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 $\chi^2$ 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. This adaptation in particular

\footnote{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 \( \chi^2 \) 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 \( \chi^2 \) 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 hypotheses on a subvector of the structural parameters for general covariance matrices but is less powerful than the improved subvector Anderson-Rubin test from Guggenberger et al. (2019) in the linear IV regression model 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 homoskedasticity to KPS or even further by extending the switching test procedure from Guggenberger et al. (2021) would extend the scope of the weak factor robust inference methods for analyzing the individual risk premia in linear asset pricing models. The KPS test would be an integral part of such extensions.

KPS or separability, which is how other fields sometimes refer to KPS, of the covariance matrix is also studied in the statistics and signal processing literature. The distance to a covariance matrix with KPS is considered in Genton (2007) and Velu and Herman (2017), while Lu and Zimmermann (2005) and Mitchell et al. (2006) analyze the likelihood ratio test of KPS of the covariance matrix of Normally distributed data. They estimate the elements of the KPS covariance matrix using a switching algorithm. Exploiting the reduced rank restriction imposed on the reordered covariance matrix by KPS is also done in Werner et al. (2008). Their results are, however, based on a complex Gaussian distribution for the data, which leads to a degrees of freedom parameter of the $\chi^2$ 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 $A$, $\text{vec}(A) = (a'_1 \ldots a'_k)' \in \mathbb{R}^{mk}$ for a $m \times k$ dimensional matrix $A = (a_1 \ldots a_k)$. 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

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

for mean zero, independently distributed random vectors $f_i \in \mathbb{R}^{kp}$, $i = 1, \ldots, n$, which satisfy $f_i = (V_i \oplus Z_i)$, with $V_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^k$ uncorrelated random vectors. The specification of $f_i$ fits, for example, a setting where $V_i$ contains the errors of a number of regression equations and $Z_i$

---

2We 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 $n$. 

4
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

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

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

$$H_0 : R = G_1 \otimes G_2,$$

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

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

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

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

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

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

with $\|\cdot\|_F$ the Frobenius or trace norm of a matrix, $\|A\|_F^2 := \text{tr}(A' A) = \text{vec}(A)' \text{vec}(A)$, for any rectangular matrix $A$. Because $\mathcal{R}(G_1 \otimes G_2)$ is a matrix of rank one, 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

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

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 $n \rightarrow \infty$. Define the distance from a KPS covariance
matrix by the Frobenius norm:

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

(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 \( \hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_{\min(p^2, k^2)} \) are the ordered singular values of \( \mathcal{R} (\hat{R}) \).

**Proof.** see the Appendix. \( \blacksquare \)

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 \( D_p \):

\[ \mathcal{R}(\hat{R}) = \mathcal{R} \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{V}_i \hat{V}_i' \otimes Z_i Z_i') \right) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \text{vec}(\hat{V}_i \hat{V}_i' \otimes Z_i Z_i') \text{vec}(Z_i Z_i')' \]
\[ = D_p R^* D_k^*, \]

(8)

with

\[ \hat{R}^* := \frac{1}{n} \sum_{i=1}^{n} \text{vec}(\hat{V}_i \hat{V}_i' \otimes Z_i Z_i')'. \]

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

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

(10)

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

\[ R^* := E \left( \frac{1}{n} \sum_{i=1}^{n} \text{vec}(V_i V_i' \otimes Z_i Z_i') \right), \]
\[ V_{R^*} := \lim_{n \to \infty} E \left( \frac{1}{n} \sum_{i=1}^{n} \left( \text{vec}(Z_i Z_i') \text{vec}(Z_i Z_i')' \otimes \text{vec}(V_i V_i') \text{vec}(V_i V_i')' \right) \right). \]

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

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

(12)

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

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

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:\(^3\)

\[
vec(\hat{G}_1) := \left( \frac{\hat{L}_{11}}{\hat{L}_{21}} \right) / \hat{L}_{11} = \begin{pmatrix} \hat{L}_{11} / \hat{L}_{21} \\ \hat{L}_{22} \end{pmatrix}^{1/2} \quad : \quad p^2 \times 1,
vec(\hat{G}_1) : = \hat{L}_2 \hat{L}_{22}^{1/2} \hat{L}_{22} (\hat{L}_{22})^{1/2} \quad : \quad p^2 \times (p^2 - 1),
vec(\hat{G}_2) : = \hat{L}_1 \hat{\sigma}_1 \left( \hat{N}_{11} : \hat{N}_{21} \right) = \hat{L}_{11} \hat{\sigma}_1 \hat{N}_{11}^{T} \quad : \quad 1 \times k^2,
vec(\hat{G}_2)_1 : = \left( \hat{N}_{22} \hat{N}_{22}^{T} \right)^{1/2} \hat{N}_{22}^{T} \quad : \quad (k^2 - 1) \times k^2,
\]
and \( A^\perp \) is the Moore-Penrose generalized inverse of a matrix \( A \). Define:

\[
\hat{A} := \left( \hat{L}_{22} \hat{L}_{22}^{1/2} \right) \hat{L}_{22} \hat{\Sigma}_2 \hat{N}_{22}^{T} \left( \hat{N}_{22} \hat{N}_{22}^{T} \right)^{-1/2} \quad : \quad (p^2 - 1) \times (k^2 - 1).
\]

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

\[
\mathcal{R}(\hat{A}) = vec(\hat{G}_1) vec(\hat{G}_2)^T + vec(\hat{G}_1) \hat{A} vec(\hat{G}_2)^T.
\]

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

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

\[
KPST := n \times vec(\hat{A})^T \left( \hat{J} \hat{V} \hat{J} \right)^{-1} vec(\hat{A}), \quad \text{where}
\hat{J} := \left[ vec(\hat{G}_2) \right] \otimes \left[ vec(\hat{G}_1) \right], \quad \hat{V} := \hat{\text{cov}}(vec(\mathcal{R}(\hat{A}))) \in \mathbb{R}^{p^2 k^2 \times p^2 k^2},
\]
and

\[
\hat{\text{cov}}(vec(\mathcal{R}(\hat{A}))) = \frac{1}{n} \sum_{i=1}^{n} \left( vec(Z_i Z_i^T) \otimes vec(\hat{V}_i \hat{V}_i^T) \right), \quad \text{where}
\hat{\text{cov}}(vec(\mathcal{R}(\hat{A}))) = \frac{1}{n} \sum_{i=1}^{n} \left( vec(Z_i Z_i^T) \otimes vec(\hat{V}_i \hat{V}_i^T) \right).
\]

\(^3\)If \( A \) is a positive semi definite \( m \times m \) symmetric matrix \( A = EL^2 E^T \), 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 \( A^{1/2} := EL^2 E^T \) and \( A^{-1/2} := E L^{-1} E^T \), where \( L_1 \) is a diagonal matrix containing the non-zero eigenvalues of \( A \) and \( E_1 \) consists of the corresponding eigenvectors.

---

7
Using $\mathcal{R}(R)$, our hypothesis of interest $H_0$ (3) is transformed into

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

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

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

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

where $\text{vech}(G_1)_\perp$ and $\text{vech}(G_2)_\perp$ are $(1/2)p(p + 1) \times ((1/2)p(p + 1) - 1)$ and $1/2k(k + 1) \times ((1/2)k(k + 1) - 1)$ dimensional matrices that contain the orthogonal complements of $\text{vech}(G_1)$ and $\text{vech}(G_2)$, $\text{vech}(G_1)'_\perp \text{vech}(G_1) = 0$, $\text{vech}(G_1)'_\perp \text{vech}(G_1)_\perp = I_{(1/2)p(p+1)-1}$, $\text{vech}(G_2)'_\perp \text{vech}(G_2) = 0$, $\text{vech}(G_2)'_\perp \text{vech}(G_2)_\perp = I_{(1/2)k(k+1)-1}$. 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 $\text{vech}(\hat{\Lambda})$ converges to a normal distributed random vector. It therefore also shows the number of restrictions tested, which equals $(1/2)k(k+1) - 1)(1/2)p(p+1) -1$, 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 $E\left(\|f_i\|^6\right) < \kappa$ 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 $H_0$, $\text{KPST} \xrightarrow{d} \chi^2_a$ with

$$a := (1/2)k(k+1) - 1)(1/2)p(p+1) -1)$$

degrees of freedom.

