# Spatial Correlation Robust Inference in Linear Regression and Panel Models

## SUPPLEMENTAL MATERIAL

Revision: September 7, 2022

# 1 World Development Indicators Data

This appendix summarizes the pre-processing of variables from the World Development Indicators database prior to their use in Section 3.

- For the 2015-2005 decadal difference dataset:
  - Transformations: The level of the series was used if (a) its description included the word *index* or included the symbol % or (b) the series included non positive values. Otherwise the logarithm of the series was used. Let $y = x_{2015} - x_{2005}$ denote the decadal difference of the transformed series
  - Outlier adjustments: If $y$ contained 4 or fewer unique values, no outlier adjustment was performed. Otherwise, observations with $|y_l - \text{median}(y)| > 5 \times \text{IQR}$ were replaced with missing values, where IQR denotes the inter-quartile range.
  - Series were discarded if: (1) they contained fewer than 100 countries, (2) had a sample kurtosis greater than 20, (3) had fewer than 10 countries with different values.
  - This resulted in a dataset with 749 series.
- For the 2006-2015 panel dataset:
  - Transformations: Same as above. Let $x_{t,l}$ denote the transformed series, $\overline{x}_l$ denote the mean for country $l$, $\overline{x}$ denote the sample mean over all $t$ and $l$, and $y_{t,l} = x_{t,l} - \overline{x}_l$ denote the country-demeaned data
  - Outlier adjustment: Observations with $|y_{t,l} - \text{median}(y)| > 5 \times \text{IQR}$ were replaced with missing values.

- $y_{t,l}$ was replaced with a missing value if $y_{\tau,l}$ was missing for any value of $\tau \in [2006, 2015]$. Thus, the panel is balanced.
- The series was discarded if it contained non-missing values for fewer than 100 countries or had a kurtosis greater than 20.
- This resulted in a dataset with 644 series.

# 2 Size control under an alternative model of conditional heteroskedasticity, additional spatial persistence and Expected Length of Confidence intervals as a function of $\overline{\rho}_{\mathbf{max}}$

## 2.1 Size control under an alternative model of conditional heteroskedasticity

The experiments summarized in Figures 2-4 were carried out, but with a conditionally heteroskedastic version of the error $e$. Let $\mathbf{\Sigma}_0$ denote the covariance matrix for $\mathbf{e}$ from the models summarized in panels (a) of Figures 2-4. The heteroskesdastic models used the covariance matrices $\mathbf{\Sigma}_{\text{Het}} = \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}'$ where $\mathbf{D}$ is a diagonal matrix whose $i$th element for the $l$th cluster is $(1 + x_{i,l})^{1/2}$, and where the regressors for each experiment are scaled so that $N^{-1}\sum_{i,l} x_{i,l}^2 = 1$. The following three figures summarize rejection frequencies for SCPC and C-SCPC in the homoskedastic models (shown previously in Figures 2-4) and in their heteroskesdastic counterparts.

