Overview

The code in this replication package conducts analyses using one data source (Hornung 2014) using R and Stata. 14 files run all of the code to generate the data for the 12 figures and 20 tables in the paper. The replicator should expect the code to run for about 12 hours.

Data Availability and Provenance Statements

The paper uses data obtained from the OPENICPSR (Hornung, 2014). OPENICPSR allows for redistribution for the purpose of replication. The archive contains the extracted data files, Stata do files in the folder "Hornung 2014 - Data". Data used in the empirical application is publicly available at the following link: https://www.openicpsr.org/openicpsr/project/112731/version/V1/view.

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☒ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

- ☒ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

Data can also be directly downloaded using https://www.openicpsr.org/openicpsr/project/112731/version/V1/view. A copy of the data is provided in the folder "Hornung 2014 – Data" as part of this archive. The data are in the public domain.

Datafile: `hornung_data_textiles.txt`

Dataset list

Data file	Source	Notes	Provided
`hornung_data_textiles.txt`	OPENICPSR	Data set used for the empirical analyses	Yes

Computational requirements

The file `0_set_up.R` installs all R packages needed for replicating all results.

Software Requirements

- Stata (code was last run with version 15)
- R 4.0.2
  - "`estrpac`" the binary of this package is included in the replication package
  - "`pbapply`" (1.5-0)
  - "rstudioapi" (0.13)
  - "sandwich" (3.0-1)

Controlled Randomness

A random seed is set in the Monte Carlo simulation files for each random sample. The results of the Monte Carlo simulations should therefore be reproducible. Same applies to the use of the bootstrap in both simulations and the empirical example.

Memory and Runtime Requirements

Summary

Approximate time needed to reproduce the analyses on a standard 2022 laptop:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☐ 8-24 hours
- ☒ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Details

The code was last run on a **2.8 GHz, Quad-Core Intel Core i7, and 16 GB RAM laptop with MacOS version 12.6**.

Description of programs/code

estrpac_0.1.0.tgz - this is a binary which can be used to install the R package ***estrpac*** which contains all routines used in simulations and empirical results.

Empirical Application

Empirical_Application_Hornung.R - R file for replicating results of the Empirical Application in Section 5. The results are presented in (Table 5).

hornung_data_textiles.txt - data set from Hornung 2014 used in **Empirical_Application_Hornung.R**.

.**hornung_data_textiles.dta** - the Stata version of the `hornung_data_textiles.txt` data set

Empirical_Application_KP_Rank_Test_Final.do - Stata do file for computing the Kleibergen & Paap F-statistic in (Table 5).

Monte Carlo Simulations

(Table and Figure labels are as in the paper and online appendix)

MC_Reg_0A.R - DGP_0A for n = 250 (Table 2) and n = 500 (Table S.5)

MC_Reg_0B.R - DGP_0B for n = 250 (Table 2) and n = 500 (Table S.5)

MC_Reg_1_250.R - DGP_1A and DGP_1B for n = 250 (Table 3)

MC_Reg_1_500.R - DGP_1A (Table S.6) and DGP_1B (Table S.7) for n = 500

MC_Reg_2_250.R - DGP_2A (Table S.1) and DGP_2B (Table S.2) for n = 250

MC_Reg_2_500.R - DGP_2A (Table S.8) and DGP_2B (Table S.9) for n = 500

MC_Reg_3_250.R - DGP_3A (Table S.3) and DGP_3B (Table S.4) for n = 250

MC_Reg_3_500.R - DGP_3A (Table S.10) and DGP_3B (Table S.11) for n = 500

MC_Reg_4A.R - DGP_4 for n = 250, n = 500, and n = 1000 (Table 4), (Table S.15) and (Figure S.13)

MC_Reg_4B.R - DGP_4 with bounded one-to-one mappings of Z for n = 250, n = 500, and n = 1000 (Table S.14)

MC_Reg_5.R - DGP_5 for n = 250 (Table S.12) and n = 500 (Table S.13). Code for sensitivity analyses to outliers

MC_Test_Relevance.R - Simlations to examine the test of the linear completeness condition for DGPs 0A, 3B, and 1B at n = 250 (Figures S.4 through S.12)

Illustrative Examples

Illustrative_Kernel_Prob.R - this file is used to generate figures 1 and 2 as well as all numbers used in Section 2.4.