b. The expression for KPST simplifies to:

$$\text{KPST} = n \times \left(\text{vech} \left(\Sigma_2\right)\right)' \left(\hat{N}_2 \otimes \hat{L}_2\right)' \hat{V} \left(\hat{N}_2 \otimes \hat{L}_2\right) \cdot \left(\text{vech} \left(\hat{\Sigma}_2\right)\right).$$

(21)

c. Define $\text{KPST}^*$ as KPST defined in (17) but with $\hat{R}^*$ defined in (9) replacing $\mathcal{R}(\hat{R})$. Then under $H_0$, $\text{KPST}^* \xrightarrow{d} \chi^2_a$. But KPST and $\text{KPST}^*$ are not numerically identical and while $\text{KPST}^*$ 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

$$\frac{(pk)^{16}}{n^{36}} \to 0,$$

(22)
we have

\[ \lim_{n,p,k \rightarrow \infty} \Pr [KPST < \chi^2_{a,1-\alpha}] \leq \alpha, \]

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

**Proof.** see the Appendix.4

Based on Theorem 2a and d, the new KPST test rejects \( H_0 \) in (3) at nominal size \( \alpha \) 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 \( \chi^2 \) limiting distribution since \( \Sigma_2 \) does not have an asymptotic normal distribution while \( vec(\hat{\Lambda}) \) does.

Theorem 2c shows that KPSTs based on \( R(\hat{R}) \) 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 \( R(\hat{R}) \) or \( \hat{R}^* \) so while the null hypotheses tested using the specifications \( R(\hat{R}) \) 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 \( \sum_{i=1}^{n} N_i \):

\[ f_i = \sum_{j=1}^{N_i} f_{ij}, \]

(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., \( E(f_{ij}f_{is}) \) is unrestricted for all \( j, s = 1, \ldots, N_i \), while observations across clusters are independent. The \( kp \times kp \) dimensional (positive semi-definite) covariance matrix of the sample moments then results as:

\[ R = \frac{1}{n} \sum_{i=1}^{n} E(f_if_i^\prime). \]

(25)

4We 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:

\[ Y_i = Z_i^0 \Pi + V_i, \quad i = 1, \ldots, n, \]  

(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

\[ \hat{f}_i = C_1 \hat{V}_i \otimes C_2 Z_i, \]  

(27)

where \( C_1 \) and \( C_2 \) are used for normalization, see also Kleibergen and Paap (2006). For example, \( C_1 C_1' = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i \hat{V}_i' \right)^{-1} \) and \( C_2 C_2' = \left( \frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \). We further set \( \Pi \) 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 \( N (0, h(Z_i) I_p) \) distribution. We consider two different specifications of \( h(Z_i) \). The first leads to homoskedasticity and has \( h(Z_i) = 1 \) while the second leads to (scalar) heteroskedasticity and has \( h(Z_i) = \|Z_i\|^2 / k \). For each case, we compute null rejection probabilities (NRPs) using 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 \( n = (pk)^{16/3} \), 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 \( p = 2 \) and \( k = 5 \) for which KPSTs using a 5% significance level have NRPs close to the size for values of \( n \) around two hundred. 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:

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>n</th>
<th>a</th>
<th>m</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>1626</td>
<td>4</td>
<td>9</td>
<td>10.0</td>
<td>5.1</td>
<td>1.0</td>
<td>9.7</td>
<td>4.4</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>14130</td>
<td>10</td>
<td>18</td>
<td>10.0</td>
<td>5.0</td>
<td>0.8</td>
<td>9.3</td>
<td>4.2</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>65536</td>
<td>18</td>
<td>30</td>
<td>9.4</td>
<td>5.0</td>
<td>0.9</td>
<td>9.7</td>
<td>4.9</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>215444</td>
<td>28</td>
<td>45</td>
<td>9.8</td>
<td>4.7</td>
<td>0.9</td>
<td>9.8</td>
<td>5.1</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>14130</td>
<td>10</td>
<td>18</td>
<td>10.2</td>
<td>5.0</td>
<td>0.9</td>
<td>10.0</td>
<td>4.7</td>
<td>0.9</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>122827</td>
<td>25</td>
<td>36</td>
<td>9.7</td>
<td>4.9</td>
<td>1.0</td>
<td>9.8</td>
<td>5.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 1: Rejection frequencies (in percentages) of KPST at various significance levels. $\chi^2_n$ 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:

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

Table 2: Rejection frequencies (in percentages) of KPST test at various significance levels. $\chi^2_n$ 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:

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

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

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

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:

$$ \mathcal{R}(R) = vec(G_1)vec(G_2)' + \frac{1}{\sqrt{n}} \mathcal{R}(A_0) $$

$$ = vec(\tilde{G}_{1,n})vec(\tilde{G}_{2,n})' + \frac{1}{\sqrt{n}} vec(\tilde{G}_{1,n})_{\perp} \Lambda_n vec(\tilde{G}_{2,n})_{\perp}', $$

with $\tilde{G}_{1,n} \in \mathbb{R}^{p \times p}$ and $\tilde{G}_{2,n} \in \mathbb{R}^{k \times k}$ 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 $\tilde{G}_{1,n}$ and $\tilde{G}_{2,n}$ 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
\[
\delta := \lim_{n \to \infty} \text{vec}(\Lambda_n)' \left[ \left( \text{vec}(G_{2,n}) \right)' \otimes \text{vec}(G_{1,n}) \right]_+ \text{cov}(\text{vec}(\mathbf{R}(\mathbf{R}_n))) \left[ \left( \text{vec}(G_{2,n}) \right)' \otimes \text{vec}(G_{1,n}) \right]_+^{-1} \text{vec}(\Lambda_n)
\]  
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 \( \chi^2_\delta(\delta) \) 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 \( \chi^2 \) 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:

\[
V_i \sim iid \begin{cases} N(0, \Omega_1), \\ N(0, \Omega_2), \\ i = 1, \ldots, [n/2] \\ N(0, Q_{zz,1}), \\ N(0, Q_{zz,2}), \\ i = [n/2] + 1, \ldots, n, 
\end{cases}
\]

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

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

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

\[
R = \frac{1}{n} \text{var} \left( \sum_{i=1}^{n}(V_i \otimes Z_i) \right) = \frac{1}{2} \text{diag} (b + c, 1 + bc, 1 + bc, b + c)
\]

\[
= \frac{I_4}{G_1 \otimes G_2} + \frac{\sigma}{\sqrt{n}} \times \text{diag} (1, -1, -1, 1),
\]

and \( G_1 = G_2 = I_2 \). Since

\[
\mathbf{R}(\text{diag}(1, -1, -1, 1)) = \begin{pmatrix}
1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{pmatrix},
\]

(35)
\[ \text{vec}(G_1)' R(\text{diag}(1, -1, -1, 1)) \text{vec}(G_2) = 0, \]
the re-arranged specification of \( R \) (30) has \( \tilde{G}_{1,n} \) and \( \tilde{G}_{2,n} \) coinciding with \( G_1 \) and \( G_2 \):

\[
R(R) = \text{vec}(G_1) \text{vec}(G_2)' + \frac{\sigma}{\sqrt{n}} \begin{pmatrix}
1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{pmatrix}
\]

\[= \text{vec}(G_1) \text{vec}(G_2)' + \frac{1}{\sqrt{n}} \text{vec}(G_1)_\perp \Lambda_n \text{vec}(G_2)'_\perp, \]

with

\[
\text{vec}(G_1)_\perp = \text{vec}(G_2)_\perp = \frac{1}{\sqrt{2}} \begin{pmatrix}
1 & 0 & 0 \\
0 & \sqrt{2} & 0 \\
0 & 0 & \sqrt{2} \\
-1 & 0 & 0
\end{pmatrix}, \quad \Lambda_n = \sigma \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}. \tag{37}
\]

In this case \( \Lambda_n \) does not depend on \( n \) because \( \tilde{G}_{1,n} \) and \( \tilde{G}_{2,n} \) 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 \( \Omega_1 \) and \( \Omega_2 \) 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

\[
(e_1 \otimes e_1)' \left[ \left( [\text{vec}(G_{2,n})]'_\perp \otimes [\text{vec}(G_{1,n})]'_\perp \right) \right. \\
\left. \text{cov}(\text{vec}(R(R))([\text{vec}(G_{2,n})]'_\perp \otimes [\text{vec}(G_{1,n})]'_\perp)) \right]^{-1} (e_1 \otimes e_1) = \frac{1}{4} \tag{38}
\]

where \( G_{i,n} = G_i = I_2 \) for \( i = 1, 2 \), and \( e_1 = (1, 0, 0) \), so, since \( \text{vec}(\Lambda_n) = 2\sigma(e_1 \otimes e_1) \), 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 $\chi^2$ critical values (red solid), and asymptotic power from Theorem 3 (blue dashed). 10000 MC replications. $\sigma$ measures deviation from KPS in Frobenius norm. Sample size is $n = 1626$.

