The Modifiable Areal Unit Problem in Political Science

Online Appendix

Contents

List of Tables

List of Figures

1 Monte Carlo Simulation with Regular Lattice Data: Data Generation Process

We designed the seed for a random sample in a matrix of 27 rows by 3 columns with the following possible variable values: 0, 1, $-$, in which $-$ represents a non-observation. The 3 columns represent the three variables present for each location, which we conceptualize as dependent variable, independent variable 1, and independent variable 2. We visualize them in the matrix format below as an example.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ - & - & - \\ - & - & 0 \\ - & - & 1 \\ - & 0 & - \\ - & 1 & - \\ 0 & - & - \\ 1 & - & - \\ - & 0 & 0 \\ - & 0 & 1 \\ - & 1 & 0 \\ - & 1 & 1 \\ 0 & - & 0 \\ 0 & - & 1 \\ 1 & - & 0 \\ 1 & - & 1 \\ 0 & 0 & - \\ 0 & 1 & - \\ 1 & 0 & - \\ 1 & 1 & - \end{bmatrix}_{27 \times 3}$$

From these 27 variations, we drew a random sample of 10,000 rows, with replacement, to create a 100 rows by 100 columns square matrix. In other words, we created a grid with an array of (1 row $\times$ 3 columns) $\times$ 100 rows $\times$ 100 columns. Using this square grid, we can generate square spatial units of various sizes (4 locations, 16 locations, 25 locations, 100 locations) to examine the effects of the scale sub-problem and shift them across the larger grid to simulate the zoning sub-problem of the MAUP. We draw using both a uniform random sample distribution, and a distribution with unknown properties.

1.1 Summary Statistics

Tables A.1, A.2, and A.3 offer the summary statistics of the first value (which we designate as the dependent variable), the second value (IV1), and the third value (IV2), in a randomly selected row. A couple of things are worth noting. First, the mean for the uniform sample hovers closely around 0.5 across all of the mappings, as expected for random data, while the mean for this variable in the non-uniform sample has a mean closer to 0.4 across all mappings. The consistent mean values with changes to the aggregation and zone are consistent with the nature of the MAUP and with the findings of previous simulation exercises (Amrhein 1995). We see differences in other properties of this variable: the minimum, maximum, and skew increase as the aggregation and zoning changes, but not in a linear fashion, and spatial autocorrelation is also inconsistent across the mappings.[1]

1.2 Correlation in Simulated Data

Figure A.1 plots the first outcome of our simulations–the Pearson correlation coefficients to estimate the strength of a linear association between each pair of variables: the first variable (DV), the second variable (IV1), and the third variable (IV2, or "control variable"). On the left side, we show results from the uniform probability simulation, and on the right side of the non-uniform probability simulation. The top figures show results when the locations are aggregated to different values (scaling) and then shifted in the eastward direction on our grid map (zoning). The bottom figures show the aggregations and shifts in a southward direction. In all figures, the farther away from the left side of the figure, the larger the scale (we move from 1 to 4 to 16 to 25 to 100 units), and the more shifted the zone from its original location.

[1]For Moran's I statistics and their p-values against the null of spatial autocorrelation, we used grid mappings to generate a matrix of inverse Euclidean distance weights.

Table A.1: Distributions of Simulated Data (Dependent Variable)

Modified Areal Units	Uniform Sampling Probability							Non-uniform Sampling Probability						
	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I
Monte Carlo Simulated Data	10,000							10,000						
1 Grid	6721	0.506	0.500	0.000	1.000	-0.024	0.000	7138	0.413	0.492	0.000	1.000	0.354	-0.000
4 Grids (Aggregated)	2483	0.510	0.331	0.000	1.000	-0.025	-0.000	2485	0.412	0.310	0.000	1.000	0.273	0.000
4 Grids (+1 Shift Eastward)	2470	0.506	0.330	0.000	1.000	-0.018	-0.001	2477	0.416	0.312	0.000	1.000	0.288	-0.000
4 Grids (+1 Shift Southward)	2469	0.507	0.325	0.000	1.000	-0.021	-0.000	2478	0.412	0.306	0.000	1.000	0.256	-0.000
16 Grids (Aggregated)	625	0.507	0.159	0.091	1.000	0.000	-0.002	625	0.415	0.152	0.000	0.875	0.160	-0.003
16 Grids (+1 Shift Eastward)	625	0.507	0.161	0.000	1.000	-0.070	-0.004	625	0.414	0.151	0.000	1.000	0.301	-0.001
16 Grids (+1 Shift Southward)	625	0.506	0.155	0.000	1.000	0.016	-0.001	625	0.414	0.146	0.000	0.875	0.230	-0.001
16 Grids (+2 Shifts Eastward)	625	0.506	0.160	0.000	0.917	-0.186	-0.003	625	0.413	0.151	0.000	0.889	0.021	-0.001
16 Grids (+2 Shifts Southward)	625	0.506	0.156	0.091	0.923	0.028	-0.002	625	0.414	0.143	0.000	0.778	0.093	-0.001
16 Grids (+3 Shifts Eastward)	625	0.506	0.158	0.000	1.000	0.046	-0.001	625	0.414	0.149	0.667	0.900	0.252	-0.002
16 Grids (+3 Shifts Southward)	625	0.507	0.157	0.000	0.900	0.067	-0.003	625	0.414	0.146	0.091	0.900	0.248	-0.000
25 Grids (Aggregated)	400	0.507	0.129	0.118	0.889	0.061	-0.002	400	0.412	0.118	0.067	0.765	0.113	-0.002
25 Grids (+1 Shift Eastward)	400	0.506	0.125	0.063	0.857	-0.069	-0.006	400	0.413	0.121	0.056	0.750	0.110	-0.002
25 Grids (+1 Shift Southward)	400	0.506	0.123	0.188	0.867	-0.061	-0.001	400	0.413	0.116	0.143	0.789	0.258	-0.002
25 Grids (+2 Shifts Eastward)	400	0.507	0.120	0.118	0.867	-0.045	-0.002	400	0.413	0.115	0.063	0.706	0.011	0.001
25 Grids (+2 Shifts Southward)	400	0.507	0.127	0.167	0.875	0.022	-0.001	400	0.414	0.115	0.150	0.750	0.270	-0.002
25 Grids (+3 Shifts Eastward)	400	0.507	0.126	0.125	0.875	-0.009	-0.006	400	0.413	0.117	0.071	0.765	0.064	-0.001
25 Grids (+3 Shifts Southward)	400	0.507	0.125	0.125	0.818	-0.036	-0.001	400	0.414	0.116	0.125	0.737	0.191	-0.002
25 Grids (+4 Shifts Eastward)	400	0.507	0.126	0.133	0.867	0.014	-0.003	400	0.413	0.116	0.125	0.733	0.050	0.001
25 Grids (+4 Shifts Southward)	400	0.507	0.127	0.111	0.833	-0.166	-0.002	400	0.413	0.113	0.125	0.737	0.209	0.000
100 Grids (Aggregated)	100	0.506	0.066	0.338	0.698	0.542	-0.016	100	0.412	0.056	0.222	0.547	0.020	-0.014
100 Grids (+1 Shift Eastward)	100	0.506	0.061	0.388	0.662	0.613	-0.015	100	0.412	0.055	0.234	0.556	0.121	-0.014
100 Grids (+1 Shift Southward)	100	0.506	0.066	0.328	0.671	0.358	-0.012	100	0.413	0.056	0.250	0.548	0.003	-0.009
100 Grids (+2 Shifts Eastward)	100	0.506	0.059	0.380	0.667	0.413	-0.013	100	0.413	0.054	0.266	0.526	-0.207	-0.008
100 Grids (+2 Shifts Southward)	100	0.506	0.068	0.333	0.708	0.532	-0.016	100	0.413	0.058	0.262	0.548	-0.071	-0.006
100 Grids (+3 Shifts Eastward)	100	0.506	0.059	0.380	0.667	0.413	-0.013	100	0.413	0.054	0.266	0.526	-0.207	-0.008
100 Grids (+3 Shifts Southward)	100	0.506	0.068	0.333	0.708	0.532	-0.016	100	0.413	0.058	0.262	0.548	-0.071	-0.006
100 Grids (+4 Shifts Eastward)	100	0.506	0.055	0.375	0.667	0.304	-0.010	100	0.413	0.055	0.284	0.541	-0.191	-0.005
100 Grids (+4 Shifts Southward)	100	0.506	0.067	0.313	0.688	0.429	-0.008	100	0.413	0.061	0.234	0.543	-0.243	-0.010
100 Grids (+5 Shifts Eastward)	100	0.506	0.055	0.351	0.661	0.039	-0.011	100	0.413	0.057	0.277	0.541	-0.300	0.000
100 Grids (+5 Shifts Southward)	100	0.506	0.068	0.338	0.689	0.326	-0.007	100	0.412	0.063	0.266	0.600	0.290	-0.005
100 Grids (+6 Shifts Eastward)	100	0.506	0.058	0.342	0.639	-0.128	-0.003	100	0.413	0.056	0.271	0.563	-0.159	-0.002
100 Grids (+6 Shifts Southward)	100	0.506	0.063	0.353	0.681	0.267	-0.003	100	0.412	0.066	0.242	0.577	0.202	-0.010
100 Grids (+7 Shifts Eastward)	100	0.506	0.060	0.333	0.646	-0.101	-0.003	100	0.413	0.058	0.271	0.559	-0.116	-0.007
100 Grids (+7 Shifts Southward)	100	0.506	0.062	0.368	0.667	0.365	-0.000	100	0.413	0.061	0.258	0.577	0.107	-0.012
100 Grids (+8 Shifts Eastward)	100	0.506	0.065	0.343	0.692	0.214	-0.003	100	0.413	0.055	0.264	0.528	-0.119	-0.009
100 Grids (+8 Shifts Southward)	100	0.506	0.060	0.389	0.672	0.391	0.008	100	0.413	0.059	0.278	0.583	0.253	-0.013
100 Grids (+9 Shifts Eastward)	100	0.506	0.067	0.365	0.692	0.323	-0.003	100	0.413	0.055	0.217	0.534	-0.204	-0.008
100 Grids (+9 Shifts Southward)	100	0.506	0.059	0.378	0.689	0.361	-0.003	100	0.412	0.055	0.281	0.568	0.499	-0.007

Note: Moran's I statistics range from -1 (perfect clustering of dissimilar values) through 0 (no spatial autocorrelation) to 1 (perfect clustering of similar values).

Table A.2: Distributions of Simulated Data (Independent Variable 1)

Modified Areal Units	Uniform Sampling Probability							Non-uniform Sampling Probability						
	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I
Monte Carlo Simulated Data	10,000							10,000						
1 Grid	6602	0.508	0.500	0.000	1.000	-0.033	0.000	6691	0.541	0.498	0.000	1.000	-0.166	0.000
4 Grids (Aggregated)	2475	0.511	0.327	0.000	1.000	-0.025	0.000	2465	0.538	0.325	0.000	1.000	-0.121	-0.000
4 Grids (+1 Shift Eastward)	2467	0.512	0.331	0.000	1.000	-0.021	0.000	2463	0.540	0.327	0.000	1.000	-0.144	0.001
4 Grids (+1 Shift Southward)	2462	0.506	0.330	0.000	1.000	-0.014	-0,000	2468	0.540	0.327	0.000	1.000	-0.149	0.000
16 Grids (Aggregated)	625	0.509	0.158	0.000	1.000	0.001	-0.000	625	0.541	0.155	0.083	1.000	-0.006	-0.000
16 Grids (+1 Shift Eastward)	625	0.511	0.155	0.091	0.917	0.083	-0.001	625	0.541	0.155	0.100	0.917	-0.128	0.001
16 Grids (+1 Shift Southward)	625	0.508	0.156	0.091	1.000	0.095	-0.000	625	0.542	0.155	0.000	1.000	-0.143	0.002
16 Grids (+2 Shifts Eastward)	625	0.508	0.157	0.000	0.923	-0.133	-0.001	625	0.540	0.156	0.083	0.909	-0.124	0.002
16 Grids (+2 Shifts Southward)	625	0.507	0.157	0.000	1.000	-0.037	-0.000	625	0.540	0.154	0.091	1.000	-0.141	0.002
16 Grids (+3 Shifts Eastward)	625	0.511	0.160	0.091	1.000	0.117	-0.001	625	0.540	0.159	0.000	1.000	-0.082	0.001
16 Grids (+3 Shifts Southward)	625	0.507	0.159	0.091	1.000	0.048	-0.001	625	0.540	0.156	0.083	0.917	-0.107	0.001
25 Grids (Aggregated)	400	0.508	0.124	0.111	0.909	0.014	0.001	400	0.540	0.121	0.154	0.875	-0.100	0.002
25 Grids (+1 Shift Eastward)	400	0.509	0.129	0.118	0.857	-0.012	0,000	400	0.541	0.123	0.167	0.867	-0.159	0.004
25 Grids (+1 Shift Southward)	400	0.508	0.124	0.133	0.882	0.035	-0.001	400	0.540	0.127	0.167	0.857	-0.328	0.002
25 Grids (+2 Shifts Eastward)	400	0.510	0.128	0.188	0.929	0.088	0.001	400	0.540	0.123	0.188	0.875	-0.211	0.006
25 Grids (+2 Shifts Southward)	400	0.508	0.125	0.167	0.857	-0.004	-0.001	400	0.540	0.122	0.167	0.857	-0.195	0.001
25 Grids (+3 Shifts Eastward)	400	0.509	0.126	0.176	0.889	0.137	-0.001	400	0.540	0.127	0.133	0.857	-0.332	0.004
25 Grids (+3 Shifts Southward)	400	0.508	0.128	0.176	0.923	0.068	-0.004	400	0.541	0.121	0.125	0.833	-0.227	0.002
25 Grids (+4 Shifts Eastward)	400	0.510	0.124	0.118	1.000	0.079	0.002	400	0.541	0.128	0.158	0.938	-0.120	0.002
25 Grids (+4 Shifts Southward)	400	0.508	0.125	0.154	0.882	0.047	-0.001	400	0.540	0.117	0.222	0.875	0.067	0.004
100 Grids (Aggregated)	100	0.508	0.058	0.377	0.692	0.141	0.015	100	0.541	0.065	0.381	0.738	0.134	-0.004
100 Grids (+1 Shift Eastward)	100	0.508	0.061	0.379	0.691	-0.103	-0,009	100	0.541	0.067	0.393	0.750	0.192	-0.003
100 Grids (+1 Shift Southward)	100	0.508	0.057	0.391	0.677	0.290	0,014	100	0.541	0.064	0.400	0.672	-0.083	0.002
100 Grids (+2 Shifts Eastward)	100	0.508	0.063	0.328	0.682	-0.236	-0.007	100	0.541	0.068	0.368	0.699	-0.232	0.001
100 Grids (+2 Shifts Southward)	100	0.509	0.060	0.373	0.682	0.217	-0.010	100	0.541	0.066	0.415	0.690	0.186	0.001
100 Grids (+3 Shifts Eastward)	100	0.508	0.063	0.328	0.682	-0.236	-0.007	100	0.541	0.068	0.368	0.699	-0.232	0.001
100 Grids (+3 Shifts Southward)	100	0.509	0.060	0.373	0.682	0.217	-0.010	100	0.541	0.066	0.415	0.690	0.186	0.001
100 Grids (+4 Shifts Eastward)	100	0.509	0.059	0.338	0.672	-0.270	-0.003	100	0.541	0.070	0.362	0.704	-0.190	-0.001
100 Grids (+4 Shifts Southward)	100	0.508	0.063	0.377	0.688	0.405	-0.014	100	0.542	0.065	0.391	0.696	0.036	-0.001
100 Grids (+5 Shifts Eastward)	100	0.509	0.061	0.323	0.672	-0.232	-0.002	100	0.541	0.069	0.343	0.692	-0.315	-0.001
100 Grids (+5 Shifts Southward)	100	0.508	0.063	0.375	0.672	0.241	-0.012	100	0.542	0.058	0.409	0.691	0.106	0.005
100 Grids (+6 Shifts Eastward)	100	0.509	0.061	0.323	0.667	-0.046	-0.002	100	0.540	0.069	0.345	0.687	-0.213	-0.001
100 Grids (+6 Shifts Southward)	100	0.508	0.062	0.373	0.644	0.076	-0.010	100	0.542	0.059	0.411	0.701	-0.011	0.011
100 Grids (+7 Shifts Eastward)	100	0.509	0.062	0.338	0.672	0.346	0.001	100	0.541	0.067	0.366	0.657	-0.344	0.009
100 Grids (+7 Shifts Southward)	100	0.508	0.062	0.375	0.641	-0.031	-0.007	100	0.542	0.059	0.386	0.676	-0.113	0.007
100 Grids (+8 Shifts Eastward)	100	0.509	0.064	0.333	0.701	0.451	-0.010	100	0.540	0.068	0.362	0.667	-0.450	0.020
100 Grids (+8 Shifts Southward)	100	0.509	0.063	0.353	0.661	-0.091	-0.003	100	0.541	0.059	0.403	0.688	-0.080	0.007
100 Grids (+9 Shifts Eastward)	100	0.509	0.059	0.358	0.703	0.499	0.005	100	0.541	0.067	0.339	0.692	-0.314	0.022
100 Grids (+9 Shifts Southward)	100	0.509	0.061	0.393	0.738	0.434	0.013	100	0.541	0.062	0.406	0.697	0.052	-0.004

Note: Moran's I statistics ranges from -1 (perfect clustering of dissimilar values) through 0 (no spatial autocorrelation) to 1 (perfect clustering of similar values).

