



Department of Economics Discussion Paper Series

A Test for Kronecker Product Structure Covariance Matrix

Patrik Guggenberger, Frank Kleibergen, Sophocles Mavroeidis

Number 962
January, 2022

A Test for Kronecker Product Structure Covariance Matrix*

Patrik Guggenberger
Department of Economics
Pennsylvania State University

Frank Kleibergen
Amsterdam School of Economics
University of Amsterdam

Sophocles Mavroeidis
Department of Economics
University of Oxford

First Version: October, 2019

Revised: January 31, 2022

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.

Keywords: covariance matrix, heteroskedasticity, invariance, Kronecker product structure, linear instrumental variables regression model, reduced rank, weak identification

JEL codes: C12, C26

*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, Associate Editor and two anonymous referees for helpful comments and Lewis McLean for research assistance.

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 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) below, and our procedure adapts the

¹Dufour 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.

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) below). 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

²Another 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.

the improved subvector Anderson-Rubin test from Guggenberger et al. (2019) when the test in the first step does not reject and using the size correct AR\AR test procedure 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 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, a KPS structure 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 and derive the asymptotic null distribution of the test statistic, which we denote as KPST. 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 , $\text{vec}(A) := (a'_1 \dots a'_k)' \in \mathbb{R}^{mk}$ 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 $\text{vech}(A)$ that vectorizes only the lower triangular part of A :

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

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

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

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

$$f_i := (V_i \otimes Z_i) \quad (2)$$

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 covariance matrix of the sample covariance between these errors and the regressors.

It follows that the covariance matrix has a block structure

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

where $R_{jl} \in \mathbb{R}^{k \times k}$, $j, l = 1, \dots, p$. Because $R_{jl} = E \left(\frac{1}{n} \sum_{i=1}^n V_{ij} V_{il} Z_i Z_i' \right) = R'_{lj}$, 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:

$$H_0 : R = G_1 \otimes G_2 \quad (4)$$

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 $p = 1$ or $k = 1$ the null is always true and from

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

⁴Normalizing $G_{1,11}$ 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.

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

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

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

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

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

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

with $\|\cdot\|_F$ the Frobenius or trace norm of a matrix, $\|A\|_F^2 := \text{tr}(A'A) = \text{vec}(A)' \text{vec}(A)$, 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

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

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 $i = 1, \dots, n$, as $n \rightarrow \infty$. Define the distance from a KPS covariance matrix by the Frobenius norm

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

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 $\mathcal{R}(\hat{R})$:

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

where $\hat{\Sigma} := \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_{\min(p^2, k^2)})$ denotes a $p^2 \times k^2$ dimensional diagonal matrix with the singular values $\hat{\sigma}_j$ ($j = 1, \dots, \min(p^2, k^2)$) 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

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

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 $\hat{\sigma}_1 \geq \dots \geq \hat{\sigma}_{\min(p^2, k^2)}$ 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:*

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

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 4 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 :

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

with

$$\hat{R}^* := \frac{1}{n} \sum_{i=1}^n \text{vech}(\hat{V}_i \hat{V}_i') \text{vech}(Z_i Z_i')'. \quad (13)$$

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 $\text{vec}(\hat{R}^*)$ satisfies a central limit theorem:

$$\sqrt{n}(\text{vec}(\hat{R}^*) - \text{vec}(R^*)) \xrightarrow{d} \psi = \text{vec}(\Psi), \quad (14)$$

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

$$\begin{aligned} R^* &:= E \left(\frac{1}{n} \sum_{i=1}^n \text{vech}(V_i V_i') \text{vech}(Z_i Z_i')' \right), \\ V_{R^*} &:= \lim_{n \rightarrow \infty} \left[E \left(\frac{1}{n} \sum_{i=1}^n (\text{vech}(Z_i Z_i') \text{vech}(Z_i Z_i')' \otimes \text{vech}(V_i V_i') \text{vech}(V_i V_i')') \right) \right. \\ &\quad \left. - E \left(\text{vec}(\frac{1}{n} \sum_{i=1}^n \text{vech}(V_i V_i') \text{vech}(Z_i Z_i')') \right) E \left(\text{vec}(\frac{1}{n} \sum_{i=1}^n \text{vech}(V_i V_i') \text{vech}(Z_i Z_i')')' \right)' \right]. \end{aligned} \quad (15)$$

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-identical distributed independent random variables holds under mild conditions, e.g. under the Liapounov's or Lindeberg's condition, see White (1984).

Define

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

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

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

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

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

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

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

where $\text{vec}(G_1)_{\perp}$ and $\text{vec}(G_2)_{\perp}$ are $p^2 \times (p^2 - 1)$ and $k^2 \times (k^2 - 1)$ dimensional matrices that contain the orthogonal complements of $\text{vec}(G_1)$ and $\text{vec}(G_2)$, $\text{vec}(G_1)'_{\perp} \text{vec}(G_1) \equiv 0$, $\text{vec}(G_1)'_{\perp} \text{vec}(G_1)_{\perp} \equiv I_{p^2-1}$, $\text{vec}(G_2)'_{\perp} \text{vec}(G_2) \equiv 0$, $\text{vec}(G_2)'_{\perp} \text{vec}(G_2)_{\perp} \equiv I_{k^2-1}$. The KPST test uses the sample analog of the last component in (18) to test H_0 . It further results from identifying $\text{vec}(G_1)$ and $\text{vec}(G_2)$ 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 $\text{vec}(\mathcal{R}(\hat{R}))$ is singular, the (degenerate) asymptotic normal distribution of $\text{vec}(\hat{\Lambda})$ 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

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

where

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

and

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

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

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

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 $\text{vec}(\hat{\Lambda})$ 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 $\text{vec}(G_1) = D_p \text{vech}(G_1)$, $\text{vec}(G_2) = D_k \text{vech}(G_2)$, the hypothesis of interest (18) can also be specified as:

$$H_0 : R^* = \text{vech}(G_1) \text{vech}(G_2)' \text{ or } H_0 : \text{vech}(G_1)'_\perp R^* \text{vech}(G_2)_\perp = 0, \quad (23)$$

where $\text{vech}(G_1)_\perp$ and $\text{vech}(G_2)_\perp$ 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 $\text{vech}(G_1)$ and $\text{vech}(G_2)$, $\text{vech}(G_1)'_\perp \text{vech}(G_1) = 0$, $\text{vech}(G_1)'_\perp \text{vech}(G_1)_\perp = I_{\frac{1}{2}p(p+1)-1}$, $\text{vech}(G_2)'_\perp \text{vech}(G_2) = 0$, $\text{vech}(G_2)'_\perp \text{vech}(G_2)_\perp = I_{\frac{1}{2}k(k+1)-1}$. 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. Therefore, the corresponding specification of $\text{vec}(\hat{\Lambda})$ converges to a Normally distributed random vector. The specification of null hypothesis in (23) allows us to easily infer the number of restrictions tested, which equals $(\frac{1}{2}k(k+1) - 1)(\frac{1}{2}p(p+1) - 1)$, but the resulting rank statistic does not equal KPST in (19). Specifically, define the SVD of $\hat{R}^* = \hat{L}^* \hat{\Sigma}^* \hat{N}^{*'}$, where

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

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

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

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

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 $i = 1, \dots, n$, as $n \rightarrow \infty$, 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.

$$KPST \xrightarrow{d} \chi_{df}^2$$

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

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

b.

