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: April 9, 2021

Abstract

We propose a test for a covariance matrix to have Kronecker Product Structure (KPS). KPS implies a reduced rank restriction on an invertible transformation of the covariance matrix and the new procedure is an adaptation of the Kleibergen and Paap (2006) reduced rank test. The main extension concerns the singularity of the covariance matrix estimator involved in the rank test which complicates the derivation of its limiting distribution. We show this limiting distribution to be χ^2 with degrees of freedom corresponding to the number of restrictions tested. Re-examining sixteen highly cited papers conducting IV regressions, we find that KPS is not rejected in 24 of 30 specifications for moderate sample sizes at the 5% nominal size.

Keywords: covariance matrix, heteroskedasticity, Kronecker product structure, linear instrumental variables regression model, reduced rank, weak identification
JEL codes: C12, C26

1 Introduction

The robustness properties of nonparametric covariance matrix estimators, like, those proposed by White (1980) against heteroskedasticity and the heteroskedasticity and autocorrelation (hac) robust

*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 Lewis McLean for research assistance.

ones by, for example, Newey and West (1987) and Andrews (1991), have enabled 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, and this fact 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, however, may lead to computational challenges for estimation of the structural parameters and makes the distribution theory of the corresponding test statistics under weak identification difficult to derive. Prominent examples of such models are the linear instrumental variables (IV) regression model and the linear factor model in asset pricing. The literature on weak-identification-robust inference in GMM models has produced a number of weak-identification-robust tests on the structural parameters based on statistics centered around the continuous updating estimator (CUE) of Hansen et al. (1996). When using a nonparametric covariance matrix estimator, the CUE objective function is often ill behaved, like, for example, being multi modal, making the CUE cumbersome to compute.

Furthermore, the favorable power properties of certain weak-identification-robust tests are only available for joint hypotheses specified on all structural parameters and not for tests specified on a subvector of the structural parameter vector which are typically what applied researchers care about most.

When we use a Kronecker Product Structure (KPS) covariance matrix estimator instead of a nonparametric covariance matrix estimator for 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 and 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). This all shows that in weakly identified settings where consistency of estimators of the structural parameters cannot be guaranteed, there is a trade-off between the robustness provided by a nonparametric covariance matrix estimator and the ease of conducting accurate statistical inference resulting from the use of a KPS covariance matrix estimator.

To help in the trade-off between robustness and accuracy of GMM estimation and inference, we develop a test for KPS of the covariance matrix of the sample moment vector of the unrestricted linear reduced form encompassing e.g. linear IV and factor asset pricing models. The procedure is based on the insight that KPS implies a reduced rank restriction on an invertible transformation of the covariance matrix, see Van Loan and Pitsianis (1993). We therefore adapt the Kleibergen and Paap (2006), henceforth KP, reduced rank statistic to test for KPS.[1] This adaptation in particular

[1] Another adaptation of the KP reduced-rank statistic is by Donald et al. (2007), who develop a test for singularity,

concerns the singularity of the covariance matrix of the sample covariance matrix estimator because of which it is not obvious if usage of the Moore-Penrose generalized inverse in the expression of the KP reduced rank statistic leads to an appropriate χ^2 limiting distribution. We therefore derive the limiting distribution of the estimator representing the reduced rank restriction and show it to be degenerate normal. We next show that the probability limit of the Moore-Penrose inverse of the covariance matrix involved in the KP rank statistic is such that it offsets this degeneracy which results in a χ^2 limiting distribution of the KPS test with degrees of freedom equal to the number of tested restrictions.

We apply the new KPS test to the different specifications of linear IV models employed in sixteen highly cited empirical studies published in top ranked economic journals. We find that for the specifications with moderate numbers of observations KPS is not rejected in 24 of 30 cases at the 5% significance level while for smaller number of observations it is rejected in 14 of 28 cases. The relatively high number of nonrejections illustrates the importance of the KPS test for applied work.

In a companion paper, Guggenberger et al. (2021), we show how the new KPS test statistic can be used as an ingredient for a pre-test for conducting size correct inference on a subvector of the structural parameter vector in the linear IV regression model with a general covariance matrix. Namely, the KPS test is used to test for a KPS of the covariance matrix of the unrestricted reduced form sample moment vector. Depending on whether the resulting value of the KPS test statistic exceeds a sample size dependent threshold, the hypothesis of interest on a subvector of the structural parameters is either tested using the improved subvector Anderson-Rubin test from Guggenberger et al. (2019) when it is below the threshold or using the size correct AR\AR test procedure from Andrews (2017) when it is above the threshold. The AR\AR procedure from Andrews (2017) is a size correct inference procedure for testing hypothezes 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 which is, however, only size correct under a KPS covariance matrix. Guggenberger et al. (2021) develop the asymptotic theory to show that the switching test procedure is asymptotically size correct 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 just weakly identified, see e.g. Kan and Zhang (1999), Kleibergen (2009) and Kleibergen and Zhan (2020), so 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.

or reduced rank, of a symmetric matrix.

Extending homoskedasiticy 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 not comparable to the one derived here.

KPS is an example of dimension reduction of a covariance matrix. Other examples 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 its asymptotic distribution. The third and fourth sections conduct simulation studies 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 a $m \times k$ 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 $\text{vech}(A)$: $\text{vech}(A) = (D_m'D_m)^{-1}D_m'\text{vec}(A)$ and $\text{vec}(A) = D_m\text{vech}(A)$.

2 Test for Kronecker Product Structure of a covariance matrix

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

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

for mean zero, independently distributed random vectors $f_i \in \mathbb{R}^{kp}$, $i = 1, \dots, n$, which satisfy $f_i = (V_i \otimes Z_i)$, with $V_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^k$ uncorrelated random vectors[2]. 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

[2]We could also allow V_i and Z_i to be correlated but this would require recentering of the sample moment vector f_i. Furthermore, R can depend on the sample size n but for simplicity of notation we do not index R by n.

contains the regressors, so that R is then the covariance matrix of the sample covariance between these errors and the regressors.

The covariance matrix has a block structure

$$R := \begin{pmatrix} R_{11} & \cdots & R_{1p} \\ \vdots & \ddots & \vdots \\ R_{p1} & \cdots & R_{pp} \end{pmatrix}, \tag{2}$$

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

$$H_0 : R = G_1 \otimes G_2, \tag{3}$$

with $G_1 \in \mathbb{R}^{p\times p}$ and $G_2 \in \mathbb{R}^{k\times k}$ symmetric positive definite matrices of which one for normalization purposes has a diagonal element equal to one (say the upper left element of G_1), against the alternative hypothesis of not having KPS. To measure the distance of the sample covariance matrix estimator below 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 (2) define

$$\mathcal{R}(A) := \begin{pmatrix} A_1 \\ \vdots \\ A_p \end{pmatrix} \in \mathbb{R}^{p^2\times k^2}, \quad \text{with } A_j := \begin{pmatrix} vec\,(A_{1j})' \\ \vdots \\ vec\,(A_{pj})' \end{pmatrix} \in \mathbb{R}^{p\times k^2}, \tag{4}$$

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

$$\mathcal{R}(G_1 \otimes G_2) = vec(G_1)vec(G_2)' \tag{5}$$

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

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

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

Consider the covariance matrix estimator

$$\widehat{R} := \frac{1}{n}\sum_{i=1}^{n} \hat{f}_i\hat{f}_i' \in \mathbb{R}^{kp\times kp} \tag{6}$$

which uses sample values, $\hat{f}_i$, of the random vectors f_i, which are assumed to converge to f_i, $\hat{f}_i = f_i + o_p(1)$, uniformly for $i = 1, \ldots, n$, as $n \to \infty$. Define the distance from a KPS covariance matrix by the Frobenius norm:

$$DS := \min_{G_1>0, G_2>0} \left\| \mathcal{R}(\hat{R}) - vec(G_1)vec(G_2)' \right\|_F, \tag{7}$$

where G_1, $G_2 > 0$ indicates that G_1 and G_2 are positive definite symmetric matrices.