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

Table 3: List of papers used in the empirical applications.

<table>
<thead>
<tr>
<th>Paper</th>
<th># specifications</th>
<th>KPS rejection</th>
<th># observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCN 10</td>
<td>2</td>
<td>none</td>
<td>moderate</td>
</tr>
<tr>
<td>Nunn 08</td>
<td>4</td>
<td>4</td>
<td>small</td>
</tr>
<tr>
<td>AJ 05</td>
<td>24</td>
<td>10</td>
<td>small</td>
</tr>
<tr>
<td>HG 10</td>
<td>2</td>
<td>2</td>
<td>huge</td>
</tr>
<tr>
<td>AGN 13</td>
<td>6</td>
<td>1</td>
<td>moderate</td>
</tr>
<tr>
<td>Yogo 04</td>
<td>22</td>
<td>5</td>
<td>moderate</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Paper</th>
<th># specific.</th>
<th>KPS rej.</th>
<th># obs.</th>
<th>clustered KPS rej.</th>
<th># clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT 11</td>
<td>8</td>
<td>6</td>
<td>large</td>
<td>5</td>
<td>moderate</td>
</tr>
<tr>
<td>AJRY 08</td>
<td>9</td>
<td>7</td>
<td>large</td>
<td>5</td>
<td>moderate</td>
</tr>
<tr>
<td>JPS 06</td>
<td>4</td>
<td>4</td>
<td>huge</td>
<td>4</td>
<td>huge</td>
</tr>
<tr>
<td>PSJM 13</td>
<td>2</td>
<td>2</td>
<td>huge</td>
<td>2</td>
<td>huge</td>
</tr>
<tr>
<td>ADH 13</td>
<td>18</td>
<td>18</td>
<td>large</td>
<td>13</td>
<td>small</td>
</tr>
<tr>
<td>AD 13</td>
<td>7</td>
<td>7</td>
<td>huge</td>
<td>7</td>
<td>small</td>
</tr>
<tr>
<td>ACJR 11</td>
<td>1</td>
<td>1</td>
<td>small</td>
<td>1</td>
<td>very small</td>
</tr>
<tr>
<td>MSS 04</td>
<td>3</td>
<td>0</td>
<td>large</td>
<td>3</td>
<td>small</td>
</tr>
<tr>
<td>V et al 12</td>
<td>6</td>
<td>2</td>
<td>moderate</td>
<td>0</td>
<td>small</td>
</tr>
</tbody>
</table>

Table 5: Summary of results of 5 percent significance 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
\[ \mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}', \] (40)
the SVD of \( \mathcal{R}(\hat{R}) \) with \( \hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_{\min(l,q)}) \) 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

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

This expression shows that it is minimized with respect to \( G_1 > 0, G_2 > 0 \) at \( \text{vec}(\hat{G}_1) = \hat{L}_1/\hat{L}_{11}, \) \( \text{vec}(\hat{G}_2) = \hat{L}_{11}\hat{\sigma}_1\hat{N}_1 \) 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: \( H_0 : \text{vec}(G_1)'\mathcal{R}(R)\text{vec}(G_2)_\perp = 0. \) We test this hypothesis using a SVD of \( \mathcal{R}(\hat{R}) : \)

\[ \mathcal{R}(\hat{R}) = \hat{L}\hat{\Sigma}\hat{N}', \]
whose elements can be specified as

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

where \( \hat{A} \) is a \( \frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1) \) dimensional matrix, \( \hat{A}'D_p'\hat{D}_p\hat{A} = I_{\frac{1}{2}p(p+1)} \), \( \hat{B} \) is a \( \frac{1}{2}k(k+1) \times \frac{1}{2}k(k+1) \) dimensional matrix, \( \hat{B}'D_k'\hat{D}_k\hat{B} = I_{\frac{1}{2}k(k+1)} \), \( \hat{\Sigma}_{22} \) 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 \( D_k \), \( D_p'\hat{D}_{p,\perp} \equiv 0 \), \( D_k'\hat{D}_{k,\perp} \equiv I_{\frac{1}{2}p(p-1)} \), \( D_k'\hat{D}_{k,\perp} \equiv 0 \) 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 "'s) and where under \( H_0 : \Sigma_{22} = 0. \)
To obtain the limit distribution of the sample analog of the parameter tested under $H_0$:

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

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

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

$$= \left[\text{vec}(G_1)_\perp + O_p(n^{-\frac{1}{2}})\right]' \left[\text{vec}(G_1)\text{vec}(G_2)' + \frac{1}{n} D_p \Psi D_k' + o_p(n^{-\frac{1}{2}})\right] \left[\text{vec}(G_2)_\perp + O_p(n^{-\frac{1}{2}})\right]$$

$$= \frac{1}{\sqrt{n}} \text{vec}(G_1)'_\perp D_p \Psi D_k' \text{vec}(G_2)_\perp + o_p(n^{-\frac{1}{2}}).$$

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

$$\text{vec}(G_1)_\perp = L_2 L_2' (L_2 L_2')^{-1/2},$$

$$L_2 = \begin{pmatrix} e_1' \frac{1}{2}p(p+1) & A_2 \\ D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix},$$

$$N_2 = \begin{pmatrix} e_1' \frac{1}{2}k(k+1) & B_2 \\ D_{2,k} B_2 & D_{2,k\perp} \end{pmatrix},$$

$$A = \begin{pmatrix} a_1 \\ A_2 \end{pmatrix},$$

$$B = \begin{pmatrix} b_1 \\ B_2 \end{pmatrix},$$

where we use that $D_p = (e_1, \frac{1}{2}p(p+1): D_{2,p}')', D_{2,p} : (p^2 - 1) \times \frac{1}{2}p(p+1)$ and $D_k = (e_1, \frac{1}{2}k(k+1): D_{2,k}')', D_{2,k} : (k^2 - 1) \times \frac{1}{2}k(k+1)$ with $e_1$ the first $i$ dimensional unity vector (i.e. the first column of $I_i$). We also use that $a_1 : \frac{1}{2}p(p+1) \times 1$, $A_2 : \frac{1}{2}p(p+1) \times (\frac{1}{2}p(p+1) - 1)$, $b_1 : \frac{1}{2}k(k+1) \times 1$, $b_2 : \frac{1}{2}k(k+1) \times (\frac{1}{2}k(k+1) - 1)$, $D_{p\perp} = (0 : D_{2,p\perp}')', D_{2,p\perp} : (p^2 - 1) \times \frac{1}{2}p(p-1)$, $D_{k\perp} = (0 : D_{2,k\perp}')', D_{2,k\perp} : (k^2 - 1) \times \frac{1}{2}k(k-1)$, 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 $B_2'D_{2,k}D_{2,k}B_2 : (\frac{1}{2}k(k+1) - 1) \times (\frac{1}{2}k(k+1) - 1)$:

