Supplemental material for "Data wrangling, computational burden, automation, robustness and accuracy in ecological inference forecasting of R×C tables"

Jose M. Pavía[1] and Rafael Romero[2]

June 2023

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 47(1).

[1] GIPEyOP, Universitat de Valencia, Valencia (Spain), email: pavia@uv.es ORCID: https://orcid.org/0000-0002-0129-726X

[2] Universidad Politécnica de Valencia, Valencia (Spain), email: rromero@eio.upv.es

S1. Supplementary material. Section 4

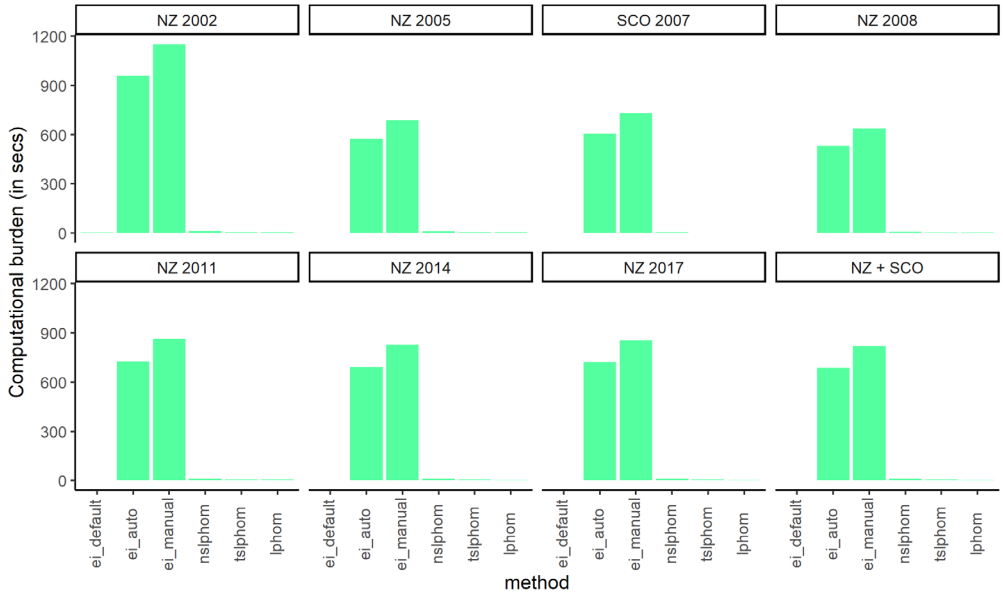

Figure S1. Graphical representation of average computation times grouped by election and algorithm in the reference database, after applying the functions `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options to the official data from the New Zealand electoral commission and the Scotland Electoral Office described in Section 3. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

S2. Supplementary material. Subsection 5.1

Table S1. Summary of the performance of the algorithms in scenarios with few units.

Country Year	NZ 2002	NZ 2005	SCO 2007	NZ 2008	NZ 2011	NZ 2014	NZ 2017	NZ + SCO
# of Elections	N = 69	N = 69	N = 73	N = 70	N = 70	N = 71	N = 71	N = 493
Avg. # of units	$\overline{I}$ = 14.0	$\overline{I}$ = 15.5	$\overline{I}$ = 15.1	$\overline{I}$ = 14.5	$\overline{I}$ = 15.3	$\overline{I}$ = 14.4	$\overline{I}$ = 14.7	$\overline{I}$ = 14.8
Avg. # of cells	$\overline{RC}$ = 39.5	$\overline{RC}$ = 23.8	$\overline{RC}$ = 35.2	$\overline{RC}$ = 23.4	$\overline{RC}$ = 26.2	$\overline{RC}$ = 27.9	$\overline{RC}$ = 24.8	$\overline{RC}$ = 28.7
				Averages of *EI* measures				
ei_default	31.51	35.43	52.05	36.80	36.10	37.98	39.54	38.59
ei_auto	31.55	35.12	50.40	36.14	34.81	36.98	37.19	37.55
ei_manual	22.34	17.50	42.52	18.81	16.92	19.62	18.76	22.47
nslphom	17.08	12.26	12.46	12.59	13.00	13.71	12.04	13.29
tslphom	18.75	13.24	14.03	13.71	14.32	14.74	13.32	14.58
lphom	19.75	13.84	14.91	14.35	14.99	15.30	13.94	15.25
				Averages of *WPE* measures				
ei_default	23.03	28.33	44.70	29.97	28.91	30.70	32.13	31.22
ei_auto	22.91	28.02	43.27	29.32	27.57	29.70	30.05	30.22
ei_manual	14.61	12.31	35.94	13.84	11.82	14.15	13.61	16.73
nslphom	10.08	7.90	7.12	8.93	8.68	8.86	7.61	8.44
tslphom	11.16	8.60	8.08	9.96	9.73	9.60	8.50	9.36
lphom	11.77	8.99	8.52	10.42	10.17	9.96	8.88	9.81
				Averages of computational burden (in secs)				
ei_default	0.39	0.25	0.33	0.23	0.27	0.27	0.24	0.28
ei_auto	188.68	120.92	163.77	109.87	130.62	130.99	117.05	137.45
ei_manual	226.41	145.12	227.34	131.93	156.83	157.28	140.66	169.57
nslphom	1.12	1.04	0.85	0.92	1.05	1.01	0.98	0.99
tslphom	0.14	0.11	0.09	0.10	0.11	0.11	0.11	0.11
lphom	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: compiled by the authors after applying the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) with different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) to the data from the New Zealand electoral commission and the Scotland Electoral Office attained after randomly merging polling units as described in Section 3. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

