Final Project: When the Model Does What it is Supposed to Do but Not What You Intend

Gerald Smith, Elena Szefer,Xiong Yi, Biljana J. Stojkova
under the supervision of Dave Campbell

August 8, 2013

Introduction

The purpose of election polling is to attempt to determine the true voting intentions of the public. However, polling data provide an imperfect measure of these intentions: each poll is subject to random measurement error, design effects and systematic bias. Poll results can vary widely over the span of an election campaign, both within polling organisations, as well as between different organisations or 'houses' as they are commonly known. This wide variation in polls makes it difficult to determine the true voting intentions of the public, especially since the variation is more likely to be due to polling methodologies rather than true public opinion volatility. By pooling together estimates from different polls, the goal is to produce a more accurate estimate which dampens the impact of the random variation between polls, as well as the systematic effects created by the various methodologies employed by polling houses, which are commonly termed 'house effects' (Jackman, 2005).

Why should we care? Despite the way the media covers election polling, they are more than just entertainment. Information from polls can have an impact on how the public actually votes, especially when it comes to so called 'strategic voting' (Pickup and Johnston, 2007). For example, in the days preceding the 2011 Canadian federal election a clear consensus emerged among pollsters, forecasters, and columnists that the Conservative would not win a majority of seats in the House of Parliament (Rosenthal, 2011). However, the Conservatives did in fact get a majority by winning 166 of the 308 local ridings across the country. The failure of the polls to reflect this possibility could have changed the way the public voted on election day. Had voters known that the Conservatives were going to win a majority government (which would allow them to pass legislative bills without needing to get votes from opposition parties) they may have switched from voting for the third or fourth place party and instead voted for the second place party in ridings where the second place party had a chance to overtake the Conservative party.

The methodological decisions made by houses will produce a systematic bias if these decisions produce a consistent error in their estimate of public voting intentions. Different terminology used in the vote intention question is one cause of this consistent error. A poll that asks respondents in an open-ended way what party they will vote for can produce a different response than a poll that asks voters to choose between a list of parties. The Canadian polling company Nanos

Research uses an open-ended approach and one consequence of this is that Nanos consistently produces lower estimates of voter support for the Green Party than other polling companies (Espey et al., 2011).

The goal of this project is to identify house effects in polling data. We attempt several methods that all build upon techniques developed by Simon Jackman (2005, 2009). We use both simulated data, as well as real data released by polling firms during the 2011 Canadian federal election.

Model and Methods

Statistical Model

In all of our methods we employ a version of the state-space model as described in Jackman (2005, 2009). This model serves three purposes: (1) pooling the polls to increase precision; (2) estimating and adjusting for the bias of any one poll; (3) tracking the trends and fluctuations in voter sentiment over the course of the election campaign. For the purposes of this project, we are interested mostly in estimating the bias of polls. Explaining the notation of the model is beneficial. Let α_t be the intended vote share for an individual party at time t, with t indexing weeks, where t=1 is the first week of the election campaign. Let i = *1,...,n* index the polls available for analysis. Each poll result is assumed to be generated as follows:

$$y_i \sim N(\mu_i, \sigma_i^2),$$

where y_i is the result of poll i. Each of the n polls is generated by organisation j_i on date t_i. σ_i is the standard error of the poll and

$$\mu_i = \alpha_{t_i} + \delta_{j_i},$$

where δ_j is the bias of polling organisation *j*, an unknown parameter to be estimated. To model change in vote intentions, we use a simple random-walk model:

$$\alpha_t \sim N(\alpha_{t-1}, \omega^2), t = 2, ..., T$$

where ω is the standard deviation of the daily changes in voting intentions (we will formulate a prior for this below) and the distribution

$$\alpha_1 \sim Uniform(0.4, 0.6)$$

initialises the random walk (that is before we see any polls, it is assumed the party's support is between 40% and 60%, we use this in our simulated data case, we change this for our real data case where we have more than two parties).