$$KPST^* \xrightarrow{d} \chi_{df}^2$$

as $n \rightarrow \infty$ (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

$$\frac{(pk)^{16}}{n^3} \rightarrow 0, \quad (26)$$

we have

$$\lim_{n,p,k \rightarrow \infty} \Pr \left[KPST < \chi_{df,1-\alpha}^2 \right] \leq \alpha, \quad (27)$$

where $\chi_{df,1-\alpha}^2$ denotes the $1 - \alpha$ quantile of a χ_{df}^2 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

$$KPST > \chi_{df,1-\alpha}^2. \quad (28)$$

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 $(pk)^6/n \rightarrow 0$ 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

$$\hat{f}_i = C_1' \hat{V}_i \otimes C_2' Z_i, \quad (29)$$

where C_1 and C_2 are the Choleski factors of the inverse of the second moments of $\hat{V}_i$ and Z_i , i.e., $C_1 C_1' = \left(\frac{1}{n} \sum_{i=1}^n \hat{V}_i \hat{V}_i' \right)^{-1}$ and $C_2 C_2' = \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \right)^{-1}$. 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 Choleski factor of $\left(\frac{1}{n} \sum_{i=1}^n AZ_i Z_i' A' \right)^{-1}$. 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 $C_{2A}' AZ_i$ is given by $T_A := C_{2A}' A C_2^{-1'}$, i.e., $C_{2A}' AZ_{A,i} = C_{2A}' AZ_i = T_A (C_2' Z_i)$. Now, observe that T_A is an orthonormal matrix, because $T_A' T_A = C_2^{-1} A' C_{2A}' C_{2A}' A C_2^{-1'}$

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

$= C_2^{-1} A' \left(\frac{1}{n} \sum_{i=1}^n A Z_i Z_i' A' \right)^{-1} A C_2^{-1'} = C_2^{-1} \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \right)^{-1} C_2^{-1'} = I_k$. 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 $\sum_{i=1}^n N_i$:

$$f_i = \sum_{j=1}^{N_i} f_{ij}, \quad (30)$$

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

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

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:

$$H_1 : R = (G_1 \otimes G_2) + \frac{1}{\sqrt{n}} A_0, \quad (32)$$

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 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), $\bar{G}_{1,n} \rightarrow G_1$ and $\bar{G}_{2,n} \rightarrow G_2$. 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:

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

with

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

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,

$$KPST \xrightarrow{d} \chi_{df}^2(\delta)$$

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

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

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

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

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:

$$Y_i = Z_i' \Pi + V_i, \quad i = 1, \dots, n, \quad (36)$$

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 $N(0, h(Z_i) I_p)$ 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 $h(Z_i) = \|Z_i\|^2 / k$. 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 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, Figures 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,

<i>Data Generating Process:</i>					homoskedastic			scalar hetero		
p	k	n	a	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 1: Rejection frequencies (in percentages) of KPST test at various significance levels. χ_{df}^2 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.

but just to $(pk)^4$, which is for $p = 2$, $k = 7$ at the bottom right hand side of Figure 1, equal to approximately 40,000, and for $p = 5$, $k = 4$ at the bottom right hand side of Figure 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. Figures 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 Figure 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 generated 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:

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

⁶When 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.

<i>Data Generating Process:</i>					homoskedastic			scalar hetero		
p	k	n	a	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: Rejection frequencies (in percentages) of KPST test at various significance levels. χ_{df}^2 critical values. $n = (pk)^4$, df : number of restrictions given in eq. (25), m : number of estimated parameters. Computed using 40,000 MC replications.

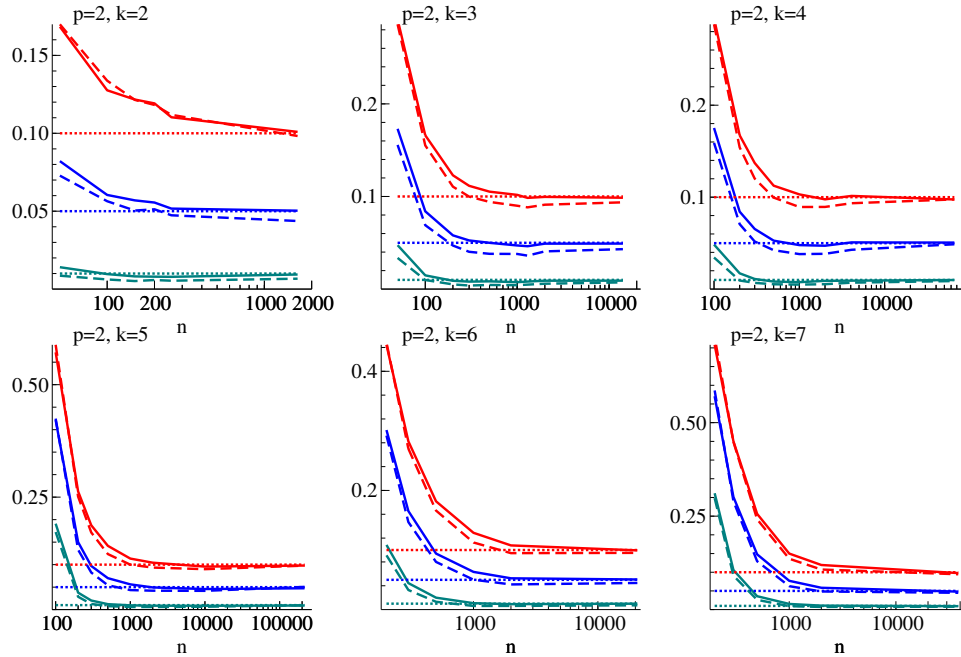


Figure 1: Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications.

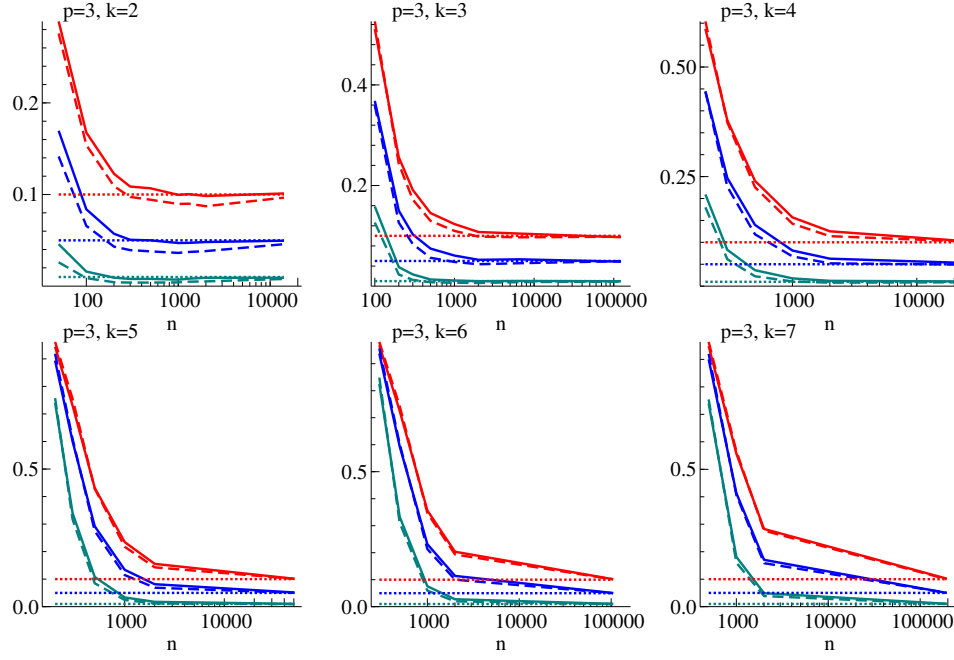


Figure 2: Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications.

