Spatial correlation

Hans Martinez

2023-07-10

1 Computing rejection probability

$\mathbb{P}_{c \geq c_{max}}(\cdot)$ computes the maximum rejection probability for a range of c values that go from c_{min} —which corresponds to $\hat{\rho}$— to c_{max}, a minimum value for which size control is checked. In the code, this value corresponds to a minimum average pairwise correlation of 0.00001 (line 478, `scpc.ado`).

The rejection probability is computed using a Gaussian quadrature and the eigenvalues of $\Omega(c)$ (pag. 12), for a given critical value.

Then, the critical value is selected such that the rejection probability is the desired, for example, $\alpha = 0.05$ (line 144, `scpc.ado`)

Follow-up questions

Follow-up questions to look into the documentation and figure out:

1. What are they reporting as standard errors. Guess sq.root of sigma hat of eq. 5
2. what are they reporting as "t" and "P>|t|"
3. what is q and why 10 is the "optimal" value
4. reverse engineer the critical value from the confidence interval

The Stata documentation is very succinct and does not provide details for the questions above. I had to figure out from the code itself.

Question 1: The `scpc` post-estimation command displays the standard error as the estimated variance using eq. 5 modified as follows $\hat{\sigma}_{SCPC}/(\sqrt{n}\sqrt{n-1})$ (line 559, scpc.ado).

Q2: They report "t", using eq. 4, but with their variance estimator. "P>|t|" is the rejection probability, eq. in pag. 13 top left. Line 110 in `scpc.ado`.

Q3: q is the number of principal components used to estimate eq. 5. There is a trade-off in choosing the optimal q. For a critical value fixed, the confidence interval shrinks for larger values of q.

Q4: It can be done. $cv = (coef - LCI)/SE$. The critical value is computed using the equation on page 13 (line 144, scpc.ado).

Intro

Open questions to address in Spatial Correlation Principal Component (SCPC) (Müller and Watson 2022b, 2022a):

1. What's the "worst case" covariance matrix? How does that work?
2. Can SCPC accommodate a vector of variables?
3. Does the method work if there exists measurement error in the locations?

2 The setting

$$y_l = x_l\beta + e_l$$

$\mathbb{E}[e_l|x_l] = 0$, and (y_l, x_l, e_l) are associated with observed spatial location $s_l \in \mathbb{R}^d$. Spatial location in time is $d = 1$, two dimensions, like in altitude and latitude, $d = 2$, and so on.

e_l is generated by a Gaussian process with covariance function $cov(e_l, e_{l'}) = \exp(-c||s_l - s_{l'}||)$, where s_l and $s_{l'}$ denote the spatial locations of e_l and $e_{l'}$, and $c > 0$ is a parameter that governs the strength of the spatial correlation. The value of c is calibrated to induce a specific average pairwise correlation $\bar{\rho} = [n(n-1)]^{-1}\sum_{l,l'\neq l} cov(e_l, e_{l'})$. In other words, $c = c_{\bar{\rho}}$ solves $[n(n-1)]^{-1}\sum_{l,l'\neq l}\exp(-c_{\bar{\rho}}||s_l - s_{l'}||) = \bar{\rho}$.

Let $\Sigma(c)$ be the covariance matrix of e_l evaluated at the sample locations, so that $\Sigma(c)_{l,l'} = \exp(-c||s_l - s_{l'}||)$, and let $\bar{\rho}(c)$ denote the resulting average pairwise correlation $\bar{\rho}(c) = [n(n-1)]^{-1}\sum_{l,l'\neq l}\Sigma(c)_{l,l'}$. If the researcher desires a test that controls size for values of $\bar{\rho}$ as large as $\bar{\rho}_{max}$, then he can choose c_{min} such that $\bar{\rho}(c_{min}) = \bar{\rho}_{max}$. Then, $\Sigma(c_{min})$ is the **worst case** covariance matrix in the sense that it induces the largest value of σ^2 among all $\Sigma(c)$ with $\bar{\rho} \leq \bar{\rho}_{max}$

3 Literature Review

3.1 Müller and Watson (2022b)

The objective of the paper is to develop a robustified version of their Spatial Correlation Principal Components (SCPC) (Müller and Watson 2022a). In particular, the authors propose modifications to deal with non-stationarities and strong dependence in finite samples (small samples).

According to the authors, the SPCPC method addresses the challenge of spatial correlation robust inference under small samples and *empirically relevant* forms of strong dependence (which ones?).

According to the authors, Conley (1999) doesn't work well in small samples because it relies on the consistency of the estimator of σ^2, while SCPC (and fixed-b type approaches) rely on the *stationarity* of u_l. Stationarity might break in practice, for example, when x_l is a dummy for treatment, and treatment is more likely in one region than another region.

The SCPC method is based on a principal component estimator of σ^2 based on a pre-specified "worst-case" exponential covariance function conditional on the observed locations.

(On going…)

4 Worst case covariance function

Müller and Watson (2022b) assume a functional form for the covariance function of the error term that depends on a single parameter. This single parameter has a unique value for a specific average pairwise correlation. The researcher has to provide before the estimation, a maximum average pairwise correlation. Then, because the average pairwise correlation is a decreasing function of the parameter of the covariance function, this parameter induces the largest value of the error variance for the given average pairwise correlation, according to the authors. In the STATA and MATLAB codes, this parameter is set to 0.03 in both, like in their paper. The authors provide instructions on how to adjust the parameter in the help section of the STATA code. In the MATLAB code, a parameter can be adjusted before estimations.

References

Conley, T G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92: 1–45.

Müller, Ulrich K, and Mark W Watson. 2022a. "SPATIAL CORRELATION ROBUST INFERENCE." *Econometrica* 90: 2901–35. https://doi.org/10.3982/ECTA19465.

———. 2022b. "Spatial Correlation Robust Inference in Linear Regression and Panel Models." *Journal of Business & Economic Statistics* 00: 1–15. https://doi.org/10.1080/07350015.2022.2127737.