Replication code for Monte Carlo Simulation

Authors: Jianfei Cao, Christian Hansen, Damian Kozbur, Lucciano Villacorta

Date: July 16, 2021

These are the replication files for the Monte Carlo simulation in "Inference for Dependent Data with Learned Clusters" by Jianfei Cao, Christian Hansen, Damian Kozbur, and Lucciano Villacorta. All codes are written in Matlab and all simulation results are obtained under Matlab R2017b (9.3.0.713579).

Structure

There are 8 folders, each of which generates results for one setting:

Folder Name	Model	Spatial Model	Number of locations	Approximate computation time (s)
sim1_ols_small_baseline	OLS	BASELINE	205	5756
sim2_ols_small_sar	OLS	SAR	205	6342
sim3_iv_small_baseline	IV	BASELINE	205	9485
sim4_iv_small_sar	IV	SAR	205	9245
sim5_ols_large_baseline	OLS	BASELINE	820	33332
sim6_ols_large_sar	OLS	SAR	820	27526
sim7_iv_large_baseline	IV	BASELINE	820	36215
sim8_iv_large_sar	IV	SAR	820	37318

All simulation results in the main text are included in sim1-4. Extra simulation results with a larger set of locations (reported in the supplementary materialS) are included in sim5-8.

For each setting, there are four subfolders, /code/, /data/, /output/, and /temp/. To generate the results for a specific setting, run /code/main.m under the corresponding folder. The results (5 csv tables and 1 pdf figure) are reported under /output/. The location data from the empirical application is stored under /data/ and temporary files are under /temp/.

The run_all.m file generates output for all eight settings, although we suggest run each setting separately.

Replicating tables and figures

To replicate tables and figures in the paper, check the needed output files in the following table:

Table	Simulation No.	Output file under /output/
Table 3: Simulation Results	sim1-4	table_simulation_results.csv
Table 4: Distribution of $\widehat{k}$	sim1-4	table_kHat_distribution.csv
Table 5: Distribution of $\widehat{\alpha}$	sim1-4	table_alphaHat_distribution.csv
Table 6: Summary: OLS - BASELINE ($N_{\text{pan}} = 205$)	sim1	table_summary_complete.pdf
Table 7: Summary: OLS - SAR ($N_{\text{pan}} = 205$)	sim2	table_summary_complete.pdf
Table 8: Summary: IV - BASELINE ($N_{\text{pan}} = 205$)	sim3	table_summary_complete.pdf
Table 9: Summary: IV - SAR ($N_{\text{pan}} = 205$)	sim4	table_summary_complete.pdf
Table 10: Clustering: OLS - BASELINE ($N_{\text{pan}} = 205$)	sim1	table_clustering_complete.pdf
Table 11: Clustering: OLS - SAR ($N_{\text{pan}} = 205$)	sim2	table_clustering_complete.pdf
Table 12: Clustering: IV - BASELINE ($N_{\text{pan}} = 205$)	sim3	table_clustering_complete.pdf
Table 13: Clustering: IV - SAR (	sim4	table_clustering_complete.pdf

$N_{\text{pan}} = 205$)		
Table 14: Summary: OLS - BASELINE ($N_{\text{pan}} = 820$)	sim5	table_summary_complete.pdf
Table 15: Summary: OLS - SAR ($N_{\text{pan}} = 820$)	sim6	table_summary_complete.pdf
Table 16: Summary: IV - BASELINE ($N_{\text{pan}} = 820$)	sim7	table_summary_complete.pdf
Table 17: Summary: IV - SAR ($N_{\text{pan}} = 820$)	sim8	table_summary_complete.pdf
Table 18: Clustering: OLS - BASELINE ($N_{\text{pan}} = 820$)	sim5	table_clustering_complete.pdf
Table 19: Clustering: OLS - SAR ($N_{\text{pan}} = 820$)	sim6	table_clustering_complete.pdf
Table 20: Clustering: IV - BASELINE ($N_{\text{pan}} = 820$)	sim7	table_clustering_complete.pdf
Table 21: Clustering: IV - SAR ($N_{\text{pan}} = 820$)	sim8	table_clustering_complete.pdf

Figure	Simulation No.	Output file under /output/
Figure 2: OLS power curves	sim1-2	power_curve_ols_baseline_small.pdf, power_curve_ols_sar_small.pdf
Figure 3: IV power curves	sim3-4	power_curve_iv_baseline_small.pdf, power_curve_iv_sar_small.pdf
Figure 4: OLS power curves ($N_{\text{pan}} = 820$)	sim5-6	power_curve_ols_baseline_large.pdf, power_curve_ols_sar_small.pdf
Figure 5: IV power curves ($N_{\text{pan}} = 820$)	sim7-8	power_curve_iv_baseline_large.pdf, power_curve_iv_sar_small.pdf