat{R})))\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) &&= O(1)\\
B_1 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes vec(G_1)_\perp'\right)\text{cov}(vec(\mathcal{R}(\hat{R})))\\
&\quad \left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) + \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\\
&\quad \text{cov}(vec(\mathcal{R}(\hat{R})))\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right) &&= O_p(n^{-\frac{1}{2}})\\
B_2 &= \left(vec(G_2)_\perp' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right)\text{cov}(vec(\mathcal{R}(\hat{R})))\\
&\quad \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)' + \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\\
&\quad \text{cov}(vec(\mathcal{R}(\hat{R})))\left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) &&= O_p(n^{-\frac{1}{2}})
\end{aligned}$$

[8] To show this, one can use the Woodbury matrix identity which implies that for invertible $m \times m$ matrices H and G, with $H + G$ also invertible: $(H+G)^{-1} = H^{-1} - H^{-1}(G^{-1} + H^{-1})^{-1}H^{-1}$.

[9] All combined we get: $O_p\left(n\left(\frac{p^2k^2}{n}\right)\left(\frac{p^2k^2}{n}\right)\frac{1}{\sqrt{n}}(k^2p^2)^2\right) = O_p\left(\left(\frac{(p^2k^2)^4}{n\sqrt{n}}\right)\right) = O_p\left(\frac{(pk)^{16}}{N^3}\right)$.

$$
\begin{aligned}
B_3 &= \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\left[\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \text{cov}(vec(\mathcal{R}(\hat{R})))\right] \\
&\quad \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)' && = O_p(n^{-\frac{1}{2}}) \\
C_1 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right) \\
&\quad \text{cov}(vec(\mathcal{R}(\hat{R})))\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) + \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right) \\
&\quad \text{cov}(vec(\mathcal{R}(\hat{R})))\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) && = O_p\left(\frac{(pk)^2}{n}\right) \\
C_2 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes vec(G_1)_\perp'\right)\text{cov}(vec(\mathcal{R}(\hat{R}))) \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right) && = O_p\left(\frac{k^4}{n}\right) \\
C_3 &= \left(vec(G_2)_\perp' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right)\text{cov}(vec(\mathcal{R}(\hat{R}))) \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) && = O_p\left(\frac{p^4}{n}\right) \\
C_4 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes vec(G_1)_\perp'\right)\text{cov}(vec(\mathcal{R}(\hat{R}))) \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) + \\
&\quad \left(vec(G_2)_\perp' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right)\text{cov}(vec(\mathcal{R}(\hat{R}))) \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right) && = O_p\left(\frac{p^2k^2}{n}\right) \\
C_5 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes vec(G_1)_\perp'\right)\left[\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \right. \\
&\quad \left.\text{cov}(vec(\mathcal{R}(\hat{R})))\right]\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) + \\
&\quad \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\left[\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \text{cov}(vec(\mathcal{R}(\hat{R})))\right] \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right) && = O_p\left(\frac{p^2k^2}{n}\right) \\
C_6 &= \left(vec(G_2)_\perp' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right)\left[\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \right. \\
&\quad \left.\text{cov}(vec(\mathcal{R}(\hat{R})))\right]\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) + \\
&\quad \left(vec(G_2)_\perp' \otimes vec(G_1)_\perp'\right)\left[\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R}))) - \text{cov}(vec(\mathcal{R}(\hat{R})))\right] \\
&\quad \left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) && = O_p\left(\frac{p^2k^2}{n}\right) \\
D_1 &= \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right) \\
&\quad \text{cov}(vec(\mathcal{R}(\hat{R})))'\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_\perp \otimes vec(G_1)_\perp\right) + \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]_\perp \otimes vec(G_1)_\perp\right)'\text{cov}(vec(\mathcal{R}(\hat{R}))) \\
&\quad \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) && = O_p\left(\frac{k^4p^4}{n\sqrt{n}}\right)
\end{aligned}
$$

Proof of Theorem 3. Note first that for $i = 1, 2$,

$$n^{1/2}(\bar{G}_{i,n} - G_i) \to g_i \tag{52}$$

as $n \to \infty$ for certain matrices g_i. This can be shown as follows. Consider $M(t) = vec(G_1)vec(G_2)' + t\mathcal{R}(A_0)$ for $t \in R$. For $|t|$ small enough, $\sigma_1(M(t))$, the largest singular value of $M(t)$, is simple and obviously $M(t)$ depends differentiably on t. By Theorems 7 and 8 in Lax (2007) it follows that for small $|t|$, $\sigma_1(M(t))$ depends differentiably on t and that there exists left and right eigenvectors corresponding to $\sigma_1(M(t))$ that depend differentiably on t.[10] From (11) we know that $vec(\bar{G}_{1,n})$ equals the normalized left eigenvector, $L_1(n^{-1/2})$ say, corresponding to $\sigma_1(M(n^{-1/2}))$. Given that $L_1(t)$ is differentiable at $t = 0$ it follows that $n^{1/2}(L_1(n^{-1/2}) - L_1(0))$ converges to some vector $vec(g_1) \in \mathbb{R}^{p^2}$. But this proves the claim for $\bar{G}_{1,n}$. The proof for $\bar{G}_{2,n}$ is identical. Intuitively, $(vec(G_1) + O(n^{-a}))(vec(G_2) + O(n^{-b}))' = vec(G_1)vec(G_2)' + O(n^{-1/2})$ implies that $a = b \geq 1/2$, see Kleibergen and Paap (2006).