Figure 2(a) with homoskesdatic and heteroskesdastic errors

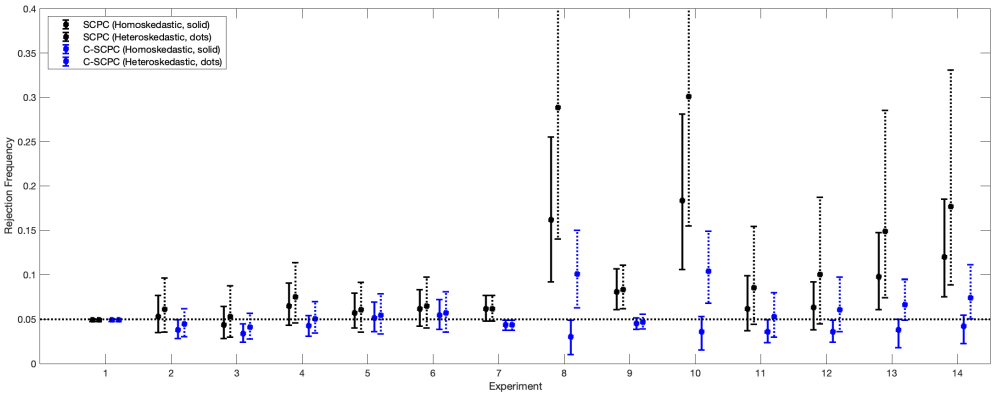



Figure 3(a) with homoskesdatic and heteroskesdastic errors

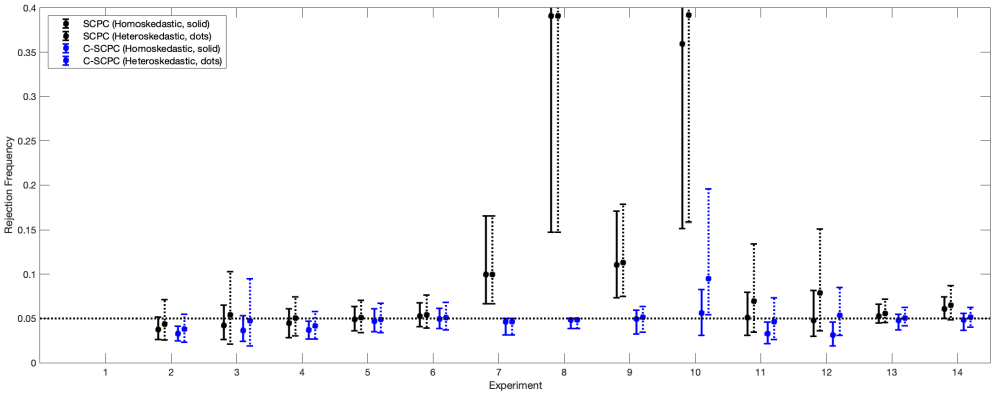
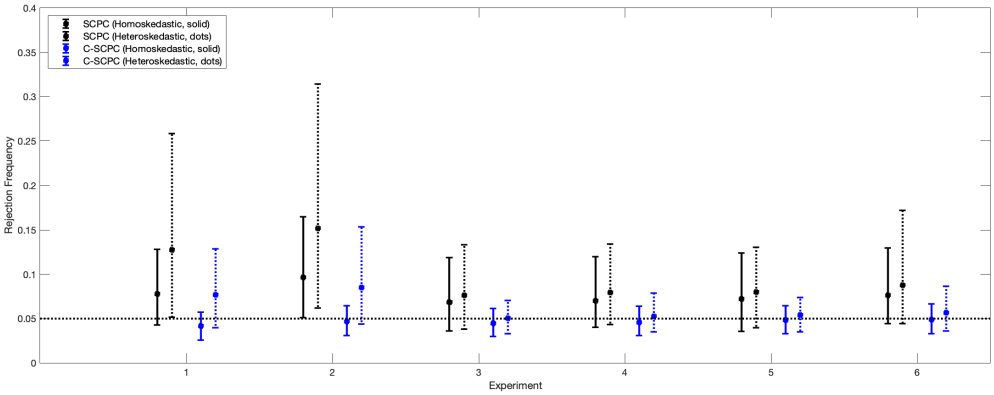



Figure 4(a) with homoskesdatic and heteroskesdastic errors



## 2.2 Size control with additional spatial persistence

The experiments summarized in panel (a) Figures 2-4 were carried out, but with errors that were more persistent. The results in the paper generated the errors $e$ from $\mathcal{G}_{exp}(c_{0.03})$ processes, and carried out the SCPC and C-SCPC tests using $c_{\max} = c_{0.03}$. In the experiments reported here the errors $e$ were generated from $\mathcal{G}_{exp}(c_{0.10})$, but the SCPC and C-SCPC tests continued to use $c_{\max} = c_{0.03}$. The following three figures summarize rejection frequencies for SCPC and C-SCPC using $c_{max} = c_{0.03}$ with the $e \sim G_{exp}(c_{0.03})$ (shown previously in Figures 2-4) and with more persistent $e \sim G_{exp}(c_{0.03})$ errors.