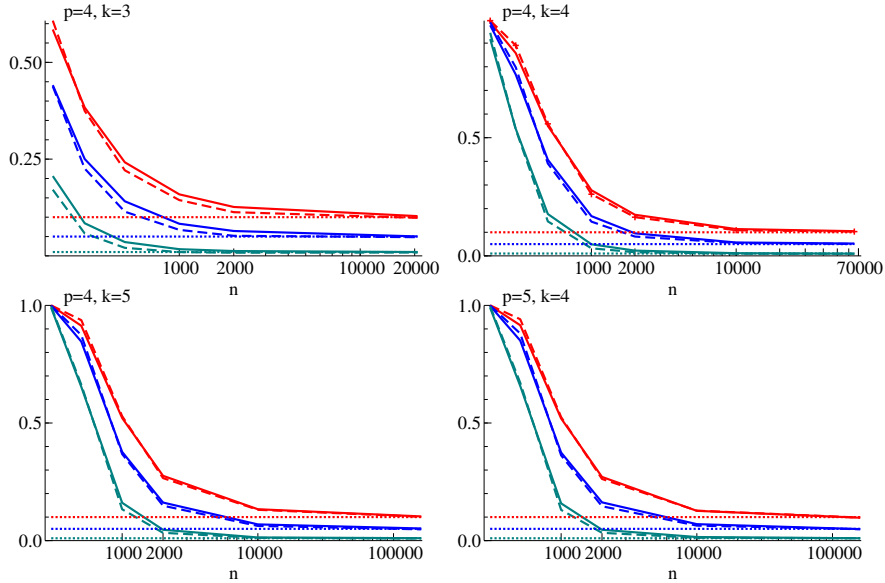


Figure 3: Null rejection probabilities of KPST test as a function of sample size n at different significance levels: 10% (red), 5% (blue) and 1% (green); and different data generating processes: homoskedastic (solid) and scalar heteroskedastic (dashed). Computed using 40,000 MC replications.

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

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

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

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

and $G_1 = G_2 = I_2$. Because

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

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

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

where

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

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

$$\begin{aligned} &(e_1 \otimes e_1)' \left([\text{vec}(G_2)]'_\perp \otimes [\text{vec}(G_1)]'_\perp \right) \\ &\text{cov} \left(\text{vec} \left(\mathcal{R}(\hat{R}) \right) \right) \left([\text{vec}(G_2)]_\perp \otimes [\text{vec}(G_1)]_\perp \right) \Big|^{-1} (e_1 \otimes e_1) = \frac{1}{4}, \end{aligned} \quad (43)$$

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

$$\delta = \frac{1}{4} \sigma^2. \quad (44)$$

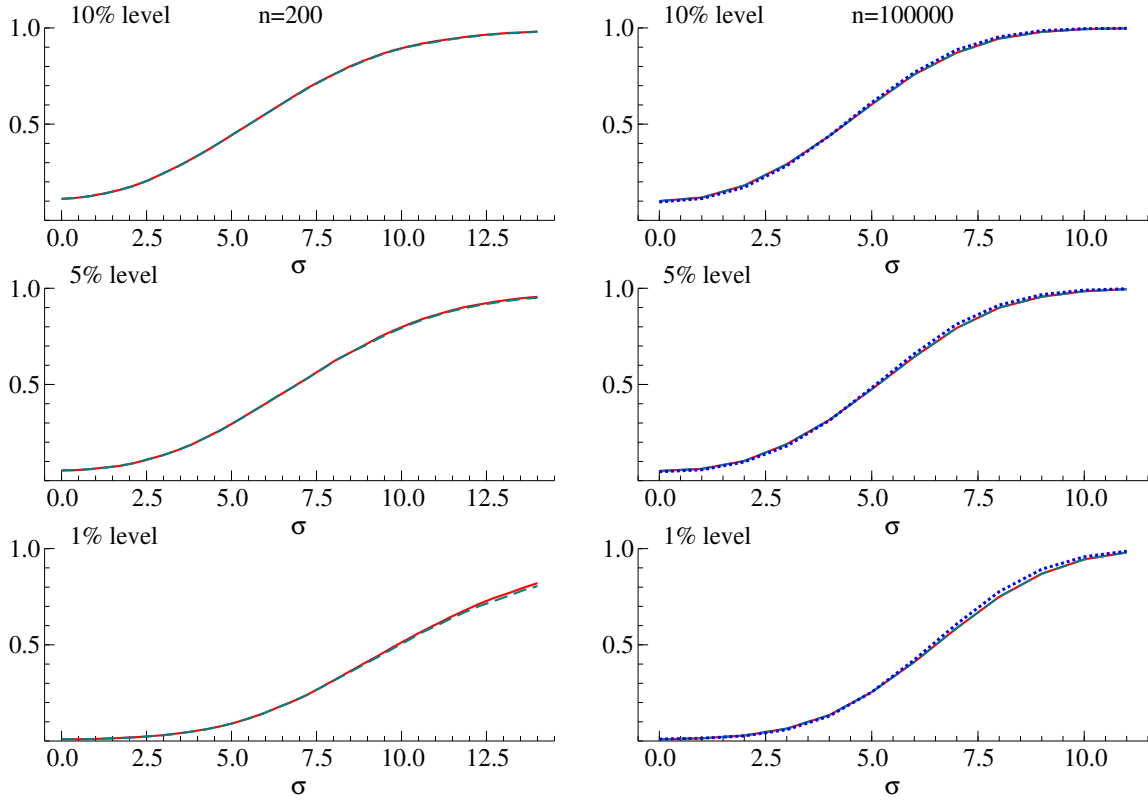


Figure 4: Power of KPST (solid red) and KPST* (dashed green) tests with sample size $n = 200$ (left) and $n = 10^5$ (right). Asymptotic approximation from Theorem 3 (dotted blue) superimposed on the right. σ measures deviation from KPS in Frobenius norm. Computed using 10,000 Monte Carlo replications.

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.

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 Figure 4. The left-hand-side graphs in Figure 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 Figure 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.

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

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.

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.

Summarising, 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.

⁷Both 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.

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

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

A Proofs

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

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

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

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

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

exist that are also symmetric and positive definite. Because $\hat{R}$ is symmetric by construction and positive definite by assumption, (46) and $\|\hat{R} - G_1 \otimes G_2\|_F = \|\mathcal{R}(\hat{R}) - \text{vec}(G_1)\text{vec}(G_2)'\|_F$ (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 $\|\mathcal{R}(\hat{R}) - \text{vec}(G_1)\text{vec}(G_2)'\|_F^2$ over $G_1 \in \mathbb{R}^{p \times p}, G_2 \in \mathbb{R}^{k \times k}$. Therefore, because the rank of $\text{vec}(\hat{G}_1)\text{vec}(\hat{G}_2)'$ is one, $\hat{B} := \text{vec}(\hat{G}_1)\text{vec}(\hat{G}_2)'$ 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 $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_{\min\{p^2, k^2\}} \geq 0$ 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 $\text{vec}(\hat{G}_1)\text{vec}(\hat{G}_2)' = \hat{\sigma}_1 \hat{L}_1 \hat{N}_1'$. There are many different ways we can define the argmins such that $\text{vec}(\hat{G}_1)$ and $\text{vec}(\hat{G}_2)$ 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 $\text{vec}(\hat{G}_1) = \hat{L}_1 / \hat{L}_{11}$, so that $\text{vec}(\hat{G}_2) = \hat{L}_{11} \hat{\sigma}_1 \hat{N}_1$, 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 Equation (22). Because $\hat{L}_{22}^{-1}(\hat{L}_{22}\hat{L}_{22}')^{1/2}$ and $\hat{N}_{22}^{-1}(\hat{N}_{22}\hat{N}_{22}')^{1/2}$ are invertible which

follows from expressions stated in the proof of Theorem 2a, KPST can be rewritten as:

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

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

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

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 “ $\hat{\cdot}$ ”) 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