Table A.3: Distributions of Simulated Data (Independent Variable 2)

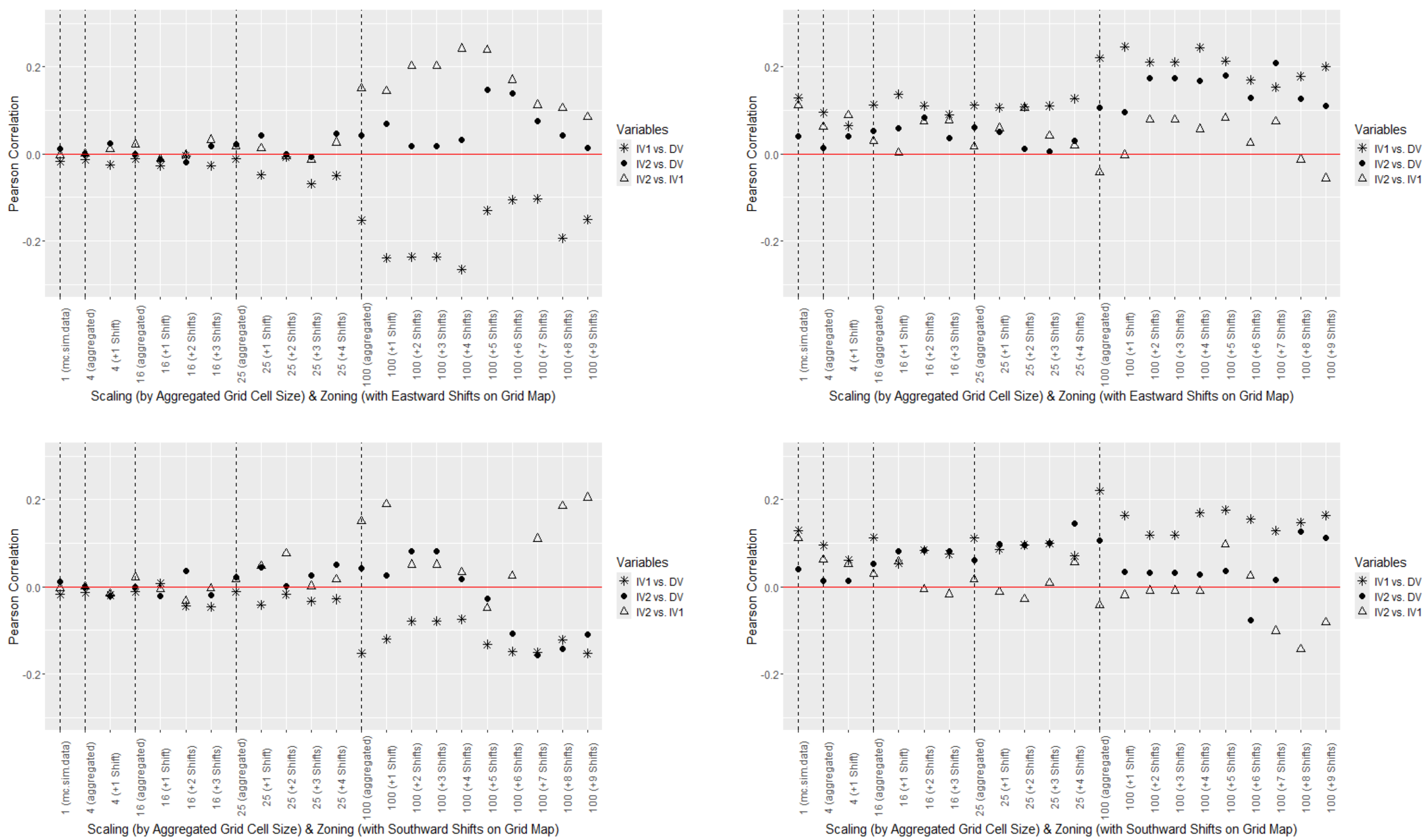
Modified Areal Units	Uniform Sampling Probability							Non-uniform Sampling Probability						
	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I	N	Mean	Std.Dev.	Min	Max	Skew	Moran's I
Monte Carlo Simulated Data	10,000							10,000						
1 Grid	6611	0.494	0.500	0.000	1.000	0.026	0.001**	6901	0.447	0.497	0.000	1.000	0.214	0.000
4 Grids (Aggregated)	2464	0.489	0.336	0.000	1.000	0.024	0.000	2476	0.441	0.317	0.000	1.000	0.175	0.001
4 Grids (+1 Shift Eastward)	2463	0.494	0.337	0.000	1.000	0.001	0.002	2472	0.449	0.322	0.000	1.000	0.151	-0.000
4 Grids (+1 Shift Southward)	2465	0.494	0.333	0.000	1.000	0.034	0.001	2477	0.445	0.324	0.000	1.000	0.174	0.000
16 Grids (Aggregated)	625	0.495	0.160	0.000	1.000	0.041	0.001	625	0.446	0.149	0.000	0.900	-0.047	0.001
16 Grids (+1 Shift Eastward)	625	0.494	0.161	0.000	1.000	0.072	0.003	625	0.447	0.153	0.000	0.889	-0.137	-0.001
16 Grids (+1 Shift Southward)	625	0.494	0.155	0.000	1.000	0.000	0.005	625	0.446	0.153	0.000	0.875	0.061	0.004
16 Grids (+2 Shifts Eastward)	625	0.493	0.159	0.071	1.000	0.103	0.002	625	0.447	0.154	0.000	0.900	-0.067	-0.001
16 Grids (+2 Shifts Southward)	625	0.494	0.158	0.100	1.000	0.226	0.001	625	0.446	0.154	0.000	1.000	0.114	0.003
16 Grids (+3 Shifts Eastward)	625	0.493	0.155	0.077	1.000	-0.027	0.005	625	0.449	0.151	0.000	0.900	-0.010	-0.001
16 Grids (+3 Shifts Southward)	625	0.495	0.160	0.077	1.000	0.111	0.005	625	0.448	0.155	0.000	0.900	0.116	-0.001
25 Grids (Aggregated)	400	0.493	0.127	0.063	0.875	-0.059	0.004	400	0.447	0.136	0.000	0.824	0.103	-0.002
25 Grids (+1 Shift Eastward)	400	0.494	0.128	0.125	0.917	0.104	0.005	400	0.447	0.128	0.063	0.824	0.102	-0.001
25 Grids (+1 Shift Southward)	400	0.494	0.127	0.118	0.923	0.129	0.005	400	0.447	0.126	0.000	0.846	0.007	0.005
25 Grids (+2 Shifts Eastward)	400	0.493	0.129	0.125	0.813	-0.006	0.004	400	0.446	0.127	0.067	0.857	0.009	-0.002
25 Grids (+2 Shifts Southward)	400	0.494	0.126	0.063	1.000	0.275	0.006	400	0.446	0.128	0.083	0.818	0.032	0.004
25 Grids (+3 Shifts Eastward)	400	0.493	0.130	0.083	0.813	-0.115	0.006	400	0.446	0.126	0.067	0.813	-0.022	-0.002
25 Grids (+3 Shifts Southward)	400	0.494	0.130	0.143	0.867	0.054	0.004	400	0.446	0.131	0.118	0.750	-0.050	-0.002
25 Grids (+4 Shifts Eastward)	400	0.494	0.130	0.118	0.875	-0.067	0.002	400	0.446	0.128	0.063	0.857	0.046	-0.003
25 Grids (+4 Shifts Southward)	400	0.492	0.123	0.176	0.833	0.037	0.009	400	0.447	0.133	0.059	0.786	-0.144	-0.002
100 Grids (Aggregated)	100	0.494	0.064	0.338	0.654	0.245	0.011	100	0.446	0.065	0.288	0.563	-0.313	-0.008
100 Grids (+1 Shift Eastward)	100	0.494	0.065	0.333	0.660	0.129	0.019	100	0.446	0.062	0.274	0.588	-0.308	-0.009
100 Grids (+1 Shift Southward)	100	0.494	0.065	0.343	0.671	0.214	0.006	100	0.447	0.061	0.292	0.582	-0.150	0.002
100 Grids (+2 Shifts Eastward)	100	0.494	0.065	0.319	0.671	0.027	0.026	100	0.446	0.063	0.267	0.638	-0.206	-0.017
100 Grids (+2 Shifts Southward)	100	0.494	0.064	0.309	0.667	-0.024	0.013	100	0.446	0.064	0.268	0.594	-0.127	0.005
100 Grids (+3 Shifts Eastward)	100	0.494	0.065	0.319	0.671	0.027	0.026	100	0.446	0.063	0.267	0.638	-0.206	-0.017
100 Grids (+3 Shifts Southward)	100	0.494	0.064	0.309	0.667	-0.024	0.013	100	0.446	0.064	0.268	0.594	-0.127	0.005
100 Grids (+4 Shifts Eastward)	100	0.494	0.067	0.286	0.662	-0.010	0.023	100	0.446	0.063	0.284	0.618	-0.085	-0.017
100 Grids (+4 Shifts Southward)	100	0.494	0.066	0.296	0.639	-0.148	0.008	100	0.446	0.061	0.260	0.569	-0.415	-0.001
100 Grids (+5 Shifts Eastward)	100	0.494	0.070	0.310	0.673	0.085	0.009	100	0.447	0.065	0.265	0.642	0.270	-0.015
100 Grids (+5 Shifts Southward)	100	0.493	0.064	0.314	0.690	0.111	0.021	100	0.447	0.063	0.229	0.592	-0.367	-0.000
100 Grids (+6 Shifts Eastward)	100	0.494	0.070	0.322	0.662	-0.100	0.001	100	0.447	0.069	0.306	0.681	0.584	-0.016
100 Grids (+6 Shifts Southward)	100	0.493	0.064	0.304	0.667	0.026	0.018	100	0.446	0.068	0.261	0.597	-0.061	-0.007
100 Grids (+7 Shifts Eastward)	100	0.494	0.068	0.354	0.682	0.166	0.001	100	0.447	0.067	0.301	0.682	0.588	-0.016
100 Grids (+7 Shifts Southward)	100	0.493	0.067	0.279	0.667	-0.018	0.001	100	0.446	0.062	0.265	0.583	-0.227	-0.004
100 Grids (+8 Shifts Eastward)	100	0.493	0.067	0.313	0.651	-0.031	-0.002	100	0.447	0.068	0.274	0.662	0.346	-0.019
100 Grids (+8 Shifts Southward)	100	0.493	0.067	0.303	0.639	-0.083	0.003	100	0.446	0.063	0.300	0.587	-0.047	-0.003
100 Grids (+9 Shifts Eastward)	100	0.494	0.066	0.328	0.644	-0.068	0.009	100	0.447	0.062	0.306	0.583	-0.135	-0.011
100 Grids (+9 Shifts Southward)	100	0.494	0.064	0.353	0.636	0.176	0.014	100	0.446	0.067	0.268	0.569	-0.392	-0.014

Note: Moran's I statistics range from -1 (perfect clustering of dissimilar values) through 0 (no spatial autocorrelation) to 1 (perfect clustering of similar values).

Figure A.1: The MAUP in Bivariate Correlation, by Sampling Probability Type

(a). Uniform Probability

(b). Non-uniform Probability



Notes: Stars represent the Pearson correlation between IV1 and the DV. Circles represent the correlation between IV2 and DV. Hollow triangles represent the correlation between IV1 and IV2. Dotted vertical lines separate different scale sizes (e.g., 1 grid cell, 4 grid cells, 16 grid cells, 100 grid cells). The x-axis tick labels show zonal changes within each set of analyses at a given scale.

In all four quadrants of Figure A.1, we can see the impacts of the MAUP. The bivariate correlations at the aggregate level are substantively affected by how we combine data despite the fact that they are based on the same underlying values of the individual locations. Starting from a correlation of approximately zero with our locations for all variables, we see the correlations change, in both positive and negative directions, as the scale and zoning change. Importantly, the effects of the MAUP do not seem to increase or decrease monotonically with scaling or zoning, although more aggregated units show more variance than smaller units. These unpredictable effects are apparent both in the uniform probability samples and in the samples based on non-uniform probabilities. Thus, depending on the set of borders we use to divide our 10,000 locations into aggregate units, we would obtain different results for the correlations between these three variables. This finding confirms existing scholarship on the MAUP that shows in a simulation setting that correlations are unstable based on changes to both scaling and zoning (Amrhein 1995; Openshaw 1984).

2 Monte Carlo Simulation with Irregular Lattice Data & Spatial Dependence

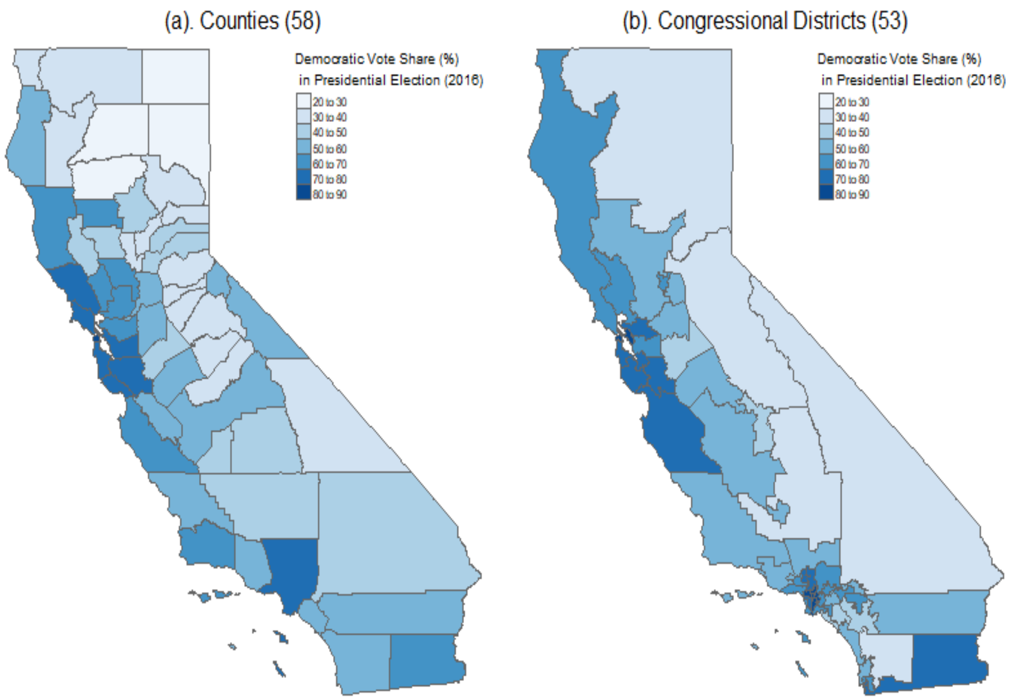
We also examine the impact of the MAUP in settings with spatially dependent real-world data across irregularly shaped units (such as administrative or electoral boundaries) (Bisbee and Zilinsky 2023; Briant et al. 2010; Fotheringham and Wong 1991; Hipp 2007; Wilson 2013). We explore the boundary data of 58 counties and 53 Congressional districts in California. We pick California because the number of counties and Congressional districts is similar, so results are less likely to be driven by differences in scale. We treat these administrative and electoral units as irregularly shaped polygons. See Figure A.2 for comparison of boundary patterns. Figure A.3 reports the 'baseline scenario' of OLS bivariate regression, which shows that changing the boundaries of the units we use (from counties to Congressional districts) can have large effects on Pearson correlations and regression results.