$$A_2'D_{2,p}D_{2,p}A_2 = L_{D_{2,p}A_2} A_{D_{2,p}A_2}^2 L_{D_{2,p}A_2}'$$

$$B_2'D_{2,p}D_{2,p}B_2 = L_{D_{2,k}B_2} A_{D_{2,k}B_2}^2 L_{D_{2,k}B_2}',$$

with $L_{D_{2,p}A_2}$ and $L_{D_{2,k}B_2}$ 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 $A_2^2_{D_{2,p}A_2}$ and $A_2^2_{D_{2,k}B_2}$ 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:

$$L_{22} = \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \\ \end{pmatrix} = \begin{pmatrix} D_{2,p} A_2 \left( L_{D_{2,p},A_2} A_{D_{2,p},A_2}^2 L'_{D_{2,p},A_2} \right)^{-\frac{1}{2}} L_{D_{2,p},A_2} A_{D_{2,p},A_2} L'_{D_{2,p},A_2} & D_{2,p\perp} \end{pmatrix},$$

$$\left( L_{22} L'_{22} \right)^{\frac{1}{2}} = \begin{pmatrix} D_{2,p} A_2 \left( L_{D_{2,p},A_2} A_{D_{2,p},A_2}^2 L'_{D_{2,p},A_2} \right)^{-\frac{1}{2}} L_{D_{2,p},A_2} & D_{2,p\perp} \end{pmatrix},$$

$$L_{22}^{-1} \left( L_{22} L'_{22} \right)^{1/2} = \begin{pmatrix} D_{2,p} A_2 \left( L_{D_{2,p},A_2} A_{D_{2,p},A_2}^2 L'_{D_{2,p},A_2} \right)^{-\frac{1}{2}} L_{D_{2,p},A_2} & D_{2,p\perp} \end{pmatrix},$$

$$L_2 L_{22}^{-1} \left( L_{22} L'_{22} \right)^{1/2} = \begin{pmatrix} D_{p} A_2 & D_{p,\perp} \end{pmatrix} \begin{pmatrix} A_2' D_{2,p}^T D_{2,p} A_2 & \frac{1}{2} A_2' D_{2,p}^T D_{2,p} \\ \frac{1}{2} A_2' D_{2,p}^T D_{2,p} & \end{pmatrix},$$

$$N_{22}^{-1} N_{22}^{-1} \left( N_{22} N'_{22} \right)^{1/2} = \begin{pmatrix} D_{2,p} A_2 & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} A_2' D_{2,p}^T D_{2,p} A_2 & \frac{1}{2} A_2' D_{2,p}^T D_{2,p} \\ \frac{1}{2} A_2' D_{2,p}^T D_{2,p} & \end{pmatrix} + o_p(1),$$

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 $L_{22}^{-1}$ results directly from the SVD of $L_{22}$ since $L_{22}$ is invertible.

Then, under $H_0$:

$$\sqrt{n}\bar{\Lambda} = \text{vec}(G_1)' \left( D_{p} \Psi D_{k} \right)' \text{vec}(G_2) \perp + o_p(1)$$

$$= (L_{22} L'_{22})^{1/2} L_{22}^{-1} \left( L_{22} L'_{22} \right)^{1/2} + o_p(1)$$

$$= \begin{pmatrix} D_{p} A_2 & D_{p,\perp} \end{pmatrix} \begin{pmatrix} \left( A_2' D_{2,p}^T D_{2,p} A_2 \right)^{-\frac{1}{2}} A_2' D_{2,p}^T D_{2,p} & \frac{1}{2} A_2' D_{2,p}^T D_{2,p} \\ \frac{1}{2} A_2' D_{2,p}^T D_{2,p} & \end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix} D_{2,p} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \left( A_2' D_{2,p}^T D_{2,p} A_2 \right)^{-\frac{1}{2}} A_2' D_{2,p}^T D_{2,p} & \frac{1}{2} A_2' D_{2,p}^T D_{2,p} \\ \frac{1}{2} A_2' D_{2,p}^T D_{2,p} & \end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix} D_{2,p} & D_{2,p\perp} \end{pmatrix} \begin{pmatrix} \left( A_2' D_{2,p}^T D_{2,p} A_2 \right)^{-\frac{1}{2}} A_2' D_{2,p}^T D_{2,p} & \frac{1}{2} A_2' D_{2,p}^T D_{2,p} \\ \frac{1}{2} A_2' D_{2,p}^T D_{2,p} & \end{pmatrix} + o_p(1)$$

with

$$\bar{\Lambda} = \left( A_2' D_{2,p}^T D_{2,p} A_2 \right)^{-\frac{1}{2}} A_2' D_{2,p}^T D_{p} \Psi D_{k} \perp D_{2,p} B_2 (B_2' D_{2,p} \perp D_{2,p} B_2)^{-\frac{1}{2}},$$

21
which is a \((\frac{1}{2}p(p + 1) - 1) \times (\frac{1}{2}k(k + 1) - 1)\) normally distributed random matrix with mean zero. The covariance matrix of \(\text{vec}(\hat{\Lambda})\) is:

\[
V_{\text{vec}(\hat{\Lambda})} = \left( (B_2' D_k' D_k B_2) - \frac{1}{2} B_{2,k}' D_k \otimes (A_2' D_p' D_p A_2)^{-\frac{1}{2}} A_2' D_p' D_p \right) V_{R^*} \times \left( (B_2' D_k' D_k B_2) - \frac{1}{2} B_{2,k}' D_k \otimes (A_2' D_p' D_p A_2)^{-\frac{1}{2}} A_2' D_p' D_p \right)'.
\]

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 \(\hat{\bar{V}}\) defined in (17). The matrix \(\hat{\bar{V}}\) 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:

\[
\left( \text{vec}(\hat{G}_2)_{\perp} \otimes \text{vec}(\hat{G}_1)_{\perp} \right)' \hat{\bar{V}} \left( \text{vec}(\hat{G}_2)_{\perp} \otimes \text{vec}(\hat{G}_1)_{\perp} \right) \rightarrow_p V_{\text{vec}(\hat{\Lambda})} \quad \text{and} \quad \left( \text{vec}(\hat{G}_2)_{\perp} \otimes \text{vec}(\hat{G}_1)_{\perp} \right)' \left( (D_k \otimes D_p) V_{R^*} (D_k \otimes D_p) \right) \left( \text{vec}(\hat{G}_2)_{\perp} \otimes \text{vec}(\hat{G}_1)_{\perp} \right) = \left( \left( \begin{array}{cc} D_k B_2 & D_{k,\perp} \\ D_{p,\perp} & D_p \end{array} \right) \otimes \left( \begin{array}{cc} A_2' D_{2,p}' D_{2,p} A_2 \end{array} \right) \right) V_{R^*}.
\]
The convergence behavior of KPST is then characterized by:

$$KPST = n \times \left[ vec(vec(G_1) \perp R(\hat{R}) vec(G_2) \perp ) \right]'$$

$$\left[ \left( vec(G_2) \perp \otimes vec(G_1) \perp \right)' \left( D_k \otimes D_p \right) V_{R^*} (D_k \otimes D_p) \left( vec(G_2) \perp \otimes vec(G_1) \perp \right) \right]^{-1}$$

$$\left[ vec(vec(G_1) \perp R(\hat{R}) vec(G_2) \perp ) \right] + o_p(1)$$

$$= vec(\Lambda)'(D_{2,k}B_2 \otimes D_{2,p}A_2) \left( \left( D_{2,k}B_2 \left( B_2' D_{2,k}^{-1} D_{2,k} B_2 \right)^{-1} D_{2,p} \right) \otimes \left( \begin{array}{ccc}
V_{vec(\Lambda)} & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{array} \right) \right)^{-1}$$

$$= vec(\Lambda)'(D_{2,k}B_2 \otimes D_{2,p}A_2)vec(\Lambda) + o_p(1) \xrightarrow{d} X_a^2,$$

with \( a = (1/2)p(p + 1) - (1/2)k(k + 1) - 1 \). 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