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

Next, we use that $\text{vec}(\hat{G}_1)_{\perp} = \text{vec}(G_1)_{\perp} + O_p(n^{-\frac{1}{2}})$, $\text{vec}(\hat{G}_2)_{\perp} = \text{vec}(G_2)_{\perp} + O_p(n^{-\frac{1}{2}})$ (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 $D_p R^* D'_k = \text{vec}(G_1) \text{vec}(G_2)'$ which holds under H_0 . Thus, under H_0

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

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

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

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,

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

where we use that $D_p = (e_{1, \frac{1}{2}p(p+1)} : D'_{2,p})'$, for $D_{2,p} : (p^2 - 1) \times \frac{1}{2}p(p+1)$ and $D_k = (e_{1, \frac{1}{2}k(k+1)} : D'_{2,k})'$, 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 $D_{p\perp} = (0 : D'_{2,p\perp})'$, 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 $D_{k\perp} = (0 : D'_{2,k\perp})'$, 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 $B'_2 D'_{2,k} D_{2,k} B_2 : (\frac{1}{2}k(k+1) - 1) \times (\frac{1}{2}k(k+1) - 1)$:

$$\begin{aligned}A'_2 D'_{2,p} D_{2,p} A_2 &= L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 L'_{D_{2p} A_2} \\ B'_2 D'_{2,k} D_{2,k} B_2 &= L_{D_{2k} B_2} \Lambda_{D_{2,k} B_2}^2 L'_{D_{2k} B_2},\end{aligned}$$

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 $\Lambda_{D_{2,p} A_2}^2$ and $\Lambda_{D_{2,k} B_2}^2$ 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 $L = (D_p A : D_{p\perp})$ which is orthonormal so $L' L = L L' = I_{p^2}$. 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 can not 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:

$$\begin{aligned}
L_{22} &= \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \\
&= \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 L'_{D_{2p} A_2})^{-\frac{1}{2}} L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2} L'_{D_{2,p} A_2} & D_{2,p\perp} \end{pmatrix} \\
&= \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 L'_{D_{2p} A_2})^{-\frac{1}{2}} L_{D_{2p} A_2} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \Lambda_{D_{2,p} A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\
&\quad \begin{pmatrix} L'_{D_{2,p} A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix}, \\
(L_{22} L'_{22})^{\frac{1}{2}} &= \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 L'_{D_{2p} A_2})^{-\frac{1}{2}} L_{D_{2p} A_2} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \Lambda_{D_{2,p} A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\
&\quad \begin{pmatrix} D_{2,p} A_2 (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 L'_{D_{2p} A_2})^{-\frac{1}{2}} L_{D_{2p} A_2} & D_{2,p\perp} \end{pmatrix}', \\
L_{22}^{-1} (L_{22} L'_{22})^{1/2} &= \begin{pmatrix} L_{D_{2p} A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \begin{pmatrix} L'_{D_{2p} A_2} (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 P'_{D_{2p} A_2})^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} \\
&= \begin{pmatrix} (L_{D_{2p} A_2} \Lambda_{D_{2,p} A_2}^2 P'_{D_{2p} A_2})^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} = \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
vec(G_1)_{\perp} &= L_2 L_{22}^{-1} (L_{22} L'_{22})^{1/2} = \begin{pmatrix} D_p A_2 & D_{p\perp} \end{pmatrix} \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}, \\
vec(G_2)_{\perp} &= N_2 N_{22}^{-1} (N_{22} N'_{22})^{1/2} = \begin{pmatrix} D_k B_2 & D_{k\perp} \end{pmatrix} \begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix},
\end{aligned} \tag{48}$$

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 :

$$\begin{aligned}
\sqrt{n}\hat{\Lambda} &= \text{vec}(G_1)'_{\perp} D_p \Psi D'_k \text{vec}(G_2)_{\perp} + o_p(1) \\
&= (L_{22} L'_{22})^{1/2} L_{22}^{-1} L'_{22} D_p \Psi D'_k N_2 N_{22}^{-1} (N_{22} N'_{22})^{1/2} + o_p(1) \\
&= \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}' \begin{pmatrix} D_p A_2 & D_{p\perp} \end{pmatrix}' D_p \Psi D'_k \\
&\quad \begin{pmatrix} D_k B_2 & D_{k\perp} \end{pmatrix} \begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} + o_p(1) \\
&= \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}' \begin{pmatrix} A'_2 D'_p D_p \Psi D'_k D_k B_2 & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad \begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} + o_p(1) \\
&= D_{2,p} A_2 \bar{\Lambda} B'_2 D'_{2,k} + o_p(1),
\end{aligned} \tag{49}$$

where

$$\bar{\Lambda} := (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_p D_p \Psi D'_k D_k B_2 (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}},$$

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

$$\begin{aligned}
V_{\text{vec}(\bar{\Lambda})} &= \left((B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_k D_k \otimes (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_p D_p \right) V_{R^*} \\
&\quad \times \left((B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_k D_k \otimes (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_p D_p \right)'.
\end{aligned}$$

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 $E(\|f_i\|^8) < \kappa$. To derive the limit of the covariance matrix estimator in the KPST statistic, the following derivations are important:

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

$$\begin{aligned}
& \left(\begin{pmatrix} D_k B_2 & D_{k\perp} \end{pmatrix} \begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} \otimes \right. \\
& \left. \begin{pmatrix} D_p A_2 & D_{p\perp} \end{pmatrix} \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} \right) \\
&= \left(\begin{pmatrix} D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix} \otimes \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \right) \\
& \left(\begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_k D_k \\ 0 \end{pmatrix} \otimes \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_p D_p \\ 0 \end{pmatrix} \right) V_{R^*} \\
& \left(\begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-\frac{1}{2}} B'_2 D'_k D_k \\ 0 \end{pmatrix} \otimes \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-\frac{1}{2}} A'_2 D'_p D_p \\ 0 \end{pmatrix} \right)' \\
& \left(\begin{pmatrix} D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix} \otimes \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \right)' \\
&= \left(\begin{pmatrix} D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix} \otimes \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \right) \begin{pmatrix} V_{vec(\bar{\Lambda})} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
& \left(\begin{pmatrix} D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix} \otimes \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \right)', \tag{50}
\end{aligned}$$

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:

$$\begin{aligned}
KPST &= n \times \left[vec(vec(G_1)'_\perp \mathcal{R}(\hat{R}) vec(G_2)_\perp) \right]' \\
& \left[(vec(G_2)_\perp \otimes vec(G_1)_\perp)' (D_k \otimes D_p) V_{R^*} (D_k \otimes D_p)' (vec(G_2)_\perp \otimes vec(G_1)_\perp) \right]^- \\
& \left[vec(vec(G_1)'_\perp \mathcal{R}(\hat{R}) vec(G_2)_\perp) \right] + o_p(1) \\
&= vec(\bar{\Lambda})' (D_{2,k} B_2 \otimes D_{2,p} A_2)' \left(\begin{pmatrix} D_{2,k} B_2 (B'_2 D'_{2,k} D_{2,k} B_2)^{-1} & D_{2,k\perp} \end{pmatrix} \otimes \right. \\
& \left. \begin{pmatrix} D_{2,p} A_2 (A'_2 D'_{2,p} D_{2,p} A_2)^{-1} & D_{2,p\perp} \end{pmatrix} \right) \begin{pmatrix} V_{vec(\bar{\Lambda})} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}^- \\
& \left(\begin{pmatrix} (B'_2 D'_{2,k} D_{2,k} B_2)^{-1} B'_2 D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} \otimes \begin{pmatrix} (A'_2 D'_{2,p} D_{2,p} A_2)^{-1} A'_2 D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} \right) \\
& (D_{2,k} B_2 \otimes D_{2,p} A_2) vec(\bar{\Lambda}) + o_p(1) \\
&= vec(\bar{\Lambda})' V_{vec(\bar{\Lambda})}^{-1} vec(\bar{\Lambda}) + o_p(1) \xrightarrow[d]{} \chi_{df}^2,
\end{aligned}$$

with $df = (\frac{1}{2}p(p+1) - 1) (\frac{1}{2}k(k+1) - 1)$. 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

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

which follows because

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

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

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

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 $\hat{R}^* = \hat{L}^* \hat{\Sigma}^* \hat{N}^{*'}$ which are related through:

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

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

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

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

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_{df}^2 .