Theorem 1 *The distance measure DS (7) 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 $\hat{\sigma}_1 \geq ... \geq \hat{\sigma}_{\min(p^2,k^2)}$ are the ordered singular values of $\mathcal{R}(\hat{R})$.*

Proof. see the Appendix. ■

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 :

$$\begin{aligned} \mathcal{R}(\hat{R}) = &\ \mathcal{R}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n (\hat{V}_i\hat{V}_i' \otimes Z_iZ_i')\right) \\ = &\ \tfrac{1}{n}\textstyle\sum_{i=1}^n vec(\hat{V}_i\hat{V}_i')vec(Z_iZ_i')' \\ = &\ D_p\hat{R}^* D_k', \end{aligned} \tag{8}$$

with

$$\hat{R}^* := \tfrac{1}{n}\textstyle\sum_{i=1}^n vech(\hat{V}_i\hat{V}_i')vech(Z_iZ_i')'. \tag{9}$$

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

$$\sqrt{n}(vec(\hat{R}^*) - vec(R^*)) \underset{d}{\rightarrow} \psi = vec(\Psi), \tag{10}$$

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

$$\begin{aligned} R^* := &\ E\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n vech(V_iV_i')vech(Z_iZ_i')'\right) \\ V_{R^*} := &\ \lim_{n\to\infty} E\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n (vech(Z_iZ_i')vech(Z_iZ_i')' \otimes vech(V_iV_i')vech(V_iV_i')')\right). \end{aligned} \tag{11}$$

In fact, we assume the slightly stronger result that, $\hat{R}^* = R^* + \frac{1}{\sqrt{n}}\Psi + o_p(n^{-\frac{1}{2}})$, holds. The central limit theorem (10) holds under mild conditions like those for central limit theorems for non-identical distributed independent random variables such as the Liapounov and Lindeberg-Feller central limit theorems, see White (1984).

We test for $\mathcal{R}(\hat{R})$ being a rank one matrix using the KP rank statistic.

To describe the KP rank statistic consider first a singular value decomposition (SVD) of $\mathcal{R}(\hat{R})$:

$$\mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}', \tag{12}$$

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

$\mathbb{R}^{p^2\times p^2}$ and $\hat{N}\in\mathbb{R}^{k^2\times k^2}$ orthonormal matrices. Decompose

$$\hat{L}:=\begin{pmatrix}\hat{L}_{11} & \hat{L}_{12}\\ \hat{L}_{21} & \hat{L}_{22}\end{pmatrix}=\left(\hat{L}_1 \vdots \hat{L}_2\right),\ \hat{\Sigma}:=\begin{pmatrix}\hat{\sigma}_1 & 0\\ 0 & \hat{\Sigma}_2\end{pmatrix},\ \hat{N}:=\begin{pmatrix}\hat{N}_{11} & \hat{N}_{12}\\ \hat{N}_{21} & \hat{N}_{22}\end{pmatrix}=(\hat{N}_1 \vdots \hat{N}_2), \tag{13}$$

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. By having $\hat{G}_1$ and $\hat{G}_2$ connected to the largest singular value, we have:[3]

$$\begin{array}{lll}
vec(\hat{G}_1):=\binom{\hat{L}_{11}}{\hat{L}_{21}}/\hat{L}_{11} & & :p^2\times 1,\\
vec(\hat{G}_1)_\perp:=\hat{L}_2\hat{L}_{22}^-(\hat{L}_{22}\hat{L}_{22}')^{1/2} & & :p^2\times(p^2-1),\\
vec(\hat{G}_2)':=\hat{L}_{11}\hat{\sigma}_1\left(\hat{N}_{11}\vdots\hat{N}_{21}'\right)=\hat{L}_{11}\hat{\sigma}_1\hat{N}_1' & & :1\times k^2,\\
vec(\hat{G}_2)_\perp':=\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}\hat{N}_{22}'^{-}\hat{N}_2' & & :(k^2-1)\times k^2,
\end{array} \tag{14}$$

and A^- is the Moore-Penrose generalized inverse of a matrix A. Define:

$$\hat{\Lambda}:=\ \left(\hat{L}_{22}\hat{L}_{22}'\right)^{-1/2}\hat{L}_{22}\hat{\Sigma}_2\hat{N}_{22}'\left(\hat{N}_{22}\hat{N}_{22}'\right)^{-1/2}\quad :(p^2-1)\times(k^2-1). \tag{15}$$

It can be shown that $\hat{\Lambda}=vec(\hat{G}_1)_\perp'\mathcal{R}(\hat{R})vec(\hat{G}_2)_\perp$, see Kleibergen and Paap (2006). Theorems 5.3-5.6 from Van Loan and Pitsianis (1993) show that if $\hat{R}$ is a symmetric positive definite matrix then so are $\hat{G}_1$ and $\hat{G}_2$ because the SVD is unique. We then have

$$\mathcal{R}(\hat{R})=vec(\hat{G}_1)vec(\hat{G}_2)'+vec(\hat{G}_1)_\perp\hat{\Lambda}vec(\hat{G}_2)_\perp'. \tag{16}$$

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

We define the statistic KPST for testing H_0 in (3) as

$$\begin{array}{l}
KPST:=n\times vec(\hat{\Lambda})'\left(\hat{J}'\hat{V}\hat{J}\right)^-vec(\hat{\Lambda}),\ \text{where}\\
\hat{J}:=\left(\left[vec(\hat{G}_2)\right]_\perp\otimes\left[vec(\hat{G}_1)\right]_\perp\right),\quad \hat{V}:=\widehat{\text{cov}}(vec(\mathcal{R}(\hat{R})))\in\mathbb{R}^{p^2k^2\times p^2k^2},
\end{array} \tag{17}$$

and

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

[3] If A is a positive semi definite $m\times m$ symmetric matrix $A=EL^2E'$, where L is a diagonal $m\times m$ matrix containing the square roots of the eigenvalues of A, and E is a $m\times m$ matrix that contains the orthonormal eigenvectors of A, define $A^{1/2}:=ELE'$ and $A^{-1/2}:=E_1L_1^{-1}E_1'$, where L_1 is a diagonal matrix containing the non-zero eigenvalues of A and E_1 consists of the corresponding eigenvectors.

Using $\mathcal{R}(R)$, our hypothesis of interest H_0 (3) is transformed into

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

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

Since $vec(G_1) = D_p vech(G_1)$, $vec(G_2) = D_k vech(G_2)$, the hypothesis of interest (18) can also be specified as:

$$H_0 : R^* = vech(G_1)vech(G_2)' \text{ or } H_0 : vech(G_1)'_{\perp}R^* vech(G_2)_{\perp} = 0, \tag{19}$$

where $vech(G_1)_{\perp}$ and $vech(G_2)_{\perp}$ are $\frac{1}{2}p(p+1) \times (\frac{1}{2}p(p+1)-1)$ and $\frac{1}{2}k(k+1) \times (\frac{1}{2}k(k+1)-1)$ dimensional matrices that contain the orthogonal complements of $vech(G_1)$ and $vech(G_2)$, $vech(G_1)'_{\perp}vech(G_1) \equiv 0$, $vech(G_1)'_{\perp}vech(G_1)_{\perp} \equiv I_{\frac{1}{2}p(p+1)-1}$, $vech(G_2)'_{\perp}vech(G_2) \equiv 0$, $vec(G_2)'_{\perp}$ $vec(G_2)_{\perp} \equiv I_{\frac{1}{2}k(k+1)-1}$. This specification of the hypothesis fits directly in the setup of the KP rank test since the covariance matrix of $\hat{R}^*$ is non-singular so the corresponding specification of $vec(\hat{\Lambda})$ converges to a normal distributed random vector. It therefore also shows the number of restrictions tested, which equals $\left(\frac{1}{2}k(k+1)-1\right)\left(\frac{1}{2}p(p+1)-1\right)$, but the resulting rank statistic does not equal KPST and tests for a KPS on a matrix which differs from R, as shown in the following Theorem.