This model is basically a Kalman filter. The first two equations above define the measurement or observational part of the model, relating the observed poll estimate to the latent target α_t, while the final two equations specify the way in which the hidden target fluxuates over time. For the simulated data, the unkown model parameters are:(1) the 33 α_t parameters (one poll from each of the 3 houses polled together each week over the 33 weekes, see the conditional distribution section beow for this special case with multiple polls at the same time period), representing the weekly changes of intended support for the first party (the second party support is simply found by subtracting the first party support from 1 since we are dealing with proportions); (2) the three house effect parameters δj specific to each polling company; and (3) ω, the standard deviation of the normal distribution that defines week-to-week volatility in the α_t. The real data case pertains to the 2011 Canadian federal election, the polling data starts on March 28, 2011 when the election was called and is broken into 5 weeks until the final election day on May 2, 2011. There are 5 $alpha_t$ parameters and 11 different house effect paramaters in the real data.

It should be noted that we also attempted to use the *filter-foward, sample backwards* (FFSB) algrithm as described in Jackman (2009). However, we had the same issue as with our original model where the alpha values grew unreasonably fast.

Prior Distributions

For the house-effects parameter δ_j we use a vague normal prior:

$$\delta_j \sim N(0, d^2),$$

with d an arbitrarily large constant. A δ_j with a 95% confidence interval between -.15 and .15 would constitute a range that includes some very large house-effects, and therefore would be a vague prior. This corresponds to a $d^2 = (.15/2)^2 = 0.005625$. Hence, a vague normal prior can be specified with any d > 0.075.

For ω, we use a uniform prior,

$$f(\omega) = \begin{Bmatrix} 100 & if\ 0 < \omega < .01 \\ 0 & otherwise \end{Bmatrix}$$

That is we assume that the day-to-day changes are not large, with the standard deviation of the daily changes being no larger than 1%. At this maximum value of $\omega = .01$, the implication is that 95% of the daily changes in α_t are no larger than plus or minus 2%. The idea being that day-to-day changes in voter sentiment will not change significantly.

Conditional Distributions for the Gibbs Sampler

It is important to briefly state the full conditional distributions we use for the Gibbs sampler since these are what gave us the most trouble and led to the model not working as intended. The Directed Acyclic Graph (DAG) is presented to explain the model and the dependence structure of all the variables in the model.

1. The conditional distribution for α is a special case since our index is weeks and so we have multiple polls in every period. Letting s index weeks, define $P_s = \left[i : t_i = s\right]$ as the subset of polls with date s, Then α_s is a normal distribtion with mean

$$\left[\left(\sum_{i\epsilon P_s} \frac{y_i - \delta_i}{\sigma_i^2}\right) + \frac{\alpha_{s-1} + \alpha_{s+1}}{\omega^2}\right] * \left[\frac{1}{\sum_{i\epsilon P_s} \sigma_i^2} + \frac{2}{\omega^2}\right]^{-1}$$

and variance

$$\left[\frac{1}{\sum_{i\epsilon P_s} \sigma_i^2} + \frac{2}{\omega^2}\right]^{-1}.$$

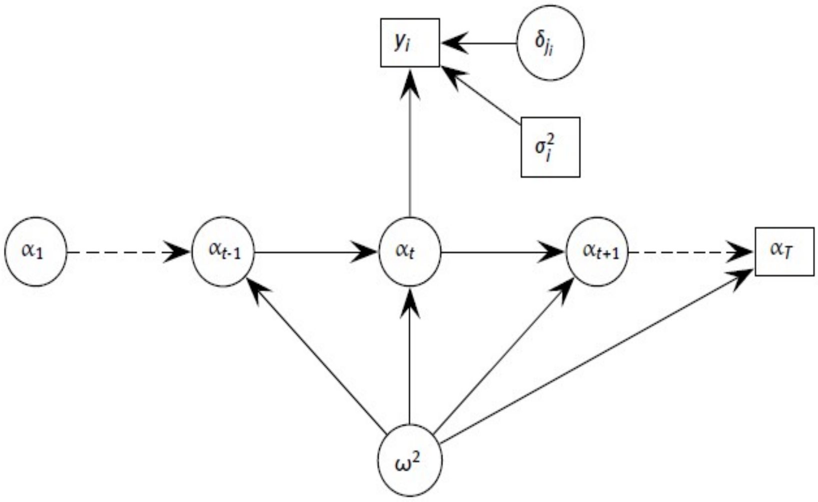


Figure 1: DAG for the Polling Data Model

2. Letting k index polling organisations, define $P_k = \left[i : j_i = k\right]$ as the set of polls published by polling organisation k The conditional distribution for δ_k is normal with mean

$$\left[\sum_{i \epsilon P_k} \frac{y_i - \alpha_{t_i}}{\sigma_i^2}\right] * \left[\frac{1}{\sum_{i \epsilon P_k} \sigma_i^2} + \frac{1}{d^2}\right]^{-1}$$

and variance

$$\left[\frac{1}{\sum_{i \epsilon P_k} \sigma_i^2} + \frac{1}{d^2}\right]^{-1}.$$