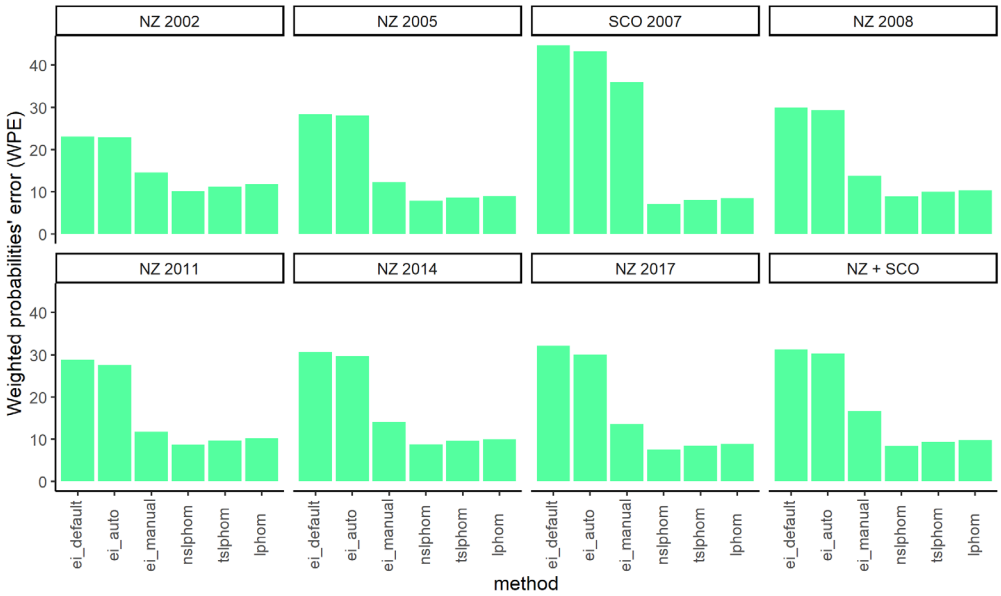

Figure S2. Graphical representation of average values of EPW error measures grouped by election and algorithm in the scenarios attained after randomly merging polling units as described in Section 3. Individual solutions have been attained with the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 1000, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument.

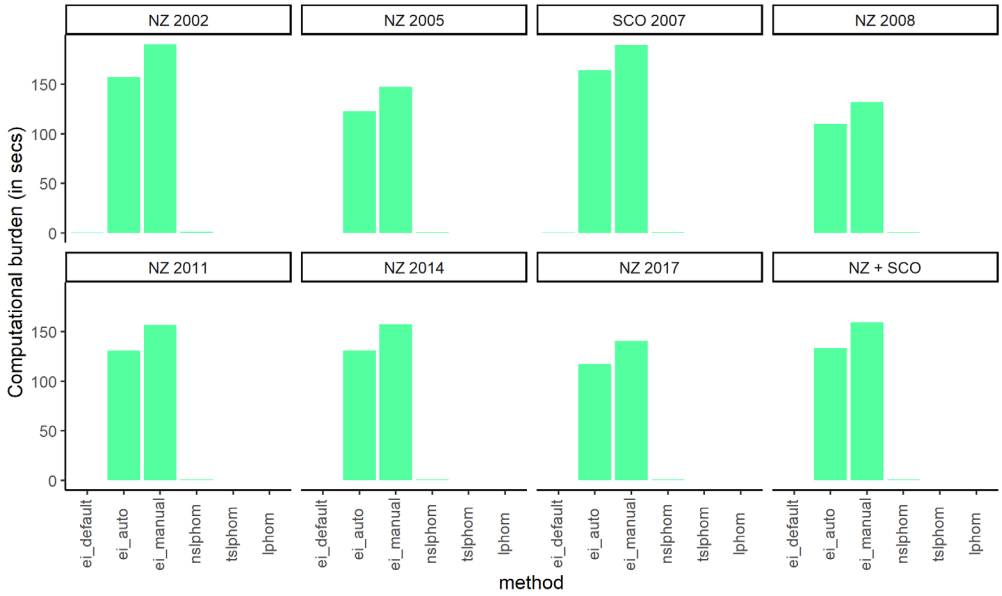

Figure S3. Graphical representation of average computation times grouped by election and algorithm in in the scenarios attained after randomly merging polling units as described in Section 3, after applying the functions `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

S3. Supplementary material. Subsection 5.2

Table S2. Summary of the performance of the algorithms in scenarios with few cells.

