Online Appendix

1 Estimation

This appendix presents the formalization of the *Metropolis-within-Gibbs-Sampler* used to fit the *Hybrid Multinomial-Dirichlet-Model* described in subsection Model Assumptions and Structure.

(I) Choose number of iterations by defining Burn-In, Sample-Size and Thinning

(II) Choose starting values $(\beta_i^{rc}, \alpha_{rc})^{(0)}$ with $(i = 1, \dots P)$, $(r = 1, \dots R)$ and $(c = 1, \dots C)$

(III) Start of algorithm with Iteration $j = 1$:

(i) Successively simulate $(\beta_i^{rc})^{(j)}$ with Full Conditional f_β for β_i^{rc}:

$f_\beta(\beta_i^{rc}) \propto (\beta_i^{rc})^{Z_i^{rc}} \times (\Theta_{c,i})^{(T_{c,i} - Z_{c,i})} \times (\beta_i^{rc})^{\alpha_{rc}-1}$

(a) Draw $(\beta_i^{rc})^*$ from $h_\beta \sim N\Big((\beta_i^{rc})^{(j-1)}, \sigma^2_{\beta_i^{rc}}\Big)$

(b) Calculate $\gamma = min\left\{1, \dfrac{f_\beta\Big((\beta_i^{rc})^*\Big)}{f_\beta\Big((\beta_i^{rc})^{(j-1)}\Big)}\right\}$

(c) Set $(\beta_i^{rc})^{(j)} = (\beta_i^{rc})^*$ with probability γ

(ii) Successively simulate $(\alpha_{rc})^{(j)}$ with Full Conditional f_α of α_{rc}:

$f_\alpha(\alpha_{rc}) \propto \dfrac{\Gamma(\sum_{c'=1}^{C} \alpha_{rc'})}{\Gamma(\alpha_{rc})} \times \prod_{i=1}^{P} (\beta_i^{rc})^{(\alpha_{rc}-1)} \times \alpha_{rc}^{\lambda_1^{rc}-1} \times exp(-\lambda_2^{rc}\alpha_{rc})$

(a) Draw $(\alpha_{rc})^*$ from $h_\alpha \sim N\Big((\alpha_{rc})^{(j-1)}, \sigma^2_{\alpha_{rc}}\Big)$

(b) Calculate $\gamma = min\left\{1, \dfrac{f_\alpha\Big((\alpha_{rc})^*\Big)}{f_\alpha\Big((\alpha_{rc})^{(j-1)}\Big)}\right\}$

(c) Set $(\alpha_{rc})^{(j)} = (\alpha_{rc})^*$ with probability γ

(iii) $(\beta_i^{rc}, \alpha_{rc})^{(j)}$ will be saved with following conditions:

- $j >$ Burn-In
- Rest of $\dfrac{j}{\text{Thinning}} = 0$

(IV) Result is Markov-chain $\left((\beta_i^{rc}, \alpha_{rc})^{(1)}, \ldots, (\beta_i^{rc}, \alpha_{rc})^{(S)}\right)$

2 First Simulation Setup and Computational Details

The population of each district i was simulated using a Poisson distribution (equation (15)). Results of election one in each district i were drawn with a Dirichlet distribution (equation (16)). We have chosen $q = 50$ to have slightly different but realistic election one results in each district.

$$N_i \sim Po(\lambda = 800) \tag{15}$$

$$\begin{aligned}(N_{CSU1,i}, \ldots, N_{NW1,i}) \sim \\ Dir(q \times (0.269, 0.244, 0.086, 0.092, 0.039, 0.034, 0.266))\end{aligned} \tag{16}$$

The voting transitions between election one and election two are based on a realistic example in the city of Munich. The expected voting probabilities displayed in Table 7 are assumed in every district in every data set. In the next step we drew row-wise voting probabilities via $(\beta_i^{rc}, \ldots, \beta_i^{rc}) \sim Dir(k \times (\alpha_{r1}, \ldots, \alpha_{rC}))$ with $(\alpha_{r1}, \ldots, \alpha_{rC})$ referring to one row in Table 7. For heterogeneous and homogeneous probabilities between districts, we chose $k_{het} = 5$ and $k_{hom} = 20$, respectively.