Figure 2(a) with $e \sim G_{exp}(c_{0.03})$ and $e \sim G_{exp}(c_{0.10})$ errors

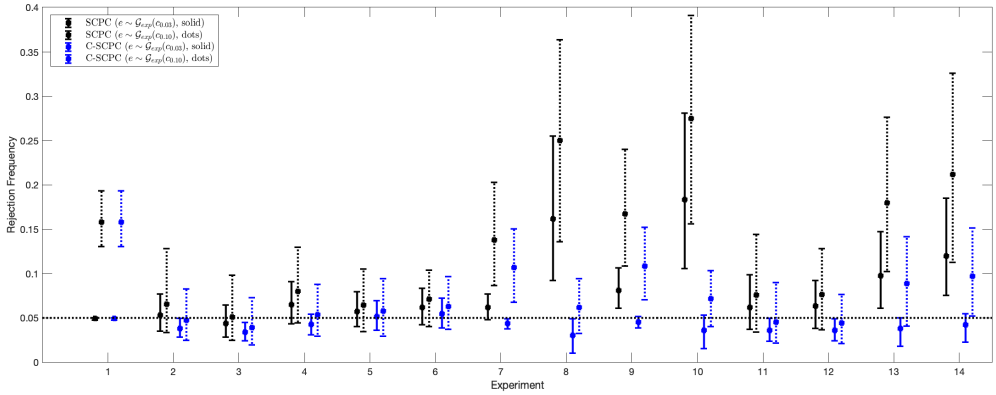



Figure 3(a) with $e \sim G_{exp}(c_{0.03})$ and $e \sim G_{exp}(c_{0.10})$ errors

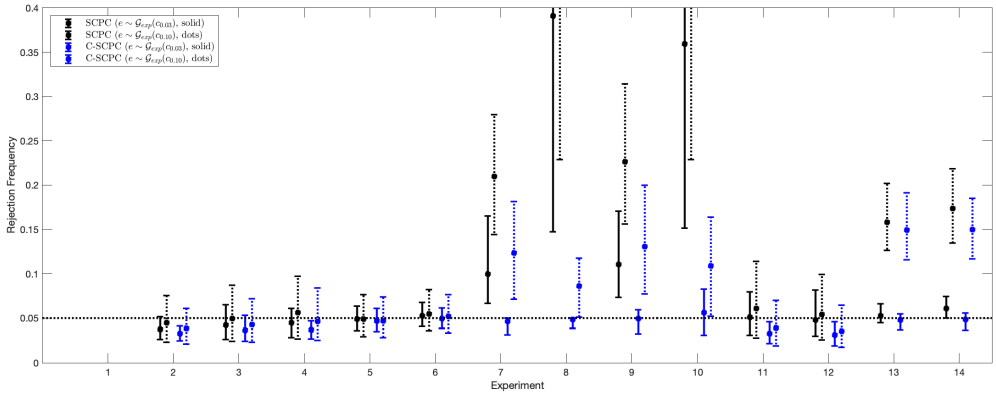
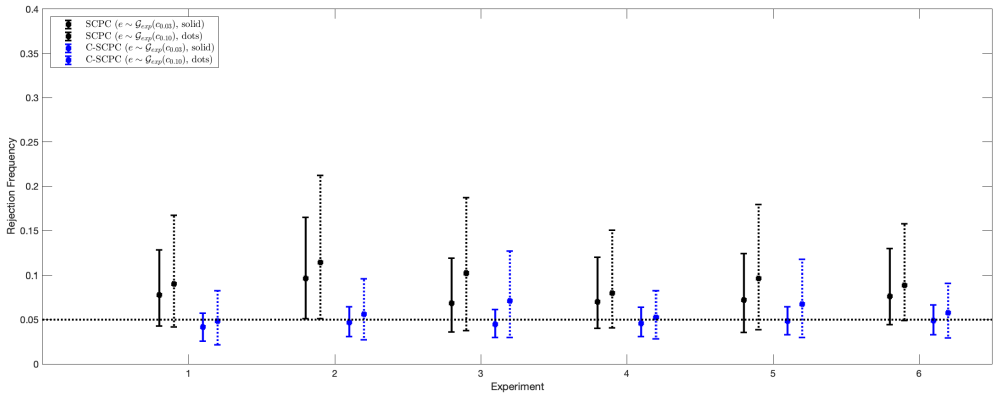



Figure 4(a) with $e \sim G_{exp}(c_{0.03})$ and $e \sim G_{exp}(c_{0.10})$ errors



## 2.3 Expected length of confidence intervals as a function of $\overline{\rho}_{\max}$

The average length (over realizations of $e_l$, conditional on the regressors) of C-SCPS confidence intervals were computed in the experiments in Figures 2(a) and 3(a). The figures below show summarize the distribution and of these lengths (across realizations of the regressors) relative to the length of the oracle confidence interval with $\mathbf{e} \sim N(0, I_N)$.