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

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

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

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \text{vec}(Q\hat{V}_i\hat{V}_i'Q') \text{vec}(H Z_i Z_i' H')' = \\
& (Q \otimes Q) D_p \left[\frac{1}{n} \sum_{i=1}^n \text{vec}(\hat{V}_i\hat{V}_i') \text{vec}(Z_i Z_i')' \right] D'_k (H \otimes H)',
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \text{vech}(Q\hat{V}_i\hat{V}_i'Q') \text{vech}(H Z_i Z_i' H')' = \\
& (D'_p D_p)^{-1} D'_p (Q \otimes Q) D_p \left[\frac{1}{n} \sum_{i=1}^n \text{vech}(\hat{V}_i\hat{V}_i') \text{vech}(Z_i Z_i')' \right] \\
& \times D'_k (H \otimes H)' D_k (D'_k D_k)^{-1}.
\end{aligned}$$

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

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

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 $\text{vec}(\hat{\Lambda})$ in (16) and its covariance matrix estimator.

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

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

for

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

In the derivation above, we use that under H_0 , $\mathcal{R}(R) = \text{vec}(G_1)\text{vec}(G_2)'$, see (6). Therefore $\text{vec}(G_1)'_\perp \mathcal{R}(R) = 0$, $\mathcal{R}(R)\text{vec}(G_2)_\perp = 0$. The limit behavior of KPST results from the limit behavior of a . We specify both $\text{vec}(G_1)_\perp$ and $\text{vec}(G_2)_\perp$, whose dimensions increase as k and p get larger, as orthonormal matrices, $\text{vec}(G_1)'_\perp \text{vec}(G_1)_\perp \equiv I_{\frac{1}{2}k(k-1)}$ and $\text{vec}(G_2)'_\perp \text{vec}(G_2)_\perp \equiv I_{\frac{1}{2}p(p-1)}$. Hence the length of each column of $\text{vec}(G_1)_\perp$ and $\text{vec}(G_2)_\perp$ 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 $\text{vec}(\hat{G}_1)_\perp - \text{vec}(G_1)_\perp$ and $\text{vec}(\hat{G}_2)_\perp - \text{vec}(G_2)_\perp$, see Kleibergen and Paap (2006). Because $\text{vec}(\hat{G}_1)_\perp$ and $\text{vec}(\hat{G}_2)_\perp$ are solved from $\mathcal{R}(\hat{R})$, it follows that $\mathcal{R}(\hat{R}) - \mathcal{R}(R)$, $\text{vec}(\hat{G}_1)_\perp - \text{vec}(G_1)_\perp$ and $\text{vec}(\hat{G}_2)_\perp - \text{vec}(G_2)_\perp$ 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:

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

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 q -th element of each of the five components in the above expression:

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

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 $O_p(n^{-\frac{1}{2}})$. This convergence rate follows because $vec(\mathcal{R}(\hat{R}) - \mathcal{R}(R))$ is $O_p(n^{-\frac{1}{2}})$ 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 $[vec(G_1)_\perp]_{il}$ $i = 1, \dots, p^2$, $j = 1, \dots, k^2$ 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 $\mathcal{R}(\hat{R}) - \mathcal{R}(R) = O_p(n^{-\frac{1}{2}})$.

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

for $m = 1 + \lfloor (q-1)/(k^2-1) \rfloor$, $l = q - (p^2-1)(m-1)$, which is of order $O_p(\frac{k^2}{n})$. This order results from the k^2 dependent components $[vec(\hat{G}_2)_\perp - vec(G_2)_\perp]_{jm}$ and $[vec(\mathcal{R}(\hat{R}) - \mathcal{R}(R))]_{(j-1)k^2+i}$ 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 $[vec(\hat{G}_2)_\perp - vec(G_2)_\perp]_{jm}$ and $[vec(\mathcal{R}(\hat{R}) - \mathcal{R}(R))]_{(j-1)k^2+i}$ are $O_p(n^{-\frac{1}{2}})$ so summing over k^2 of them and multiplying through results in $O_p(\frac{k^2}{n})$. The additional weights $[vec(G_1)_\perp]_{il}$, $i = 1, \dots, p^2$, 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.

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

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

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

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.

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

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:

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

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

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

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 $O_p\left(\frac{k^4 p^4}{n\sqrt{n}}\right)$. 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 $k^2 p^2$ components in $vec(\hat{\Lambda})$, we obtain a slightly stronger condition than just for $\hat{\Lambda}$:

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

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 $p^2 k^2$ components.⁹ The rate that would result from $\hat{\Lambda}$ is $\frac{(pk)^{12}}{n^3} \rightarrow 0$. The convergence rate is in between the rate implied by Newey and Windmeijer (2009) which would be $\frac{k^4 p^4}{n}$ for

⁸To 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: $(H + G)^{-1} = H^{-1} - H^{-1}(G^{-1} + H^{-1})^{-1}H^{-1}$.

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

$\hat{\Lambda}$ and $\frac{k^6 p^6}{n}$ for convergence of the test statistic which is slightly stricter than our rate of $\frac{(pk)^{16}}{n^3} \rightarrow 0$.

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 $O_p\left(\frac{k^4 p^4}{n\sqrt{n}}\right)$ is less than the $O_p\left(\frac{p^2 k^2}{n}\right)$ that results from some of the C components.

$$\begin{aligned}
A_1 &= (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp}) cov(vec(\mathcal{R}(\hat{R}))) (vec(G_2)_{\perp} \otimes vec(G_1)_{\perp}) &= O(1) \\
B_1 &= \left(\left[vec(\hat{G}_2)_{\perp} - vec(G_2)_{\perp} \right]' \otimes vec(G_1)'_{\perp} \right) cov(vec(\mathcal{R}(\hat{R}))) \\
&\quad (vec(G_2)_{\perp} \otimes vec(G_1)_{\perp}) + (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp}) \\
&\quad 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) cov(vec(\mathcal{R}(\hat{R}))) \\
&\quad (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp})' + (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp}) \\
&\quad 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}}) \\
B_3 &= (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp}) \left[\widehat{cov}(vec(\mathcal{R}(\hat{R}))) - cov(vec(\mathcal{R}(\hat{R}))) \right] \\
&\quad (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp})' &= 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 cov(vec(\mathcal{R}(\hat{R}))) (vec(G_2)_{\perp} \otimes vec(G_1)_{\perp}) + (vec(G_2)'_{\perp} \otimes vec(G_1)'_{\perp}) \\
&\quad 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) 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) 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) 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) 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^2 k^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{cov}(vec(\mathcal{R}(\hat{R}))) - \right.
\end{aligned}$$

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

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

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

as $n \rightarrow \infty$ for certain matrices g_i . This can be shown as follows. Consider $M(t) = \text{vec}(G_1)\text{vec}(G_2)' + t\mathcal{R}(A_0)$ 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 $\text{vec}(\bar{G}_{1,n})$ 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 $\text{vec}(g_1) \in \mathbb{R}^{p^2}$. But this proves the claim for $\bar{G}_{1,n}$. The proof for $\bar{G}_{2,n}$ is identical. Intuitively, $(\text{vec}(G_1) + O(n^{-a}))(\text{vec}(G_2) + O(n^{-b}))' = \text{vec}(G_1)\text{vec}(G_2)' + O(n^{-1/2})$ implies that $a = b \geq 1/2$, see Kleibergen and Paap (2006).