Table 7: Expected voting probabilities in simulation study

Election 1 \ Election 2	CSU	SPD	TheLeft	Greens	Other parties	Nonvoter	
CSU	.80	.02	.03	.03	.05	.07	N_{CSU_1}
SPD	.04	.79	.03	.03	.05	.06	N_{SPD_1}
TheLeft	.02	.09	.45	.01	.18	.25	$N_{TheLeft\_1}$
FDP	.37	.06	.01	.01	.52	.03	N_{FDP_1}
Greens	.02	.17	.03	.71	.04	.03	N_{Greens_1}
Other parties	.03	.06	.03	.01	.82	.05	N_{OTH_1}
Nonvoter	.03	.04	.03	.03	.04	.83	N_{NV_1}
	T_{CSU_2}	T_{SPD_2}	$T_{TheLeft\_2}$	T_{Greens_2}	T_{OTH_2}	T_{NV_2}	N

We fit the model with the *Hybrid Multinomial-Dirichlet-Model* as implemented in `eiwild`. After running tests with some data sets to examine convergence in trade-off with running time, we ran all data sets with Markov chain with 1 300 000 iterations. With a burn-in of 50 000 and thinning of 1 250, calculation of probabilities was done with a left-over sample of 1 000.

3 Simulation Study I: Additional Consideration

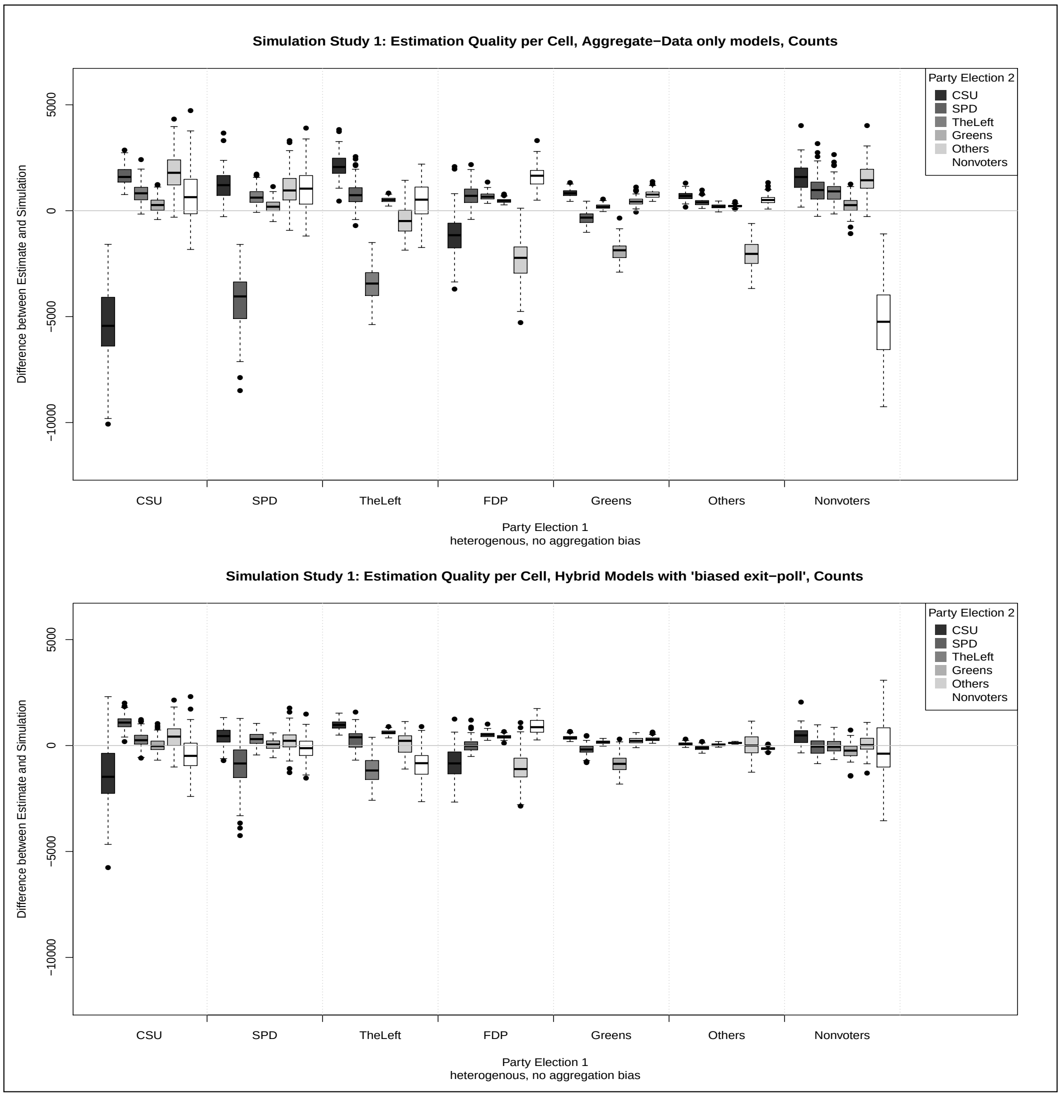

Figure S1: Results of the first part of the simulation study. Cell-specific distance from the true voting numbers for each party combination. y-axis displays the difference between the estimated number of voters and the value in the simulated individual level data for each of the 100 data sets and x-axis displays the considered voter transition. The upper figure shows the difference using the estimates from the aggregate data only model, the lower figure using the estimates from the hybrid model with biased exit-poll. Reading example: Using the aggregate data only model for the estimation, the median of the bias for the loyal CSU voters of all 100 simulated data sets is around -5500 voters.

4 Second Simulation Setup and Computational Details

The population in each district i was simulated using a Poisson distribution (equation (17)). Results of election one in each district i were results of the German Federal Election (GFE) 2009 in the city of Munich and were drawn with a Dirichlet distribution (equation (18)). We have chosen $q = 50$ to have slightly different but realistic election one results in each district. Before we assumed voting transitions and calculated election two results, we split the population in three sub-populations: voters who behave loyally (1), are more willing to change parties (2) and some middle-ground between (1) and (2). The sub-population distribution in each district was drawn by a Dirichlet distribution, which assumes in expectation equal distribution (equation (19)).

