How to Conduct Inference with Spatial Dependence in Stata

NAME   Damian Clarke   NAME   NAME   NAME
Department of Economics
University of Chile
dclarke@fen.uchile.cl

Guido Imbens
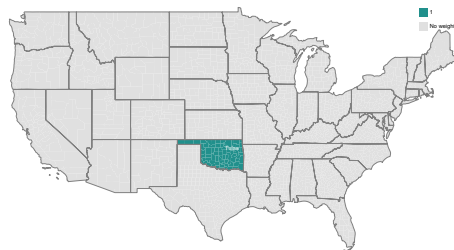Graduate School of Business
Stanford University
imbens@stanford.edu

Abstract. In this paper...
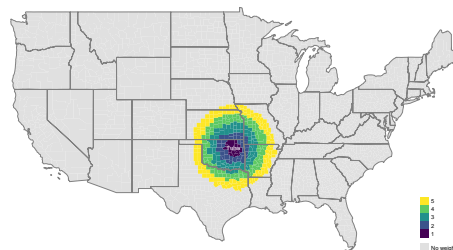
Keywords: spatial dependence, inference.

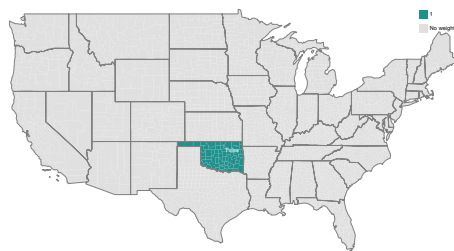# 1   Introduction

DellaVigna et al. (2025)
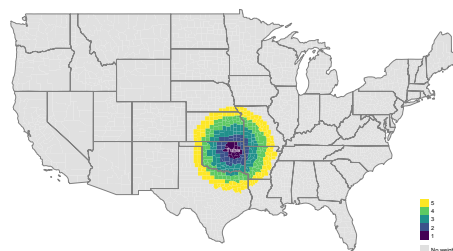
Conley (1999)

(a) Clustering by State

(b) Conley

(c) SCPC

(d) TMO

Figure 1: Spatial Standard Error Procedures

2   Methods

Consider an equation of the form . . .

Moulton (1986)

Müller and Watson (2022)

3   The tmo command

3.1   Syntax

tmo, cmd(*cmdline*) x(*varname*) ylist(*varname*) idvar(*varname*) $\left[\,\text{options}\,\right]$

3.2 Options

cmd(*cmdline*) *cmdline* is the command that produces the regression of interest. tmo currently supports regressions using `regress`, `reghdfe`, `ivreg2`, or `ivreghdfe`.

x(*varname*) Regressor of interest in `cmd` for which to estimate TMO standard errors. tmo *estimates the standard error for only this declared independent variable.*

ylist(*varlist*) *List of auxiliary outcomes to use in* tmo.

idvar(*varname*) Location identifier variable; must be unique (within each time period for panel case).

misslimit(*#*) Limit for proportion of observations allowed to be missing for auxiliary outcomes. Auxiliary outcomes missing more than `misslimit` are not used. Must be between [0,1]; default is 0.1.

Panel Setting

timevar(*varname*) Time identifier variable; must be declared for panel case.

Distance-based Settings

latitude(*varname*) Latitude variable in signed decimal degrees.

longitude(*varname*) Longitude variable in signed decimal degrees.

distthreshold(*#*) Distance threshold in miles to allow for arbitrary correlation between pairs of locations that are `distthreshold` or fewer miles apart. Combines tmo *with a Conley adjustment using a uniform kernel. Requires* `latitude` *and* `longitude`.

Saving Figures and Tables

filesuffix(*str*) Folder path and base filename for saving figures and results. Required for plot or save options.

savedyad Save Stata data file with correlation and contribution to standard error for each location pair.

plotq Save plot of optimal threshold estimator.

plothist Save plot for histogram of correlations between locations.

plothistnbins(*#*) Number of bins for histogram of correlations (default 10000).

plotse Save plot for standard error estimates across thresholds.

saveplot seest Save Stata data file with standard error estimates across thresholds.

saveest Save results in `r()` to Stata data file.

Custom Threshold

`threshold`(#) Set custom threshold instead of using the optimal threshold from the interquartile range method. Must be between [0,1].

`thresholdoff` Turns off the tmo *adjustment entirely.*

SCPC Options

`scpc_cmd` (*cmdline*) Command for regression of interest before applying SCPC adjustment. Combines tmo with the SCPC method from Müller and Watson (2022).

`scpc_uncond` Turns on the unconditional SCPC inference setting.

Returned Objects

`sdid` stores the following in `e()`:

Scalars:

`e(ATT)`   Average Treatment Effect on the Treated

The matrices `e(b)` and `e(V)` are included to facilite the exportation of results from `sdid` with routines such as `estout`.