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

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

Namely, let $\text{vec}(G_1)_\perp = (v_2, \dots, v_{p^2})$. Clearly $(\text{vec}(\bar{G}_{1,n}), v_2, \dots, v_{p^2})$ will be of full rank for all n large

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

enough and $vec(\bar{G}_{1,n})_{\perp}$ can be obtained as the last $p^2 - 1$ vectors by performing Gram-Schmidt orthogonalization to $(vec(\bar{G}_{1,n}), v_2, \dots, v_{p^2})$. 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

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

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

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

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 $\sqrt{n} vec(\hat{\Lambda}) \rightarrow_d N(a_0, V_{\Lambda})$ and thus $KPST \rightarrow_d \chi_{df}^2(\delta)$ as claimed.

References

Acemoglu, D., D. Cantoni, S. Johnson, and J. A. Robinson (2011). The Consequences of Radical Reform: The French Revolution. *American Economic Review* 101(7), 3286–3307.

- Acemoglu, D. and S. Johnson (2005). Unbundling Institutions. *Journal of Political Economy* 113(5), 949–995.
- Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008). Income and Democracy. *American Economic Review* 98(3), 808–42.
- Alesina, A., P. Giuliano, and N. Nunn (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics* 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 discussion papers 3005, Cowles Foundation for Research in Economics, Yale University.
- Autor, D. H. and D. Dorn (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* 103(5), 1553–97.
- Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review* 103(6), 2121–68.
- Chen, Q. and Z. Fang (2019). Improved inference on the rank of a matrix. *Quantitative Economics* 10, 1787–1824.
- Dahl, G. B. and L. Lochner (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review* 102(5), 1927–56.
- Donald, S. G., N. Fortuna, and V. Pipiras (2007). On Rank Estimation in Suymmetric Matrices: The Case of Indefinite Matrix Estimators. *Journal of Econometrics* 23, 1217–1232.
- Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* 73(4), 1351–1365.
- Duranton, G. and M. A. Turner (2011). The Fundamental Law of Road Congestion: Evidence from US Cities. *American Economic Review* 101(6), 2616–52.
- Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* 18, 681–695.
- Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019). A more powerful subvector Anderson Rubin test in linear instrumental variable regression. *Quantitative Economics* 10(2), 487–526.
- Guggenberger, P., F. Kleibergen, and S. Mavroeidis (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, 1029–1054.
- Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262–280.

- Hansford, T. G. and B. T. Gomez (2010). Estimating the Electoral Effects of Voter Turnout. *American Political Science Review* 104(2), 268–288.
- Johnson, D. S., J. A. Parker, and N. S. Souleles (2006). Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review* 96(5), 1589–1610.
- Kan, R. and C. Zhang (1999). Two-pass tests of asset pricing models with useless factors. *Journal of 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. *Journal of Econometrics* 149, 149–173.
- Kleibergen, F. (2021). Efficient size correct subset inference in homoskedastic linear instrumental variables regression. *Journal of Econometrics* 221(1), 78–96.
- Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1), 97–126.
- Kleibergen, F. and Z. Zhan (2020). Robust inference for consumption-based asset pricing. *The Journal of Finance* 75(1), 507–550.
- Lax, P. D. (2007). *Linear Algebra and its Applications* (2nd ed.). Wiley-Interscience.
- Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics* 40, 1024–1060.
- Ledoit, O. and M. Wolf (2015). Spectrum Estimation: a unified approach for covariance estimation and PCA in large dimensions. *Journal of Multivariate Analysis* 139, 360–384.
- Ledoit, O. and M. Wolf (2018). Optimal estimation of a large-dimensional covariance matrix under Stein’s loss. *Bernoulli* 24, 3791–3832.
- Lu, N. and D. L. Zimmermann (2005). The likelihood ratio test for a separable covariance matrix. *Statistics and Probability Letters* 73, 449–457.
- Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* 112(4), 725–753.
- Mitchell, M., M. Genton, and M. Gumpertz (2006). A likelihood ratio test for separability of covariance. *Journal of Multivariate Analysis* 97, 1025–1043.
- Newey, W. and F. Windmeijer (2009). GMM with many weak moment conditions. *Econometrica* 77(3), 687–719.
- Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Nunn, N. (2008). The Long-Term Effects of Africa’s Slave Trades. *The Quarterly Journal of Economics* 123(1), 139–176.

- Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013). Consumer Spending and the Economic Stimulus Payments of 2008. *American Economic Review* 103(6), 2530–53.
- Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam. *American Economic Review* 100(1), 557–71.
- Van Loan, C. and N. Pitsianis (1993). Approximation with kronecker products. In *Linear algebra for large scale and real-time applications*, NATO Adv. Sci. Inst. Ser. E Appl. Sci. 232, pp. 293–314. Kluwer Academic Publishers.
- Velu, R. and K. Herman (2017). Separable Covariance Matrices and Kronecker Approximations. *Procedia Computer Science* 108, 1019–1029.
- Voors, M. J., E. E. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. Van Soest (2012). Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review* 102(2), 941–64.
- Werner, K., M. Jansson, and P. Stoica (2008). On estimation of covariance matrices with Kronecker product structure. *IEEE Transactions of Signal Processing* 56, 478–491.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4), 817–38.
- White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press.
- Yogo, M. (2004). Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak. *Review of Economics and Statistics* 86(3), 797–810.

Acronym	Paper
ACJR 11	Acemoglu et al. (2011)
AD 13	Autor and Dorn (2013)
ADG 13	Autor et al. (2013)
AGN 13	Alesina et al. (2013)
AJ 05	Acemoglu and Johnson (2005)
AJRY 08	Acemoglu et al. (2008)
DT 11	Duranton and Turner (2011)
HG 10	Hansford and Gomez (2010)
JPS 06	Johnson et al. (2006)
MSS 04	Miguel et al. (2004)
Nunn 08	Nunn (2008)
PSJM 13	Parker et al. (2013)
TCN 10	Tanaka et al. (2010)
V et al 12	Voors et al. (2012)
Yogo 04	Yogo (2004)

Table 5: List of papers used in the empirical applications.

B Supplementary Appendix: Detailed empirical results

Tables 6 and 7 give detailed empirical results in the applications considered, with non-clustered and clustered data, respectively. The acronyms refer to the different papers listed in Table 5.

Table 6: Applications of KPST.

Paper	Specif.	Y	Z	p	k	n	KPST	p val
TCN 10	T5.P2.C1	Value function curvature, Income	Rainfall, Head of Household Cannot Work (dummy variable)	2	2	181	4.944	0.293
	T5.P2.C2	Value function curvature, Relative Income, Mean Income	Rainfall, Head of Household Cannot Work (dummy variable)	3	2	181	14.859	0.137
Nunn 08	T4.C1	Log income in 2000, Slave exports	Atlantic distance, Indian distance, Saharan distance, Red Sea distance	2	4	52	32.307	0.02
	T4.C2	Log income in 2000, Slave exports, (X: Colonization effect)	Atlantic distance, Indian distance, Saharan distance, Red Sea distance	2	4	52	30.922	0.029
	T4.C3	Log income in 2000, Slave exports, (X: Col. effect, geographical controls)	Atlantic distance, Indian distance, Saharan distance, Red Sea distance	2	4	52	34.597	0.011
	T4.C4	Log income in 2000, Slave exports, (X: Col. effect, geographical controls)	Atlantic distance, Indian distance, Saharan distance, Red Sea distance	2	4	42	28.263	0.058
AJ 05	T4.P1.C1	Log GDP per capita, legal formalism, constraint on executive	English legal origin, settler mortality	3	2	51	8.18	0.611
	T4.P1.C2	Log GDP per capita, legal formalism, constraint on executive	English legal origin, population density 1500	3	2	60	25.969	0.004
	T4.P1.C3	Log GDP per capita, constraint on executive, procedural complexity	English legal origin, settler mortality	3	2	60	5.574	0.85
	T4.P1.C4	Log GDP per capita, constraint on executive, number of procedures	English legal origin, settler mortality	3	2	61	10.916	0.364
	T4.P1.C5	Log GDP per capita, legal formalism, average protection against risk of expropriation	English legal origin, settler mortality	3	2	51	7.075	0.718