Country Year	NZ 2002	NZ 2005	SCO 2007	NZ 2008	NZ 2011	NZ 2014	NZ 2017	NZ + SCO
# of Elections	N = 69	N = 69	N = 73	N = 70	N = 70	N = 71	N = 71	N = 493
Avg. # of units	$\overline{I}$ = 83.2	$\overline{I}$ = 81.8	$\overline{I}$ = 70.2	$\overline{I}$ = 84.1	$\overline{I}$ = 85.7	$\overline{I}$ = 81.2	$\overline{I}$ = 101.9	$\overline{I}$ = 84.0
Avg. # of cells	$\overline{RC}$ = 7.6	$\overline{RC}$ = 8.8	$\overline{RC}$ = 9.7	$\overline{RC}$ = 8.6	$\overline{RC}$ = 7.7	$\overline{RC}$ = 7.6	$\overline{RC}$ = 8.6	$\overline{RC}$ = 8.4
Averages of *EI* measures								
ei_default	11.89	18.84	21.60	19.85	15.04	16.02	18.40	17.41
ei_auto	15.36	23.12	26.70	24.12	20.13	20.16	23.20	21.87
ei_manual	8.57	9.73	7.01	9.56	7.05	7.36	8.34	8.22
nslphom	7.58	9.11	6.93	9.47	7.30	7.72	8.21	8.04
tslphom	9.25	10.34	8.82	11.36	9.02	9.32	9.75	9.69
lphom	10.10	11.28	9.82	12.76	10.09	10.20	10.74	10.71
Averages of *WPE* measures								
ei_default	9.77	14.88	18.14	15.98	12.39	13.35	15.04	14.25
ei_auto	12.90	18.33	22.80	19.82	16.96	17.22	19.37	18.24
ei_manual	6.88	7.87	5.72	7.70	5.63	5.92	6.85	6.64
nslphom	6.19	7.12	5.41	7.54	5.91	6.19	6.33	6.38
tslphom	7.57	8.37	7.25	9.29	7.30	7.47	7.86	7.87
lphom	8.26	9.14	8.08	10.44	8.18	8.18	8.67	8.70
Averages of computational burden (in secs)								
ei_default	0.31	0.40	0.41	0.41	0.37	0.33	0.47	0.39
ei_auto	135.70	175.55	169.84	181.63	162.63	147.71	206.62	171.36
ei_manual	162.75	210.59	226.42	217.91	194.99	177.02	247.83	205.59
nslphom	3.90	4.84	3.45	4.95	4.76	4.38	6.78	4.74
tslphom	0.93	0.84	0.46	0.98	1.01	0.89	1.01	0.85
lphom	0.49	0.59	0.05	0.87	0.85	0.65	0.95	0.64

Source: compiled by the authors after applying the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) with different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) to the data from the New Zealand electoral commission and the Scotland Electoral Office attained after merging in Others the election options not surpassing 20% of the vote. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

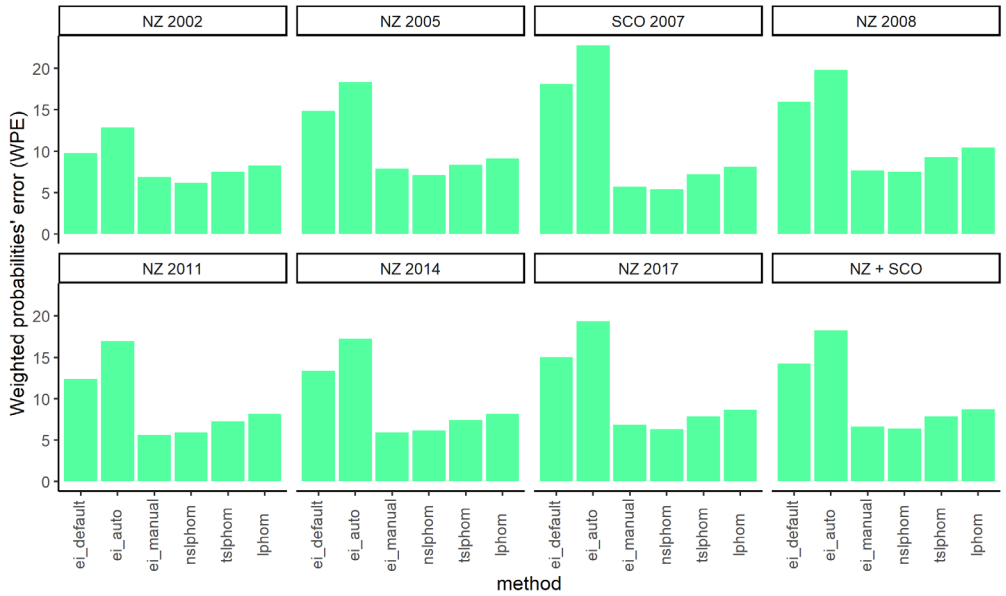

Figure S4. Graphical representation of average values of EPW error measures grouped by election and algorithm in the scenarios attained after merging in Others the election options not surpassing 20% of vote. Individual solutions have been attained with the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 1000, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument.