4   Examples

```
. use ../example/county_differences
. qui ds fips stfips PIN_persincpc_d EDU_college_d, not
. local ylist `r(varlist)´
.
. tmo, cmd(reg PIN_persincpc_d EDU_college_d i.stfips, r) x(EDU_college_d) ylist(`ylist´) i(fips)
Linear Regression with TMO                          Number of obs =  3,060
                                                    R-squared     = 0.3434
                                                    Adj R-squared = 0.3434
                                                    Root MSE      = 0.1409
```

PIN_persinc~d	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
EDU_college_d	.1745893	.0197081	8.86	0.000	.1359466	.213232

```
                                      Optimal threshold      0.389
                            % of off-diag in SE est.      0.497
              % >= threshold (excl. clusters/Conley)      0.497
                                          # outcomes     90.000
                                  Degrees of freedom     59.986
```
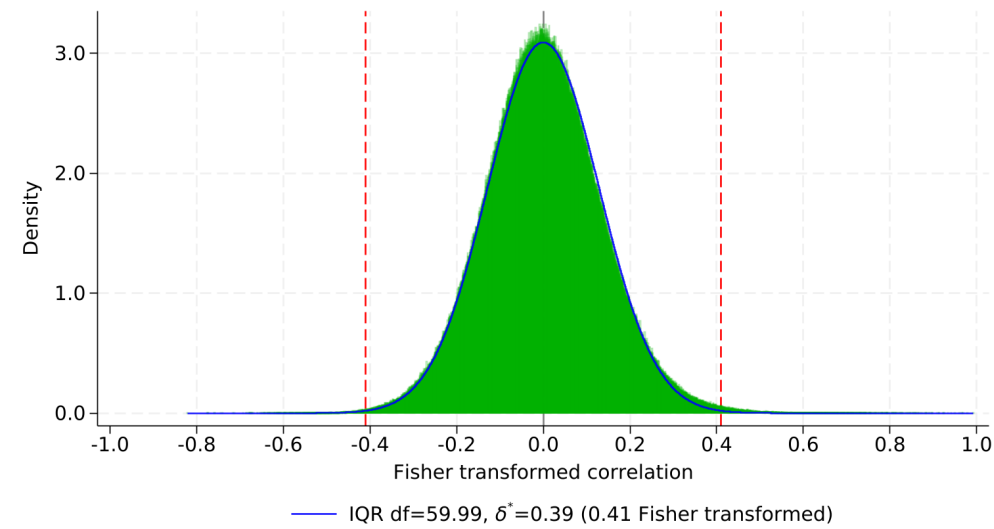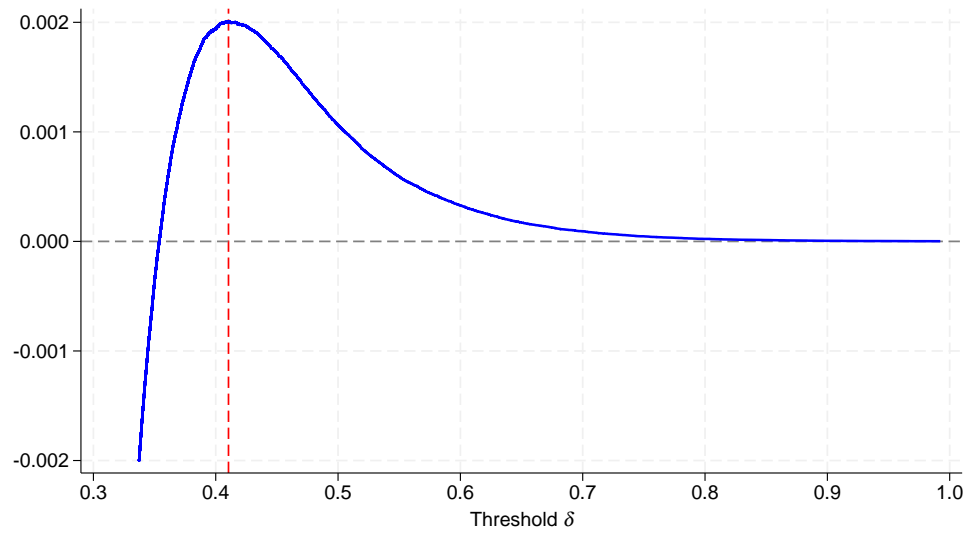
IQR df=59.99, δ^*=0.39 (0.41 Fisher transformed)

Figure 2: Correlations between residuals across US counties

Figure 3: Optimal Thresholding

5 Conclusion

Acknowledgments

We are grateful to . . .

6 References

Conley, T. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* 92(1): 1–45.

DellaVigna, S., G. Imbens, W. Kim, and D. M. Ritzwoller. 2025. Using Multiple Outcomes to Adjust Standard Errors for Spatial Correlation. https://arxiv.org/abs/2504.13295.

Moulton, B. R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Müller, U. K., and M. W. Watson. 2022. Spatial Correlation Robust Inference. *Econometrica* 90(6): 2901–2935.

About the authors

Susan Athey is the Economics of Technology Professor at Stanford Graduate School of Business.

Damian Clarke is an Associate Professor at The Department of Economics of The Universidad de Chile, a Research Fellow at IZA and an Associate at the Millennium Institute for Market Imperfections and Public Policy and CAGE, Warwick.

Guido Imbens is the Applied Econometrics Professor and Professor of Economics at Stanford Graduate School of Business.

Daniel Pailañir is a Senior Analyst at Ministry of Economics, Development and Tourism of Chile.