Theorem 2 *In parts* ***a.***, ***c.***, *and* ***d.*** *assume* $E\left(\|f_i\|^8\right) < \kappa$ *for some* $\kappa < \infty$, $\hat{f}_i = f_i + o_p(1)$, *uniformly for* $i = 1, \ldots, n$, *as* $n \to \infty$, *and that the central limit theorem in (10) holds. Then:*
a. *Under* H_0, *KPST* $\underset{d}{\to} \chi^2_a$ *with*

$$a := \left(\tfrac{1}{2}k(k+1)-1\right)\left(\tfrac{1}{2}p(p+1)-1\right) \tag{20}$$

degrees of freedom.
b. *The expression for KPST simplifies to:*

$$KPST = \ n \times \left(vec\left(\hat{\Sigma}_2\right)\right)' \left[(\hat{N}_2 \otimes \hat{L}_2)'\hat{V}\left(\hat{N}_2 \otimes \hat{L}_2\right)\right]^{-} \left(vec\left(\hat{\Sigma}_2\right)\right). \tag{21}$$

c. *Define KPST* as KPST defined in (17) but with* $\hat{R}^*$ *defined in (9) replacing* $\mathcal{R}(\hat{R})$. *Then under* H_0, *KPST** $\underset{d}{\to} \chi^2_a$. *But KPST and KPST* are not numerically identical and while KPST* is not invariant to orthonormal transformations of the data in* $\hat{V}_i$ *and* Z_i, *KPST is invariant.*
d. *Under* H_0, *for sequences* p, k, n *that satisfy*

$$\tfrac{(pk)^{16}}{n^3} \to 0, \tag{22}$$

we have

$$\lim_{n,p,k\to\infty} \Pr\left[KPST < \chi^2_{a,1-\alpha}\right] \leq \alpha, \tag{23}$$

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

Proof. see the Appendix.[4] ■

Based on Theorem 2a and d, the new KPST test rejects H_0 in (3) at nominal size α if

$$KPST > \chi^2_{a,1-\alpha}.$$

Theorem 2b 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 since $\hat{\Sigma}_2$ does not have an asymptotic normal distribution while $vec(\hat{\Lambda})$ does.

Theorem 2c shows that KPSTs based on $\mathcal{R}(\hat{R})$ and $\hat{R}^*$ are not identical. This difference results since Wald statistics, like KPST, are in general not invariant to non-linear transformations. KPST conducts a test using $\hat{\Lambda}$ which is a non-linear transformation of $\mathcal{R}(\hat{R})$ or $\hat{R}^*$ so while the null hypotheses tested using the specifications $\mathcal{R}(\hat{R})$ and $\hat{R}^*$ are equivalent, Wald tests of these hypotheses are not.

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 slightly less strict than those in Newey and Windmeijer (2009). They prove the validity of the asymptotic approximation of test statistics where the number of observations grows faster than the cube of the number of moment restrictions. The number of moment restrictions here is proportional to $(pk)^2$ so their rate would be $(pk)^6/n \to 0$ which is more restrictive than the rate in (22).

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

$$f_i = \sum_{j=1}^{N_i} f_{ij}, \tag{24}$$

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

$$R = \frac{1}{n} \sum_{i=1}^n E(f_i f_i'). \tag{25}$$

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

3 Simulation study

We evaluate the accuracy of the limiting distribution in Theorem 2 to set critical values for testing for KPS. We do so in a small simulation experiment using the linear regression model:

$$Y_i = Z_i'\Pi + V_i, \qquad i = 1, \ldots, n, \tag{26}$$

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. The test statistic results from the moment vector

$$\hat{f}_i = C_1'\widehat{V}_i \otimes C_2'Z_i, \tag{27}$$

where C_1 and C_2 are used for normalization, see also Kleibergen and Paap (2006). For example, $C_1C_1' = \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_i\widehat{V}_i'\right)^{-1}$ and $C_2C_2' = \left(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\right)^{-1}$. We further set Π to zero (which is without loss of generality since KPST uses the residual vectors) and generate the Z_i's independently from $N(0, I_k)$ distributions and V_i given Z_i independently from a $N\left(0, h\left(Z_i\right) I_p\right)$ distribution. We consider two different specifications of $h(Z_i)$. The first leads to homoskedasticity and has $h(Z_i) = 1$ while the second leads to (scalar) heteroskedasticity and has $h\left(Z_i\right) = \left\|Z_i\right\|^2 / k$. For each case, we compute null rejection probabilities (NRPs) using a nominal significance level of 5%. The NRPs are computed using 40.000 Monte Carlo replications for the KPST that uses the asymptotic critical values resulting 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 1-3 show the NRPs as a function of the sample size n for smaller sample sizes than in Tables 1, 2 for different settings of p and k. Depending on the value of the latter, the NRPs are close to the nominal level for values of n much smaller than $(pk)^4$. For larger values of pk, we therefore do not, like, for the smaller values of pk, show the rejection frequencies all the way up to $n = (pk)^{\frac{16}{3}}$, i.e. the value indicated by Theorem 2d, but just to $(pk)^4$, which is for $p = 2$, $k = 7$ at the bottom right hand side of Figure 1, equal to approximately 40.000, and for $p = 5$, $k = 4$ at the bottom right hand side of Figure 3 equal to 160.000 (note that the horizontal axis is in log-scale). In many cases, the NRPs are still much closer to the 5% significance level 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 when when $p = 2$ and $k = 5$ for which KPSTs using a 5% significance level have NRPs close to the size for values of n around two hunderd. Figures 1-3 also show that the KPS test generally over rejects for small n, which suggests that failure to detect deviations from KPS in small or moderate samples is not likely due to an inflated type 2 error.

Data Generating Process:					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 at various significance levels. χ^2_a critical values. $n = (pk)^{16/3}$, a: number of restrictions given in eq. (20), m: number of estimated parameters. Computed using 40.000 MC replications.

Data Generating Process:					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^2_a critical values. $n = (pk)^4$, a: number of restrictions given in eq. (20), m: number of estimated parameters. Computed using 40.000 MC replications.

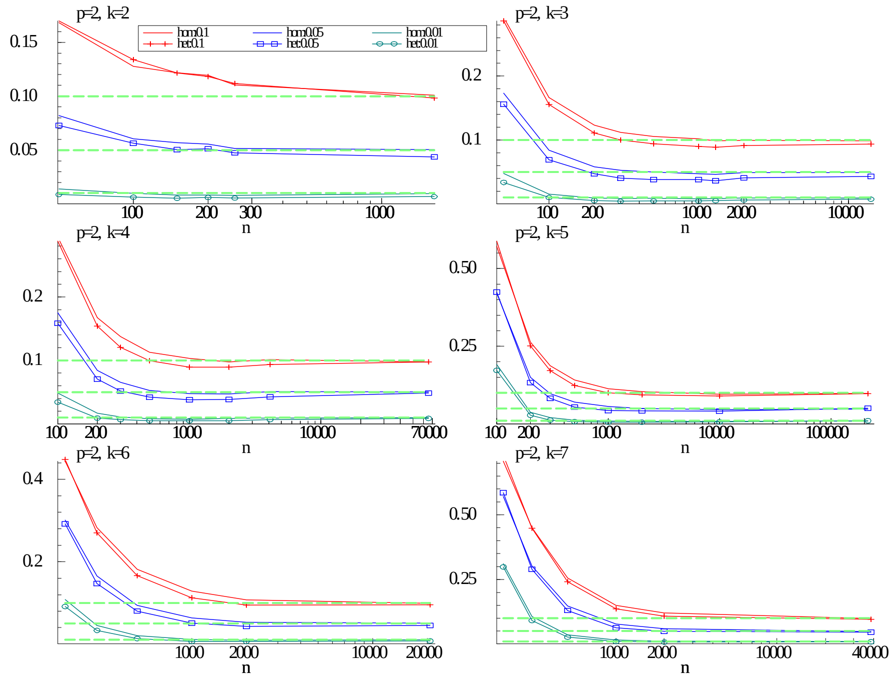

Figure 1: Null rejection probabilities of KPST at different sample sizes and DGPs, hom: conditional homoskedastic; het: scalar heteroskedastic. Computed using 40.000 MC replications.