We also build on this 'baseline scenario' by drawing a random sample, iterating from the real value range of three socioeconomic measures – Democratic presidential vote share, median household income, and population share working from home. We chose these two independent variables because the first (median household income) is expected to be highly correlated with Democratic presidential vote share (subfigure a) but the second (population share working from home) is not (subfigure c).

We apply a Monte Carlo simulation of random sampling over 100 trials. These models account for spatial lag dependence (Figure A.4) and spatial error dependence (Figure A.5) with alternative specifications of neighboring connectivity weights: K-nearest, distance-decay, and queen contiguity.

As discussed in the main text, the results show the MAUP is present in all models, including the baseline scenario and the models accounting for spatial lag dependence and spatial error dependence. The simulation results are indicated with point estimates and standard error bars. The highlighted results in the figures are those for which values for the county are significantly different from those calculated for the Congressional district.

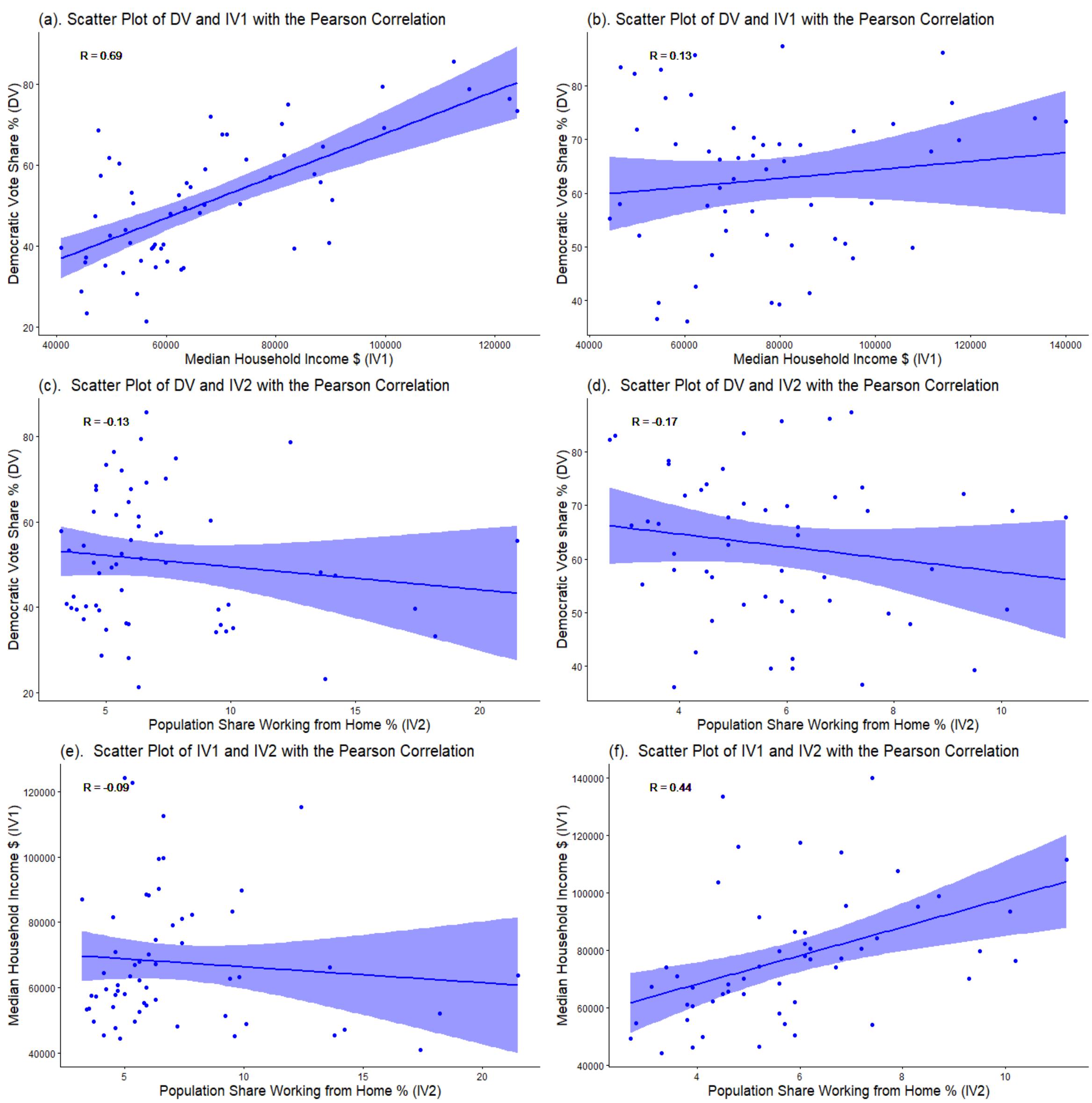
Figure A.2: Application of Irregular Lattice Data (Geographic Polygons: California)



Note: Democratic vote share for the presidential election of 2016 for California counties is available from Amlani and Algara (2021).

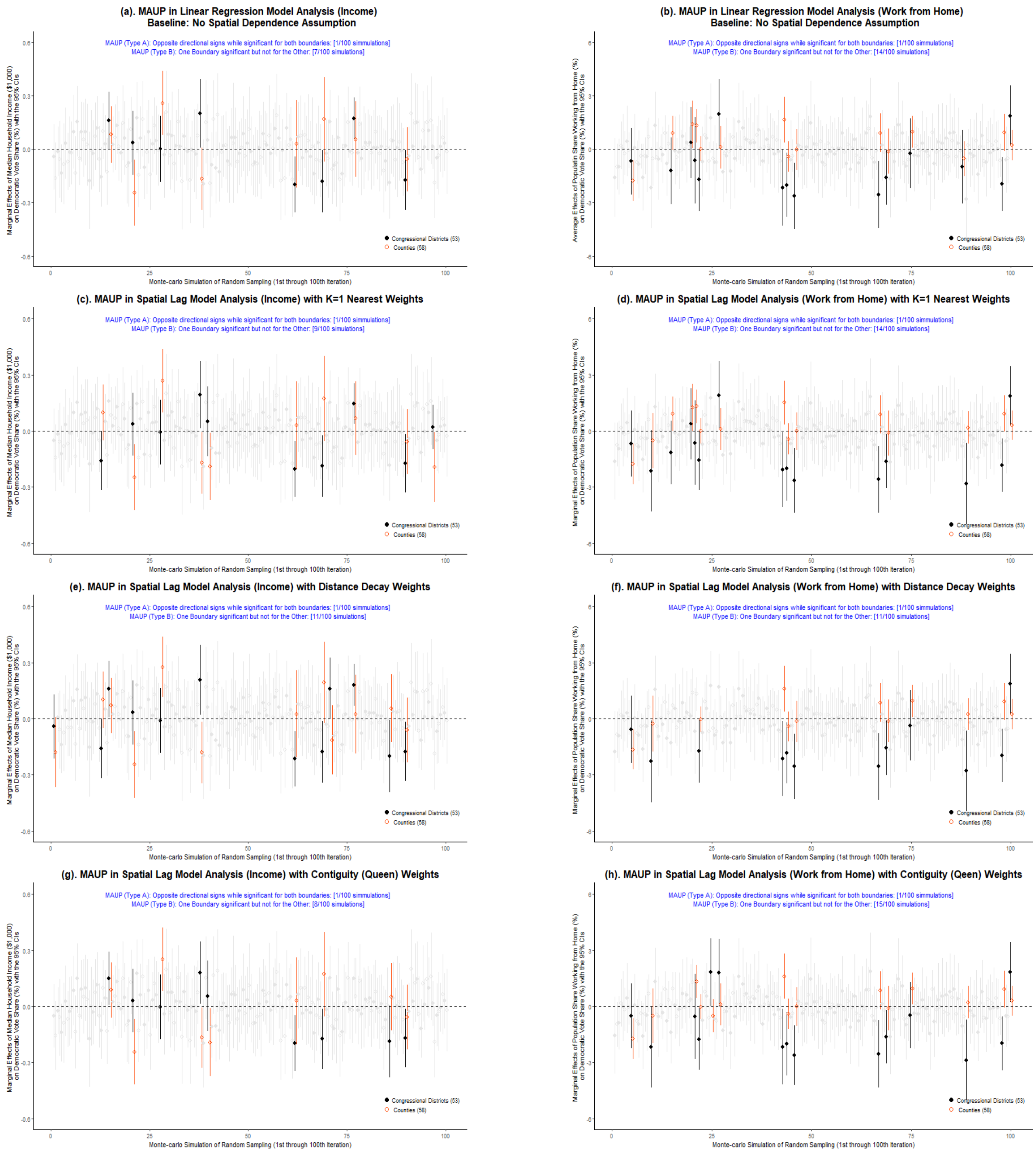
Figure A.3: Unit Selection and Changes in Pearson Correlation

Counties | Congressional Districts



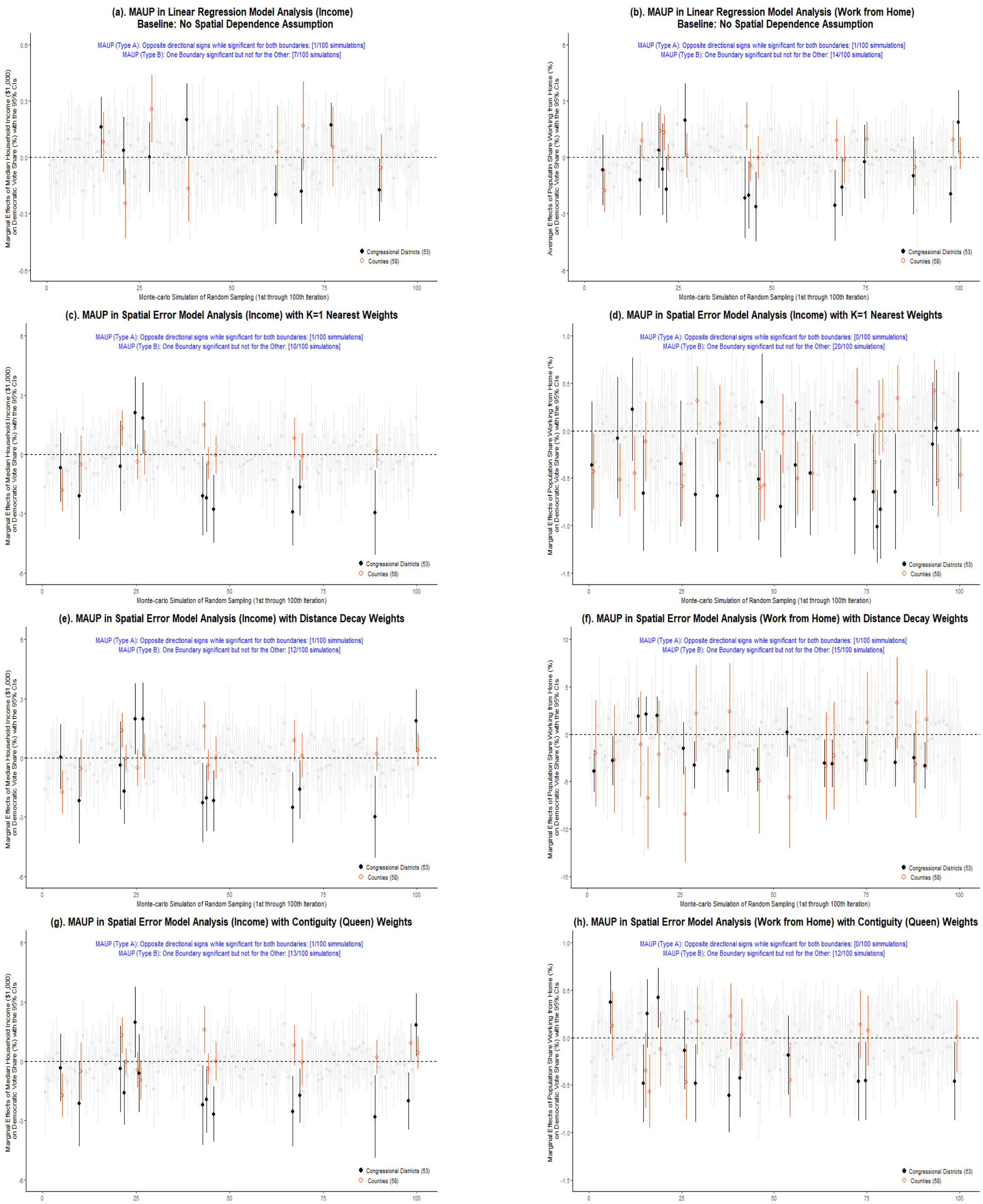
Notes: Each graph marks an OLS bivariate regression fit with the confidence intervals (purple bands). We then develop a multivariate model for which we regress Democratic presidential vote share (DV) on two socioeconomic factors: median household income (IV1) and population share working from home (IV2). The county-level and Congressional district-level data for these socioeconomic characteristics are 5-year estimates (2015-2019) available from the American Community Survey from the US Census.

Figure A.4: Monte Carlo Analysis of the MAUP in Spatial Lag Model Regression with Neighboring Connectivity Weights



Notes: Our analysis treats California's democratic vote share in the 2016 presidential election as a function of median household income (IV1) and population share working from home (IV2) at the county and Congressional district levels. All weight matrices for regression analysis are row-standardized.

Figure A.5: Monte Carlo Analysis of the MAUP in Spatial Error Model Regression with Neighboring Connectivity Weights



3 Replications: Results of Reanalysis and Discussion

Our replications focus on the consequences of scaling and zoning changes on results and inferences from analyses. As we discuss in the main text, scholars may face complications in choosing appropriate spatial units for analyses due to imprecision in theories or limits to data availability. To show the issues at stake in these choices, we re-analyze two prominent articles using different areal units and explore the implications for the inferences drawn from their findings.

3.1 Unit Scale and the Determinants of Support for Brexit

Colantone and Stanig (2018) argue that support for Brexit should be understood as a reaction by individuals to "the general economic situation of their region" (p.201), especially the extent of import competition. Regions that are "left behind areas of globalization" (p.203) see higher support for Brexit. Using the shock from Chinese imports as an index for the impact of globalization (and an instrumental variable approach in some models) the authors show that "support for Leave is systematically higher in regions that are falling behind in relative terms" (p.204).

