Spatial correlation

Hans Martinez

2023-07-13

1 Power of the test using SCPC

Müller and Watson (2022b) discuss briefly the power of the tests using their variance estimator.

1.1 Conditional vs. Unconditional SCPC (p.8)

The authors argue that the size-adjusted power (average length of the confidence intervals) of SCPC and C-SCPC (conditional) are identical because both methods are based on the same t-statistic and only differ in their critical values. The authors argue that the differences in the average confidence interval lengths of both methods are small. They based their argument on an unreported comparison realized for only one experiment design (p. 8).

The authors acknowledge that the size-adjusted Kernel method is "somewhat" more efficient than SCPC and C-SCPC. The authors argue however that this is not a reason to prefer the Kernel method because in practice the "size-adjustment that adjust for the larger bias in the $\hat{\sigma}^2_K$" is not feasible.

1.2 Trade-off in choosing q (p.4)

The authors discuss that there is a trade-off between the number of principal components q and the expected length of the 95% confidence intervals. In particular, for a fixed critical value, the expected length of the confidence interval falls as q increases, but larger q requires larger critical values to control coverage. The authors suggest that minimizing the expected 95% confidence interval length under the iid benchmark yields a value of q that works well for a range of values of c.

1.3 Expected length of confidence intervals as a function of $\bar{\rho}_{max}$ (p.11,12)

The authors also discuss what happens to the rejection frequencies and confidence intervals if the researcher misjudges the spatial correlation $\bar{\rho}_{max}$.

2 Summary of the `scpc` stata output

Description of the stata output from the post-estimation command `scpc`:

- Coefficient
- S.E.: The `scpc` post-estimation command displays the standard error as the SCPC estimated variance using eq. 5 modified as follows $\hat{\sigma}_{SCPC}/(\sqrt{n}\sqrt{q})$ (line 559, `scpc.ado`).
- t: The command reports the t-statistic, using eq. 4, but using the SCPC variance estimator (eq. 5)
- "P>|t|": This value corresponds to the maximum rejection probability for the adjusted t-statistic. The maximum rejection probability, $\mathbb{P}_{c \geq c_{max}}(\cdot)$, is the maximum of the rejection probabilities computed for a range of c values, that go from c_{min} to c_{max}. c_{min} is provided by the researcher as a maximum average pairwise correlation, $\bar{\rho}$. The default is $\bar{\rho} = 0.03$. c_{max} is the maximum value for which size control is checked. In the code, c_{max} corresponds to a minimum average pairwise correlation of 0.00001 (line 478, `scpc.ado`).
- Confidence interval: The `scpc` command computes the CI using the critical value that corresponds to a 5% maximum rejection probability. For the range of c values $\in [c_{min}, c_{max}]$, the algorithm searches for the critical value that corresponds to a maximum rejection probability $\alpha = 0.05$, the desred size of the test.

3 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`)

4 Replication (updated)

Here is the second try in replicating Table 1 of Müller and Watson (2022b). In all models, locations are $s_l \sim U(0,1)$ and $\beta_0 = 0$. In all models $e_l \sim \mathcal{G}_{exp}(c_{0.030})$ except model 3. In model 1, $x_l = 1$. In model 2, $x_l \sim \mathcal{G}_{exp}(c_{0.03}/2)$ and $e_l \sim \mathcal{G}_{exp}(c_{0.03}/2)$. In model 3, x_l is a step function with $x_l = -0.15$ for the 85% of the locations closest to $s = 0$, and $x_l = 0.85$ for the remaining locations closest to $s = 1$. In model 4, x_l follows a demeaned random walk.

I ran 1,000 simulations for each model. I used 250 observations for each iteration. One location per observation. The authors do not specify clustering at all.

The rejection frequencies are the percentage of times that the t-statistic was above the critical value of 5%. For the heteroskedastic robust (HR), the normal standard critical values were used. For the SCPC, the critical values were estimated as in Müller and Watson (2022b) using the accompanying post-estimation stata command `scpc, cvs`. The HR method uses the `robust` method in Stata. The SCPC uses the post-estimation command `scpc` provided by the authors.

The rejection frequencies in Table 1 are now closer to what Müller and Watson (2022b) show.

Table 1: Rejection frequencies

Model	HR	SCPC	Mueller_HR	Mueller_SCPC
1	0.23	0.02	0.51	0.05
2	0.30	0.04	0.52	0.08
3	0.24	0.02	0.52	0.15
4	0.22	0.01	0.50	0.08

4.1 Drawing from $\mathcal{G}_{exp}(c)$

To draw from $\mathcal{G}_{exp}(c_{0.03})$, first I estimated the covariance matrix $\Sigma(c)$ using the covariance function $cov(e_l, e_{l'}) = \exp(-c||s_l - s_{l'}||)$ for a value of c. Then, I used the covariance matrix to draw from a multivariate normal distribution with mean zero and covariance matrix $\Sigma(c)$.

5 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 $\bar{\rho} \leq \bar{\rho}_{max}$

6 Literature Review

6.1 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…)

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.