Figure 2(a): Expected length of 95% confidence interval (relative to oracle)

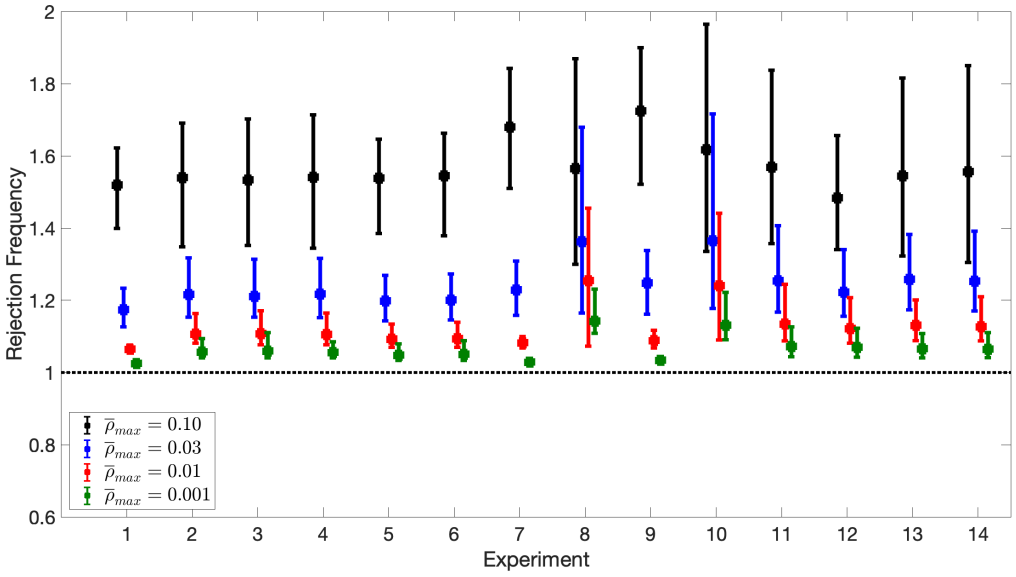
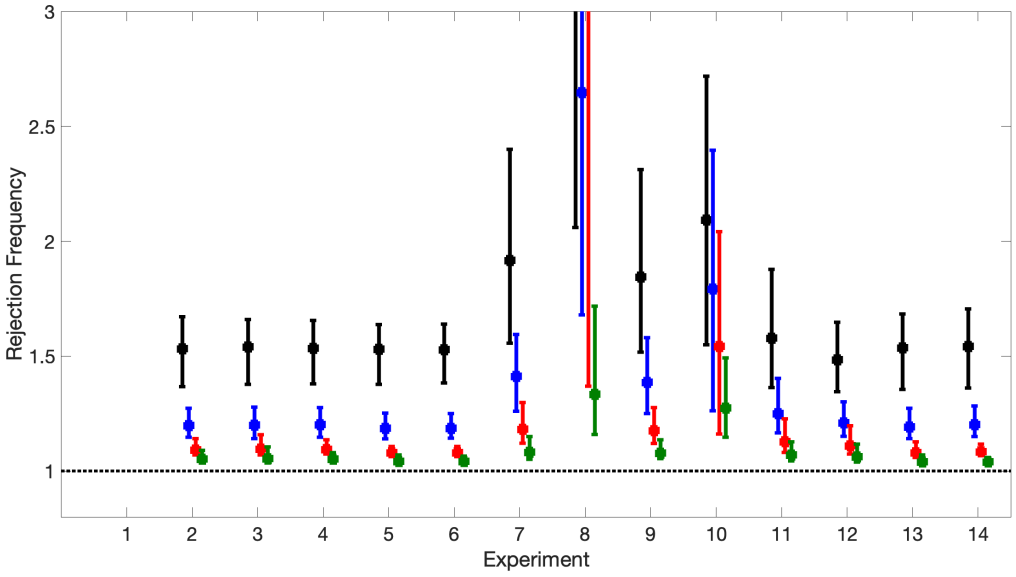



Figure 3(a): Expected length of 95% confidence interval (relative to oracle)




Notes: The bars show the 5th through 95th quantiles of the distribution of the average length of confidence intervals conditional on the regressors. Values are relative the length of the oracle confidence interval in the model with $\mathbf{e} \sim$ *i.i.d.* $N(0, I_N)$.