The spatial unit used to operationalize 'region' in the original analysis is Level 3 of the Eurostat NUTS (Nomenclature of Units for Territorial Statistics) framework, which is intended to capture "small regions for specific diagnoses" (p.204).[2] This spatial unit plausibly captures the concept of economic 'region' that the authors have in mind, although they do not align in most cases with formal political or administrative unit in the United Kingdom (Schraff et al. 2023). Yet it is not definitively the only spatial scale appropriate to explore the dynamics of sociotropic economic voting at the heart of their argument. We suggest that a somewhat larger scale could also be a plausible operationalization of the economic region. Indeed, the authors themselves argue that the objective impact of

[2] The European Union's NUTS system identifies three hierarchically nested levels: NUTS-1 (large regions), NUTS-2 (basic regions for regional development policies, equivalent to US states), and NUTS-3 (smaller regions, equivalent to US counties but typically smaller). Whether these units are administrative or politically meaningful varies by country.

import competition is not limited to specific industries but to "entire communities" (p.203), and the boundaries of these communities may plausibly be larger than the NUTS-3 regions. The broader scholarship on the sociotropic determinants of support for Brexit has explored these relationships at the larger NUTS-2 level (Arnorsson and Zoega 2018; Los et al. 2017) and both the NUTS-3 and NUTS-2 levels (Huggins 2018; Brautzsch and Holtemöller 2021).[3]

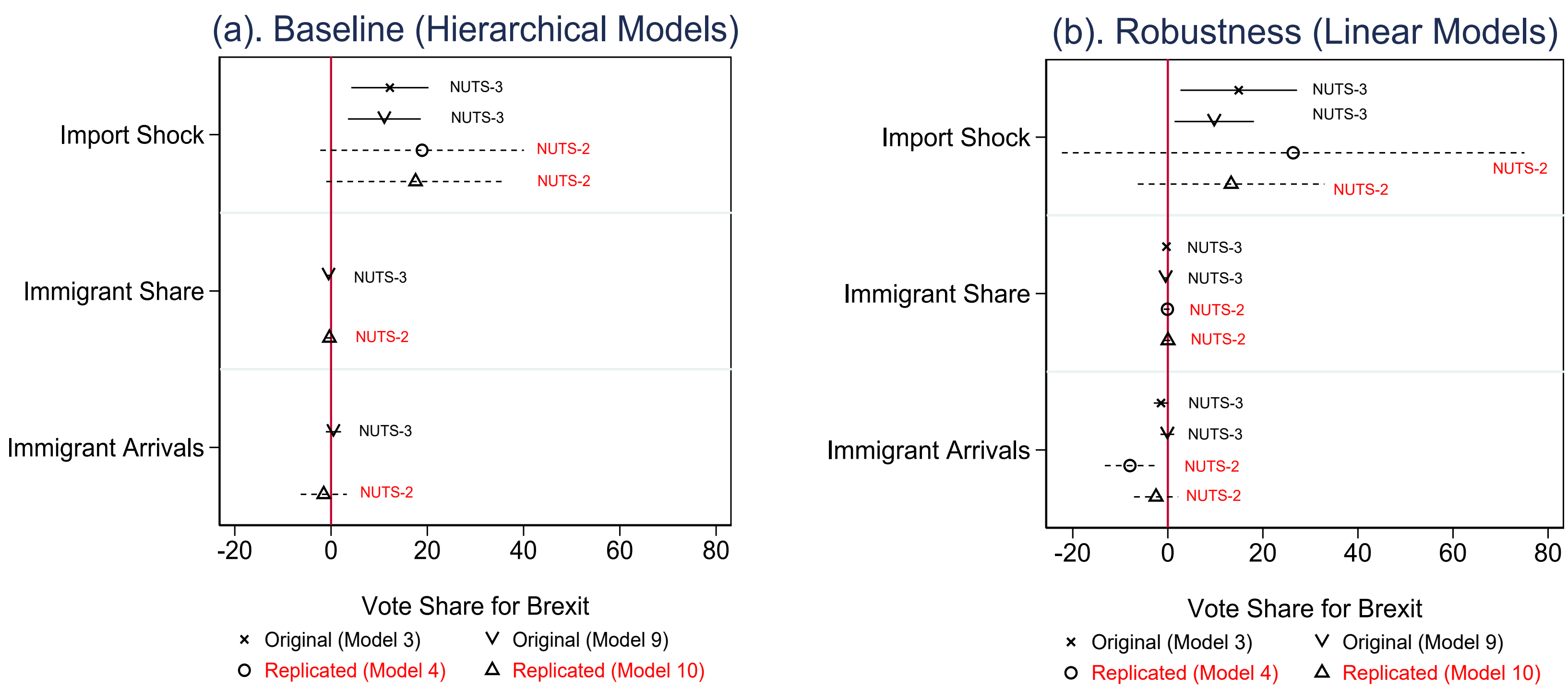
We therefore suggest that the NUTS-2 level also provides a plausibly valid spatial mapping of the 'region' affected by an import shock, in terms of both objective economic effects and voters' perceptions. The NUTS 2 borders align with political constituencies in the UK (Schraff et al. 2023). In our reanalysis, we show that the results generated by analyzing support for Brexit at this larger scale would lead us to different inferences about its determinants. Since both of these spatial operationalizations of the region are, we suggest, plausibly valid, inconsistency across them should call into question what we actually know about the determinants of support for Brexit.

The central finding in Colantone and Stanig (2018) is that an import shock in a region is associated with increased support for the 2016 Brexit referendum. This finding is robust to the inclusion of a wide range of regional control variables (p.211). Figure A.6 shows the core results of our reanalysis, using their data but aggregating to the NUTS-2 level. We find that the baseline results shown in Figure A.6(a) do not substantively change when we move to a larger scale, though some have weaker statistical significance. However, our re-analyses of the robustness checks that account for a wide range of regional characteristics display some important differences; these are shown in Figure A.6(b). We find that the import shock variable loses significance in two of the six models from the original paper.

We also explore results for the effect of immigration on support for Brexit. Here, we find even stronger evidence for Colantone and Stanig's argument for the primacy of import competition over

[3]While we explore the possibility that a larger administrative boundary might provide a valid spatial mapping, there is no reason to believe that the spatial units on which people base their subjective evaluations must be bounded by administrative borders. Many additional spatial mappings might therefore be plausibly valid.

Figure A.6: Region-level Reanalysis Results (Colantone and Stanig 2018)



Notes: Replicated results for core variables of Colantone and Stanig (2018-210-11) with 95% CIs. For subfigure (a), additional information is available from OA Table A.4. For subfigure (b), refer to OA Table A.5.

Table A.4: Region-level Results (Colantone and Stanig 2018)

VARIABLES	Model 1 CS†	Model 2 MAUP	Model 3 CS†	Model 4 MAUP	Model 5 CS†	Model 6 MAUP	Model 7 CS†	Model 8 MAUP	Model 9 CS†	Model 10 MAUP	Model 11 CS†	Model 12 MAUP
Import Shock	**12.233**	**38.739**	**12.225**	**18.919**	**12.965**	**40.838**	**12.085**	**28.902**	**11.073**	**17.564**	**12.299**	**33.074**
	(4.763)	(18.546)	(4.091)	(10.793)	(4.543)	(18.558)	(3.890)	(8.023)	(3.861)	(9.472)	(3.726)	(10.276)
Immigrant Share							**-0.490**	-0.077	**-0.513**	-0.338	**-0.491**	-0.114
							(0.165)	(0.456)	(0.155)	(0.400)	(0.154)	(0.400)
Immigrant Arrivals							-0.066	-2.427	0.496	-1.519	-0.058	-2.128
							(0.741)	(2.677)	(0.801)	(2.445)	(0.691)	(2.362)
Model	Linear	Linear	Hierarchical	Hierarchical	IV	IV	Linear	Linear	Hierarchical	Hierarchical	IV	IV
NUTS-1 Fixed Effects	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
NUTS-2 Random Intercepts	No	No	Yes	No	No	No	No	No	Yes	No	No	No
NUTS-1 Random Intercepts	No	No	No	Yes	No	No	No	No	No	Yes	No	No
Clustered Std. Err NUTS Level	Yes NUTS-2	Yes NUTS-1	No	No	Yes NUTS-2	Yes NUTS-1	Yes NUTS-2	Yes NUTS-1	No	No	Yes NUTS-2	Yes NUTS-1
Observations	167	39	167	39	167	39	167	39	167	39	167	39
R-squared	0.573	0.210			0.573	0.209	0.646	0.449			0.646	0.447
Kleibergen-Paaap F Statistic					662.7	1669					614	2134
Number of groups			39	11					39	11		
Dependent Variable:	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)	NUTS-3 Leave Share (%)	NUTS-2 Leave Share (%)

Notes:
† Colantone and Stanig (2018), Table 1. See page 210.
Emboldened digits are statistically significant coefficient estimates.
Constant estimates are not reported here.

Table A.5: Region-level Robustness (Colantone and Stanig 2018)

VARIABLES	Model 1 CS†	Model 2 MAUP §	Model 3 CS†	Model 4 MAUP §	Model 5 CS†	Model 6 MAUP §	Model 7 CS†	Model 8 MAUP §	Model 9 CS†	Model 10 MAUP §	Model 11 CS†	Model 12 MAUP §
Import Shock	**9.391**	**25.534**	**14.920**	26.372	**9.460**	**23.460**	**10.592**	**33.210**	**9.765**	13.322	**7.997**	**22.431**
	(3.858)	(9.616)	(6.061)	(21.839)	(4.084)	(10.752)	(4.075)	(14.478)	(4.125)	(8.814)	(4.011)	(7.089)
Immigrant Share	**-0.328**	-0.048	**-0.282**	-0.080	**-0.592**	0.194	**-0.617**	0.153	**-0.462**	0.039	**-0.529**	-0.453
	(0.130)	(0.377)	(0.123)	(0.350)	(0.178)	(0.375)	(0.183)	(0.421)	(0.163)	(0.336)	(0.147)	(0.437)
Immigrant Arrivals	-1.141	**-6.421**	**-1.434**	**-7.963**	-0.083	**-5.082**	0.025	-4.692	-0.102	-2.471	0.309	0.273
	(0.822)	(2.529)	(0.751)	(2.373)	(0.777)	(2.342)	(0.809)	(2.640)	(0.713)	(2.086)	(0.652)	(2.563)
EU Accession Immigrants (2001)	**-12.045**	8.388	-10.301	**37.572**								
	(5.824)	(8.944)	(8.104)	(15.608)								
EU Accession Immigrants Growth (2001-2011)	**1.527**	**4.023**	**2.431**	2.872								
	(0.549)	(1.277)	(1.286)	(3.187)								
EU Accession Immigrants * Import Shocks			-15.685	**-130.073**								
			(34.567)	(58.350)								
EU Accession Immigrants Growth * Import Shock			-1.831	11.860								
			(3.745)	(8.047)								
Fiscal Cuts					**0.022**	0.014	0.014	-0.011				
					(0.006)	(0.010)	(0.013)	(0.031)				
Cancer Treated in 62_days					-0.591	**-7.036**	-0.503	**-7.324**				
					(0.596)	(1.442)	(0.616)	(1.513)				
Public Employment Growth					0.813	-2.590	**0.910**	-2.681				
					(0.519)	(1.782)	(0.536)	(1.711)				
Fiscal Cuts * Import Shock							0.028	0.089				
							(0.031)	(0.084)				
EU Economic Dependence									**0.683**	**1.195**		
									(0.384)	(0.433)		
Change in Relative Income vs. Median Region											**-0.225**	**-0.422**
											(0.059)	(0.122)
Model	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
NUTS-1 Fixed Effects	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Clustered Std. Err (NUTS)	NUTS-2	NUTS-1	NUTS-2	NUTS-1	NUTS-2	NUTS-1	NUTS-2	NUTS-1	NUTS-2	NUTS-1	NUTS-2	NUTS-1
Observations	167	39	167	39	167	39	167	39	167	39	167	39
R-squared	0.677	0.521	0.680	0.549	0.698	0.678	0.700	0.683	0.659	0.590	0.692	0.559
Dependent Variables: Regional Vote Share (%)	NUTS-3	NUTS-2	NUTS-3	NUTS-2	NUTS-3	NUTS-2	NUTS-3	NUTS-2	NUTS-3	NUTS-2	NUTS-3	NUTS-2

Notes:
† Colantone and Stanig (2018), Table 2.
Emboldened digits are statistical significant coefficient estimates.
Constant estimates are not reported here.

immigration as a predictor for anti-Brexit public sentiments than what they produced at the NUTS-3 level, shown in Figure A.6(b). Whereas Colantone and Stanig found that increased immigrant share has a negative and significant (yet minuscule) impact on support for Brexit – NUTS-3 areas with more foreign-born residents as a share of the local population in 2015 were less supportive of Brexit - analysis at the NUTS-2 level finds that immigrant share has no effect on support for 'Leave' once the import shock is taken into consideration. Moreover, against the literature that emphasizes the role of immigration in driving support for Brexit, our results at the NUTS-2 level for immigrant arrivals (the inflow of immigrant workers as a share of the total working-age population of the region in 2015) actually show that immigrant arrivals have a negative effect on support for Brexit. This relationship was found to be insignificant in the original, NUTS-3 level analysis.

Our claim is not that we have uncovered the true relationships between import competition, immigration, and support for Brexit. Nor do we suggest that Colantone and Stanig's findings are incorrect. Instead, we have shown that changing the scale at which economic and social context is operationalized alters some of the findings - in particular, it is unclear whether the import shock is a consistent predictor of Brexit support, or whether immigrant arrivals drove support for leaving the EU. Unless we have a clear justification for why a particular scale is more valid than another for operationalizing the spatial concept at the heart of the theory, the conclusions we should reach about the causes of support for Brexit are less clear.

A skeptical reader might suggest that the changes we observe in our reanalysis are simply the result of the aggregation effect: since we have re-analyzed at a larger spatial scale, estimates are less precise simply by virtue of the fact that we have a smaller number of observations, and this makes us more likely to find null relationships. Yet in some cases, we actually found stronger statistical relationships when we analyzed at a larger scale: this was the case for the finding that immigrant arrivals in the region are negatively associated with support for Brexit.