Second, note that $vec(\bar{G}_{i,n})_\perp$ for $i = 1, 2$ can be specified such that

$$vec(\bar{G}_{i,n})_\perp - vec(G_i)_\perp = O(n^{-1/2}). \tag{53}$$

Namely, let $vec(G_1)_\perp = (v_2, \ldots, v_{p^2})$. Clearly $(vec(\bar{G}_{1,n}), v_2, \ldots, v_{p^2})$ will be of full rank for all n large enough and $vec(\bar{G}_{1,n})_\perp$ can be obtained as the last $p^2 - 1$ vectors by performing Gram–Schmidt orthogonalization to $(vec(\bar{G}_{1,n}), v_2, \ldots, v_{p^2})$. E.g., the first column of $vec(\bar{G}_{1,n})_\perp$ (before normalizing its length to one) then equals

$$\begin{aligned}
& v_2 - [vec(\bar{G}_{1,n})'v_2]vec(\bar{G}_{1,n})/\|vec(\bar{G}_{1,n})\| \\
&= v_2 - [(vec(G_1) + n^{-1/2}g_1 + o(n^{-1/2}))'v_2]vec(\bar{G}_{1,n})/\|vec(\bar{G}_{1,n})\| \\
&= v_2 - [(n^{-1/2}g_1 + o(n^{-1/2}))'v_2]vec(\bar{G}_{1,n})/\|vec(\bar{G}_{1,n})\| \\
&= v_2 + O(n^{-1/2}),
\end{aligned}$$

where in the first equality we use (52) and in the second equality $vec(G_1)'v_2 = 0$. Continuing further with Gram–Schmidt orthogonalization with the other columns of $vec(\bar{G}_{1,n})_\perp$ yields the desired result (53). The result for $vec(G_2)_\perp$ is established analogously.

Next, by the definition of Λ_n in (34), we have

$$\lim_{n\to\infty} \sqrt{n}\Lambda_n = \lim_{n\to\infty} \sqrt{n}vec(\bar{G}_{1,n})'_\perp \left(vec(G_1)vec(G_2)' + \frac{1}{\sqrt{n}}\mathcal{R}(A_0) \right) vec(\bar{G}_{2,n})_\perp = a_0,$$

where for the last equality we use (53). Namely,

$$\begin{aligned}
& \sqrt{n}vec(\bar{G}_{1,n})'_\perp vec(G_1)vec(G_2)'vec(\bar{G}_{2,n})_\perp \\
&= \sqrt{n}vec(G_1)_\perp + O(1))'vec(G_1)vec(G_2)'(vec(G_2)_\perp + O(n^{-1/2}) \\
&= O(1)vec(G_1)vec(G_2)'O(n^{-1/2}) \\
&= O(n^{-1/2}).
\end{aligned}$$

Recall that KPST is defined as a quadratic form in $\sqrt{n}vec(\hat{\Lambda})$. To derive the limiting distribution of the latter quantity, note first that its asymptotic variance is given by

$$\begin{aligned}
V_\Lambda \coloneqq & \lim_{n\to\infty} \left[vec(\bar{G}_{2,n})'_\perp \otimes vec(\bar{G}_{1,n})'_\perp \right] cov\left(\mathcal{R}\left(\hat{R}\right) \right) \left[vec(\bar{G}_{2,n})_\perp \otimes vec(\bar{G}_{1,n})_\perp \right] \\
= & \left(vec(G_2)'_\perp \otimes vec(G_1)'_\perp \right) (D_k \otimes D_p) V_{R^*} (D_k \otimes D_p)' \left(vec(G_2)_\perp \otimes vec(G_1)_\perp \right).
\end{aligned}$$

The limiting distribution of $\sqrt{n}vec(\hat{\Lambda})$ under local to KPS alternatives can be derived along the same lines as under KPS, which was done in the proof of Theorem 2a, see (49). It follows that $\sqrt{n}vec(\hat{\Lambda}) \to_d N(a_0, V_\Lambda)$ and thus $KPST \to_d \chi^2_{df}(\delta)$ as claimed.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2022.01.005.

References

Acemoglu, D., Cantoni, D., Johnson, S., Robinson, J.A., 2011. The consequences of radical reform: the french revolution. Amer. Econ. Rev. 101 (7), 3286–3307.
Acemoglu, D., Johnson, S., 2005. Unbundling institutions. J. Polit. Econ. 113 (5), 949–995.
Acemoglu, D., Johnson, S., Robinson, J.A., Yared, P., 2008. Income and democracy. Amer. Econ. Rev. 98 (3), 808–842.
Alesina, A., Giuliano, P., Nunn, N., 2013. On the origins of gender roles: women and the plough. Q. J. Econ. 128 (2), 469–530.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Andrews, D.W.K., 2017. Identification-robust subvector inference. Cowles Foundation for Research in Economics, Yale University.
Autor, D.H., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the US labor market. Amer. Econ. Rev. 103 (5), 1553–1597.