Continued on next page

Table 6 – continued from previous page

Paper	Specif.	Y	Z	p	k	n	KPST	p val
	T4.P1.C6	Log GDP per capita, legal formalism, private property	English legal origin, settler mortality	3	2	52	8.646	0.566
	T4.P2.C1	Investment-GDP ratio, legal formalism, constraint on executive	English legal origin, settler mortality	3	2	51	13.068	0.22
	T4.P2.C2	Investment-GDP ratio, legal formalism, constraint on executive	English legal origin, population density 1500	3	2	60	36.298	0
	T4.P2.C3	Investment-GDP ratio, constraint on executive, procedural complexity	English legal origin, settler mortality	3	2	61	16.838	0.078
	T4.P2.C4	Investment-GDP ratio, constraint on executive, number of procedures	English legal origin, settler mortality	3	2	62	14.82	0.139
	T4.P2.C5	Investment-GDP ratio, legal formalism, average protection against risk of expropriation	English legal origin, settler mortality	3	2	51	13.75	0.185
	T4.P2.C6	Investment-GDP ratio, legal formalism, private property	English legal origin, settler mortality	3	2	52	8.582	0.572
	T5.P1.C1	Private credit, legal formalism, constraint on executive	English legal origin, settler mortality	3	2	51	9.296	0.504
	T5.P1.C2	Private credit, legal formalism, constraint on executive	English legal origin, population density 1500	3	2	60	31.406	0.001
	T5.P1.C3	Private credit, constraint on executive, procedural complexity	English legal origin, settler mortality	3	2	60	13.721	0.186
	T5.P1.C4	Private credit, constraint on executive, number of procedures	English legal origin, settler mortality	3	2	61	11.605	0.312
	T5.P1.C5	Private credit, legal formalism, average protection against risk of expropriation	English legal origin, settler mortality	3	2	51	12.206	0.272

Continued on next page

Table 6 – continued from previous page

Paper	Specif.	Y	Z	p	k	n	KPST	p val
	T5.P1.C6	Private credit, legal formalism, private property	English legal origin, settler mortality	3	2	52	19.304	0.037
	T5.P2.C1	Stock market capitalization, legal formalism, constraint on executive	English legal origin, settler mortality	3	2	50	19.178	0.038
	T5.P2.C2	Stock market capitalization, legal formalism, constraint on executive	English legal origin, population density 1500	3	2	59	19.405	0.035
	T5.P2.C3	Stock market capitalization, constraint on executive, procedural complexity	English legal origin, settler mortality	3	2	59	34.566	0
	T5.P2.C4	Stock market capitalization, constraint on executive, number of procedures	English legal origin, settler mortality	3	2	59	28.06	0.002
	T5.P2.C5	Stock market capitalization, legal formalism, average protection against risk of expropriation	English legal origin, settler mortality	3	2	50	35.531	0
	T5.P2.C6	Stock market capitalization, legal formalism, private property	English legal origin, settler mortality	3	2	51	21.344	0.019
HG 10	T1.C2	Democratic vote share, turnout, turnout * partisan composition, turnout * Republican incumbent	Rainfall, rainfall*partisan composition, rainfall*Republican incumbent	4	3	27401	507.919	0
	T1.C3	Democratic vote share, turnout, turnout * partisan composition, turnout * Republican incumbent	Rainfall, rainfall*partisan composition, rainfall*Republican incumbent	4	3	27401	457.962	0
AGN 13	T8.P3.C1	Female LF participation, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	160	6.191	0.185

Continued on next page

Table 6 – continued from previous page

Paper	Specif.	Y	Z	p	k	n	KPST	p val
	T8.P3.C2	Female LF participation, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	160	4.939	0.294
	T8.P3.C3	Share firm ownership female, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	122	3.586	0.465
	T8.P3.C4	Share firm ownership female, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	122	6.785	0.148
	T8.P3.C5	Share political position female, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	140	9.29	0.054
	T8.P3.C6	Share political position female, Traditional plough use	Plough-neg. environment, Plough-pos. environment	2	2	140	10.982	0.027
Yogo 04	AUL	cons growth, risk-free rtn	Twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio	2	4	114	16.628	0.549
		cons growth, stk mkt rtn		2	4	114	22.879	0.195
	CAN	cons growth, risk-free rtn		2	4	115	24.078	0.152
		cons growth, stk mkt rtn		2	4	115	32.528	0.019
	FRA	cons growth, risk-free rtn		2	4	113	28.015	0.062
		cons growth, stk mkt rtn		2	4	113	25.608	0.109
	GER	cons growth, risk-free rtn		2	4	79	25.452	0.113
		cons growth, stk mkt rtn		2	4	79	31.24	0.027
	ITA	cons growth, risk-free rtn		2	4	106	18.266	0.438
		cons growth, stk mkt rtn		2	4	106	25.889	0.102
	JAP	cons growth, risk-free rtn		2	4	114	22.835	0.197
		cons growth, stk mkt rtn		2	4	114	16.132	0.583
	NTH	cons growth, risk-free rtn		2	4	86	20.969	0.281
		cons growth, stk mkt rtn		2	4	86	21.762	0.243
	SWD	cons growth, risk-free rtn		2	4	116	18.967	0.394
		cons growth, stk mkt rtn		2	4	116	29.714	0.04

Continued on next page

Table 6 – continued from previous page

Paper	Specif.	Y	Z	p	k	n	KPST	p val
	SWT	cons growth, risk-free rtn		2	4	91	14.889	0.67
		cons growth, stk mkt rtn		2	4	91	43.768	0.001
	UK	cons growth, risk-free rtn		2	4	115	30.148	0.036
		cons growth, stk mkt rtn		2	4	115	19.94	0.336
	US	cons growth, risk-free rtn		2	4	114	18.478	0.425
		cons growth, stk mkt rtn		2	4	114	22.373	0.216

Specification T: table; P: panel; C: column.

Table 7: Applications of cluster KPST.

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
<i>AJRY 08</i>										
T5.C5	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, Democracy in t-1	2	2	891	23.86	0.000	134	20.204	0.001
T5.C7	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, labour share of income	2	2	471	21.85	0.000	98	6.037	0.303
T5.C8.S1	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, democracy in t-1	2	2	471	17.21	0.002	98	13.500	0.019
T5.C8.S2	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, democracy in t-2 X: democracy in t-2, t-3	2	2	471	14.96	0.005	98	11.738	0.039
T5.C8.S3	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, democracy in t-3 X: democracy in t-1, t-3	2	2	471	6.83	0.145	98	4.388	0.495
T5.C9	Freedom House measure of democracy, Log GDP per capita in t-1	Savings rate in t-2, t-3 X: democracy in t-1, t-2	2	2	796	12.14	0.016	125	18.960	0.002
T6.C5	Freedom House measure of democracy, Log GDP per capita in t-1	Trade-weighted (tw) log GDP in t-1, democracy in t-1	2	2	796	4.71	0.318	125	12.970	0.024