In the analyses shown in Figure A.6, both dependent and independent variables were unit aggregates. Examining the individual level analysis in Colantone and Stanig (2018) allows us to see how spatially aggregated independent variables (*Import Shock*, *Immigrant Share*, and *Immigrant Arrivals*) impact individual-level Brexit support. The upper left quadrant of Table A.6 shows how the unit af-

Table A.6: Individual-level Results (Colantone and Stanig 2018)

	Binary Dependent Variable: 1 = The respondent declares the intention to vote for the Leave option. 0 = Otherwise.						Categorical Dependent Variable The respondent's perception: Immigration change (l= Getting a lot lower, 5=Getting a lot higher).		
Independent Variables	Model 1 CS†	Model 2 MAUP	Model 3 MAUP	Model 4 CS‡	Model 5 MAUP	Model 6 MAUP	Model 7 CS§	Model 8 MAUP	Model 9 MAUP
Regional Level Indicators	*NUTS-3*	*NUTS-2*	*NUTS-1*	*NUTS-3*	*NUTS-2*	*NUTS-1*	*NUTS-3*	*NUTS-2*	*NUTS-1*
Import Shock	**0.246**	0.338	**1.062**	**0.322**	0.310	**1.012**	**0.125**	0.168	0.463
	(0.104)	(0.210)	(0.527)	(0.119)	(0.217)	(0.574)	(0.064)	(0.121)	(0.284)
Immigrant Share	-0.006	-0.010	0.064	-0.006	-0.010	0.065	**0.008**	**0.010**	**0.044**
	(0.005)	(0.011)	(0.058)	(0.005)	(0.010)	(0.059)	(0.003)	(0.005)	(0.020)
Immigrant Arrivals	0.011	-0.027	-0.423	0.012	-0.024	-0.427	**-0.055**	**-0.067**	**-0.292**
	(0.024)	(0.051)	(0.399)	(0.024)	(0.051)	(0.400)	(0.014)	(0.030)	(0.138)
Individual Level Indicators									
Age (15-93)	**0.014**	**0.014**	**0.014**	**0.015**	**0.015**	**0.016**	**0.012**	**0.013**	**0.013**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)
Gender (1=Male)	**-0.050**	**-0.050**	**-0.048**	**-0.048**	-0.048	**-0.046**	**0.055**	**0.055**	**0.056**
	(0.028)	(0.030)	(0.026)	(0.028)	(0.030)	(0.027)	(0.012)	(0.012)	(0.012)
1. Education (GCSE D-G)	-0.097	-0.095	-0.087	-0.107	-0.104	-0.097	**-0.055**	**-0.054**	**-0.054**
	(0.085)	(0.081)	(0.085)	(0.085)	(0.081)	(0.088)	(0.033)	(0.033)	(0.033)
2. Education (GCSE A*-C)	**-0.186**	**-0.187**	**-0.182**	**-0.197**	**-0.197**	**-0.192**	**-0.069**	**-0.069**	**-0.070**
	(0.059)	(0.060)	(0.068)	(0.059)	(0.059)	(0.071)	(0.024)	(0.024)	(0.024)
3. Education (A-level)	**-0.449**	**-0.451**	**-0.444**	**-0.455**	**-0.455**	**-0.448**	**-0.284**	**-0.284**	**-0.285**
	(0.059)	(0.060)	(0.063)	(0.059)	(0.059)	(0.064)	(0.025)	(0.025)	(0.025)
4. Education (Undergraduate)	**-0.729**	**-0.730**	**-0.728**	**-0.738**	**-0.737**	**-0.735**	**-0.473**	**-0.475**	**-0.479**
	(0.059)	(0.070)	(0.064)	(0.059)	(0.069)	(0.066)	(0.023)	(0.023)	(0.023)
5. Education (Postgraduate)	**-1.072**	**-1.073**	**-1.080**	**-1.082**	**-1.084**	**-1.091**	**-0.648**	**-0.655**	**-0.663**
	(0.066)	(0.065)	(0.070)	(0.065)	(0.064)	(0.070)	(0.028)	(0.028)	(0.028)
Retired				0.028	-0.147	-0.175			
				(0.078)	(0.092)	(0.156)			
Interaction									
Retired * Import Shock				**-0.407**	0.122	0.211			
				(0.200)	(0.253)	(0.429)			
Model	Probit	Probit	Probit	Probit	Probit	Probit	Hierarchical	Hierarchical	Hierarchical
NUTS-1 Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Random Intercepts	No	No	No	No	No	No	NUTS-3	NUTS-2	NUTS-1
Std. Err. Clustered by NUTS	NUTS-3	NUTS-2	NUTS-1	NUTS-3	NUTS-2	NUTS-1	No	No	No
Number of Groups							167	39	11
Observations	16,331	16,331	16,331	16,331	16,331	16,331	20,623	20,623	20,623

Notes:
Standard errors in parentheses
Emboldened digits are statistically significant coefficient estimates.
GSCE stands for General Certificate of Secondary Education.
† Colantone and Stanig (2018). Duplicates of Table 3, Model 4 See page 213 for details.
‡ Colantone and Stanig (2018). Duplicates of Table 4, Model 1. See page 215 for details.
§ Colantone and Stanig (2018). Duplicates of Table 5, Model 3. See page 216 for details.
Constant is not repolled here.

fects the statistical significance of the result: the import shock variable is significant in all original models at the NUTS-3 level, but insignificant across all models for reanalysis at the NUTS-2 level. Yet as we move up to an even larger scale and operationalize this variable at the NUTS-1 level, it becomes significant once again. This is a further demonstration that the effects of unit scale on regression results are not monotonic or predictable. Moreover, the fact that variables can become significant at larger scales is further evidence that our results do not simply reflect the aggregation effect. Another instance of a finding that changes in surprising ways is also highlighted in the table: the interaction term *Retired*Import Shock*, which measures the impact of the import shock on working-aged people, while significant in the expected (negative) direction at the NUTS-3 level, is positive although not significant at the NUTS-2 and NUTS-1 levels.

As we discussed above, the authors' theory can accommodate several spatial units as plausibly valid scales at which to measure the impact of the import shock. While the NUTS-3 unit might be a reasonable "approximate" unit to capture individuals' sense of economic community, so might an even smaller level, if such data were available. So might the NUTS-2 or NUTS-1 level if people view their economy in broader terms. Because we cannot say whether the NUTS-3 results provide a more accurate account of the theory, the fact that findings vary across scales casts doubt on the inferences we can draw about the causes of support for Brexit.

3.2 Unit Scale and the Moderating Effects of Neighborhood Context on Social Rewards for Participation

The MAUP can also alter regression results via its effects on control variables even when the IV and DV are individual-level variables. We explore this effect in our examination of Anoll (2018), showing what happens when we operationalize a moderating variable at a different spatial scale. Anoll theorizes that the social rewards of participation are moderated by the composition of the local community. More specifically, she explores how the proportion of co-racials increases the social rewards to Black participation, and how the foreign-born (either naturalized or noncitizens) participation of local communities has the same impact on Latino participation. She uses the 5-digit zip code as a measure of the survey respondents' residential community. We replicate Anoll's regression results

(Table 3, p.504) using the somewhat larger scale 3-digit zip code area as an alternative operationalization of the local community. While many of our conclusions are similar, we find that some of the results are not consistent across these different spatial scales.

Laudably, Anoll provides an explicit and plausible justification for why she chose the zip code for mapping the boundaries of the local community. She bases it on work by Velez and Wong (2017) from which she concludes that "across an array of measures, zip code most closely approximates people's perceptions of the racial composition of their local community" (Anoll 2018, p.503, fn. 13). It is true that in terms of capturing perceptions of racial composition, Velez and Wong show that the zip code outperforms both user-defined boundary measures like that developed by Wong et al. (2012) and a measure based on people's reports of locations they frequently visited and that the zip code (or ZCTA; the census-defined zip code tabulation area) also outperforms the county. But our suggestion is that while this logic may justify the choice of the zip code over the county (which is too large), it does not provide a dispositive reason to prefer the standard zip code over other all alternative specifications of local context.

We submit that it is therefore reasonable to operationalize local context at the 3-digit zip code level, which is larger than the 5-digit zip code but far smaller than the county in most cases. To do this, we use a list of 3-digit US postal zip code prefixes that identify the names of the sectional center facilities (SCF). This is a plausible alternative way to operationalize community at a larger scale since each SCF serves local addresses whose five-digit codes begin with the same three digits. Our local contexts are larger than Anoll's, but not so implausibly large as to make them inappropriate for the concept of 'community.'

Table A.7 reports our results of replicating Anoll's analyses at the 5-digit zip code scale, and reanalyzing them at the 3-digit zip code area scale. Many of the results are similar, but there are several differences of note, which are highlighted in bold. The first can be seen by comparing the results of Models 3 and 4. Here, while Anoll finds (in Model 3) that the social rewards to Black Americans for participation in a rally are substantively and significantly affected by the proportion of co-racials in a neighborhood, we do not find the same effect using a spatially larger definition of neighborhood. While the size of the coefficient does not change substantially when we move from the 5-digit to the

Table A.7: Neighborhood Context Effects on Social Rewards for Voting and Political Rally Attendance (Anroll 2018)

	Samples: Blacks & Whites				Samples: Latinos & Whites				Samples: Latinos & Whites			
	(1) Voters		(2) Rally Attenders		(3) Voters		(4) Rally Attenders		(5) Voters		(6) Rally Attenders	
	5-Zip	3-Zip	5-Zip	3-Zip	5-Zip	3-Zip	5-Zip	3-Zip	5-Zip	3-Zip	5-Zip	3-Zip
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Intercept	**0.763**	**0.757**	**0.697**	**0.705**	**0.767**	**0.786**	**0.739**	**0.751**	**0.731**	**0.746**	**0.713**	**0.715**
	(0.014)	(0.016)	(0.015)	(0.017)	(0.016)	(0.017)	(0.017)	(0.019)	(0.016)	(0.018)	(0.018)	(0.020)
White	0.000	0.004	-0.015	-0.024	-0.020	**-0.033**	**-0.044**	**-0.045**	0.025	0.012	-0.013	-0.006
	(0.012)	(0.014)	(0.013)	(0.015)	(0.015)	(0.017)	(0.016)	(0.018)	(0.016)	(0.018)	(0.017)	(0.019)
Income	-0.000	-0.000	**-0.004**	**-0.005**	0.000	-0.000	**-0.004**	**-0.004**	0.000	0.000	**-0.004**	**-0.004**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Education	**-0.010**	**-0.009**	-0.001	-0.001	**-0.008**	**-0.007**	-0.003	-0.003	**-0.008**	**-0.007**	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Male	**-0.038**	**-0.040**	**-0.014**	**-0.016**	**-0.043**	**-0.044**	**-0.020**	**-0.021**	**-0.044**	**-0.045**	**-0.021**	**-0.022**
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Prop.Black.Zip	**0.062**	**0.091**	**0.058**	0.052								
	(0.023)	(0.037)	(0.025)	(0.039)								
Prop.Black.Zip*White	**-0.075**	**-0.087**	0.033	0.041								
	(0.045)	(0.050)	(0.048)	(0.054)								
Prop.Latino.Zip					-0.010	-0.058	-0.001	-0.039				
					(0.031)	(0.043)	(0.033)	(0.046)				
Prop.Latino.Zip*White					**0.083**	0.078	-0.000	-0.030				
					(0.045)	(0.052)	(0.048)	(0.055)				
Prop.Foreign-Born Zip									**0.144**	0.109	**0.116**	0.120
									(0.053)	(0.070)	(0.057)	(0.075)
Prop.Foreign-Born*White									**-0.163**	-0.133	**-0.175**	**-0.239**
									(0.071)	(0.086)	(0.076)	(0.091)
Numbers of Obs.	2,463	2,311	2,463	2,311	1,981	1,838	1,981	1,838	1,981	1,838	1,981	1,838
RMSE	0.178	0.179	0.189	0.191	0.191	0.194	0.204	0.207	0.191	0.194	0.204	0.207
R^2	0.029	0.029	0.022	0.022	0.024	0.022	0.025	0.029	0.025	0.025	0.028	0.031
Mean of DV	0.716	0.716	0.660	0.660	0.712	0.712	0.661	0.661	0.712	0.712	0.661	0.661
SD of DV	0.177	0.177	0.186	0.186	0.175	0.175	0.185	0.185	0.175	0.175	0.185	0.185

Notes: Emboldened digits are statistically significant weighted OLS estimates. The social reward is calculated by averaging likability and respectability scores and standardizing from 0-1 with 1 being “extremely likable and respectable.” The excluded category is minority respondents-either Black or Latino. 5-zip is the unit used in the original study (Anroll 2018, 503).

3-digit zip code, a move to the larger set of boundaries around 'neighborhood' increases the standard error, and makes the result statistically insignificant.

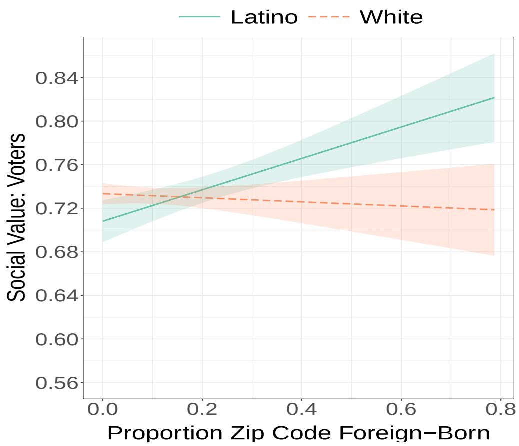
We find additional differences from Anoll's results when we explore the impact of community composition on the rewards to participation for Latinos. In Models 9 and 11, Anoll finds at the 5-digit zip code level that the proportion of foreign-born citizens has a large, positive, and significant effect on the social rewards to both voting and rally participation. By contrast, our analysis at the 3-digit zip code level (Models 10 and 12) finds that these effects are not statistically significant. This is an instance of how a contextual variable subject can affect our inferences about other predictors.

We also highlight an instance in which the MAUP impacts the moderating effects created by interaction variables included in the model – how a respondent's race interacts with her neighborhood characteristics to affect social incentives to participate in political action. We replicate the interaction terms in Anoll (2018, p.505) in Figure A.7, using the 3-digit zip code areas (compared to the 5-digit zip code in the original analysis). The figures, a subset of the original, show the interaction of White or Latino race/ethnicity with the percentage of racial groups in the individual's zip code, on that individual's likelihood to vote (top two figures) or attend a political rally (bottom two figures).

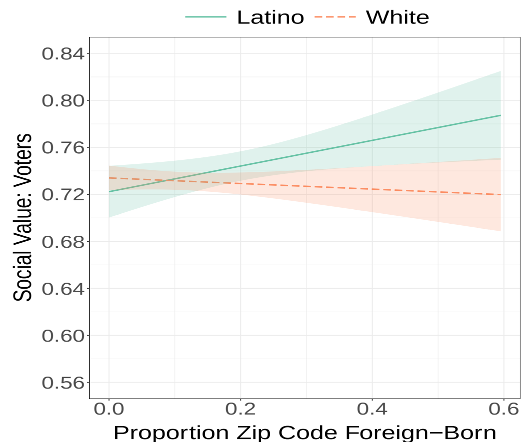
We see a contrast between Figures A.7(a) with A.7(b), in which the effect of context disappears at a larger scale. Figure A.7(a) captures relative predicted probabilities of perceiving social rewards to participation (with 80% confidence intervals, as in the original paper), for Latinos and Whites. Here, Anoll suggests that as the proportion of foreign-born citizens in the zip code area increases, White respondents' willingness to participate in voting remains nearly constant whereas the attitudes toward voting of Latinos become more positive; this is said to be a result of their embeddedness in a community "filled with group members joined by their experiences with immigration" (Anoll 2018, p.503). Evidence for the finding is the diverging slopes between Latinos and Whites in Figure A.7(a). However, our reanalysis at the 3-digit zip code level shown in Figure A.7(b) finds these differences between Latinos and Whites to be insignificant. Comparing Figures A.7(c) with A.7(d) shows another contrast. In these figures, contextual differences in social rewards for voting participation and political rally attendance for members of minority groups are more pronounced when the community is operationalized on a larger scale.

Figure A.7: Moderating Effect of Neighborhood Context on Social Rewards (Anroll 2018)

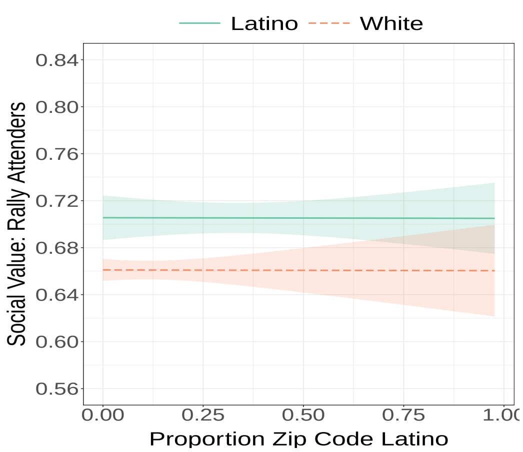
(a). Context: Foreign-Born (5-Zip)



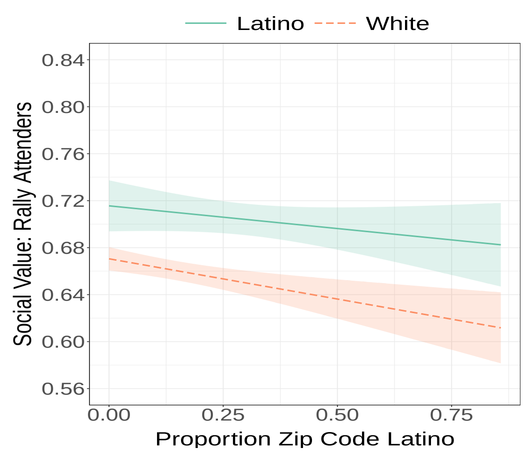
(b). Context: Foreign-Born (3-Zip)



(c). Context: Latino (5-Zip)



(d). Context: Latino (3-Zip)



4 Exploring the MAUP in "Flagship" Journals

In the main text of the paper, we argued that the most compelling way scholars can address the threats to inference posed by the MAUP is to justify the spatial unit they choose as uniquely valid, and thus rule out any findings obtained by aggregating with a different spatial mapping as irrelevant. We identified three grounds on which validity can be justified: (1) the logic of the theory in terms of the spatial mapping at which treatment is assigned (theoretical logic), (2) empirical evidence from the research setting about that spatial mapping (empirical evidence), and (3) empirical or theoretical evidence about the spatial mapping associated with a particular theorized causal mechanism (causal mechanism).