3. For ω we work with $\tau = g(\omega) = \omega^{-2}$. The uniform prior over ω implies that $\omega^2 < 0.0001$ and in turn, $\tau > 10000$. Also, the uniform prior on ω implies the following prior on τ:

$$f(\tau) = \frac{1}{2}\tau^{-3/2}p(\tau),$$

where $p(\tau) = 100$ if $\tau > 10000$ and 0 otherwise, giving us

$$f(\tau) \propto \tau^{-3/2}, \tau > 10000.$$

Then the conditional distribution of τ is a Gamma distribution with parameters (T-4)/2 and $\frac{1}{2}\sum t = 2^T(\alpha_t - \alpha_{t-1})^2$ along with an indicator for $\tau > 10000$. We can then transform the sampled τ back into $\omega = \tau^{-1/2}$ (see Jackman appendix for more details on derivations, 2005b).

JAGS

JAGS, also known as "Just Another Gibbs Sampler", is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation. JAGS works efficiently but it only outputs the resulting samples without giving any information about the posterior distributions. JAGS is a blackbox style software and this motivated us to develop our own Gibbs sampler to validate the results from JAGS.

Our Modified Sampler

The standard Gibbs sampling algorithm is displayed below in figure 2. In this standard algorithm,α are depending on the previous state and the next state. They are updated at each iteration. We found out in this case,the $\alpha'\mathbf{s}$ are increasing without bound. So instead of using this method, we developed an adaptive Gibbs sampler by anchoring the next state α_{t+1} to always be sampled from the first iteration. This is similar to anchoring to the actual election result and prevented our α's from growing unreasonably large. The modified algorithm is displayed in figure 3.

- Choose the initial value of $\boldsymbol{\alpha}^0, \boldsymbol{\delta}^0, \mathbf{w}^0$
- Sample $\alpha_t^{(i)}$ from $[\alpha_t^{(i)} \mid \alpha_{t-1}^{(i)}, \alpha_{t+1}^{(i-1)}, y, w^{(i-1)}, \boldsymbol{\delta}^{(i-1)}]$
- After α_1 ,α_2,.....α_{33} all have been updated, start sample $\boldsymbol{\delta}, \mathbf{w}$
- Sample w^i from [$w^i \mid \mathrm{y}, \boldsymbol{\alpha}^{\mathrm{i}}$]
- Sample $\boldsymbol{\delta}^{\mathbf{i}}$ from [$\boldsymbol{\delta}^i \mid \mathrm{y}, \boldsymbol{\alpha}^{\mathrm{i}}, w^i$]

Figure 2: Standard Gibbs Sampling Algorithm

- Choose the initial value of $\boldsymbol{\alpha}^0, \boldsymbol{\delta}^0, \mathbf{w}^0$
- Sample $\alpha_t^{(i)}$ from $[\alpha_t^{(i)} \mid \alpha_{t-1}^{(i)}, \alpha_{t+1}^{(1)}, y, w^{(i-1)}, \boldsymbol{\delta}^{(i-1)}]$
- After α_1 ,α_2,.....α_{33} all have been updated, start sample $\boldsymbol{\delta}, \mathbf{w}$
- Sample w^i from [$w^i \mid \mathrm{y}, \boldsymbol{\alpha}^{\mathrm{i}}$]
- Sample $\boldsymbol{\delta}^{\mathbf{i}}$ from [$\boldsymbol{\delta}^i \mid \mathrm{y}, \boldsymbol{\alpha}^{\mathrm{i}}, w^i$]

Figure 3: Anchored Gibbs Sampling Algorithm

Simulated Data

We simulate polling data from 33 weekly surveys conducted by 3 organizations (giving 99 total polls), to keep things simple we assumed a political system of 2 parties. It is also assumed that each organization has a house effect when presenting the polling results. The house effects are 0.02, -0.04 and 0.005 for House 1, House 2 and House 3 respectively. A house effect greater than