$$N_i \sim Po(\lambda = 1500) \tag{17}$$

$$\begin{aligned}(N_{CSU1,i}, \dots, N_{NW1,i}) \sim \\ Dir(q \times (0.236, 0.142, 0.129, 0.128, 0.099, 0.266))\end{aligned} \tag{18}$$

$$(Pop1, Pop2, Pop3) \sim Dir(q \times (1/3, 1/3, 1/3)) \tag{19}$$

The expected voting transitions for 'normal' sub-population are displayed in Table 8. The diagonal of the other two sub-populations were slightly adjusted up or down.

Table 8: Expected voting probabilities of 'normal' sub-population in second simulation study

Election 1 \ Election 2	CSU	SPD	FDP	Greens	Other parties	Nonvoter	
CSU	.80	.02	.03	.03	.05	.07	N_{CSU_1}
SPD	.04	.79	.02	.03	.06	.06	N_{SPD_1}
FDP	.37	.06	.37	.01	.16	.03	N_{FDP_1}
Greens	.02	.18	.01	.71	.06	.02	N_{Greens_1}
Other parties	.03	.07	.01	.01	.77	.11	N_{OTH_1}
Nonvoter	.03	.05	.03	.03	.06	.80	N_{NV_1}
	T_{CSU_2}	T_{SPD_2}	T_{FDP_2}	T_{Greens_2}	T_{OTH_2}	T_{NV_2}	N

Similarly to the first simulation, we draw row-wise voting probabilities via $(\beta_i^{rc}, \ldots, \beta_i^{rc}) \sim Dir(k \times (\alpha_{r1}, \ldots, \alpha_{rC}))$ with $(\alpha_{r1}, \ldots, \alpha_{rC})$ referring to one row in Table 8. For heterogeneity, we chose k=5. After calculating each voter transition table, we combined the three sub-populations to get the voting transition table of one city. The column totals gave us the election two results.

As in the first simulation, we fit the model with the *Hybrid Multinomial-Dirichlet-Model* as implemented in `eiwild`. The Markov chain had 1 400 000 iterations with a burn-in of 150 000 and thinning of 1 250. The calculation of probabilities was done with a left-over sample of 1 000. We slightly adjusted the Burn-In to better fit the model with 600 districts.

5 Computational Details of Voter Transition Estimation between Two Elections in Munich

Computational details of the application to real data (section Voter Transitions Between Two Elections in Munich): We fit the '(Hybrid) Multinomial-Dirichlet-Model' with `eiwild_v0.6.4` (eiwild 2014) on `R_v3.0.2` (R 2008). The prior distribution was *Gamma*$(\lambda_1 = 4, \lambda = 2)$, which is the most uninformative prior without defining cell-specific prior distributions (subsection Introducing Prior Knowledge). Convergence of the Markov-Chains or rather the empiric measures was achieved with 2 700 000 iterations (Burn-in = 200 000, thinning = 1 250). We calculated the probabilities with the left-over sample-size of 2 000. The running-time of the model was approximately 21 hours with a 2-threaded 2.15Ghz CPU.