In this section of the Appendix, we discuss the *APSR* and *AJPS* articles we coded to show that in cases where scholars use these grounds to justify their chosen spatial mapping, these justifications provide a convincing response to the MAUP. In some cases, these justifications were explicit; we strongly recommend this as a best practice that complements discussions of other research design choices. In many papers, the reader must identify on their own the information provided in the paper that can help them to assess whether we were concerned about the MAUP as a threat to inference. We do our best in the examples that follow to assemble these justifications on the authors' behalf. Conversely, where scholars provide no justification that the spatial units they choose are uniquely valid, we are able to identify additional spatial mappings beyond those used that are plausibly valid. Since the effects of the MAUP on findings are potentially sizable and significant, these studies face the possibility that analyses conducted at these plausibly valid alternative mappings would lead to different inferences about the relationships under study.

After discussing a range of examples in the text, we review of all of the *APSR* papers in which the MAUP potentially poses a threat to inference in Table A.8. We then elaborate on the process by which we code these papers and identify which ones we see as potentially vulnerable to concerns about the MAUP.

4.1 Theoretical Justifications

4.1.1 Theoretical Logic

- Healy et al. (2017) explore the relative importance of pocketbook and sociotropic conditions on voting. In principle, one could assess the aggregate economic outcomes relevant to sociotropic evaluations using any number of spatial mappings, but the authors make clear (p.775) that the literature on which the paper builds has reached a consensus that the national level is the appropriate aggregate unit. Thus, although the paper is not explicit in justifying the use of the country as the aggregate spatial unit in sociotropic voting, it is the uniquely valid unit for analysis based on the authors' theory.

- Blom-Hansen et al. (2016) explore the relationship between jurisdiction size and the costs of service provision. While variation in these costs of public services could in principle be studied at a wide range of spatial scales, a study of the effect of jurisdiction size on that outcome can only be studied via comparison across the jurisdictions that provide those services. In this case, the authors compare via a difference-in-differences analysis the evolution of service costs of new jurisdictions created via amalgamation, which captures the effects of increased size, with those left untouched by that institutional change, where any change in costs over time is not a function of size. Their unit is the jurisdiction of service provision.

- Gulzar and Pasquale (2017) explore the determinants of the performance of local bureaucrats, focusing on the implementation of a guaranteed rural employment program in India. This outcome could, of course, be explored at many spatial mappings, ranging from the individual bureaucrat upwards. Interested in the impact of political oversight, the authors ask whether performance is affected by whether a bureaucrat is overseen by a single politician or by multiple politicians. Variation in this is a function of whether the bureaucratic area (a "block") is contained within the constituency of a single politician, or split across multiple political jurisdictions. The question of the effects of this variation itself, then, specifies a spatial mapping that is appropriate - the authors compare across bureaucratic "blocks," coding whether a bureaucratic “block" splits jurisdictions or does not.

- Lindgren et al. (2019) ask whether an education initiative in Sweden affects voter turnout. Voter turnout could be studied at many levels of aggregation, but the treatment in this case - a secondary education reform rolled out in the early 1990s - is assigned at the municipal level, so it is the only plausibly valid spatial mapping to be used to assess its effect. The logic here is an instance of a point seen most clearly in experimental research: when studying the impact of a treatment, analysis is most appropriately conducted at the spatial mapping at which the treatment is assigned.

4.1.2 Empirical Evidence

- Larreguy et al. (2016) examine how the number of polling stations within an electoral precinct affects turnout, and therefore the effectiveness of party machines, in Mexico. In principle, party machines' effectiveness could be a function of concentration of polling stations within any of a number of larger spatial aggregates - it depends on how the party machine is constructed. They claim that party brokers are “typically designated to the electoral precinct, and possess detailed knowledge of the vote intentions of the local population” (p.163) and provide detailed evidence (pp.163-4, including Figure 1 and citations to other studies) about how these brokers demonstrate their effectiveness to the party at the precinct level. This evidence about how broker-party relations operate provides justification for the precinct as the uniquely appropriate spatial mapping to assess variation in the effectiveness of party machines in the Mexican case.

- Touchton et al. (2017) explore the impact of participatory institutions on infant mortality in Brazil and argue that the effects of these institutions work through the administrative level at which services are delivered. In Brazil, this level is the municipality, and this justifies the municipal-level comparison of infant mortality outcomes (which can be operationalized at many spatial scales) as the only plausibly valid operationalization in this case.

- Similarly, the theory that Tajima et al. (2018) develop about diversity and public good provision requires the identification of two spatial mappings for operationalization: what they call the "user level" at which people experience public goods, and the "decision level" at which policy is made. In their setting of Indonesia, these are the village and the district, respectively, and therefore these are the two uniquely valid spatial mappings at which analysis should take place. The authors provide detailed justification of this component of their research design in the paper (645) and its online appendices.

4.1.3 Causal Mechanism

- Rundlett and Svolik (2016) develop a theory to account for the supra-optimal level of electoral fraud in authoritarian elections and apply it to the 2012 Russian presidential election. In looking for evidence of such fraud, one could examine results at various spatial mappings at which votes could plausibly be aggregated and manipulated - from the precinct to the region. The theory is centered on the incentives of individual agents responsible for delivering the fraud at the local level, and the mechanisms related to their decision-making, as outlined in the formal model, operate at that local level. In the Russian case, as discussed in the paper (pp.190-1) the kind of fraud being studied was executed locally at the precinct level. Thus, both theoretical and empirical evidence about the spatial scale at which the theorized mechanisms operated justifies the spatial mapping chosen for analysis, and rules out as irrelevant any hypothetically different results obtained from analyses that use larger spatial aggregates.

4.2 Justification

Where scholars do not justify the validity of the spatial units they use, and alternative ones can be identified, concerns about the MAUP remain. This is the case both when scholars provide no justification at all of the spatial units they use, where they justify the spatial mapping used based on data availability alone, or when they justify the spatial unit they use as valid but do not rule out plausibly valid alternatives.

4.2.1 No Justification

In a striking number of instances, scholars provide no justification for the validity of the spatial mappings that they use. This is most commonly seen, at least in our coding of articles, in studies where characteristics of spatial aggregates are included as control variables in analyses focused on assessing individual-level relationships. We provide three examples of this drawn from our review of *APSR* articles, but we identified many more as well.

- Davenport (2016) includes neighborhood-level contextual variables in her study of political attitudes. These are measured at the zip code level, but no justification is provided for this spatial mapping. As we argue in the text, the MAUP impacts inferences on all variables in a model; the issue here, then, is not just that the inferences drawn about the impact of these control variables might change were 'neighborhood' to be operationalized using a different spatial mapping, but that findings on all variables might change.

- Tezcür (2016) studies individual-level determinants of participation in the Kurdish rebellion. The analysis includes contextual variables; these are operationalized at the district level without any justification for this choice.

- In some papers, contextual variables enter as mediators rather than moderators. Valenzuela and Michelson (2016), for example, argue that community-level resources mediate the relationship between identity strength and turnout for Latinos. In their experiment, they measure this mediating variable with information on median household income (p.620) at the community level. But there is no discussion in the paper of the possible spatial mappings that might be valid operationalizations of community - the authors indicate that they use cities (such as Montebello), unincorporated sub-county regions (East Los Angeles) and three "communities" in Texas. It is unclear whether these are the appropriate spatial units to capture the authors' concept of "community-level variation in identity strength" (p.618); we are given no justification for the choice to measure community resources at this level rather than a smaller scale (such as the census tract), larger scale such as the county, or even non-contiguous units based on social networks.

It is also the case, however, that scholars conducting aggregate-level analyses sometimes choose their spatial mappings without any attempt to justify the validity of these choices.

- Ritter and Conrad (2016) study the relationship between dissent and repression in the United States and Africa. To get around the endogeneity inherent in that relationship in the form of preventive repression and strategic self-censorship, they use rainfall as an instrumental variable for observed dissent, arguing that it affects dissent but not repression (p.90). Analysis in the United States is conducted at the state level, and in Africa variation across provinces is explored. There is no reason given why individual instances of repression or dissent in a given day should be studied at these scales, rather than any of a number of plausible alternatives.

- As we mention briefly in the paper, studies that use grid-based data fall into this category: scholars who use grid-cells justify this choice precisely because these spatial units are exogenous to any political or social process; i.e., because they are arbitrary in scale and zonation. But, because there are no theoretical grounds for why a particular grid mapping is used, it follows that the MAUP may pose serious concerns for the validity of findings in these studies. Given that our simulations using gridded data show limited reliability across grid mappings, we suggest that there is reason to believe grid-based analyses may be particularly vulnerable to the MAUP. Though grid-based data is more commonly used in some research communities than others, some examples we encountered in our review of articles include Abramson and Carter (2016), Abramson and Carter (2016), Harris and Posner (2019), and Ahmed and Stasavage (2020).

4.2.2 Data Availability

In some instances, scholars justify the choice of spatial unit based on data availability. The availability of data on key variables is an important constraint on our research design choices - it may be possible to aggregate data up to an alternative spatial unit, but it is much harder to move to a smaller scale. Thus, scholars may be bound to conduct analyses at particular spatial units even if those are not the only plausibly valid mappings - the spatial unit at which data exists may be one of several plausibly valid alternatives, or it may be the closest one can get to the valid mapping. Under either of these two circumstances, the MAUP poses an important threat to the inferences drawn from findings: were one able to obtain data at a different valid spatial mapping, one might arrive at different conclusions about the relationship under study.

For example, Brooke and Ketchley (2018) explore the determinants of formation of Muslim Brotherhood branches in interwar Egypt. Because their focus is on contextual characteristics at the moment of formation in the interwar period, they require data about local characteristics. This data is only available in the 1937 census (p.380) where it is aggregated to the district and the sub-district level. Based on a preliminary analysis that shows most variation is at a smaller scale than the district level, the authors turn to analyzing the sub-district census data, which is the only available spatial mapping at which their analysis can be conducted.

4.2.3 Units that are Not Uniquely Valid

In this instance, scholars justify a particular spatial mapping as plausibly valid for the relevant variable, but do not rule out other mappings that are equally plausibly valid. In our view, this is an incomplete response to the MAUP, since it does not buttress the inferences from a scholars' analyses against the possibility that findings obtained using alternative plausibly valid mappings might differ.

- For example, Anoll (2018) uses the zip code as the spatial mapping for neighborhood-level variables in her analysis of how the ethnic/racial context moderates social rewards to participation. She justifies this spatial mapping as valid by citing Velez and Wong (2017), who compare several standard operationalizations of local community in American politics and find that the zip code better approximates the local community respondents have in mind than does the county. Yet these two choices do not exhaust the range of plausible operationalizations of this concept. Thus, the fact that Anoll makes a case that the zip code is more plausibly valid than the county does not provide sufficient grounds for ruling out other plausibly valid alternatives. As we show in our replication of Anoll in the Appendix Section 3.2, some of her results do in fact change in substantively important ways when an alternative and (we argue) plausibly valid spatial mapping is used.

- Sexton (2016) studies the impact of development aid on insurgent violence in the Afghan conflict, theorizing that the relationship is moderated by the extent of control an incumbent has in a locality. He chooses the district as the spatial aggregate unit for analysis, and then uses the presence of a US military battalion in a district as an indicator of control. This measure is justified by the claim (p.737) that "a battalion has the resources to control the district around them" and the district is a plausible spatial unit to which one could aggregate both development aid and insurgent violence. But other plausible units for these key variables also exist, and control could also be measured at other scales, including both those determined by administrative boundaries, and non-jurisdictional units. Thus, although Sexton's analysis uses plausibly valid units, they are not uniquely valid - and the fact that plausibly valid alternatives exist raises the threats to inference posed by the MAUP.

4.3 Explaining our Coding Approach

To code the *APSR* and *AJPS* articles, we began by limiting our coding exercise to articles that included an empirical component, excluding those that were purely theoretical, normative, or methodological, and where at least one variable was a characteristic of a spatial aggregate unit. Here we erred on the side of inclusion, and further considered any paper where any one variable characterized a spatial aggregate unit. As we discuss further in the text of the paper, some of the more important and more poorly understood effects of the MAUP emerge in settings where analyses include contextual variables as moderators. On the other hand, we took a conservative approach to assessing how widespread the MAUP is in existing research and did not consider the choices of spatial unit in the clustering of standard errors, or in controls for geographic characteristics such as urban/rural where distinct definitions for coding might have sizable impacts on findings (Nemerever and Rogers 2021).

In the five volumes of articles in the *APSR* that we coded (2016-2020), there were 221 papers that included empirical analyses, and 124 of these (56%) included a spatial aggregate unit in at least one component of their analyses. In the *AJPS* for the same years, 114 articles included a spatial aggregate unit. Considering these articles further, two of the authors coded whether in each paper a plausibly valid alternative to the spatial unit used could be identified. This entailed both searching the paper for a justification for the spatial units chosen (along the lines detailed earlier in this portion of the Appendix) and considering whether the concepts deployed in the theory, the research setting, and information about the variables and the mechanisms could eliminate all alternative spatial mappings for all variables included in the analyses. We worked together to reach a consensus in the rare cases where our initial codings differed.

Table A.8 explains our justification for each of the 52 *APSR* articles where we believe analysis could have equally plausibly been conducted at a different spatial scale. These are papers where we think the MAUP could potentially be a threat to inferences. Our brief descriptions of each also lay out responses, if any, that the authors take in response to the MAUP, including most commonly, demonstrations of reliability of findings at an alternative spatial mapping. In our sample from the *APSR*, then, 43% of the papers that include an aggregate spatial unit, and nearly 1/4 of all empirical papers, may be affected by the MAUP. The rate in the *AJPS* is similar: of the 251 papers, 114 incorporated a spatial aggregate unit, and in 59 of these (52%) we identified a potential alternative spatial mapping that could be valid for analysis. It may be the case that in many of these papers, authors can plausibly rule out the alternative spatial mappings we have identified as invalid for their analyses; our point is that doing so would improve our confidence and that it should be standard practice to discuss whether and how one has addressed the MAUP when it is a potential concern.