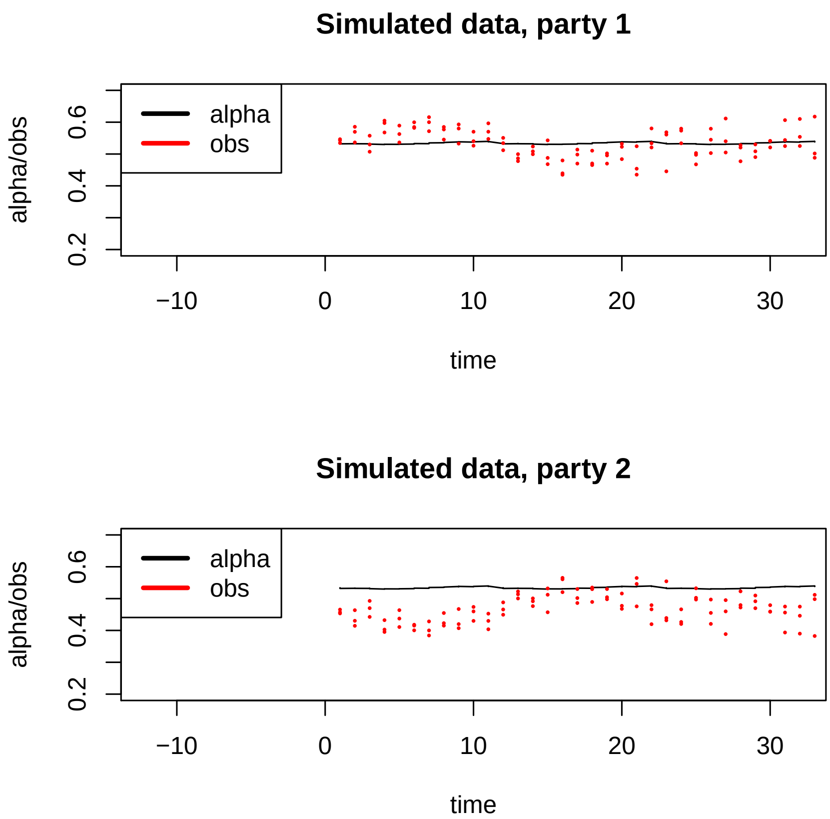


Figure 4: Plots of our simulated data for two parties

0 means a polling organisation is over-estimating the true vote intentions for that party and underestimating the other party. The standard error for each poll, σ_i, was drawn from a normal distribution with mean 0.04 and standard deviation 0.0001. Finally, ω was set to be 0.001.

Computational Comparison Study Using Simulated Data

Since JAGS is just a black box, we do not know how does it actually work. It is written in C++ and consequently is very efficient. However, for our particular problem, accuracy of results from JAGS depend on our model assumptions. In order for JAGS to run for our problem, we supply state space model and prior distributions for all parameters and JAGS do the rest. It looks in all connections between parameters and ultimately derive the posteriors in order to estimate parameters. For the simulated data that we were using for testing, it seems like JAGS does not estimate exact house effects that we plugged in our simulated data.

So we thought we could develop our own sampler that will detect the bias in election polls more accurately and more stable than JAGS. This model is developed according to the Simon Jackman's paper ("Appendix to Accompany 'Pooling the Polls Over an Election Campaign' Australian Journal of Political Science, 2005, V40(4): 499 - 517").

Our data are simulated such that they follow the state space model presented in the paper.

Original Gibbs sampler that finds house effects from our simulated data according to the Jackman's model, presented in the previous section does not give reasonable results at all. .R Code files of Our original sampler are in the folder "original model".

We conducted extensive debugging and diagnostic checks on Our sampler. However, nothing seems to be quite reasonable. Through our debugging we found out that Our sampler is sensitive to our well behaved simulated data. Namely, Our sampler depends heavily on the marginal error of our observations as well as on variance of the prior distribution of house effects.

So we found that due to the big variance of the prior distribution of the house effects, Our sampler cannot detect the small bias that we plugged in our data.

Based on this findings we modified Our sampler. So instead of using d=0.08 (variance of the house effects prior effects), we used much smaller variance 0.0005. And we also set up priors for each house effects. That is we modified Our sampler to work under the assumption that we know the results in advance.