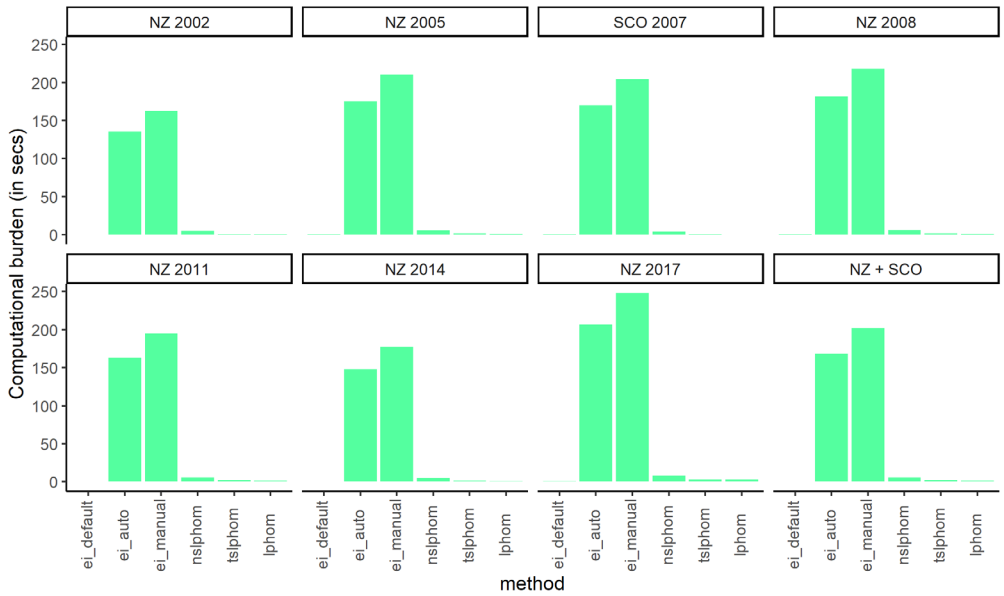

Figure S5. Graphical representation of average computation times grouped by election and algorithm in in the scenarios attained after merging in Others the election options not surpassing 20% of vote, after applying the functions `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

S3. Supplementary material. Subsection 5.3

Table S3. Summary of the performance of the algorithms in scenarios with few units and cells.

Country Year	NZ 2002	NZ 2005	SCO 2007	NZ 2008	NZ 2011	NZ 2014	NZ 2017	NZ + SCO
# of Elections	N = 69	N = 69	N = 73	N = 70	N = 70	N = 71	N = 71	N = 493
Avg. # of units	$\overline{I}$ = 14.6	$\overline{I}$ = 15.6	$\overline{I}$ = 14.4	$\overline{I}$ = 15.4	$\overline{I}$ = 15.2	$\overline{I}$ = 15.2	$\overline{I}$ = 15.6	$\overline{I}$ = 15.2
Avg. # of cells	$\overline{RC}$ = 7.6	$\overline{RC}$ = 8.8	$\overline{RC}$ = 9.7	$\overline{RC}$ = 8.6	$\overline{RC}$ = 7.7	$\overline{RC}$ = 7.6	$\overline{RC}$ = 8.6	$\overline{RC}$ = 8.4
Averages of *EI* measures								
ei_default	26.97	29.64	39.95	29.69	28.99	29.83	31.63	31.02
ei_auto	29.99	31.22	40.64	29.60	29.51	33.30	31.01	32.24
ei_manual	20.80	14.12	35.02	14.16	14.90	15.64	12.79	18.29
nslphom	10.57	11.09	9.12	11.46	9.39	9.77	10.33	10.24
tslphom	11.16	11.86	9.97	12.38	10.11	10.53	11.14	11.01
lphom	11.55	12.44	10.36	13.13	10.65	11.08	11.64	11.54
Averages of *WPE* measures								
ei_default	22.90	25.01	35.24	25.34	25.19	25.51	26.99	26.66
ei_auto	25.56	26.44	35.95	25.33	25.65	28.89	26.65	27.84
ei_manual	17.66	11.30	30.88	11.43	12.48	12.92	10.43	15.38
nslphom	8.64	8.78	6.83	9.15	7.48	7.76	8.10	8.10
tslphom	9.11	9.42	7.48	9.98	8.07	8.35	8.84	8.74
lphom	9.40	9.84	7.77	10.58	8.50	8.78	9.22	9.14
Averages of computational burden (in secs)								
ei_default	0.07	0.09	0.10	0.09	0.08	0.07	0.09	0.08
ei_auto	34.92	41.80	46.85	40.05	35.19	34.55	40.92	39.23
ei_manual	41.87	50.15	56.21	48.06	42.23	41.44	49.10	47.06
nslphom	0.66	0.81	0.38	0.77	0.79	0.77	0.79	0.71
tslphom	0.10	0.09	0.05	0.10	0.09	0.09	0.09	0.09
lphom	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Source: compiled by the authors after applying the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) with different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021) to the data from the New Zealand electoral commission and the Scotland Electoral Office attained after merging in Others the election options not surpassing 20% of the vote and randomly merging polling units as described in Section 3. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