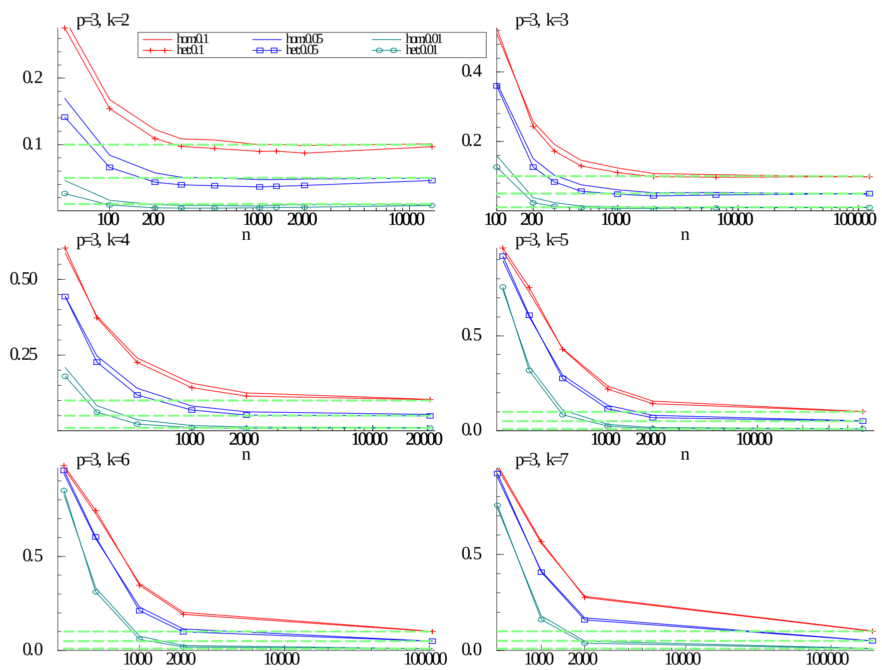

Figure 2: Null rejection probabilities of KPST at different sample sizes and DGPs, hom: conditional homoskedastic; het: scalar heteroskedastic. Computed using 40.000 MC replications.

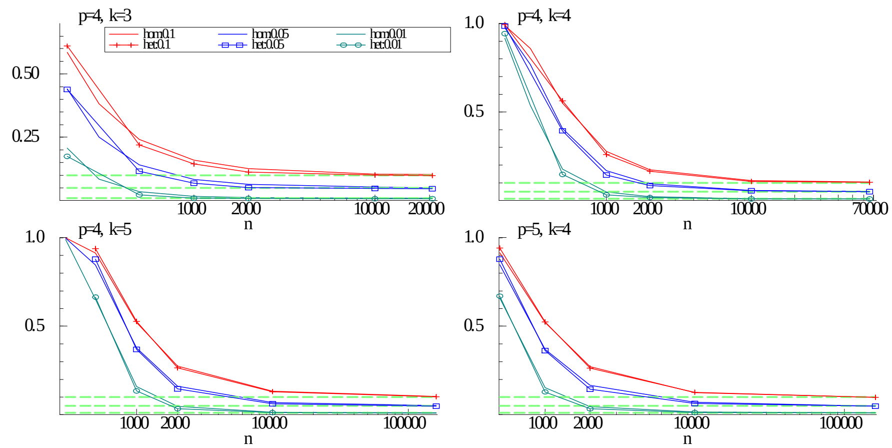

Figure 3: Null rejection probabilities of KPST at different sample sizes and DGPs, hom: conditional homoskedastic; het: scalar heteroskedastic. Computed using 40.000 MC replications.

4 Power

To analyze the power of the KPST test, we analyze settings where the covariance matrix of the moments $R \in \mathbb{R}^{kp\times kp}$ is local to KPS:

$$R = (G_1 \otimes G_2) + \tfrac{1}{\sqrt{n}}A_0, \tag{28}$$

where $G_1 \in \mathbb{R}^{p\times p}$ and $G_2 \in \mathbb{R}^{k\times k}$ are symmetric positive definite matrices and

$$A_0 := \begin{pmatrix} A_{11} & \cdots & A_{1p} \\ \vdots & \ddots & \vdots \\ A_{p1} & \cdots & A_{pp} \end{pmatrix} \in \mathbb{R}^{kp\times kp} \tag{29}$$

is a fixed symmetric matrix, where $A_{ij} \in \mathbb{R}^{k\times k}$ for $i, j = 1, \ldots, p$. The re-arranged matrix $\mathcal{R}(R)$ used to pin down the KPS is:

$$\begin{aligned} \mathcal{R}(R) = &\quad vec(G_1)vec(G_2)' + \tfrac{1}{\sqrt{n}}\mathcal{R}(A_0) \\ = &\quad vec(\bar{G}_{1,n})vec(\bar{G}_{2,n})' + \tfrac{1}{\sqrt{n}}vec(\bar{G}_{1,n})_\perp \Lambda_n vec(\bar{G}_{2,n})'_\perp, \end{aligned} \tag{30}$$

with $\bar{G}_{1,n} \in \mathbb{R}^{p\times p}$ and $\bar{G}_{2,n} \in \mathbb{R}^{k\times k}$ symmetric positive definite matrices potentially different from G_1 and G_2 but converging to them as n goes to infinity.[5] The decomposition in the last line of (30) is identical to the one in (16).

[5] The specification in (30) results from a generic SVD of $\mathcal{R}(A_0)$. For $\bar{G}_{1,n}$ and $\bar{G}_{2,n}$ to equal G_1 and G_2, one needs $vec(G_1)'\mathcal{R}(A_0)vec(G_2) = 0$ which we do not assume but do in the next example.

Theorem 3 *Assume that*

$$\delta := \lim_{n\to\infty} \quad vec(\Lambda_n)' \left[(\left[vec(\bar{G}_{2,n})\right]_{\perp}' \otimes \left[vec(\bar{G}_{1,n})\right]_{\perp}') \right. \\ \left. cov(vec(\mathcal{R}(\hat{R}_n))(\left[vec(\bar{G}_{2,n})\right]_{\perp} \otimes \left[vec(\bar{G}_{1,n})\right]_{\perp}) \right]^{-} vec(\Lambda_n) \tag{31}$$

exists. Then, under local to KPS sequences of covariance matrices as in (28) and for mean zero, independently distributed random vectors $f_i \in \mathbb{R}^{kp}$ with finite eighth moments, KPST has a $\chi^2_a(\delta)$ limiting distribution as $n \to \infty$ (with k, p fixed).

Proof. Follows directly from the proof of Theorem 2 in the Appendix. ■

Power simulation 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 (26). The two dimensional vectors containing the regressors Z_i and errors V_i are simulated according to:

$$V_i \sim iid \begin{cases} N(0, \Omega_1), \\ N(0, \Omega_2), \end{cases} Z_i \sim iid \begin{cases} N(0, Q_{zz,1}), & i = 1, ..., [n/2] \\ N(0, Q_{zz,2}), & i = [n/2] + 1, ..., n, \end{cases} \tag{32}$$

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

$$b := \frac{1}{2}\frac{\sigma}{\sqrt{n}} - \frac{1}{2}\sqrt{\frac{\sigma}{\sqrt{n}}\left(\frac{\sigma}{\sqrt{n}} + 8\right)} + 1, \quad c := \frac{1}{2}\frac{\sigma}{\sqrt{n}} + \frac{1}{2}\sqrt{\frac{\sigma}{\sqrt{n}}\left(\frac{\sigma}{\sqrt{n}} + 8\right)} + 1, \tag{33}$$

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

$$\begin{aligned} R = & \ \tfrac{1}{n} var\left(\textstyle\sum_{i=1}^{n}(V_i \otimes Z_i)\right) = \tfrac{1}{2} diag\,(b + c, 1 + bc, 1 + bc, b + c) \\ = & \ \underbrace{I_4}_{G_1 \otimes G_2} + \tfrac{\sigma}{\sqrt{n}} \times diag\,(1, -1, -1, 1), \end{aligned} \tag{34}$$

and $G_1 = G_2 = I_2$. Since

$$\mathcal{R}(diag\,(1, -1, -1, 1)) = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \tag{35}$$

$vec(G_1)'\mathcal{R}(diag\,(1,-1,-1,1))vec(G_2) = 0$, the re-arranged specification of R (30) has $\bar{G}_{1,n}$ and $\bar{G}_{2,n}$ coinciding with G_1 and G_2 :

$$\begin{aligned} \mathcal{R}(R) = & \quad vec(G_1)vec(G_2)' + \tfrac{\sigma}{\sqrt{n}} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \\ = & \quad vec(G_1)vec(G_2)' + \tfrac{1}{\sqrt{n}} vec(G_1)_\perp \Lambda_n vec(G_2)'_\perp, \end{aligned} \tag{36}$$

with

$$vec(G_1)_\perp = vec(G_2)_\perp = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ -1 & 0 & 0 \end{pmatrix}, \ \Lambda_n = \sigma \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{37}$$

In this case Λ_n does not depend on n because $\bar{G}_{1,n}$ and $\bar{G}_{2,n}$ coincide with G_1 and G_2. For $\sigma = 0$, R has KPS, so the null hypothesis in (3) holds. For the limiting case of $\sigma = \sqrt{n} : b = 0$, so Ω_1 and Ω_2 are singular.