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

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

$$KPST = n \times \left[ vec(vec(\hat{G}_1) \perp R(\hat{R}) vec(\hat{G}_2) \perp ) \right]'$$

$$\left[ \left( vec(\hat{G}_2) \perp \otimes vec(\hat{G}_1) \perp \right) \hat{V}_R \left( vec(\hat{G}_2) \perp \otimes vec(\hat{G}_1) \perp \right) \right]^{-1}$$

$$\left[ vec(vec(\hat{G}_1) \perp R(\hat{R}) vec(\hat{G}_2) \perp ) \right]$$

$$= \left( vec(\hat{L}_2' \hat{L}_2 \hat{\Sigma} \hat{N}' \hat{N}_2) \right)' \left( \hat{N}_{22}^{-1} \left( \hat{N}_{22} \hat{N}_{22}' \right)^{1/2} \otimes \hat{L}_{22}^{-1} \left( \hat{L}_{22} \hat{L}_{22}' \right)^{1/2} \right)$$

$$\left[ \left( \left( \hat{N}_{22} \hat{N}_{22}' \right)^{1/2} \hat{N}_{22}^{-1} \otimes \left( \hat{L}_{22} \hat{L}_{22}' \right)^{1/2} \hat{L}_{22}^{-1} \right) \left( \hat{N}_2 \otimes \hat{L}_2 \right)' (D_k \otimes D_p) \hat{V}_R \right.$$}

$$\left. (D_k \otimes D_p)' \left( \hat{N}_2 \otimes \hat{L}_2 \right) \left( \hat{N}_{22}^{-1} \left( \hat{N}_{22} \hat{N}_{22}' \right)^{1/2} \otimes \hat{L}_{22}^{-1} \left( \hat{L}_{22} \hat{L}_{22}' \right)^{1/2} \right) \right]^{-1}$$

$$\left( \left( \hat{N}_{22} \hat{N}_{22}' \right)^{1/2} \hat{N}_{22}^{-1} \otimes \left( \hat{L}_{22} \hat{L}_{22}' \right)^{1/2} \hat{L}_{22}^{-1} \right) \left( vec(\hat{L}_2' \hat{L}_2 \hat{\Sigma} \hat{N}' \hat{N}_2) \right)$$

23
c. We show that if we replaced \( \mathcal{R}(\hat{R}) = D_p \hat{R}^* D'_k \), with

\[
\tilde{R} = D_p(D'_pD_p)^{-\frac{1}{2}} \hat{R}^*(D'_kD_k)^{-\frac{1}{2}} D'_k
\]

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 \( \tilde{R} = \tilde{L} \tilde{\Sigma} \tilde{N} \) and \( \hat{R}^* = \hat{L} \hat{\Sigma}^* \hat{N}^* \) which are related through:

\[
\begin{align*}
\tilde{L} &= \begin{pmatrix} D_p(D'_pD_p)^{-\frac{1}{2}} & \hat{L}_R^* & D_p,\perp \end{pmatrix} \\
\tilde{\Sigma} &= \begin{pmatrix} \hat{\Sigma}_R^* & 0 \\
0 & 0 \end{pmatrix} \\
\tilde{N} &= \begin{pmatrix} D_k(D'_kD_k)^{-\frac{1}{2}} & \hat{N}_R^* & D_k,\perp \end{pmatrix}.
\end{align*}
\]

To show that KPST using \( \tilde{R} \), indicated by KPST\( \tilde{R} \), equals KPST*, we analyze KPST\( \tilde{R} \):

\[
KPST_{\tilde{R}} = n \times \left[ \begin{array}{c}
vec(\vec(G_1)'_{\perp} \tilde{R} \vec(G_2)_{\perp}) \\
\left[ (\vec(G_2)'_{\perp} \otimes \vec(G_1)'_{\perp}) \tilde{V}_{\tilde{R}} (\vec(G_2)_{\perp} \otimes \vec(G_1)_{\perp}) \right]^{-1} \\
\left[ \vec(\vec(G_1)'_{\perp} \tilde{R} \vec(G_2)_{\perp}) \right]
\end{array} \right]'
\]
which is the KPST expression using $\hat{R}^*$ so it differs from KPST and where we also used that:

$$
\hat{V}_R = \left( D_k(D'_kD_k)^{-\frac{1}{2}} \otimes \frac{1}{2} \right) \hat{V}_R^* \left( D_k(D'_kD_k)^{-\frac{1}{2}} \otimes \frac{1}{2} \right)^t
$$

$$
\tilde{L}_2 = \left( D_p(D'_pD_p)^{-\frac{1}{2}} \tilde{L}_{2,R^*} : D_{p,\perp} \right)
$$

$$
\tilde{N}_2 = \left( D_k(D'_kD_k)^{-\frac{1}{2}} \tilde{N}_{2,R^*} : D_{k,\perp} \right).
$$

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 $\chi^2_p$.

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

$$vec(Q\hat{V}_i\hat{V}_i'(Q') = (Q \otimes Q)vec(\hat{V}_i\hat{V}_i')$$

$$ = (Q \otimes Q)D_pvec(\hat{V}_i\hat{V}_i')$$

$$vech(Q\hat{V}_i\hat{V}_i'(Q') = (D'_pD_p)^{-1}D'_p(Q \otimes Q)D_pvec(\hat{V}_i\hat{V}_i'),$$

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

$$\frac{1}{n} \sum_{i=1}^{n} vec(Q\hat{V}_i\hat{V}_i'(Q')vec(GZ_iZ_i'G')'$$

$$(Q \otimes Q)D_p \left[ \frac{1}{n} \sum_{i=1}^{n} vec(\hat{V}_i\hat{V}_i')vec(Z_iZ_i') \right] D_k'(G \otimes G)'$$

and

$$\frac{1}{n} \sum_{i=1}^{n} vech(Q\hat{V}_i\hat{V}_i'(Q')vech(GZ_iZ_i'G')'$$

$$(D'_pD_p)^{-1}D'_p(Q \otimes Q)D_p \left[ \frac{1}{n} \sum_{i=1}^{n} vech(\hat{V}_i\hat{V}_i')vech(Z_iZ_i') \right]$$

$$\times D_k'(G \otimes G)'D_k(D'_kD_k)^{-1}.$$
\[
\begin{align*}
&= (\text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp})' \text{vec}(\mathcal{R}(R)) + \left( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes I_{p^2-1} \right)'
\text{vec}(\mathcal{R}(R)) + \\
&\left[ \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes \text{vec}(G_1)_{\perp} \right]' \text{vec}(\mathcal{R}(R)) + \\
&\left[ \left[ \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right] \otimes \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right]' \text{vec}(\mathcal{R}(R)) + \\
&\left( \left[ \text{vec}(G_1)_{\perp} \right] \otimes \left[ \text{vec}(G_1)_{\perp} - \text{vec}(G_1)_{\perp} \right] \right)' \text{vec}(\mathcal{R}(R)) + \\
&\left[ \left[ \text{vec}(G_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right]' \text{vec}(\mathcal{R}(R)) + \\
&= a + b + c,
\end{align*}
\]

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

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

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

From (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 \( \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \) and \( \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \), see Kleibergen and Paap (2006). Since \( \text{vec}(\hat{G}_1)_{\perp} \) and \( \text{vec}(\hat{G}_2)_{\perp} \) are solved from \( \mathcal{R}(\hat{R}), \mathcal{R}(\hat{R}) - \mathcal{R}(R), \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \) and \( \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \) are all jointly dependent. In a limiting sequence where the dimensions \( p \) and \( k \) jointly
increase with the sample size \( n \), we then have the following convergence rates:

1. \[
\left( \text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp} \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) = O_p \left( n^{-\frac{1}{2}} \right)
\]

2. \[
\left( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes \text{vec}(G_1)_{\perp} \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) = O_p \left( \frac{k_2^2}{n} \right)
\]

3. \[
\left( \text{vec}(G_2)_{\perp} \otimes \left[ \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right] \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) = O_p \left( \frac{p^2k_2^2}{n} \right)
\]

4. \[
\left( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes \left[ \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right] \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) = O_p \left( \frac{(p^2k_2^2)}{n} \right)
\]

5. \[
\left( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right] \otimes \left[ \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \right] \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) = O_p \left( \frac{p^2k_2^2}{n} \right).
\]