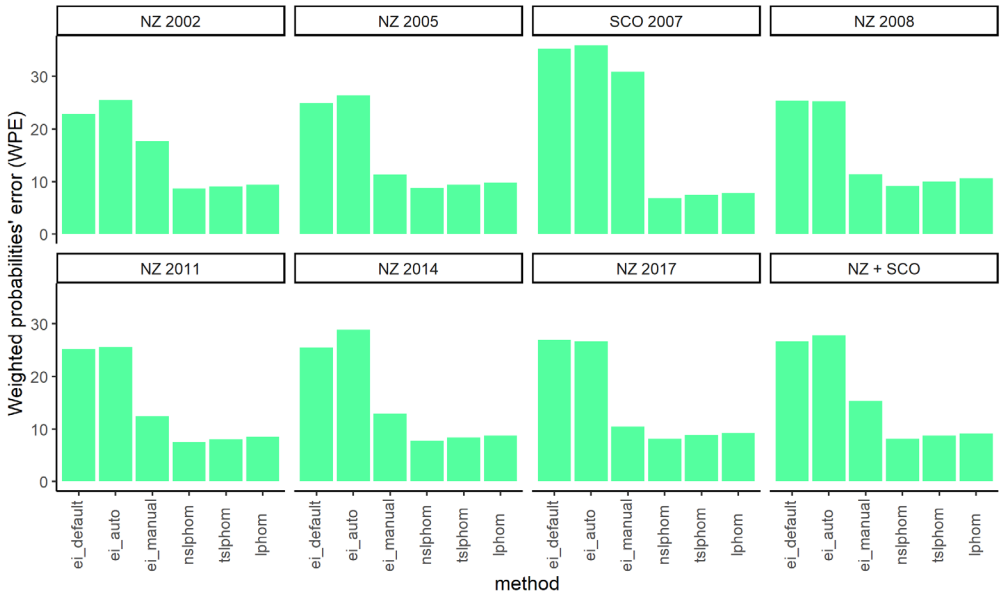

Figure S6. Graphical representation of average values of EPW error measures grouped by election and algorithm in the scenarios attained after merging in Others the election options not surpassing 20% of vote and randomly merging polling units as described in Section 3. Individual solutions have been attained with the function `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 1000, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument.

Figure S7. Graphical representation of average computation times grouped by election and algorithm in the scenarios attained after merging in Others the election options not surpassing 20% of vote and randomly merging polling units as described in Section 3, after applying the functions `ei.MD.bayes` of the R package `eiPack` (Olivia et al., 2020) using three different specifications and the functions `lphom`, `tslphom` and `nslphom` of the R package `lphom` (Pavía and Romero, 2021b) with default options. The outcomes labelled as `ei_default`, `ei_auto` and `ei_manual` have been attained after using `ei.MD.bayes` with, respectively, (i) default options, (ii) the output of the function `tuneMD` (with default options) as tune.list argument and default options for the rest of its arguments and (iii) sample = 1000, thin = 100, burnin = 100000 and the output of function `tuneMD` with ntunes = 10 and totaldraws = 100000 as tune.list argument. The computations have been performed on a desktop computer with a CPU processor Intel® Core™ i7-4930K (6 cores) 3.40GHz and 32GB of RAM.

S5. Supplementary material. Subsection 6.1

Table S4. Statistical summary of the distributions of EI and WPE errors.

	EI		WPE	
	ei_manual	nslphom	ei_manual	nslphom
Number of observations	1972	1972	1972	1972
Mean	14.60	10.21	11.32	7.26
Median	10.34	9.58	7.46	6.78
Standard deviation	10.80	4.61	9.52	3.41
Minimum	0.27	0.85	0.27	0.65
Maximum	55.64	48.45	49.03	31.35

Source: compiled by the authors.

Table S5. Statistical summary of the explicative variables.