We compute the power function for the KPST test at three significance levels 10%, 5% and 1% for a sample of size $n = 1626 \approx (kp)^{16/3}$, using 10^4 Monte Carlo replications. The results are reported in Figure 4. Alongside the simulated power curve, Figure 4 also shows its asymptotic approximation that results from Theorem 3. The non-centrality parameter of this asymptotic approximation results from noting that

$$\begin{aligned} &(e_1 \otimes e_1)' \Big[(\big[vec(\bar{G}_{2,n}) \big]'_\perp \otimes \big[vec(\bar{G}_{1,n}) \big]'_\perp) \\ &\mathrm{cov}(vec(\mathcal{R}(\hat{R}))(\big[vec(\bar{G}_{2,n}) \big]_\perp \otimes \big[vec(\bar{G}_{1,n}) \big]_\perp) \Big]^- (e_1 \otimes e_1) = \tfrac{1}{4} \end{aligned} \tag{38}$$

where $\bar{G}_{i,n} = G_i = I_2$ for $i = 1, 2$, and $e_1 = (1, 0, 0)$, so, since $vec\,(\Lambda_n) = 2\sigma(e_1 \otimes e_1)$, the noncentrality parameter is

$$\delta = \frac{1}{4}\sigma^2. \tag{39}$$

We see that the asymptotic approximation to the power of the KPST is reasonable, if somewhat optimistic, especially at the 1% level of significance.

5 Empirical applications

We investigate whether KPS covariance matrices are relevant for applied work by analyzing the covariance matrices of estimators in published empirical studies to see if they satisfy KPS. We have therefore taken sixteen highly cited papers conducting instrumental variables 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

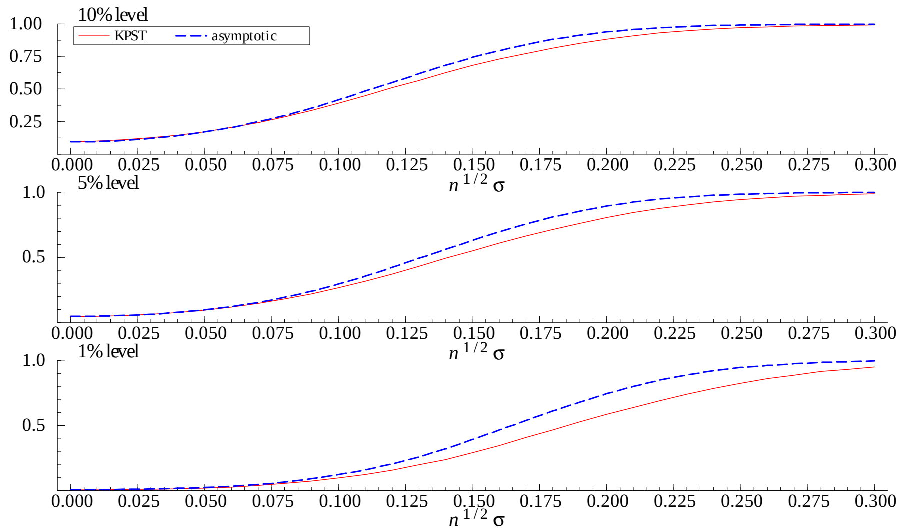

Figure 4: Power of KPST with χ^2 critical values (red solid), and asymptotic power from Theorem 3 (blue dashed). 10000 MC replications. σ measures deviation from KPS in Frobenius norm. Sample size is $n = 1626$.

instruments.[6] The involved papers, and the acronyms we use to refer to them, are listed in Table 3. Tables 6 and 7 in the Supplementary Appendix report our KPS test results for the hundred sixteen different specifications we analyzed. Table 6 does so for the studies using independent data while Table 7 lists the results for studies with clustered data. Since these tables are rather extensive, Tables 4 and 5 report a summary of our findings on the KPS tests.

Table 4, 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 fifty eight different specifications using independent data reported in Table 4, we reject KPS at the 5% significance level for about one third of them: twenty two.

Table 5, summarizing our results for papers using clustered data, shows that for the fifty eight different specifications with clustered data, we reject KPS at the 5% significance level for forty eight specifications when using the unrestricted covariance matrix estimator (6) and for forty when using the clustered covariance matrix estimator (25). 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.

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

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

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 3: List of papers used in the empirical applications.

Paper	# specifications	KPS rejection	# observations
TCN 10	2	none	moderate
Nunn 08	4	4	small
AJ 05	24	10	small
HG 10	2	2	huge
AGN 13	6	1	moderate
Yogo 04	22	5	moderate

Table 4: Summary of results of 5 percent significance KPST tests for specifications in papers using independent observations

Paper	# specific.	KPS rej.	# obs.	clustered KPS rej.	# clusters
DT 11	8	6	large	5	moderate
AJRY 08	9	7	large	5	moderate
JPS 06	4	4	huge	4	huge
PSJM 13	2	2	huge	2	huge
ADH 13	18	18	large	13	small
AD 13	7	7	huge	7	small
ACJR 11	1	1	small	1	very small
MSS 04	3	0	large	3	small
V etal 12	6	2	moderate	0	small

Table 5: Summary of results of 5 percent signficance KPST tests for specifications in papers using clustered observations

6 Conclusions

We propose a test for a covariance matrix of a vector of moment equations to have a KPS. The test is an extension of the KP 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 mostly 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 our 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. The two-step procedure is constructed such that the overall size of the test is controlled. Another 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 commonly present.

Appendix

A Proofs

Proof of Theorem 1: Let

$$\mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}', \tag{40}$$

the SVD of $\mathcal{R}(\hat{R})$ with $\hat{\Sigma} = diag(\hat{\sigma}_1 \dots \hat{\sigma}_{\min(l,q)})$ a $l \times q$ dimensional diagonal matrix with the singular values ordered non-increasingly on the main diagonal, and $\hat{L}$ and $\hat{N}$ $l \times l$ and $q \times q$ dimensional orthonormal matrices, we have

$$\left\| \mathcal{R}(\hat{R}) - vec(G_1)vec(G_2)' \right\|_F^2 = \sum_{i=1}^{\min(l,q)} \hat{\sigma}_i^2 - 2vec(G_1)'\hat{L}\hat{\Sigma}\hat{N}'vec(G_2) + \\ vec(G_1)'vec(G_1)vec(G_2)'vec(G_2).$$

This expression shows that it is minimized with respect to $G_1 > 0, G_2 > 0$ at $vec(\hat{G}_1) = \hat{L}_1/\hat{L}_{11}$, $vec(\hat{G}_2) = \hat{L}_{11}\hat{\sigma}_1\hat{N}_1$ with $\hat{L}_1$ and $\hat{N}_1$ the first columns of $\hat{L}$ and $\hat{N}$ resp. and further defined in (13). (So $(DS)^2 = \sum_{i=2}^{\min(l,q)} \hat{\sigma}_i^2$. Theorem 5.8 in Van Loan and Pitsianis (1993) states that if $\hat{R}$ is symmetric positive definite then symmetric positive definite matrices $\hat{G}_1$ and $\hat{G}_2$ exist that minimize the Frobenius norm. Because the SVD is unique, $\hat{G}_1$ and $\hat{G}_2$ must then be symmetric positive definite, see also Lemma 2 in Guggenberger et al. (2021).

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

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

whose elements can be specified as

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

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 "ˆ"'s) and where under $H_0 : \Sigma_{22} = 0$.