[10] The results in Lax (2007) are formulated for square matrices and spectral decompositions but immediately translate to rectangular matrices M and their SVDs by considering $M'M$.

Autor, D.H., Dorn, D., Hanson, G.H., 2013. The China syndrome: local labor market effects of import competition in the United States. Amer. Econ. Rev. 103 (6), 2121–2168.
Chen, Q., Fang, Z., 2019. Improved inference on the rank of a matrix. Quant. Econ. 10 (4), 1787–1824.
Donald, S.G., Fortuna, N., Pipiras, V., 2007. On rank estimation in suymmetric matrices: The case of indefinite matrix estimators. J. Econometrics 23 (6), 1217–1232.
Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73 (4), 1351–1365.
Duranton, G., Turner, M.A., 2011. The fundamental law of road congestion: evidence from US cities. Amer. Econ. Rev. 101 (6), 2616–2652.
Genton, M.G., 2007. Separable approximatioms of space-time covariance matrices. Environmetrics 18, 681–695.
Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2019. A more powerful subvector Anderson Rubin test in linear instrumental variable regression. Quant. Econ. 10 (2), 487–526.
Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2021. A powerful subvector Anderson Rubin test in linear instrumental variables regression with conditional heteroskedasticity. arXiv preprint arXiv:2103.11371.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.
Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. J. Bus. Econom. Statist. 14, 262–280.
Hansford, T.G., Gomez, B.T., 2010. Estimating the electoral effects of voter turnout. Am. Political Sci. Rev. 104 (2), 268–288.
Johnson, D.S., Parker, J.A., Souleles, N.S., 2006. Household expenditure and the income tax rebates of 2001. Amer. Econ. Rev. 96 (5), 1589–1610.
Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. J. Finance 54 (1), 203–235.
Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73 (4), 1103–1123.
Kleibergen, F., 2009. Tests of risk premia in linear factor models. J. Econometrics 149 (2), 149–173.
Kleibergen, F., 2021. Efficient size correct subset inference in homoskedastic linear instrumental variables regression. J. Econometrics 221 (1), 78–96.
Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. J. Econometrics 133 (1), 97–126.
Kleibergen, F., Zhan, Z., 2020. Robust inference for consumption-based asset pricing. J. Finance 75 (1), 507–550.
Lax, P.D., 2007. Linear Algebra and Its Applications, second ed. Wiley-Interscience.
Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. Ann. Statist. 40, 1024–1060.
Ledoit, O., Wolf, M., 2015. Spectrum estimation: a unified approach for covariance estimation and PCA in large dimensions. J. Multivariate Anal. 139, 360–384.
Ledoit, O., Wolf, M., 2018. Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli 24, 3791–3832.
Lu, N., Zimmermann, D.L., 2005. The likelihood ratio test for a separable covariance matrix. Statist. Probab. Lett. 73, 449–457.
Miguel, E., Satyanath, S., Sergenti, E., 2004. Economic shocks and civil conflict: an instrumental variables approach. J. Polit. Econ. 112 (4), 725–753.
Mitchell, M., Genton, M., Gumpertz, M., 2006. A likelihood ratio test for separability of ccovariance. J. Multivariate Anal. 97, 1025–1043.
Newey, W.K., West, K.D., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Newey, W., Windmeijer, F., 2009. GMM with many weak moment conditions. Econometrica 77 (3), 687–719.
Nunn, N., 2008. The long-term effects of Africa's slave trades. Q. J. Econ. 123 (1), 139–176.
Parker, J.A., Souleles, N.S., Johnson, D.S., McClelland, R., 2013. Consumer spending and the economic stimulus payments of 2008. Amer. Econ. Rev. 103 (6), 2530–2553.
Tanaka, T., Camerer, C.F., Nguyen, Q., 2010. Risk and time preferences: linking experimental and household survey data from Vietnam. Amer. Econ. Rev. 100 (1), 557–571.
Van Loan, C., Pitsianis, N., 1993. Approximation with kronecker products. In: Linear Algebra for Large Scale and Real-Time Applications. In: NATO Adv. Sci. Inst. Ser. E Appl. Sci. 232, Kluwer Academic Publishers, pp. 293–314.
Velu, R., Herman, K., 2017. Separable covariance matrices and kronecker approximations. Procedia Comput. Sci. 108, 1019–1029.
Voors, M.J., Nillesen, E.E., Verwimp, P., Bulte, E.H., Lensink, R., Van Soest, D.P., 2012. Violent conflict and behavior: a field experiment in burundi. Amer. Econ. Rev. 102 (2), 941–964.
Werner, K., Jansson, M., Stoica, P., 2008. On estimation of covariance matrices with kronecker product structure. IEEE Trans. Signal Process. 56, 478–491.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.
White, H., 1984. Asymptotic Theory for Econometricians. Academic Press.
Yogo, M., 2004. Estimating the elasticity of intertemporal substitution when instruments are weak. Rev. Econ. Stat. 86 (3), 797–810.