	I	HET	JK	JKratio	Chi2	var.Part	var.Cand	std.Part	std.Cand
Number of observations	1972	1972	1972	1972	1972	1972	1972	1972	1972
Mean	49.447	3.377	17.612	1.091	1674.3	0.3976	0.3546	15.033	19.023
Standard deviation	71.073	2.491	11.365	0.243	3541.57	0.4611	0.3849	5.902	6.183
Minimum	10.0	0.74	4.0	0.5	3.51	0.0013	0.0018	0.02	0.97
Maximum	705.0	23.19	56.0	2.0	24074.5	2.7707	2.476	38.65	36.52

Source: compiled by the authors.

Table S6. Correlation matrix of the explicative variables.

	I	HET	JK	JKratio	Chi2	var.Part	var.Cand	std.Part	std.Cand
I		0.5111	0.0497	-0.0042	-0.0512	0.4582	0.4527	-0.0126	0.0082
HET	0.5111		0.0815	-0.1401	-0.1039	0.3298	0.4070	-0.0388	-0.0063
JK	0.0497	0.0815		0.3379	-0.4455	0.5746	0.5556	0.2786	0.1942
JKratio	-0.0042	-0.1401	0.3379		-0.1850	0.3577	0.0327	0.2203	0.2830
Chi2	-0.0512	-0.1039	-0.4455	-0.1850		-0.2636	-0.2514	-0.1191	-0.0374
var.Part	0.4582	0.3298	0.5746	0.3577	-0.2636		0.8618	0.2386	0.2248
var.Cand	0.4527	0.4070	0.5556	0.0327	-0.2514	0.8618		0.2237	0.1601
std.Part	-0.0126	-0.0388	0.2786	0.2203	-0.1191	0.2386	0.2237		0.6250
std.Cand	0.0082	-0.0063	0.1942	0.2830	-0.0374	0.2248	0.1601	0.6250	

Source: compiled by the authors.

Table S7. Estimated model for EI errors attained with `nslphom`.

	Response variable: EI		Algorithm: `nslphom`	
Variable	Estimate	Std. error	t-value	p-value
Constant	10.2110	0.07868	129.775	< 0.0001
I	-1.2926	0.10366	-12.470	< 0.0001
JK	1.5147	0.30700	4.934	< 0.0001
JKratio	0.7677	0.33451	2.295	0.0218
HET	2.4321	0.09557	25.448	< 0.0001
Chi2	-0.7034	0.11338	-6.204	< 0.0001
var.Part	-0.4160	0.21321	-1.951	0.0511
var.Cand	-1.8088	0.20716	-8.731	< 0.0001
std.Part	0.4046	0.10427	3.880	< 0.0001
std.Cand	-0.6001	0.10342	-5.803	< 0.0001
Adjusted R^2 (%)	42.48	Standard residual error		3.49

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S8. Estimated model for EI errors attained with `ei_manual`.

	Response variable: EI		Algorithm: `ei_manual`	
Variable	Estimate	Std. error	t-value	p-value
Constant	14.5998	0.2021	72.25	< 0.0001
I	-1.9574	0.2662	-7.353	< 0.0001
JK	2.9153	0.7884	3.698	0.0002
JKratio	0.6548	0.8591	0.762	0.4460
HET	0.9372	0.2454	3.818	0.0001
Chi2	-1.2736	0.2912	-4.374	< 0.0001
var.Part	-3.0448	0.5476	-5.561	< 0.0001
var.Cand	-1.1346	0.532	-2.133	0.0330
std.Part	-1.813	0.2678	-6.771	< 0.0001
std.Cand	-2.5661	0.2656	-9.661	< 0.0001
Adjusted R^2 (%)	30.97	Standard residual error		8.97

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S9. Estimated model for the difference of EI errors.

	Response variable: differences of EI			
Variable	Estimate	Std. error	t-value	p-value
Constant	4.3888	0.2093	20.973	< 0.0001
I	-0.6648	0.2757	-2.411	0.0160
JK	1.4006	0.8165	1.715	0.0864
JKratio	-0.1129	0.8896	-0.127	0.8991
HET	-1.4949	0.2542	-5.882	< 0.0001
Chi2	-0.5702	0.3015	-1.891	0.0588
var.Part	-2.6288	0.5670	-4.636	< 0.0001
var.Cand	0.6742	0.5509	1.224	0.2212
std.Part	-2.2176	0.2773	-7.997	< 0.0001
std.Cand	-1.9660	0.2750	-7.148	< 0.0001
Adjusted R^2 (%)	21.44	Standard residual error		9.29

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S10. Estimated model for WPE errors attained with `nslphom`.

	Response variable: WPE		Algorithm: `nslphom`	
Variable	Estimate	Std. error	t-value	p-value
Constant	7.2631	0.0650	111.667	< 0.0001
I	-0.9977	0.0857	-11.643	< 0.0001
JK	-0.6129	0.2538	-2.415	0.0158
JKratio	1.5018	0.2765	5.431	< 0.0001
HET	1.7783	0.0790	22.510	< 0.0001
Chi2	-0.1495	0.0937	-1.595	0.1108
var.Part	-0.2221	0.1763	-1.260	0.2078
var.Cand	-1.2098	0.1713	-7.064	< 0.0001
std.Part	0.2740	0.0862	3.178	0.0015
std.Cand	-0.3760	0.0855	-4.399	< 0.0001
Adjusted R^2 (%)	28.44	Standard residual error		2.88

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S11. Estimated model for WPE errors attained with `ei_manual`.