Continued on next page

Table 7 – continued from previous page

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
T6.C7	Freedom House measure of democracy, Log GDP per capita in t-1	tw log GDP in t-1, tw democracy in t-1	2	2	796	10.18	0.037	125	11.808	0.038
T6.C9	Freedom House measure of democracy, Log GDP per capita in t-1	tw log GDP in t-1, t-2	2	2	796	12.83	0.012	125	12.121	0.033
<i>JPS 06</i>										
T4.P1.C5	Dollar change in strict non-durables, rebate in t+1, t	I (rebate t+1), I (rebate t)	3	2	12730	1062.30	0.000	6253	386.388	0.000
T4.P1.C6	Dollar change in non-durable goods, rebate in t+1, t	I (rebate t+1), I (rebate t)	3	2	12730	1062.05	0.000	6253	377.982	0.000
T4.P2.C5	Dollar change in strict non-durables, rebate in t+1, t, t-1	I (rebate t+1), I (rebate t), I (rebate t-1)	4	3	15022	1635.13	0.000	6295	1128.150	0.000
T4.P2.C6	Dollar change in non-durable goods, rebate in t+1, t, t-1	I (rebate t+1), I (rebate t), I (rebate t-1)	4	3	15022	1666.13	0.000	6295	1140.060	0.000
<i>PSJM 13</i>										
T4.P1.C5	Nondurable spending, ESP by check, ESP by electronic transfer	I (ESP by check), I (ESP by electronic transfer)	3	2	17281	457.30	0.000	8038	314.724	0.000

Continued on next page

Table 7 – continued from previous page

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
T4.P1.C6	All spending, ESP by check, ESP by electronic transfer	I (ESP by check), I (ESP by electronic transfer)	3	2	17281	458.98	0.000	8038	288.445	0.000
<i>ADH 13</i>										
T10.P3.C1	Δ mfg empl, Δ trade US-China net input pw (nipw)	Δ trade other-China, Δ net input other-China	2	2	1444	20.00	0.001	48	27.125	0.000
T10.P3.C2	Δ nonmfg empl, Δ trade US-China nipw	Δ trade other-China, Δ net input other-China	2	2	1444	22.95	0.000	48	24.312	0.000
T10.P3.C3	Δ mfg log wage, Δ trade US-China nipw	Δ trade other-China, Δ net input other-China	2	2	1444	31.27	0.000	48	19.553	0.002
T10.P3.C4	Δ mfg log wage, Δ trade US-China nipw	Δ trade other-China, Δ net input other-China	2	2	1444	19.40	0.001	48	22.269	0.000
T10.P3.C5	Δ nonmfg log wage, Δ trade US-China nipw	Δ trade other-China, Δ net input other-China	2	2	1444	100.88	0.000	48	10.514	0.062
T10.P3.C6	Δ log transfers, Δ trade US-China nipw	Δ trade other-China, Δ net input other-China	2	2	1444	21.82	0.000	48	16.716	0.005
T10.P4.C1	Δ mfg empl, Δ US-China net imports pw	Δ trade other-China, Δ net exports other-China	2	2	1444	16.52	0.002	48	10.187	0.070
T10.P4.C2	Δ nonmfg empl, Δ US-China net imp pw	Δ trade other-China, Δ net exports other-China	2	2	1444	18.44	0.001	48	10.014	0.075
T10.P4.C3	Δ mfg log wage, Δ US-China net imp pw	Δ trade other-China, Δ net exports other-China	2	2	1444	37.44	0.000	48	13.290	0.021
T10.P4.C4	Δ nonmfg log wage, Δ US-China net imp pw	Δ trade other-China, Δ net exports other-China	2	2	1444	11.21	0.024	48	11.072	0.050

Continued on next page

Table 7 – continued from previous page

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
T10.P4.C5	Δ log transfers, Δ US-China net imp pw	Δ trade other-China, Δ net exports other-China	2	2	1444	41.77	0.000	48	9.138	0.104
T10.P4.C6	Δ avg household inc, Δ US-China net imp pw	Δ trade other-China, Δ net exports other-China	2	2	1444	18.08	0.001	48	13.395	0.020
T10.P6.C1	Δ mfg empl, Δ net trade factor (ntf) US-China	Δ ntf other-China, Δ net export factor (nef) other-China	2	2	1444	16.57	0.002	48	14.213	0.014
T10.P6.C2	Δ nonmfg empl, Δ ntf US-China	Δ ntf other-China, Δ nef other-China	2	2	1444	43.88	0.000	48	15.611	0.008
T10.P6.C3	Δ mfg log wage, Δ ntf US-China	Δ ntf other-China, Δ nef other-China	2	2	1444	24.54	0.000	48	12.087	0.034
T10.P6.C4	Δ nonmfg log wage, Δ ntf US-China	Δ ntf other-China, Δ nef other-China	2	2	1444	10.81	0.029	48	18.869	0.002
T10.P6.C5	Δ log transfers, Δ ntf US-China	Δ ntf other-China, Δ nef other-China	2	2	1444	15.56	0.004	48	16.692	0.005
T10.P6.C6	Δ avg household inc, Δ ntf US-China	Δ ntf other-China, Δ nef other-China	2	2	1444	16.46	0.002	48	29.073	0.000
<i>AD 13</i>										
T5.P2.C1	Growth of service employment, Share of routine employment (t-1)	1950 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	141.50	0.000	48	57.891	0.000

Continued on next page

Table 7 – continued from previous page

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
T5.P2.C2	Growth of service employment, Share of routine employment (t-1)	1951 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	122.97	0.000	48	41.735	0.000
T5.P2.C3	Growth of service employment, Share of routine employment (t-1)	1952 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	140.57	0.000	48	52.603	0.000
T5.P2.C4	Growth of service employment, Share of routine employment (t-1)	1953 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	118.33	0.000	48	47.893	0.000
T5.P2.C5	Growth of service employment, Share of routine employment (t-1)	1954 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	106.08	0.000	48	47.248	0.000

Continued on next page

Table 7 – continued from previous page

Specif.	<i>Y</i>	<i>Z</i>	<i>p</i>	<i>k</i>	<i>n</i>	KPST	p val	<i>n_c</i>	KPST _{<i>c</i>}	p val
T5.P2.C6	Growth of service employment, Share of routine employment (t-1)	1955 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	146.81	0.000	48	43.400	0.000
T5.P2.C7	Growth of service employment, Share of routine employment (t-1)	1956 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.‘	2	3	2166	101.50	0.000	48	32.647	0.002
<i>ACJR 11</i> T6.P.3.C2	Urbanization in Germany, reform index	French presence in 1850, 1875 and 1900	2	3	74	12.74	0.239	13	112.422	0.000
<i>MSS 04</i> T4.C5	Civil conflict >25 deaths, Economic growth rate (t)	Current and lagged rainfall	3	2	743	10.30	0.414	41	31.022	0.003
T4.C6	Civil conflict >25 deaths, Economic growth rate (t)	Current and lagged rainfall	3	2	743	5.18	0.879	41	37.682	0.000
T4.C7	Civil conflict >1000 deaths, Economic growth rate (t)	Current and lagged rainfall	3	2	743	5.35	0.867	41	42.052	0.000

Continued on next page

Table 7 – continued from previous page

Specif.	Y	Z	p	k	n	KPST	p val	n_c	KPST _c	p val
<i>V et al 12</i>										
T3.C6	Degree of altruism scale, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	278	9.45	0.051	35	8.054	0.153
T4.C6	Risk preference, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	213	12.28	0.015	35	1.349	0.930
T5.C6	Discount rate, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	266	6.69	0.153	35	5.622	0.345
T6.C4	Degree of altruism scale, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	212	6.36	0.174	35	6.931	0.226
T6.C5	Risk preference, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	158	8.88	0.064	35	6.860	0.231
T6.C6	Discount rate, Percentage dead in attacks	Distance to Bujumbura (log), Altitude (log)	2	2	205	2.34	0.673	35	4.451	0.487

Specification: T: table; P: panel; C: column. n_c : number of clusters, KPST_c: cluster KPST statistic.