Table A.8: The MAUP as a Threat to Inference in Empirical Articles, *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
110 (1)	52-67	Davenport (2016)	Neighborhood-level study of political attitudes; zip code used to operationalize neighborhood with no justification provided.
110 (1)	85-99	Ritter and Conrad (2016)	Subnational variation in repression, using rainfall to instrument for observed dissent. Variation across provinces in Africa and US states. No explanation for use of these units versus alternatives.
110 (1)	127-147	Braun (2016)	Religious minorities and resistance to genocide–"local minority churches were more likely to rescue Jews" during Holocaust. Uses various spatial units to operationalize localities in terms of religious dominance, unclear what unit is used to code whether individual churches represent local minorities.
110 (2)	247-264	Tezcür (2016)	Individual-level determinants of participation in Kurdish rebellion. Contextual variables at the district level without justification.
110 (2)	325-341	Rogowski (2016)	US county-level variation in PGP (post offices) and partisan alignment in Congress. No theoretical justification of county unit.
110 (4)	615-630	Valenzuela and Michelson (2016)	Community level resources mediate relationship between identity strength (ethnic vs national) and turnout for Latinos. Many possible spatial units could be used to operationalize 'community.'
110 (4)	675-698	Abramson and Carter (2016)	Grid square as unit of analysis. Grids are designed to have arbitrary scale and zoning.
110 (4)	731-749	Sexton (2016)	The effect of aid on insurgent violence in Afghanistan depends on whether aid is administered in locations controlled by the ISAF, or contested areas. Uses district as spatial unit of analysis with no justification; it is unclear why ISAF control should be measured at that spatial scale versus others.
111 (2)	322-337	Hale and Colton (2017)	Uses community size quintile in some of its analyses, but does not address the possibility that community could be operationalized in alternate spatial aggregate units.
111 (3)	439-459	Reese et al. (2017)	Does the Islamic calendar affect militant violence? Subnational evidence based on local variation in weather conditions and local "societal disapproval" of violence on holidays, but multiple spatial units are plausible for each of these elements of the theory.

Continued

Table A.8. *Continued*, the MAUP in *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
111 (4)	755-770	Lerman et al. (2017)	Role of partisanship in individual level variation in uptake of US Affirmative Care Act (ACA). Includes US county-level correlation between ACA enrollment and Obama vote share, but this relationship could be studied at other levels of aggregation.
111 (4)	801-818	Charnysh and Finkel (2017)	Did proximity to Treblinka affect the life trajectories of Poles in the area and their long-term political affiliation with anti-Semitic parties? Uses the smallest administrative unit for which data are available, but other units are also theoretically appropriate.
112 (1)	167-185	Clinton and Sances (2018)	Changes in registration and turnout across US counties following Medicaid expansion in some states. Unclear that county is the only appropriate level of analysis for this question.
112 (2)	201-218	Colantone and Stanig (2018)	See discussion and replication in main text of paper.
112 (2)	339-357	Garfias (2018)	Effects of price shocks that weaken economic elites on changes in state capacity across municipalities in Mexico. Since the power of elites is de facto, it is unclear why their impact has to operate at the municipal level rather than at larger or sub-municipal scale.
112 (2)	376-394	Brooke and Ketchley (2018)	How economic and state infrastructure shaped the early contexts of Muslim Brotherhood activism in Egypt. Variation is assessed across census sub-districts and districts; unclear whether or why these are the only appropriate units of analysis.
112 (3)	473-493	Hankinson (2018)	How views about housing construction in city and in neighborhood vary for homeowners vs. renters. Uses zip code level data to study neighborhoods, without justification.
112 (3)	494-508	Anoll (2018)	See replication and discussion in main text of paper.
112 (4)	742-757	Enos and Gidron (2018)	Lab-in-the-field conducted in 20 locations in Israel that vary demographically; this is used as a control in some analyses but unclear what spatial unit is appropriate for operationalizing local demography.

Continued.

Table A.8. *Continued*, the MAUP in *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
112 (4)	971-995	Guardado (2018)	Worse quality governors see long-term negative development outcomes, measured at the district level, with contemporary districts fitted into historical colonial provinces. The colonial province unit is theoretically motivated; contemporary district level is chosen to match development outcomes due to data availability; could also be tested with other spatial units smaller than historical colonial province.
112 (4)	1050-1066	Selb and Munzert (2018)	Were campaign effects involved in Hitler's rise to power? Community-level difference-in-differences analysis comparing voting results in places that were and were not visited during campaigns. Areal unit chosen based only on data availability of election statistics (community level in some models, counties or county boroughs in others). Other units plausible.
113 (1)	123-139	Harris and Posner (2019)	Do MPs target supporters within district for goods provision? Uses grid cells around polling locations to code sub-district political orientation. Grid cells are designed to have arbitrary scale and zoning.
113 (2)	293-310	Weaver (2019)	Evolution of news coverage about lynching. U.S. county level measure of rail line as measure of exposure to outside media. County centroids used to calculate proximity and connectedness. Unclear whether or why the county is the only appropriate level of analysis.
113 (2)	372-384	Martin and McCrain (2019)	Uses "designated market area," a TV specific geographic unit of analysis. The unit is not justified other than data are available at that unit; theory suggests that other units of analysis are plausible.
113 (2)	405-422	Fouka (2019)	State-level analysis of effects of anti-German sentiment on behavior of immigrants. Unclear why state is chosen beyond data availability.
113 (2)	423-441	Charnysh et al. (2019)	Use of municipality level data of immigration transfers, justified as the smallest unit of analysis available and small enough to imply social interaction, but presumably there are other plausible alternative local-level units.
113 (2)	442-455	Hangartner et al. (2019)	Individual level survey data of Greek islands exposed to refugees. Island unit of analysis justified as exposure to refugees, use of a distance measure from island centroid to Turkey as (instrumental) independent variable. Other spatial units could plausibly be used to assess exposure to refugees.

Continued.

Table A.8. *Continued*, the MAUP in *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
113 (2)	456-474	Maxwell (2019)	Does moving to cities change immigration attitudes, or does lifelong exposure change immigration attitudes? Exposure operationalized by composition of neighborhood, Swiss commune, and city, but no justification of these choices or discussion of alternative units.
113 (2)	475-498	Arias et al. (2019)	Precinct level data on municipal elections, social network data at the precinct level. No justification of precinct as the appropriate unit of analysis to operationalize social network except for their small size; other possible units can be identified.
113 (2)	499-516	Larsen et al. (2019)	Voting precinct used as spatial unit representing "local housing markets." Precincts justified because they are small and do not overlap with local media markets, no theoretical justification for precinct being the "local economy" or discussion of alternative possible units. Includes robustness checks for individual level models with "local economy" defined as circles of varying radii of housing markets.
113 (2)	552-568	Reuter and Szakonyi (2019)	Uses region-level election outcomes, justified because regions are the "key fora" of rent-seeking and spoil-sharing. They mention other possible units of analysis but do not study them.
113 (2)	569-583	Rozenas and Zhukov (2019)	Impact of Stalin's repression on loyalty to Moscow in Ukraine. Unit of analysis is 1933-era Ukrainian rayon, a 2nd level administrative subdivision, justified only by data availability. Other units possible.
113 (3)	658-673	Hall et al. (2019)	US county level analysis of "community encouragement" for joining the Confederacy; other ways to operationalize community possible.
113 (3)	710-726	Cantú (2019)	Multiple levels of data (e.g., state, district) included in the regressions, without discussion of which contextual variables should be measured at the district vs. state level and why. Dependent variable is collected at the polling place level, aggregated to the electoral district to match other data availability.
113 (3)	727-742	Ascencio and Rueda (2019)	Use of controls (average number of years of schooling) at different levels of aggregation (municipality) than the main IV and DV (precinct), without theoretical justification. Presumably polling station level or precinct-level education data is not available.

Continued.

Table A.8. *Continued*, the MAUP in *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
113 (4)	1012-1028	Enos et al. (2019)	Precinct level analysis of effect of LA Riots on political behavior, coded based on distance from the riot. Not clear that precinct is the only spatial unit one could use, no discussion of alternatives.
113 (4)	1029-1044	Hager et al. (2019)	Neighborhood-level (primary sampling unit based on 250m x 250m grid cells) exposure to riots; presumably other ways to operationalize neighborhood are possible. Grids cells are designed to have arbitrary scale and zoning.
114 (2)	309-325	Gade (2020)	The effects of checkpoints on resistance depend on whether they are located within villages (divide up communities), or between them, leaving those communities whole but isolated. No discussion of how community borders are located spatially; implicit assumption that village = community; alternative spatial units could be appropriate.
114 (2)	326-341	Allen et al. (2020)	Survey analysis of effects of US military base presence on political attitudes. Binary variable of US military facility located within respondents' province/region. No justification for that unit.
114 (2)	426-442	Thal (2020)	US state-level data on income and regional cost of living (unit is state level, with value collected from major metropolitan areas within the state) used as controls without justifications of these aggregations.
114 (2)	443-455	Trounstine (2020)	Main analysis uses city level because it decides land use zoning. Part of the paper analyzes how neighborhood racial composition affects land use preferences. Uses census tract (aggregates of electoral precincts) to operationalize neighborhood; alternative operationalizations possible.
114 (2)	486-501	Cruz et al. (2020)	Effects of social networks on public goods provision at the village level within municipalities. Unclear whether other spatial units might also be appropriate.
114 (2)	502-518	Ahmed and Stasavage (2020)	Grid cell analysis of local economic conditions, to explain society-level existence of central and local councils. Grids are designed to have arbitrary scale and zoning.
114 (2)	573-590	Homola and Pereira (2020)	Electoral effects of location of concentration camps. Electoral district level data for district level analysis combined with distance from Nazi establishment measured at the smallest regional identifiers (Kreis, Gemeinde). No justification for use of these aggregate spatial units rather than alternatives.

Continued.

Table A.8. *Continued*, the MAUP in *APSR*, Vol. 110-114

Vol (Issue)	Pages	Authors (Year)	Notes
114 (3)	638-659	Wasow (2020)	Effects of protest on voting in nearby US counties. Unclear why county-level vote shares are the appropriate unit of analysis; no discussion of possible alternatives.
114 (3)	660-676	de Benedictis-Kessner and Warshaw (2020)	County level data as measure of "local" economy. No justification for unit choice or consideration of alternatives.
114 (4)	1055-1070	Bucchianeri (2020)	Intermixing of city and county council analysis, no justification of unit other than they are "American Local Government."
114 (4)	1071-1085	Challú et al. (2020)	Precinct level analysis of electoral returns, not the only plausible unit.
114 (4)	1213-1229	Yoder (2020)	US city council unit of analysis to understand the effect of home ownership on political participation. City is an appropriate level of analysis for effects via property values but other units are plausible.
114 (4)	1230-1246	Gulzar et al. (2020)	Use of village level not justified theoretically.
114 (4)	1335-1342	Lehmann and Masterson (2020)	Sorting into control and treatment group determined by "community" measures based on altitude. Use of altitude is theoretically justified but community boundaries are not defined or theoretically justified.
114 (4)	1359-1365	Hazlett and Mildenberger (2020)	Use of census block group as spatial unit without theoretical justification.
114 (4)	1375-1385	Hassell et al. (2020)	Examines effect of school shooting on democratic accountability, using US county level data in most cases without clear justification for unit. Explores some parts of analysis using alternative spatial units.

5 Addressing the MAUP: Theoretical Precision and Empirical Strategies

Existing scholarship recommends that scholars address the MAUP by showing that results are reliable at a different mapping. We explore the performance of those reliability checks in the main text; here we provide more detail on that exercise.

5.1 Aggregation to Larger Scale

We began by following the most common practice and identifying settings in which we could aggregate up from one of the mappings we produced to another spatial unit in which the initial units were nested. We then compared the two sets of results generated from regression analyses using these two spatial units as the units of analysis and assessed whether our results were reliable in terms of whether each of the two independent variables maintained the same sign and significance. We were able to identify 15 such opportunities to assess reliability in which all units in the original mapping could be aggregated into larger units and conducted these analyses for each of the two types (i.e., uniform sampling vs. non-uniform sampling) of individual-level data that we generated. Below in Table A.9 we list the 15 sets of comparisons conducted across spatial units:

Table A.9: 15 Opportunities to Address Reliability

Original mapping	Robustness check mapping
100 units	25 units
100 units	4 units
16 units	4 units
100 units 2 eastward shifts	25 units 1 eastward shifts
100 units 4 eastward shifts	25 units 2 eastward shifts
100 units 6 eastward shifts	25 units 3 eastward shifts
100 units 8 eastward shifts	25 units 4 eastward shifts
100 units 2 southward shifts	25 units 1 southward shifts
100 units 4 southward shifts	25 units 2 southward shifts
100 units 6 southward shifts	25 units 3 southward shifts
100 units 8 southward shifts	25 units 4 southward shifts
100 units 5 eastward shifts	4 units 1 eastward shifts
100 units 5 southward shifts	4 units 1 southward shifts
16 units 2 eastward shifts	4 units 1 eastward shifts
16 units 2 southward shifts	4 units 1 southward shift

5.2 Monte Carlo Simulation with Regular Lattice Data: Results Different Zonation

In principle, scholars could also investigate whether their results are robust to different zonation. Practically, aggregating data to multiple such spatial units is only possible where data are available at a scale small enough relative to the units being used, which is rarely the case since when such disaggregated data are available, scholars conduct analyses at that scale.[4] But it is worth exploring how our simulated data perform in terms of reliability in response to the zoning sub-problem since this might be relevant in a setting where the size of a spatial aggregate unit is unambiguously defined, but its borders are up for debate.

For the purposes of this exercise, we took each of the grid mappings that we generated, and compared results to the adjacent mappings at the same scale; that is, to the mappings generated with shifts east/west and north/south by one unit. There are 80 such comparisons that can be made for each of the two sets of individual-level data, as listed in Table A.10. Each entry in the list identifies the original unit used and the adjacent zonation used as a robustness check.

5.3 Results

The results for these analyses are shown below in Tables A.11 through A.13, using the same format as in the main paper, which shows results for the effect of IV2 on the DV. Table A.11 supplements Table 1 from the main text that focuses on the variable we arbitrarily labeled 'IV2' and shows the results of 'IV1' for alternative spatial aggregates and both of the sampling methods used to generate the individual-level data. When examining the off-diagonal cells (bolded) in the bottom-right quadrant, under non-uniform sampling, we find that only 5% of all pairs of spatial mappings, (2+2)/80, produced inconsistent results. This is starkly different from the nearly 33% cases we found in Table 1 from the main text. This contrast reveals surprising and unexpected variation in the severity of the reliability problem across two variables that were randomly generated using the same underlying process. In our

[4]See the discussion in the main text about the problems with automatically "going smaller" as a response to ambiguity about the appropriate spatial scale for analysis.