To obtain the limit distribution of the sample analog of the parameter tested under H_0 :

$$\hat{\Lambda} = \quad vec(\hat{G}_1)'_{\perp}\mathcal{R}(\hat{R})vec(\bar{G}_2)_{\perp},$$

we use that $vec(\hat{G}_1)_{\perp} = vec(G_1)_{\perp} + O_p(n^{-\frac{1}{2}})$, $vec(\hat{G}_2)_{\perp} = vec(G_2)_{\perp} + O_p(n^{-\frac{1}{2}})$, which holds under our imposed conditions, see Kleibergen and Paap (2006), so under H_0 :

$$\begin{aligned}
\hat{\Lambda} = &\ vec(\hat{G}_1)'_{\perp}\mathcal{R}(\hat{R})vec(\hat{G}_2)_{\perp} \\
= &\ \left[vec(G_1)_{\perp} + O_p(n^{-\frac{1}{2}})\right]' \left[vec(G_1)vec(G_2)' + \tfrac{1}{\sqrt{n}}D_p\Psi D_k' + o_p(n^{-\frac{1}{2}})\right] \left[vec(G_2)_{\perp} + O_p(n^{-\frac{1}{2}})\right] \\
= &\ \tfrac{1}{\sqrt{n}}vec(G_1)'_{\perp}D_p\Psi D_k' vec(G_2)_{\perp} + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

To construct the limit distribution of $\hat{\Lambda}$, we use that

$$\begin{aligned}
vec(G_1)_{\perp} = &\ L_2L_{22}^{-}(L_{22}L_{22}')^{1/2}, & vec(G_2)_{\perp} = &\ N_2N_{22}^{-}(N_{22}N_{22}')^{1/2}, \\
L_2 = &\ \begin{pmatrix} e'_{1,\frac{1}{2}p(p+1)}A_2 & 0 \\ D_{2,p}A_2 & D_{2,p\perp} \end{pmatrix}, & N_2 = &\ \begin{pmatrix} e'_{1,\frac{1}{2}k(k+1)}B_2 & 0 \\ D_{2,k}B_2 & D_{2,k\perp} \end{pmatrix}, \\
L_{22} = &\ \begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp} \end{pmatrix}, & N_{22} = &\ \begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp} \end{pmatrix}, \\
A = &\ \begin{pmatrix} a_1 & A_2 \end{pmatrix}, & B = &\ \begin{pmatrix} b_1 & B_2 \end{pmatrix},
\end{aligned}$$

where we use that $D_p = (e_{1,\frac{1}{2}p(p+1)} \vdots D'_{2,p})'$, $D_{2,p} : (p^2 - 1) \times \frac{1}{2}p(p+1)$ and $D_k = (e_{1,\frac{1}{2}k(k+1)} \vdots D'_{2,k})'$, $D_{2,k} : (k^2 - 1) \times \frac{1}{2}k(k+1)$ with $e_{1,i}$ the first i dimensional unity vector (i.e. the first column of I_i). We also use that $a_1 : \frac{1}{2}p(p+1) \times 1$, $A_2 : \frac{1}{2}p(p+1) \times (\frac{1}{2}p(p+1) - 1)$, $b_1 : \frac{1}{2}k(k+1) \times 1$, $B_2 : \frac{1}{2}k(k+1) \times (\frac{1}{2}k(k+1) - 1)$, $D_{p\perp} = (0 \vdots D'_{2,p\perp})'$, $D_{2,p\perp} : (p^2 - 1) \times \frac{1}{2}p(p-1)$, $D_{k\perp} = (0 \vdots D'_{k,p\perp})'$, $D_{k,p\perp} : (k^2 - 1) \times \frac{1}{2}k(k-1)$, where the specifications of $D_{p\perp}$ and $D_{k\perp}$ result from those of D_p and D_k.

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

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

with $L_{D_{2p}A_2}$ and $L_{D_{2k}B_2}$ orthonormal $(\frac{1}{2}p(p+1)-1) \times (\frac{1}{2}p(p+1)-1)$ and $(\frac{1}{2}k(k+1)-1) \times (\frac{1}{2}k(k+1)-1)$ dimensional matrices and $\Lambda^2_{D_{2,p}A_2}$ and $\Lambda^2_{D_{2,k}B_2}$ diagonal $(\frac{1}{2}p(p+1) - 1) \times (\frac{1}{2}p(p+1) - 1)$ and $(\frac{1}{2}k(k+1) - 1) \times (\frac{1}{2}k(k+1) - 1)$ dimensional matrices with the squared singular values in non-increasing order on the diagonal. The above spectral decomposition feature in the SVDs of

L_{22}, and N_{22}, using which we can specify:

$$
\begin{aligned}
L_{22} = & \begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp} \end{pmatrix} \\
= & \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2}\Lambda_{D_{2,p}A_2}L'_{D_{2,p}A_2} & D_{2,p\perp} \end{pmatrix} \\
= & \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \Lambda_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\
& \begin{pmatrix} L'_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix}, \\
(L_{22}L'_{22})^{\frac{1}{2}} = & \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \Lambda_{D_{2,p}A} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \\
& \begin{pmatrix} D_{2,p}A_2(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}L'_{D_{2p}A_2})^{-\frac{1}{2}}L_{D_{2p}A_2} & D_{2,p\perp} \end{pmatrix}', \\
L^-_{22}\,(L_{22}L'_{22})^{1/2} = & \begin{pmatrix} L_{D_{2,p}A_2} & 0 \\ 0 & I_{\frac{1}{2}p(p-1)} \end{pmatrix} \begin{pmatrix} L'_{D_{2p}A_2}(L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}P'_{D_{2p}A_2})^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} \\
= & \begin{pmatrix} (L_{D_{2p}A_2}\Lambda^2_{D_{2,p}A_2}P'_{D_{2p}A_2})^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix} = \begin{pmatrix} (A'_2D'_{2,p}D_{2,p}A_2)^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}, \\
L_2L^-_{22}\,(L_{22}L'_{22})^{1/2} = & \begin{pmatrix} D_pA_2 & D_{p,\perp} \end{pmatrix} \begin{pmatrix} (A'_2D'_{2,p}D_{2,p}A_2)^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}, \\
N_2N^-_{22}\,(N_{22}N'_{22})^{1/2} = & \begin{pmatrix} D_kB_2 & D_{k,\perp} \end{pmatrix} \begin{pmatrix} (B'_2D'_{2,k}D_{2,k}B_2)^{-\frac{1}{2}}B'_2D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix},
\end{aligned}
$$

where in the third line of the first equation we have the three components of the SVD and we note that the specification of the Moore-Penrose generalized inverse L^-_{22} results directly from the SVD of L_{22} since L_{22} is invertible.

Then, under H_0:

$$
\begin{aligned}
\sqrt{n}\hat{\Lambda} = & \; vec(G_1)'_\perp D_p\Psi D'_k vec(G_2)_\perp + o_p(1) \\
= & \; (L_{22}L'_{22})^{1/2}L^-_{22}L'_2D_p\Psi D'_kN_2N^-_{22}\,(N_{22}N'_{22})^{1/2} + o_p(1) \\
= & \begin{pmatrix} (A'_2D'_{2,p}D_{2,p}A_2)^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}' \begin{pmatrix} D_pA_2 & D_{p,\perp} \end{pmatrix}' D_p\Psi D'_k \\
& \begin{pmatrix} D_kB_2 & D_{k,\perp} \end{pmatrix} \begin{pmatrix} (B'_2D'_{2,k}D_{2,k}B_2)^{-\frac{1}{2}}B'_2D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} + o_p(1) \\
= & \begin{pmatrix} (A'_2D'_{2,p}D_{2,p}A_2)^{-\frac{1}{2}}A'_2D'_{2,p} \\ D'_{2,p\perp} \end{pmatrix}' \begin{pmatrix} A'_2D'_pD_p\Psi D'_kD_kB_2 & 0 \\ 0 & 0 \end{pmatrix} \\
& \begin{pmatrix} (B'_2D'_{2,k}D_{2,k}B_2)^{-\frac{1}{2}}B'_2D'_{2,k} \\ D'_{2,k\perp} \end{pmatrix} + o_p(1) \\
= & \; D_{2,p}A_2\bar{\Lambda}B'_2D'_{2,k} + o_p(1),
\end{aligned}
\tag{41}
$$

with

$$
\bar{\Lambda} = \; (A'_2D'_{2,p}D_{2,p}A_2)^{-\frac{1}{2}}A'_2D'_pD_p\Psi D'_kD_kB_2(B'_2D'_{2,k}D_{2,k}B_2)^{-\frac{1}{2}},
$$