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:

\[
1. \left[ \left( \text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp} \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_q = \sum_{i=1}^{p^2} \sum_{j=1}^{k_2^2} \left[ \text{vec}(G_2)_{\perp} \right]_{jm} \left[ \text{vec}(G_1)_{\perp} \right]_{ij} \left[ \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_{(j-1)k_2^2+i},
\]

for \( m = 1 + \lfloor (q-1)/(k_2^2 - 1) \rfloor \), \( l = q - (p^2 - 1)(m - 1) \), with \( [b] \) the integer function of a scalar \( b \), which is of order \( O_p \left( n^{-\frac{1}{2}} \right) \). This convergence rate results since \( \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \) is \( O_p \left( n^{-\frac{1}{2}} \right) \) and \( \text{vec}(G_1)_{\perp} \) and \( \text{vec}(G_2)_{\perp} \) are both orthonormal matrices. The sum of the weights \( \left[ \text{vec}(G_2)_{\perp} \right]_{jm} \) and \( \left[ \text{vec}(G_1)_{\perp} \right]_{ij} \) \( i = 1, \ldots, p^2, j = 1, \ldots, k_2^2 \) 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 \( \mathcal{R}(\hat{R}) - \mathcal{R}(R) = O_p \left( n^{-\frac{1}{2}} \right) \).

\[
2. \left[ \left( \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right) \otimes \text{vec}(G_1)_{\perp} \right)' \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_q = \sum_{i=1}^{p^2} \sum_{j=1}^{k_2^2} \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right]_{jm} \left[ \text{vec}(G_1)_{\perp} \right]_{ij} \left[ \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_{(j-1)k_2^2+i},
\]

for \( m = 1 + \lfloor (q-1)/(k_2^2 - 1) \rfloor \), \( l = q - (p^2 - 1)(m - 1) \), which is of order \( O_p \left( \frac{k_2^2}{n} \right) \). This order results from the \( k_2^2 \) dependent components \( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right]_{jm} \) and \( \left[ \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_{(j-1)k_2^2+i} \) that we sum over and that the sum of the weights in the summation is proportional to \( k_2^2 \). Each of the (dependent) components in \( \left[ \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \right]_{jm} \) and \( \left[ \text{vec} \left( \mathcal{R}(\hat{R}) - \mathcal{R}(R) \right) \right]_{(j-1)k_2^2+i} \) are \( O_p(n^{-\frac{1}{2}}) \) so summing over \( k_2^2 \) of them and multiplying through results in \( O_p \left( \frac{k_2^2}{n} \right) \). The additional weights \( \left[ \text{vec}(G_1)_{\perp} \right]_{ij}, i = 1, \ldots, p^2 \), are again such that their sum is finite so it does not grow with
the sequence of \( k \) and \( p \) because \( \text{vec}(G_1) \perp \) is orthonormal. Hence, they do not affect the convergence rate.

\[
3. \left( \left[ \text{vec}(G_2) \right] \otimes \left[ \text{vec}(G_1) - \text{vec}(G_1) \right] \right) \text{vec} \left( R(R) - R(R) \right) _q
= \sum_{i=1}^{p^2} \sum_{j=1}^{k^2} \left[ \text{vec}(G_2) \right] _{jm} \left[ \text{vec}(G_1) - \text{vec}(G_1) \right] _{il} \left[ \text{vec} \left( R(R) - R(R) \right) \right] _{(j-1)k^2+i},
\]

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

\[
4. \left( \left[ \text{vec}(G_2) - \text{vec}(G_2) \right] \otimes \left[ \text{vec}(G_1) - \text{vec}(G_1) \right] \right) \text{vec} \left( R(R) - R(R) \right) _q
= \sum_{i=1}^{p^2} \sum_{j=1}^{k^2} \left[ \text{vec}(G_2) - \text{vec}(G_2) \right] _{jm} \left[ \text{vec}(G_1) - \text{vec}(G_1) \right] _{il} \left[ \text{vec} \left( R(R) - R(R) \right) \right] _{(j-1)k^2+i},
\]

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

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

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

For the estimator of the covariance matrix of \( \hat{A} \), we further have

\[
\left( \left[ \text{vec}(G_2) \right]_{\perp} \otimes \left[ \text{vec}(G_1) \right]_{\perp} \right) \text{cov} \left( \text{vec}(R(R)) \right) \left( \left[ \text{vec}(G_2) \right]_{\perp} \otimes \left[ \text{vec}(G_1) \right]_{\perp} \right) = \left( \left[ \text{vec}(G_2) + \text{vec}(G_2) - \text{vec}(G_2) \right] \otimes \left[ \text{vec}(G_1) + \text{vec}(G_1) - \text{vec}(G_1) \right] \right) \text{cov} \left( \text{vec}(R(R)) \right) \text{cov} \left( \text{vec}(R(R)) \right) - \text{cov} \left( \text{vec}(R(R)) \right) = \left( \text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp} \right) \text{cov} \left( \text{vec}(R(R)) \right) \left( \text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp} \right) + U = A_1 + B_1 + B_2 + B_3 + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + D_1 + \ldots
\]

where below we show that the maximal convergence rates besides the zero-th order component
are \( O_p(n^{-\frac{1}{2}}) \), \( O_p\left(\frac{(\log k)^2}{n}\right) \), \( O_p\left(\frac{k^4}{n}\right) \), \( O_p\left(\frac{p^4}{n}\right) \) 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 \( \text{vec}(\hat{\Lambda}) \), we obtain a slightly stronger condition than just for \( \hat{\Lambda} \):

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

which results from the \( O_p(n^{-\frac{1}{2}}) \) components from the inverse of the covariance matrix estimator paired with the \( O_p\left(\frac{(\log k)^2}{n}\right) \) 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 \( \frac{(pk)^{12}}{n^3} \to 0 \). The convergence rate is in between the rate implied by Newey and Windmeijer (2006) which would be \( \frac{k^4}{n} \) for convergence of the test statistic which is slightly stricter than our rate of \( \frac{(pk)^{16}}{n^3} \to 0 \).

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

\[
A_1 = \left( \text{vec}(G_2)'_\perp \otimes \text{vec}(G_1)'_\perp \right) \text{cov} \left( \text{vec}(\mathcal{R}(\hat{\Lambda})) \right) \left( \text{vec}(G_2)_\perp \otimes \text{vec}(G_1)_\perp \right) = O(1)
\]

\[
B_1 = \left[ \text{vec}(\hat{G}_2)_\perp - \text{vec}(G_2)_\perp \right]' \otimes \text{vec}(G_1)_\perp \text{cov} \left( \text{vec}(\mathcal{R}(\hat{\Lambda})) \right) \left( \text{vec}(G_2)_\perp \otimes \text{vec}(G_1)_\perp \right) + \text{vec}(G_2)'_\perp \otimes \text{vec}(G_1)'_\perp \text{cov} \left( \text{vec}(\mathcal{R}(\hat{\Lambda})) \right) \left( \text{vec}(\hat{G}_2)_\perp - \text{vec}(G_2)_\perp \right) \otimes \text{vec}(G_1)_\perp = O_p(n^{-\frac{1}{2}})
\]

\[
B_2 = \left( \text{vec}(G_2)_\perp \otimes \left[ \text{vec}(\hat{G}_1)_\perp - \text{vec}(G_1)_\perp \right]' \right) \text{cov} \left( \text{vec}(\mathcal{R}(\hat{\Lambda})) \right) \left( \text{vec}(G_2)_\perp \otimes \text{vec}(G_1)_\perp \right) + \text{vec}(G_2)'_\perp \otimes \text{vec}(G_1)'_\perp \text{cov} \left( \text{vec}(\mathcal{R}(\hat{\Lambda})) \right) \left( \text{vec}(\hat{G}_2)_\perp \otimes \left[ \text{vec}(\hat{G}_1)_\perp - \text{vec}(G_1)_\perp \right] \right) = O_p(n^{-\frac{1}{2}})
\]

\[
B_3 = \left( \text{vec}(G_2)_\perp \otimes \text{vec}(G_1)_\perp \right) \left[ \text{vec}(\text{cov}(\mathcal{R}(\hat{\Lambda}))) - \text{cov}(\text{vec}(\mathcal{R}(\hat{\Lambda}))) \right] \left( \text{vec}(G_2)'_\perp \otimes \text{vec}(G_1)'_\perp \right) = O_p(n^{-\frac{1}{2}})
\]