	Response variable: WPE		Algorithm: `ei_manual`	
Variable	Estimate	Std. error	t-value	p-value
Constant	11.3220	0.1838	61.591	< 0.0001
I	-1.5012	0.2422	-6.199	< 0.0001
JK	-0.0235	0.7172	-0.033	0.9739
JKratio	2.0344	0.7815	2.603	0.0093
HET	0.1746	0.2233	0.782	0.4342
Chi2	-0.6517	0.2649	-2.460	0.01400
var.Part	-2.6846	0.4981	-5.389	< 0.0001
var.Cand	-0.2001	0.4840	-0.413	0.6794
std.Part	-1.8953	0.2436	-7.781	< 0.0001
std.Cand	-2.1921	0.2416	-9.072	< 0.0001
Adjusted R^2 (%)	26.52	Standard residual error		8.16

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S12. Estimated model for the difference of WPE errors.

	Response variable: differences of WPE			
Variable	Estimate	Std. error	t-value	p-value
Constant	4.0589	0.1898	21.390	< 0.0001
I	-0.5035	0.2500	-2.014	0.0442
JK	0.5894	0.7404	0.796	0.4261
JKratio	0.5326	0.8067	0.660	0.5092
HET	-1.6037	0.2305	-6.958	< 0.0001
Chi2	-0.5022	0.2734	-1.837	0.0664
var.Part	-2.4625	0.5142	-4.789	< 0.0001
var.Cand	1.0097	0.4996	2.021	0.0434
std.Part	-2.1693	0.2515	-8.627	< 0.0001
std.Cand	-1.8160	0.2494	-7.281	< 0.0001
Adjusted R^2 (%)	21.65	Standard residual error		8.43

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S13. Estimated model, including quadratic terms, for EI errors attained with `nslphom`.

	Response variable: EI		Algorithm: `nslphom`	
Variable	Estimate	Std. error	t-value	p-value
Constant	9.1444	0.1819	50.260	< 0.0001
I	-2.5800	0.2240	-11.516	< 0.0001
JK	-0.7991	0.5281	-1.513	0.1304
JKratio	1.9110	0.4605	4.150	< 0.0001
HET	2.6114	0.2180	11.980	< 0.0001
Chi2	-1.8055	0.2182	-8.273	< 0.0001
var.Part	-1.4430	0.3801	-3.796	0.0002
var.Cand	-1.0536	0.3836	-2.747	0.0061
std.Part	0.5580	0.1086	5.137	< 0.0001
std.Cand	-0.7311	0.1131	-6.465	< 0.0001
I^2	0.1613	0.0289	5.592	< 0.0001
JK^2	0.3622	0.1487	2.436	0.0149
$JKratio^2$	0.1513	0.1541	0.982	0.3262
HET^2	-0.0191	0.0432	-0.443	0.6579
$Chi2^2$	0.2983	0.0584	5.112	< 0.0001
$var.Part^2$	0.3062	0.0848	3.610	0.0003
$var.Cand^2$	0.1198	0.0841	1.425	0.1543
$std.Part^2$	-0.1483	0.0623	-2.381	0.0174
$std.Cand^2$	-0.1645	0.0645	-2.549	0.0109
Adjusted R^2 (%)	46.75	Standard residual error		3.36

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S14. Estimated model, including quadratic terms, for EI errors attained with `ei_manual`.

	Response variable: EI		Algorithm: `ei_manual`	
Variable	Estimate	Std. error	t-value	p-value
Constant	12.6149	0.4436	28.440	< 0.0001
I	-3.6161	0.5462	-6.621	< 0.0001
JK	7.3892	1.2875	5.739	< 0.0001
JKratio	-2.6013	1.1226	-2.317	0.0206
HET	-2.6809	0.5314	-5.045	< 0.0001
Chi2	-1.5151	0.5320	-2.848	0.0045
var.Part	-4.3150	0.9268	-4.656	< 0.0001
var.Cand	-1.0651	0.9352	-1.139	0.2549
std.Part	-2.4747	0.2649	-9.344	< 0.0001
std.Cand	-1.9347	0.2757	-7.018	< 0.0001
I^2	0.3310	0.0703	4.706	< 0.0001
JK^2	-1.4664	0.3625	-4.046	< 0.0001
$JKratio^2$	-0.3420	0.3757	-0.910	0.3627
HET^2	0.8446	0.1053	8.025	< 0.0001
$Chi2^2$	-0.0214	0.1423	-0.150	0.8806
$var.Part^2$	1.1351	0.2068	5.489	< 0.0001
$var.Cand^2$	0.6584	0.2050	3.213	0.0013
$std.Part^2$	-0.1092	0.1519	-0.719	0.4722
$std.Cand^2$	0.9558	0.1573	6.077	< 0.0001
Adjusted R^2 (%)	42.41	Standard residual error		8.20

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

Table S15. Estimated model, including quadratic terms, for the difference of EI errors.