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

$$\begin{aligned} V_{vec(\bar{\Lambda})} = & \left((B_2'D_k'D_kB_2)^{-\frac{1}{2}}B_2'D_k'D_k \otimes (A_2'D_p'D_pA_2)^{-\frac{1}{2}}A_2'D_p'D_p\right)V_{R^*} \\ & \times\left((B_2'D_k'D_kB_2)^{-\frac{1}{2}}B_2'D_k'D_k \otimes (A_2'D_p'D_pA_2)^{-\frac{1}{2}}A_2'D_p'D_p\right)'. \end{aligned}$$

The above shows that the limit behavior 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 so their 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 (17). The matrix $\hat{V}$ contains summands of eighth order products of f_i and the weak law of large numbers holds by the assumption that $E\left(\|f_i\|^8\right)<\kappa$. The convergence of the covariance matrix estimator involved in KPST is characterized by:

$$\begin{aligned}
&\left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right)'\hat{V}\left(vec(\hat{G}_2)_\perp \otimes vec(\hat{G}_1)_\perp\right) && \underset{p}{\rightarrow} \\
&(vec(G_2)_\perp \otimes vec(G_1)_\perp)'(D_k\otimes D_p)V_{R^*}(D_k\otimes D_p)\left(vec(G_2)_\perp \otimes vec(G_1)_\perp\right) && = \\
&\left(\begin{pmatrix} D_kB_2 & D_{k,\perp}\end{pmatrix}\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}'\end{pmatrix}\otimes\right. \\
&\left.\begin{pmatrix} D_pA_2 & D_{p,\perp}\end{pmatrix}\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}'\end{pmatrix}\right)'(D_k\otimes D_p)V_{R^*}(D_k\otimes D_p)' \\
&\left(\begin{pmatrix} D_kB_2 & D_{k,\perp}\end{pmatrix}\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_{2,k}' \\ D_{2,k\perp}'\end{pmatrix}\otimes\right. \\
&\left.\begin{pmatrix} D_pA_2 & D_{p,\perp}\end{pmatrix}\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_{2,p}' \\ D_{2,p\perp}'\end{pmatrix}\right) && = \\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right) \\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \\ 0\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p \\ 0\end{pmatrix}\right)V_{R^*} \\
&\left(\begin{pmatrix}(B_2'D_{2,k}'D_{2,k}B_2)^{-\frac{1}{2}}B_2'D_k'D_k \\ 0\end{pmatrix}\otimes\begin{pmatrix}(A_2'D_{2,p}'D_{2,p}A_2)^{-\frac{1}{2}}A_2'D_p'D_p \\ 0\end{pmatrix}\right)' \\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)' && = \\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)\begin{pmatrix} V_{vec(\bar{\Lambda})} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0\end{pmatrix} \\
&\left(\begin{pmatrix} D_{2,k}B_2 & D_{2,k\perp}\end{pmatrix}\otimes\begin{pmatrix} D_{2,p}A_2 & D_{2,p\perp}\end{pmatrix}\right)'
\end{aligned}$$

The convergence behavior of KPST is then characterized by:

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

with $a = \left(\frac{1}{2}p(p+1)-1\right)\left(\frac{1}{2}k(k+1)-1\right)$. The first part on the top line results from the convergence behavior of $vec(\hat{G}_1)_{\perp}$ and the second equality results from (41) and $vec(\hat{G}_2)_{\perp}$. The expression in the third equality follows since

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

b. Because $\hat{L}_{22}^-(\hat{L}_{22}\hat{L}_{22}')^{1/2}$ and $\hat{N}_{22}^-\left(\hat{N}_{22}\hat{N}_{22}'\right)^{1/2}$ are (almost surely) invertible, KPST can be rewritten as:

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

$$
\begin{aligned}
&= n \times \left(vec\left(\hat{\Sigma}_2\right)\right)' \left[(\hat{N}_2 \otimes \hat{L}_2)'\hat{V}\left(\hat{N}_2 \otimes \hat{L}_2\right)\right]^{-} \left(vec\left(\hat{\Sigma}_2\right)\right) \\
&= n \times \left(vec\left(\left(\begin{array}{cc} \hat{\Sigma}_{22} & 0 \\ 0 & 0 \end{array}\right)\right)\right)' \left[(\hat{N}_2 \otimes \hat{L}_2)'(D_k \otimes D_p)\hat{V}_{\hat{R}^*_N}\right. \\
&\qquad \left.(D_k \otimes D_p)'\left(\hat{N}_2 \otimes \hat{L}_2\right)\right]^{-} \left(vec\left(\left(\begin{array}{cc} \hat{\Sigma}_{22} & 0 \\ 0 & 0 \end{array}\right)\right)\right) \\
&= n \times vec\left(\hat{\Sigma}_{22}\right)' \left(\left(\begin{array}{cc} I_{\frac{1}{2}k(k+1)-1} & 0 \end{array}\right) \otimes \left(\begin{array}{cc} I_{\frac{1}{2}p(p+1)-1} & 0 \end{array}\right)\right) \\
&\qquad \left[\left(\left(\begin{array}{cc} D_k'D_k\hat{B}_2 & 0 \end{array}\right) \otimes \left(\begin{array}{cc} D_p'D_p\hat{A}_2 & 0 \end{array}\right)\right)' \hat{V}_{\hat{R}^*} \left(\left(\begin{array}{cc} D_k'D_k\hat{B}_2 & 0 \end{array}\right) \otimes \left(\begin{array}{cc} D_p'D_p\hat{A}_2 & 0 \end{array}\right)\right)\right]^{-} \\
&\qquad \left(\left(\begin{array}{cc} I_{\frac{1}{2}k(k+1)-1} & 0 \end{array}\right) \otimes \left(\begin{array}{cc} I_{\frac{1}{2}p(p+1)-1} & 0 \end{array}\right)\right)' vec\left(\hat{\Sigma}_{22}\right) \\
&= n \times vec\left(\hat{\Sigma}_{22}\right)' \left[\left(\hat{B}_2'D_k'D_k \otimes \hat{A}_2'D_p'D_p\right)\hat{V}_{\hat{R}^*}\left(D_k'D_k\hat{B}_2 \otimes D_p'D_p\hat{A}_2\right)\right]^{-1} vec\left(\hat{\Sigma}_{22}\right).
\end{aligned}
$$

c. We show that if we replaced $\mathcal{R}(\hat{R}) = D_p\hat{R}^*D_k'$, 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, we obtain KPST*, where the latter is defined as the statistic KPST when we use $\hat{R}^*$ instead of $\mathcal{R}(\hat{R})$. To show this, we use SVDs of $\bar{R} = \bar{L}\bar{\Sigma}\bar{N}'$ and $\hat{R}^* = \hat{L}_{R^*}\hat{\Sigma}_{R^*}\hat{N}_{R^*}'$ which are related through:

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

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

$$
\begin{aligned}
KPST_{\bar{R}} = \; & n \times \left[vec(vec(\bar{G}_1)_\perp'\bar{R}vec(\bar{G}_2)_\perp)\right]' \\
& \left[\left(vec(\bar{G}_2)_\perp' \otimes vec(\bar{G}_1)_\perp'\right)\hat{V}_{\bar{R}}\left(vec(\bar{G}_2)_\perp \otimes vec(\bar{G}_1)_\perp\right)\right]^{-} \\
& \left[vec(vec(\bar{G}_1)_\perp'\bar{R}vec(\bar{G}_2)_\perp)\right]
\end{aligned}
$$

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

which is the KPST expression using $\hat{R}^*$ so it differs from KPST and where we also used that:

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

Since $\hat{R}^*$ has the non-degenerate limiting distribution (10), the limiting distribution of KPST using $\hat{R}^*$ directly results from Kleibergen and Paap (2006) and is also χ^2_a.