---

\(^7\)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: \( O_p\left(\frac{(\log k)^2}{n}\right) \left( \frac{p^2k^2}{n} \right) \frac{1}{\sqrt{n}} (k^2p^2)^2 = O_p\left(\frac{(\log k)^2}{n^3} \right) = O_p\left(\frac{(pk)^{16}}{n^3} \right) \).
\[ C_1 = \left( \begin{bmatrix} \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \end{bmatrix}' \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix}' \right) \]

\[
\text{cov}(\text{vec}(R(\hat{R}))) \left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \text{vec}(G_1)_{\perp} \end{bmatrix} + \begin{bmatrix} \text{vec}(G_2)'_{\perp} \otimes \text{vec}(G_1)'_{\perp} \end{bmatrix} \right)
\]

\[
= O_p \left( \frac{(pk)^2}{n} \right)
\]

\[ C_2 = \left( \begin{bmatrix} \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \end{bmatrix}' \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix}' \right) \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{k^4}{n} \right)
\]

\[ C_3 = \left( \begin{bmatrix} \text{vec}(G_2)'_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix}' \end{bmatrix} \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{p^4}{n} \right)
\]

\[ C_4 = \left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \text{vec}(G_2)'_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)'_{\perp} - \text{vec}(G_1)'_{\perp} \end{bmatrix} \end{bmatrix} \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{p^2k^2}{n} \right)
\]

\[ C_5 = \left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix} \end{bmatrix} \right) \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
\left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \text{vec}(G_2)'_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)'_{\perp} - \text{vec}(G_1)'_{\perp} \end{bmatrix} \end{bmatrix} \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{p^2k^2}{n} \right)
\]

\[ C_6 = \left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix} \end{bmatrix} \right) \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
\left( \begin{bmatrix} \text{vec}(G_2)_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \text{vec}(G_2)'_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)'_{\perp} - \text{vec}(G_1)'_{\perp} \end{bmatrix} \end{bmatrix} \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{p^2k^2}{n} \right)
\]

\[ D_1 = \left( \begin{bmatrix} \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \end{bmatrix}' \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix}' \right) \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
\left( \begin{bmatrix} \text{vec}(\hat{G}_2)_{\perp} - \text{vec}(G_2)_{\perp} \end{bmatrix}' \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)_{\perp} - \text{vec}(G_1)_{\perp} \end{bmatrix}' \right) + \begin{bmatrix} \text{vec}(\hat{G}_2)'_{\perp} \otimes \begin{bmatrix} \text{vec}(\hat{G}_1)'_{\perp} - \text{vec}(G_1)'_{\perp} \end{bmatrix} \end{bmatrix} \text{cov}(\text{vec}(R(\hat{R})))
\]

\[
= O_p \left( \frac{k^4p^4}{n\sqrt{n}} \right)
\]

References


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

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

Continued on next page
<table>
<thead>
<tr>
<th>Paper</th>
<th>Specif.</th>
<th>Y</th>
<th>Z</th>
<th>$p$</th>
<th>$k$</th>
<th>$n$</th>
<th>KPST</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5.P1.C6</td>
<td>Private credit, legal formalism, private property</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>52</td>
<td>19.304</td>
<td>0.037</td>
</tr>
<tr>
<td>T5.P2.C1</td>
<td>Stock market capitalization, legal formalism, constraint on executive</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>50</td>
<td>19.178</td>
<td>0.038</td>
</tr>
<tr>
<td>T5.P2.C2</td>
<td>Stock market capitalization, legal formalism, constraint on executive</td>
<td></td>
<td>English legal origin, population density 1500</td>
<td>3</td>
<td>2</td>
<td>59</td>
<td>19.405</td>
<td>0.035</td>
</tr>
<tr>
<td>T5.P2.C3</td>
<td>Stock market capitalization, constraint on executive, procedural complexity</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>59</td>
<td>34.566</td>
<td>0</td>
</tr>
<tr>
<td>T5.P2.C4</td>
<td>Stock market capitalization, constraint on executive, number of procedures</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>59</td>
<td>28.06</td>
<td>0.002</td>
</tr>
<tr>
<td>T5.P2.C5</td>
<td>Stock market capitalization, legal formalism, average protection against risk of expropriation</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>50</td>
<td>30.225</td>
<td>0</td>
</tr>
<tr>
<td>T5.P2.C6</td>
<td>Stock market capitalization, legal formalism, private property</td>
<td></td>
<td>English legal origin, settler mortality</td>
<td>3</td>
<td>2</td>
<td>51</td>
<td>21.344</td>
<td>0.019</td>
</tr>
<tr>
<td>HG 10</td>
<td>T1.C2</td>
<td>Democratic vote share, turnout, turnout * partisan composition, turnout * Republican incumbent</td>
<td>Rainfall, rainfall<em>partisan composition, rainfal</em>Republican incumbent</td>
<td>4</td>
<td>3</td>
<td>27401</td>
<td>507.919</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>T1.C3</td>
<td>Democratic vote share, turnout, turnout * partisan composition, turnout * Republican incumbent</td>
<td>Rainfall, rainfall<em>partisan composition, rainfal</em>Republican incumbent</td>
<td>4</td>
<td>3</td>
<td>27401</td>
<td>457.962</td>
<td>0</td>
</tr>
<tr>
<td>AGN 13</td>
<td>T8.P3.C1</td>
<td>Female LF participation, Traditional plough use</td>
<td>Plough-neg. environment, Plough-pos. environment</td>
<td>2</td>
<td>2</td>
<td>160</td>
<td>6.191</td>
<td>0.185</td>
</tr>
</tbody>
</table>

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

Continued on next page
Table 6 – continued from previous page

<table>
<thead>
<tr>
<th>Paper</th>
<th>Specif.</th>
<th>$Y$</th>
<th>$Z$</th>
<th>$p$</th>
<th>$k$</th>
<th>$n$</th>
<th>KPST</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWT</td>
<td>cons growth, risk-free rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>91</td>
<td>14.889</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>cons growth, stk mkt rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>91</td>
<td>43.768</td>
<td>0.001</td>
</tr>
<tr>
<td>UK</td>
<td>cons growth, risk-free rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>115</td>
<td>30.148</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>cons growth, stk mkt rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>115</td>
<td>19.94</td>
<td>0.336</td>
</tr>
<tr>
<td>US</td>
<td>cons growth, risk-free rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>114</td>
<td>18.478</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td>cons growth, stk mkt rtn</td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>114</td>
<td>22.373</td>
<td>0.216</td>
</tr>
</tbody>
</table>

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

Continued on next page
Table 7 – continued from previous page

<table>
<thead>
<tr>
<th>Specif.</th>
<th>Y</th>
<th>Z</th>
<th>p</th>
<th>k</th>
<th>n</th>
<th>KPST</th>
<th>p val</th>
<th>n_c</th>
<th>KPST_c</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T6.C7</td>
<td>Freedom House measure of democracy, Log GDP per capita in t-1</td>
<td>tw log GDP in t-1, tw democracy in t-1</td>
<td>2</td>
<td>2</td>
<td>796</td>
<td>10.18</td>
<td>0.037</td>
<td>125</td>
<td>11.808</td>
<td>0.038</td>
</tr>
<tr>
<td>T6.C9</td>
<td>Freedom House measure of democracy, Log GDP per capita in t-1</td>
<td>tw log GDP in t-1, t-2</td>
<td>2</td>
<td>2</td>
<td>796</td>
<td>12.83</td>
<td>0.012</td>
<td>125</td>
<td>12.121</td>
<td>0.033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JPS 06</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.P1.C5</td>
</tr>
<tr>
<td>T4.P1.C6</td>
</tr>
<tr>
<td>T4.P2.C5</td>
</tr>
<tr>
<td>T4.P2.C6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSJM 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.P1.C5</td>
</tr>
</tbody>
</table>