Instructions to Replicators

- Download the data files referenced above.
- Run each program file to obtain results as indicated above. Each .R file is self-contained.
- No other programme is needed.

List of tables and programs

The provided code reproduces:

- ☒ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Table #	Program	Line Number	Output file	Note
Table 1	n.a. (no programme)	n.a.	n.a.	Characterisation of Kernels
Table 2	MC_Reg_0A.R, MC_Reg_0B.R	156 154	n.a.	DGP_{0A} and $DGP_{0B}, n = 250$
Table 3	MC_Reg_1_250.R	223, 272	n.a.	DGP_{1A} and $DGP_{1B}, n = 250$
Table 4	MC_Reg_4A.R	200, 283, 364	n.a.	DGP4, $n \in \{250,500,1000\}$
Table 5	Empirical_Application_Hornung.R, Empirical_Application_KP_Rank_Test_Final.do	167, 184, 190	n.a.	Empirical Application
Table S.1	MC_Reg_2_250.R	200	n.a.	$DGP_{2A}, n = 250$
Table S.2	MC_Reg_2_250.R	249	n.a.	$DGP_{2B}, n = 250$
Table S.3	MC_Reg_3_250.R	191	n.a.	$DGP_{3A}, n = 250$
Table S.4	MC_Reg_3_250.R	244	n.a.	$DGP_{3B}, n = 250$
Table S.5	MC_Reg_0A.R, MC_Reg_0B.R	195 194	n.a.	$DGP_{0A}, n = 500$
Table S.6	MC_Reg_1_500.R	220	n.a.	$DGP_{1A}, n = 500$
Table S.7	MC_Reg_1_500.R	268	n.a.	$DGP_{1B}, n = 500$
Table S.8	MC_Reg_2_500.R	201	n.a.	$DGP_{2A}, n = 500$

Table S.9	MC_Reg_2_500.R	251	n.a.	$DGP_{2B}, n = 500$
Table S.10	MC_Reg_3_500.R	193	n.a.	$DGP_{3A}, n = 500$
Table S.11	MC_Reg_3_500.R	243	n.a.	$DGP_{3B}, n = 500$
Table S.12	MC_Reg_5.R	193	n.a.	$DGP_5, n = 250$
Table S.13	MC_Reg_5.R	243	n.a.	$DGP_5, n = 500$
Table S.14	MC_Reg_4B.R	199, 283, 364	n.a.	DGP_4 with bounded one-to-one mappings
Table S.15	MC_Reg_4A.R	219, 301, 383		Running Time $DGP_4, n \in \{250, 500, 1000\}$
Figure #				
Figure 1	Illustrative_Kernel_Prob.R	51	Fig_Kernel_Var.pdf	Standard Deviation of Kernels
Figure 2	Illustrative_Kernel_Prob.R	117	Fig_GMDC.pdf	GMDC of Kernels
Figure S.4	MC_Test_Relevance.R	212, 221, 230	Fig_PowerCurve0A_01.pdf, Fig_PowerCurve0A_05.pdf, Fig_PowerCurve0A_10.pdf	Power Curve for the LC test - DGP_{0A}
Figure S.8	MC_Test_Relevance.R	272, 283, 294	Fig_PowerCurve1B_01.pdf, Fig_PowerCurve1B_05.pdf, Fig_PowerCurve1B_10.pdf	Power Curve for the LC test - DGP_{1B}
Figure S.12	MC_Test_Relevance.R	239, 250, 261	Fig_PowerCurve3B_01.pdf, Fig_PowerCurve3B_05.pdf, Fig_PowerCurve3B_10.pdf	Power Curve for the LC test - DGP_{3B}
Figure S.13	MC_Reg_4A.R	447	Fig_Power_Slope.pdf	DGP_4, Power curve for the slope parameter

References

Hornung, E. (2014). Replication data for: Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. Data retrieved from OPENICPSR, 2019-10-11, https://doi.org/10.3886/E112731V1.