To show (non-) invariance to orthonormal transformations of $\hat{V}_i$ and Z_i, we consider a $p \times p$

dimensional orthonormal matrix Q using which we rotate $\hat{V}_i$ to become $Q\hat{V}_i$ so

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

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

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

and

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

Hence, since Q and G are orthonormal, this implies that the different components of the SVD decomposition of $\mathcal{R}(\hat{R})$ in (12) with the transformed $\hat{R}$ become $(Q\otimes Q)\hat{L}, \hat{\Sigma}$ and $(G\otimes G)\hat{N}$. Since these rotations also transform the covariance matrix $\hat{V}$ to $(Q\otimes Q)\hat{V}(G\otimes G)'$, it immediately follows from the expression in KPST in (21) that KPST is invariant to rotations of V_i and Z_i. This is, however, not so for KPST* because $(D_p'D_p)^{-1}D_p'(Q\otimes Q)D_p$ and $(D_k'D_k)^{-1}D_k'(G\otimes G)'D_k$ are not orthonormal so the singular vectors that result from the SVD of transformed $\hat{R}^*$ are not mere multiplications of the singular vectors of untransformed $\hat{R}^*$ by $(D_p'D_p)^{-1}D_p'(Q\otimes Q)D_p$ and $(D_k'D_k)^{-1}D_k'(G\otimes G)'D_k$. This transformation thus also alters the singular values which implies that KPST* is not invariant to orthonormal transformations of $\hat{V}_i$ and Z_i.

d. Under H_0 and joint limit sequences of k, p and n, we have to consider all components of $vec(\hat{\Lambda})$ and its covariance matrix estimator. We can then specify $vec(\hat{\Lambda})$ as in (15):

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

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

where we used that $\mathcal{R}(R) = vec(G_1)vec(G_2)'$, which results under H_0 and because of (5), so $vec(G_1)'_\perp \mathcal{R}(R) = 0$, $\mathcal{R}(R)vec(G_2)_\perp = 0$, and

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

The limit behavior of KPST results from the limit behavior of a where we note that we specified both $vec(G_1)_\perp$ and $vec(G_2)_\perp$, whose dimensions increase as k and p get larger, as orthonormal matrices, $vec(G_1)'_\perp vec(G_1)_\perp \equiv I_{\frac{1}{2}k(k-1)}$ and $vec(G_2)'_\perp vec(G_2)_\perp \equiv I_{\frac{1}{2}p(p-1)}$. Hence the length of each column of $vec(G_1)_\perp$ and $vec(G_2)_\perp$ equals one and does not change when k and/or p increase.

From (10), it follows that $\mathcal{R}(\hat{R}) - \mathcal{R}(R) = O_p(n^{-\frac{1}{2}})$, and so the same holds for the convergences rates of $vec(\hat{G}_1)_\perp - vec(G_1)_\perp$ and $vec(\hat{G}_2)_\perp - vec(G_2)_\perp$, see Kleibergen and Paap (2006). Since $vec(\hat{G}_1)_\perp$ and $vec(\hat{G}_2)_\perp$ are solved from $\mathcal{R}(\hat{R})$, $\mathcal{R}(\hat{R}) - \mathcal{R}(R)$, $vec(\hat{G}_1)_\perp - vec(G_1)_\perp$ and $vec(\hat{G}_2)_\perp - vec(G_2)_\perp$ are all jointly dependent. In a limiting sequence where the dimensions p and k jointly increase with the sample size n, we then have the following convergence rates:

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

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:

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

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

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

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

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

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

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

which is of order $O_p(\frac{(pk)^2}{n})$. 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 $O_p(\frac{(pk)^2}{n})$ convergence rate.

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

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

For the limit behavior of $\sqrt{n}\hat{\Lambda}$ to just result from 1, so the limit behavior of KPST remains unaffected, it is then sufficient to have joint limit sequences that satisfy:

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

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

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

where below we show that the maximal convergence rates besides the zero-th order component are $O_p(n^{-\frac{1}{2}})$, $O_p\left(\frac{(pk)^2}{n}\right)$, $O_p\left(\frac{k^4}{n}\right)$, $O_p\left(\frac{p^4}{n}\right)$ and $O_p\left(\frac{k^4p^4}{n\sqrt{n}}\right)$. All these rates appear in an identical manner in the inverse of the estimator of the covariance matrix.[7] When taking the resulting inverse and accounting for the summations over the k^2p^2 components in $vec(\hat{\Lambda})$, we obtain a slightly stronger condition than just for $\hat{\Lambda}$:

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

which results from the $O_p(n^{-\frac{1}{2}})$ components from the inverse of the covariance matrix estimator paired with the $O_p(\frac{(pk)^2}{n})$ components from $\hat{\Lambda}$ corrected for the multiplication by n and the double summation over p^2k^2 components.[8] 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 (2006) which would be $\frac{k^4p^4}{n}$ for $\hat{\Lambda}$ and $\frac{k^6p^6}{n}$ for convergence of the test statistic which is slightly stricter than our rate of $\frac{(pk)^{16}}{n^3} \rightarrow 0$.

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 since we just showed that they do not lead to the largest error rate because the $O_p\left(\frac{k^4p^4}{n\sqrt{n}}\right)$ is less than the $O_p\left(\frac{p^2k^2}{n}\right)$ that results from some of the C components.

$$\begin{array}{rll}
A_1 = & (vec(G_2)'_\perp \otimes vec(G_1)'_\perp)\,\mathrm{cov}(vec(\mathcal{R}(\hat{R})))\,(vec(G_2)_\perp \otimes vec(G_1)_\perp) & = O(1) \\
B_1 = & \left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right]' \otimes vec(G_1)'_\perp\right)\mathrm{cov}(vec(\mathcal{R}(\hat{R}))) & \\
& (vec(G_2)_\perp \otimes vec(G_1)_\perp) + (vec(G_2)'_\perp \otimes vec(G_1)'_\perp) & \\
& \mathrm{cov}(vec(\mathcal{R}(\hat{R})))\left(\left[vec(\hat{G}_2)_\perp - vec(G_2)_\perp\right] \otimes vec(G_1)_\perp\right) & = O_p(n^{-\frac{1}{2}}) \\
B_2 = & \left(vec(G_2)'_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]'\right)\mathrm{cov}(vec(\mathcal{R}(\hat{R}))) & \\
& (vec(G_2)'_\perp \otimes vec(G_1)'_\perp)' + (vec(G_2)'_\perp \otimes vec(G_1)'_\perp) & \\
& \mathrm{cov}(vec(\mathcal{R}(\hat{R})))\left(vec(G_2)_\perp \otimes \left[vec(\hat{G}_1)_\perp - vec(G_1)_\perp\right]\right) & = O_p(n^{-\frac{1}{2}}) \\
B_3 = & (vec(G_2)'_\perp \otimes vec(G_1)'_\perp)\left[\widehat{\mathrm{cov}}(vec(\mathcal{R}(\hat{R}))) - \mathrm{cov}(vec(\mathcal{R}(\hat{R})))\right] & \\
& (vec(G_2)'_\perp \otimes vec(G_1)'_\perp)' & = O_p(n^{-\frac{1}{2}})
\end{array}$$

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

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

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

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, 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.

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. *Econometric Theory* **23**, 1217–1232.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* **73**, 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 approximatioms 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. *Quantitive Economics* **10**, 487–526.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2021). A powerful Subvector Anderson Rubin Test in Linear Instrumental Variables Regression with Conditional Heteroskedasticity. *Working Paper*.

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**, 203–235.

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**, 346–372.

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. *Journal of Finance* **75**, 507–550.

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. *Bernouilli* **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 ccovariance. *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.

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.

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, raninfall*Republican incumbent	4	3	27401	507.919	0
	T1.C3	Democratic vote share, turnout, turnout * partisan composition, turnout * Republican incumbent	Rainfall, rainfall*partisan composition, raninfall*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**
AJRY 08										
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 X: democracy in t-2, t-3	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-1, 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-2	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	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
JPS 06										
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
PSJM 13										
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
ADH 13										
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
AD 13										
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.	Y	Z	p	k	n	**KPST**	**p val**	n_c	**KPST**$_c$	**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
ACJR 11										
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
MSS 04										
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**
V etal 12										
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	18.69	0.028	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.