Table A.10: 80 Robustness Checks on 80 Mapped Pairs

Units	The adjacent zonation used
4 units	Original and 1 eastward shift
	1 eastward shifts and original
	Original and 1 southward shifts
	1 southward shifts and original
16 units	Original and 1 eastward shift
	1 eastward shifts and original
	1 eastward shifts and 2 eastward shifts
	2 eastward shifts and 1 eastward shifts
	2 eastward shifts and 3 eastward shifts
	3 eastward shifts and 2 eastward shifts
	3 eastward shifts and original
	Original and 3 eastward shifts
	Original and 1 southward shifts
	1 southward shifts and original
	1 southward shift and 2 southward shifts
	2 southward shifts and 1 southward shifts
	2 southward shifts and 3 southward shifts
	3 southward shifts and 2 southward shifts
	3 southward shifts and original
	Original and 3 southward shifts
25 units	Original and 1 eastwards shift
	1 eastward shifts and original
	1 eastward shifts and 2 eastward shifts
	2 eastward shifts and 1 eastward shifts
	2 eastward shifts and 3 eastward shifts
	3 eastward shifts and 2 eastward shifts
	3 eastward shifts and 4 eastward shifts
	4 eastward shifts and 3 eastward shifts
	4 eastward shifts and original
	Original and 4 eastward shifts
	Original and 1 southward shifts
	1 southward shifts and original
	1 southward shift and 2 southward shifts
	2 southward shifts and 1 southward shifts
	2 southward shifts and 3 southward shifts
	3 southward shifts and 2 southward shifts
	3 southward shifts and 4 southward shifts
	4 southward shifts and 3 southward shifts
	4 southward shifts and original
	Original and 4 southward shifts

Continued

Table A.10. Robustness Checks on 80 Mapped Pairs (*Continued*)

Units	The adjacent zonation used
100 units	Original and 1 eastward shift
	1 eastward shifts and original
	1 eastward shifts and 2 eastward shifts
	2 eastward shifts and 1 eastward shifts
	2 eastward shifts and 3 eastward shifts
	3 eastward shifts and 2 eastward shifts
	3 eastward shifts and 4 eastward shifts
	4 eastward shifts and 3 eastward shifts
	4 eastward shifts and 5 eastward shifts
	5 eastward shifts and 4 eastward shifts
	5 eastward shifts and 6 eastward shifts
	6 eastward shifts and 5 eastward shifts
	6 eastward shifts and 7 eastward shifts
	7 eastward shifts and 6 eastward shifts
	7 eastward shifts and 8 eastward shifts
	8 eastward shifts and 7 eastward shifts
	8 eastward shifts and 9 eastward shifts
	9 eastward shifts and 8 eastward shifts
	9 eastward shifts and original
	Original and 9 eastward shifts
	Original and 1 southward shifts
	1 southward shifts and original
	1 southward shifts and 2 southward shifts
	2 southward shifts and 1 southward shifts
	2 southward shifts and 3 southward shifts
	3 southward shifts and 2 southward shifts
	3 southward shifts and 4 southward shifts
	4 southward shifts and 3 southward shifts
	4 southward shifts and 5 southward shifts
	5 southward shifts and 4 southward shifts
	5 southward shifts and 6 southward shifts
	6 southward shifts and 5 southward shifts
	6 southward shifts and 7 southward shifts
	7 southward shifts and 6 southward shifts
	7 southward shifts and 8 southward shifts
	8 southward shifts and 7 southward shifts
	8 southward shifts and 9 southward shifts
	9 southward shifts and 8 southward shifts
	9 southward shifts and original
	Original and 9 southward shifts

Table A.11: Robustness Checks (Effects of IV1 on DV)

	Initial Result			
	Uniform Sampling		*Non-uniform Sampling*	
Robustness Check	Not significant	Significant	Not significant	Significant
Scale				
Significant	**0**	0	**7**	0
Not significant	13	**2**	8	**0**
Zoning				
Significant	**4**	2	**2**	4
Not significant	70	**4**	72	**2**

Notes: All regression outputs are based on Tables A.9 (15 exercises for scaling) and A.11 (80 exercises for zoning). The number in each cell indicates the count of analyses that fall into the conditions listed. Bold digits denote the off-diagonal cells in which the unpredictable effects of the MAUP are more concerning. We do not have any pairs where both results are significant but the sign is the opposite.

Table A.12: Robustness Checks across Changes in Scale

	Initial Result			
	Uniform Sampling		*Non-uniform Sampling*	
Robustness Check	Not significant	Significant	Not significant	Significant
IV1				
Significant	**0**	0	**7**	0
Not significant	13	**2**	8	**0**
IV2				
Significant	**0**	0	**4**	9
Not significant	11	**4**	2	**0**

Note: All regression outputs are based on Tables A.9 (15 exercises) and A.10 (80 exercises).

Table A.13: Robustness Checks across Changes in Zonation

	Initial Result			
	Uniform Sampling		*Non-uniform Sampling*	
Robustness Check	Not significant	Significant	Not significant	Significant
IV1				
Significant	**4**	2	**2**	4
Not significant	70	**4**	72	**2**
IV2				
Significant	**4**	8	**14**	36
Not significant	64	**4**	16	**14**

*Note:*All regression outputs are based on Tables A.9 (15 exercises) and A.10 80cases (80 exercises).

view, the finding that the effects of the MAUP are so unpredictable casts doubt on reliability checks alone as a sufficient response in most research settings.

To explore the results of this exercise in an alternative manner, we put both IV1 and IV2 together in Table A.12 (sorted for assessing robustness to different scales) and Table A.13 (sorted for assessing robustness to different zonings). The overall pattern in Table A.12 that focuses on changes in scale is that most results were reliable. But notably, the fact that the off-diagonal cell (bolded) results for IV2 are less consistently reliable than those for IV1 is a first indication of the unpredictable effects of the MAUP. The third and fourth columns in this table show the results using the individual-level data generated from an unknown sampling distribution are much less reliable than scale changes, and once again, IV2 performs worse than IV1 in terms of reliability. We present similar results from the exercises for different zonings in Table A.13. Here, IV1 again consistently produces more reliable results than IV2, while the mode of sampling does not have a systemic effect on reliability.

In all, this exercise shows that the reliability of results across changed border mappings does not follow any predictable or consistent pattern. This confirms the need that we emphasize in the text for scholars to focus on theoretical validity rather than reliability as the core of efforts to address the MAUP.

References

Abramson, S, and D. Carter. 2016. "The Historical Origins of Territorial Disputes". *American Political Science Review 110*(4): 675–698.

Ahmed, A. T, and D. Stasavage. 2020. "Origins of Early Democracy". *American Political Science Review 114*(2): 502–518.

Allen, M, M. Flynn, C. M. Machain, and A. Stravers. 2020. "Outside the Wire". *American Political Science Review 114*(2): 326–341.

Amrhein, C. 1995. "Searching for the Elusive Aggregation Effect". *Environment and Planning A 27*(1): 105–119.

Anoll, A. P. 2018. "What Makes a Good Neighbor?". *American Political Science Review 112*(3): 494–508.

Arias, E, P. Balán, H. Larreguy, J. Marshall, and P. Querubín. 2019. "Information Provision, Voter Coordination, and Electoral Accountability". *American Political Science Review 113*(2): 475–498.

Arnorsson, A, and G. Zoega. 2018. "On the Causes of Brexit". *European Journal of Political Economy 55*: 301–323.

Ascencio, S. J, and M. R. Rueda. 2019. "Partisan Poll Watchers and Electoral Manipulation". *American Political Science Review 113*(3): 727–742.

Bisbee, J, and J. Zilinsky. 2023. "Geographic Boundaries and Local Economic Conditions Matter for Views of the Economy". *Political Analysis 31*(2): 288–294.

Blom-Hansen, J, K. Houlberg, S. Serritzlew, and D. Treisman. 2016. "Jurisdiction size and local government policy expenditure". *American Political Science Review 110*(4): 812–831.

Braun, R. 2016. "Religious Minorities and Resistance to Genocide". *American Political Science Review 110*(1): 127–147.

Brautzsch, H.-U, and O. Holtemöller. 2021. "International Trade Barriers and Regional Employment". *Journal of Economic Structures 10*(1): 1–25.

Briant, A, P.-P. Combes, and M. Lafourcade. 2010. "Dots to boxes". *Journal of Urban Economics 67*(3): 287–302.

Brooke, S, and N. Ketchley. 2018. “Social and Institutional Origins of Political Islam”. *American Political Science Review 112*(2): 376–394.

Bucchianeri, P. 2020. “Party Competition and Coalitional Stability”. *American Political Science Review 114*(4): 1055–1070.

Cantú, F. 2019. “The Fingerprints of Fraud”. *American Political Science Review 113*(3): 710–726.

Challú, C, E. Seira, and A. Simpser. 2020. “The Quality of Vote Tallies”. *American Political Science Review 114*(4): 1071–1085.

Charnysh, V et al.. 2019. “Diversity, Institutions, and Economic Outcomes”. *American Political Science Review 113*(2): 423–441.

Charnysh, V, and E. Finkel. 2017. “The Death Camp El Dorado”. *American Political Science Review 111*(4): 801–818.

Clinton, J. D, and M. W. Sances. 2018. “The Politics of Policy”. *American Political Science Review 112*(1): 167–185.

Colantone, I, and P. Stanig. 2018. “Global Competition and Brexit”. *American Political Science Review 112*(2): 201–218.

Cruz, C, J. Labonne, and P. Querbin. 2020. “Social Network Structures and the Politics of Public Goods Provision”. *American Political Science Review 114*(2): 486–501.

Davenport, L. D. 2016. “Beyond Black and White”. *American Political Science Review 110*(1): 52–67.

de Benedictis-Kessner, J, and C. Warshaw. 2020. “Accountability for the Local Economy at All Levels of Government in United States Elections”. *American Political Science Review 114*(3): 638–659.

Enos, R, A. Kaufman, and M. Sands. 2019. “Can Violent Protest Change Local Policy Support?”. *American Political Science Review 113*(4): 1012–1028.

Enos, R. D, and N. Gidron. 2018. “Exclusion and Cooperation in Diverse Societies”. *American Political Science Review 112*(4): 742–757.

Fotheringham, A. S, and D. W. Wong. 1991. “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis”. *Environment and Planning A 23*(7): 1025–1044.

Fouka, V. 2019. “How Do Immigrants Respond to Discrimination?”. *American Political Science*

Review 113(2): 405–422.

Gade, E. K. 2020. “Social Isolation and Repertoires of Resistance”. *American Political Science Review 114*(2): 309–325.

Garfias, F. 2018. “Elite Competition and State Capacity Development”. *American Political Science Review 112*(2): 339–357.

Guardado, J. 2018. “Office-selling, Corruption, and Long-term Development in Peru”. *American Political Science Review 112*(4): 971–995.

Gulzar, S, N. Hass, and B. Pasquale. 2020. “Does Property Ownership Lead to Participation in Local Politics?”. *American Political Science Review 114*(4): 1213–1229.

Gulzar, S, and B. J. Pasquale. 2017. “Politicians, bureaucrats, and development”. *American Political Science Review 111*(1): 162–183.

Hager, A, K. Krakowski, and M. Schaub. 2019. “Ethnic Riots and Prosocial Behavior”. *American Political Science Review 113*(4): 1029–1044.

Hale, H. E, and T. J. Colton. 2017. “Who Defects? Unpacking a Defection Cascade from Russia’s Dominant Party 2008-12”. *American Political Science Review 111*(2): 322.

Hall, A. B, C. Huff, and S. Kuriwaki. 2019. “Wealth, Slaveownership, and Fighting for the Confederacy”. *American Political Science Review 113*(3): 658–673.

Hangartner, D, E. Dinas, M. Marbach, K. Matakos, and D. Xefteris. 2019. “Does Exposure to the Refugee Crisis Make Natives More Hostile?”. *American Political Science Review 113*(2): 442–455.

Hankinson, M. 2018. “When Do Renters Behave Like Homeowners?”. *American Political Science Review 112*(3): 473–493.

Harris, J. A, and D. N. Posner. 2019. “(Under What Conditions) Do Politicians Reward Their Supporters?”. *American Political Science Review 113*(1): 123–139.

Hassell, H, J. Holbein, and M. Baldwin. 2020. “Mobilize for Our Lives?”. *American Political Science Review 114*(4): 1375–1385.

Hazlett, C, and M. Mildenberger. 2020. “Wildfire Exposure Increases Pro-environment Voting within Democratic but not Republican Areas”. *American Political Science Review 114*(4): 1359–1365.

Healy, A. J, M. Persson, and E. Snowberg. 2017. “Digging into the pocketbook”. *American Political*

Science Review 111(4): 771–785.

Hipp, J. R. 2007. “Block, Tract, and Levels of Aggregation”. *American sociological review 72*(5): 659–680.

Homola, J, and M. Pereira, Miguel M.and Tavits. 2020. “Legacies of the Third Reich”. *American Political Science Review 114*(2): 573–590.

Huggins, C. 2018. “Did EU Regional Spending Affect the Brexit Referendum?”. *Regional Studies, Regional Science 5*(1): 388–397.

Larreguy, H, J. Marshall, and P. Querubin. 2016. “Parties, Brokers, and Voter Mobilization”. *American Political Science Review 110*(1): 160–179.

Larsen, M. V, F. Hjorth, P. T. Dinesen, and K. M. Sønderskov. 2019. “When Do Citizens Respond Politically to the Local Economy?”. *American Political Science Review 113*(2): 499–516.

Lehmann, C. M, and D. T. Masterson. 2020. “Does Aid Reduce Anti-refugee Violence?”. *American Political Science Review 114*(4): 1335–1342.

Lerman, A. E, M. L. Sadin, and S. Trachtman. 2017. “Policy Uptake as Political Behavior”. *American Political Science Review 111*(4): 755–770.

Lindgren, K.-O, S. Oskarsson, and M. Persson. 2019. “Enhancing Electoral Equality”. *American Political Science Review 113*(1): 108–122.

Los, B, P. McCann, J. Springford, and M. Thissen. 2017. “The Mismatch Between Local Voting and the Local Economic Consequences of Brexit”. *Regional studies 51*(5): 786–799.

Martin, G. J, and J. McCrain. 2019. “Local News and National Politics”. *American Political Science Review 113*(2): 372–384.

Maxwell, R. 2019. “Cosmopolitan Immigration Attitudes in Large European Cities”. *American Political Science Review 113*(2): 456–474.

Nemerever, Z, and M. Rogers. 2021. “Measuring the Rural Continuum in Political Science”. *Political Analysis 29*(3): 267–286.

Openshaw, S. 1984. “Ecological Fallacies and the Analysis of Areal Census Data”. *Environment and Planning A 16*(1): 17–31.

Reese, M. J, K. G. Ruby, and R. A. Pape. 2017. “Days of Action or Restraint?”. *American Political*

Science Review 111(3): 439–459.

Reuter, O. J, and D. Szakonyi. 2019. "Elite Defection under Autocracy". *American Political Science Review 113*(2): 552–568.

Ritter, H. E, and C. Conrad. 2016. "Preventing and Responding to Dissent". *American Political Science Review 110*(1): 85–99.

Rogowski, J. C. 2016. "Presidential Influence in an Era of Congressional Dominance". *American Political Science Review 110*(2): 325–341.

Rozenas, A, and Y. Zhukov. 2019. "Mass Repression and Political Loyalty". *American Political Science Review 113*(2): 569–583.

Rundlett, A, and M. W. Svolik. 2016. "Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud". *American Political Science Review 110*(1): 180–197.

Schraff, D, I. Vergioglou, and B. B. Demirci. 2023. "The European NUTS-level Election Dataset". *Party Politics 29*(3): 570–579.

Selb, P, and S. Munzert. 2018. "Examining a Most Likely Case for Strong Campaign Effects". *American Political Science Review 112*(4): 1050–1066.

Sexton, R. 2016. "Aid as a Tool Against Insurgency". *American Political Science Review 110*(4): 731–749.

Tajima, Y, K. Samphantharak, and K. Ostwald. 2018. "Ethnic Segregation and Public Goods". *American Political Science Review 112*(3): 637–653.

Tezcür, G. M. 2016. "Ordinary People, Extraordinary Risks". *American Political Science Review 110*(2): 247–264.

Thal, A. 2020. "The Desire for Social Status and Economic Conservatism among Affluent Americans". *American Political Science Review 114*(2): 426–442.

Touchton, M, N. B. Sugiyama, and B. Wampler. 2017. "Democracy at Work". *American Political Science Review 111*(1): 68–82.

Trounstine, J. 2020. "The Geography of Inequality". *American Political Science Review 114*(2): 443–455.

Valenzuela, A. A, and M. R. Michelson. 2016. "Turnout, Status, and Identity". *American Political*

Science Review 110(4): 615–630.

Velez, Y. R, and G. Wong. 2017. "Assessing Contextual Measurement Strategies". *The Journal of Politics 79*(3): 1084–1089.

Wasow, O. 2020. "Agenda Seeding". *American Political Science Review 114*(3): 638–659.

Weaver, M. 2019. "'Judge Lynch' in the Court of Public Opinion". *American Political Science Review 113*(2): 293–310.

Wilson, R. 2013. "Changing Geographic Units and the Analytical Consequences: An Example of Simpson's Paradox". *Cityscape 15*(2): 289–304.

Wong, C, J. Bowers, T. Williams, and K. D. Simmons. 2012. "Bringing the Person Back In". *The Journal of Politics 74*(4): 1153–1170.

Yoder, J. 2020. "Does Property Ownership Lead to Participation in Local Politics?". *American Political Science Review 114*(4): 1213–1229.