Continued on next page
Table 7 – continued from previous page

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

Continued on next page
<table>
<thead>
<tr>
<th>Specif.</th>
<th>$Y$</th>
<th>$Z$</th>
<th>$p$</th>
<th>$k$</th>
<th>$n$</th>
<th>KPST</th>
<th>p val</th>
<th>$n_c$</th>
<th>KPST$_c$</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T10.P4.C5</td>
<td>Δ log transfers, Δ US-China net imp pw</td>
<td>Δ trade other-China, Δ net exports other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>41.77</td>
<td>0.000</td>
<td>48</td>
<td>9.138</td>
<td>0.104</td>
</tr>
<tr>
<td>T10.P4.C6</td>
<td>Δ avg household inc, Δ US-China net imp pw</td>
<td>Δ trade other-China, Δ net exports other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>18.08</td>
<td>0.001</td>
<td>48</td>
<td>13.395</td>
<td>0.020</td>
</tr>
<tr>
<td>T10.P6.C1</td>
<td>Δ mfg empl, Δ net trade factor (ntf) US-China</td>
<td>Δ ntf other-China, Δ net export factor (nef) other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>16.57</td>
<td>0.002</td>
<td>48</td>
<td>14.213</td>
<td>0.014</td>
</tr>
<tr>
<td>T10.P6.C2</td>
<td>Δ nonmfg empl, Δ ntf US-China</td>
<td>Δ ntf other-China, Δ nef other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>43.88</td>
<td>0.000</td>
<td>48</td>
<td>15.611</td>
<td>0.008</td>
</tr>
<tr>
<td>T10.P6.C3</td>
<td>Δ mfg log wage, Δ ntf US-China</td>
<td>Δ ntf other-China, Δ nef other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>24.54</td>
<td>0.000</td>
<td>48</td>
<td>12.087</td>
<td>0.034</td>
</tr>
<tr>
<td>T10.P6.C4</td>
<td>Δ nonmfg log wage, Δ ntf US-China</td>
<td>Δ ntf other-China, Δ nef other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>10.81</td>
<td>0.029</td>
<td>48</td>
<td>18.869</td>
<td>0.002</td>
</tr>
<tr>
<td>T10.P6.C5</td>
<td>Δ log transfers, Δ ntf US-China</td>
<td>Δ ntf other-China, Δ nef other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>15.56</td>
<td>0.004</td>
<td>48</td>
<td>16.692</td>
<td>0.005</td>
</tr>
<tr>
<td>T10.P6.C6</td>
<td>Δ avg household inc, Δ ntf US-China</td>
<td>Δ ntf other-China, Δ nef other-China</td>
<td>2</td>
<td>2</td>
<td>1444</td>
<td>16.46</td>
<td>0.002</td>
<td>48</td>
<td>29.073</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**AD 13**

<table>
<thead>
<tr>
<th>Specif.</th>
<th>$Y$</th>
<th>$Z$</th>
<th>$p$</th>
<th>$k$</th>
<th>$n$</th>
<th>KPST</th>
<th>p val</th>
<th>$n_c$</th>
<th>KPST$_c$</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5.P2.C1</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1950 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>141.50</td>
<td>0.000</td>
<td>48</td>
<td>57.891</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Specif.</th>
<th>Y</th>
<th>Z</th>
<th>p</th>
<th>k</th>
<th>n</th>
<th>KPST</th>
<th>p val</th>
<th>n_c</th>
<th>KPST_c</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5.P2.C2</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1951 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>122.97</td>
<td>0.000</td>
<td>48</td>
<td>41.735</td>
<td>0.000</td>
</tr>
<tr>
<td>T5.P2.C3</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1952 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>140.57</td>
<td>0.000</td>
<td>48</td>
<td>52.603</td>
<td>0.000</td>
</tr>
<tr>
<td>T5.P2.C4</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1953 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>118.33</td>
<td>0.000</td>
<td>48</td>
<td>47.893</td>
<td>0.000</td>
</tr>
<tr>
<td>T5.P2.C5</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1954 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>106.08</td>
<td>0.000</td>
<td>48</td>
<td>47.248</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Continued on next page
Table 7 – continued from previous page

<table>
<thead>
<tr>
<th>Specif.</th>
<th>Y</th>
<th>Z</th>
<th>p</th>
<th>k</th>
<th>n</th>
<th>KPST</th>
<th>p val</th>
<th>n_c</th>
<th>KPST_c</th>
<th>p val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5.P2.C6</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1955 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>146.81</td>
<td>0.000</td>
<td>48</td>
<td>43.400</td>
<td>0.000</td>
</tr>
<tr>
<td>T5.P2.C7</td>
<td>Growth of service employment, Share of routine employment (t-1)</td>
<td>1956 employment share by commuting zone excluding those corresponding to observation: 1980; 1990;2000.¹</td>
<td>2</td>
<td>3</td>
<td>2166</td>
<td>101.50</td>
<td>0.000</td>
<td>48</td>
<td>32.647</td>
<td>0.002</td>
</tr>
<tr>
<td>ACJR 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6.P.3.C2</td>
<td>Urbanization in Germany, reform index</td>
<td>French presence in 1850, 1875 and 1900</td>
<td>2</td>
<td>3</td>
<td>74</td>
<td>12.74</td>
<td>0.239</td>
<td>13</td>
<td>112.422</td>
<td>0.000</td>
</tr>
<tr>
<td>MSS 04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4.C5</td>
<td>Civil conflict &gt;25 deaths, Economic growth rate (t)</td>
<td>Current and lagged rainfall</td>
<td>3</td>
<td>2</td>
<td>743</td>
<td>10.30</td>
<td>0.414</td>
<td>41</td>
<td>31.022</td>
<td>0.003</td>
</tr>
<tr>
<td>T4.C6</td>
<td>Civil conflict &gt;25 deaths, Economic growth rate (t)</td>
<td>Current and lagged rainfall</td>
<td>3</td>
<td>2</td>
<td>743</td>
<td>5.18</td>
<td>0.879</td>
<td>41</td>
<td>37.682</td>
<td>0.000</td>
</tr>
<tr>
<td>T4.C7</td>
<td>Civil conflict &gt;1000 deaths, Economic growth rate (t)</td>
<td>Current and lagged rainfall</td>
<td>3</td>
<td>2</td>
<td>743</td>
<td>5.35</td>
<td>0.867</td>
<td>41</td>
<td>42.052</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Continued on next page
Table 7 – continued from previous page

<table>
<thead>
<tr>
<th>Specif.</th>
<th>$Y$</th>
<th>$Z$</th>
<th>$p$</th>
<th>$k$</th>
<th>$n$</th>
<th>KPST</th>
<th>$p$ val</th>
<th>$n_c$</th>
<th>KPST$_c$</th>
<th>$p$ val</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3.C6</td>
<td>Degree of altruism scale, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>278</td>
<td>9.45</td>
<td>0.051</td>
<td>35</td>
<td>8.054</td>
<td>0.153</td>
</tr>
<tr>
<td>T4.C6</td>
<td>Risk preference, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>213</td>
<td>12.28</td>
<td>0.015</td>
<td>35</td>
<td>1.349</td>
<td>0.930</td>
</tr>
<tr>
<td>T5.C6</td>
<td>Discount rate, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>266</td>
<td>6.69</td>
<td>0.153</td>
<td>35</td>
<td>5.622</td>
<td>0.345</td>
</tr>
<tr>
<td>T6.C4</td>
<td>Degree of altruism scale, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>212</td>
<td>6.36</td>
<td>0.174</td>
<td>35</td>
<td>6.931</td>
<td>0.226</td>
</tr>
<tr>
<td>T6.C5</td>
<td>Risk preference, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>158</td>
<td>18.69</td>
<td>0.028</td>
<td>35</td>
<td>6.860</td>
<td>0.231</td>
</tr>
<tr>
<td>T6.C6</td>
<td>Discount rate, Percentage dead in attacks</td>
<td>Distance to Bujumbura (log), Altitude (log)</td>
<td>2</td>
<td>2</td>
<td>205</td>
<td>2.34</td>
<td>0.673</td>
<td>35</td>
<td>4.451</td>
<td>0.487</td>
</tr>
</tbody>
</table>

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