	Response variable: differences of EI			
Variable	Estimate	Std. error	t-value	p-value
Constant	3.4705	0.4681	7.414	< 0.0001
I	-1.0362	0.5764	-1.798	0.0724
JK	8.1883	1.3586	6.027	< 0.0001
JKratio	-4.5123	1.1846	-3.809	0.0001
HET	-5.2923	0.5608	-9.437	< 0.0001
Chi2	0.2903	0.5614	0.517	0.6051
var.Part	-2.8720	0.9780	-2.937	0.0034
var.Cand	-0.0115	0.9869	-0.012	0.9907
std.Part	-3.0327	0.2795	-10.851	< 0.0001
std.Cand	-1.2036	0.2909	-4.137	< 0.0001
I^2	0.1697	0.0742	2.286	0.0224
JK^2	-1.8285	0.3825	-4.781	< 0.0001
$JKratio^2$	-0.4933	0.3964	-1.244	0.2135
HET^2	0.8637	0.1111	7.777	< 0.0001
$Chi2^2$	-0.3197	0.1501	-2.129	0.0333
$var.Part^2$	0.8289	0.2182	3.799	0.0002
$var.Cand^2$	0.5386	0.2163	2.490	0.0128
$std.Part^2$	0.0391	0.1603	0.244	0.8073
$std.Cand^2$	1.1203	0.1660	6.750	< 0.0001
Adjusted R^2 (%)	31.94	Standard residual error		8.65

Source: compiled by the authors. All the explicative variables have been standardised before fitting the model to make easier the comparisons of coefficients.

S6. Supplementary material. Subsection 6.2

Table S16. Estimated model for absolute value errors (×100) attained with `nslphom` estimating p_{jk}.

	Response variable: AVE		Algorithm: `nslphom`	
Variable	Estimate	Std. error	t-value	p-value
Constant	0.4912	0.10044	4.89055	> 0.0000
$p_{jk} \times 100$	0.7425	0.00838	88.6334	> 0.0000
$p_{jk}^2 \times 100$	-0.0080	0.00010	-78.3135	> 0.0000

Source: compiled by the authors.

Table S17. Estimated model for absolute value errors (×100) attained with `ei_manual` estimating p_{jk}.

	Response variable: AVE		Algorithm: `ei_manual`	
Variable	Estimate	Std. error	t-value	p-value
Constant	4.2505	0.12285	34.5993	> 0.0000
$p_{jk} \times 100$	0.1884	0.01025	18.3907	> 0.0000
$p_{jk}^2 \times 100$	-0.0003	0.00012	-2.27001	0.0232

Source: Compiled by the authors.

S7. Supplementary material. Subsection 6.3

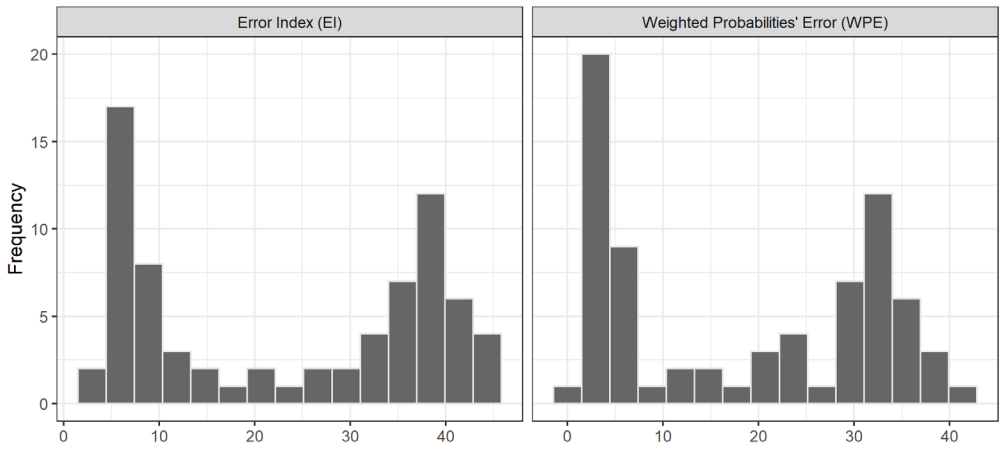

Figure S8. Histograms of the distributions of the error measures (EI left panel and WPE right panel) associated to the solutions attained after applying `ei.MD.bayes` using the `ei_manual` specification and `nslphom` to the in the 73 data of Scotland included in the reference database.