We also found that state parameters estimated from Our sampler were highly correlated and non-converging. So we tried many methods such as forward filter backward sample method, and Jackman's method which anchored the alphas to an election result. These methods that we tried are discussed more in depth in the Insights section. After trying different methods, our findings suggest that without anchoring alphas to the election results, they grow without bound.

We run Our modified sampler and compare it with JAGS. Both samplers run for 100 000 iterations. Then we conduct diagnostic checks for the parameters estimated from both samplers: Our modified sampler and JAGS sampler.

.R Code files of Our modified sampler are in the folder "anchored alphas - diagnostic".

Efficiency

In order for JAGS state parameters to converge it has to run for a long time. Namely,the paper that we used says that in order for state parameters to converge they needed to run JAGS sampler for $25 * 10^6$ iterations.We tried JAGS sampler for 10^6 and it finishes in less than three hours. Our sampler was running for 15 hours and it could not produce 10^6 samples. Therefore we did not develop more efficient sampler than JAGS sampler. This was expected as JAGS is written in C++. For the given time frame for this project, we could not try to rewrite Our sampler in C++, so we are leaving this for the future research. So we reduced Samples to 100 000 for our comparison study.

Accuracy

Diagnostic checks reveal that state parameters from both samplers hardly achieve convergence. But Our modified sampler shows much better convergence of the state parameters, as well as very low auto-correlation when compared to those obtained from Our original sampler.

Not achieving the convergence from both samplers could be due to the small sample size.

House Effects from both samplers converge.

As regards accuracy of the estimated house effects, it turns out that JAGS sampler does not estimate the exact bias that we plugged in our data. It underestimates house effects.

Namely, it turns out that JAGS is also sensitive to our simulated data. For the testing purposes, we changed marginal error of the observations and under the assumption that house effects from both parties add up to zero, JAGS underestimates our bias more heavily.

However, this setup of the simulated data is quite good for Our modified sampler. So Our modified sampler returns much closer results to the bias that we plugged in our simulated data, for both parties. Our sampler also estimates approximately correctly the bias of the second party, which is negative of the bias of the first party. Alphas of the second party are 1-corresponding alpha of the first party, which is also correct.

Diagnostic Tests

Our main result are House Effects. Our simulated data consist of results for election polls conducted by 3 houses in 33 weeks time period. Only two political parties are considered. Each of the samplers returns set of chains for 3 house effects, 33 chains for our state parameters and one chain for the state parameters variance, for each of the two parties. We run Raftery-Lewis and Geweke diagnostic for house effects and state chains from both samplers for each party.

More details about running JAGS and Our sampler and diagnostic

The main file where both samplers are run and all the diagnostic is performed is $FinalProject\_MainCode.R$

We perform diagnostics on saved results from the previous run that was conducted for 10^5 iterations.

Next we run diagnostic for JAGS and Our modified sampler. We also run Raftery-Lewis diagnostic and Gewake for house effects and alphas.

	M(JAGS,partyI)	I(JAGS,partyI)	M(OurS.,partyI)	I(OurS.,partyI)
H.E.1	1.00	1.00	1.00	0.98
H.E.2	2.00	0.95	1.00	1.01
H.E.3	1.00	1.01	2.00	1.05

Table 1: Raftery Lewis diagnostic results for Party I from JAGS and Our modified sampler

	M(JAGS,partyII)	I(JAGS,partyII)	M(OurS.,partyII)	I(OurS.,partyII)
H.E.1	1.00	1.00	2.00	0.97
H.E.2	2.00	0.96	1.00	1.01
H.E.3	1.00	1.01	1.00	1.01

Table 2: Raftery Lewis diagnostic results for Party II, from JAGS and Our modified sampler

Gewake, trace and ACF plots for the last 3 alphas:

ACF plots of the last 3 alphas from both samplers, for the first party:

ACF plots of the last 3 alphas from both samplers, for the first party:

Diagnostic summary

Raftery-Lewis results tables show that House Effects from all three houses estimated from both samplers do not have significant amount of burning iterations. Moreover, the inflation factor is close to one in all cases, which means that all House Effects chains from both samplers converge. Autocorrelation functions for house effects show that there is no significant autocorrelation in all of the three house effects chains. Also, autocorrelation functions for the last 3 alphas exhibit no autocorrelation in alpha chains. This was not the case with alpha chains obtained from our original sampler. The reason that we create significant amount of autocorrelation in alpha chains

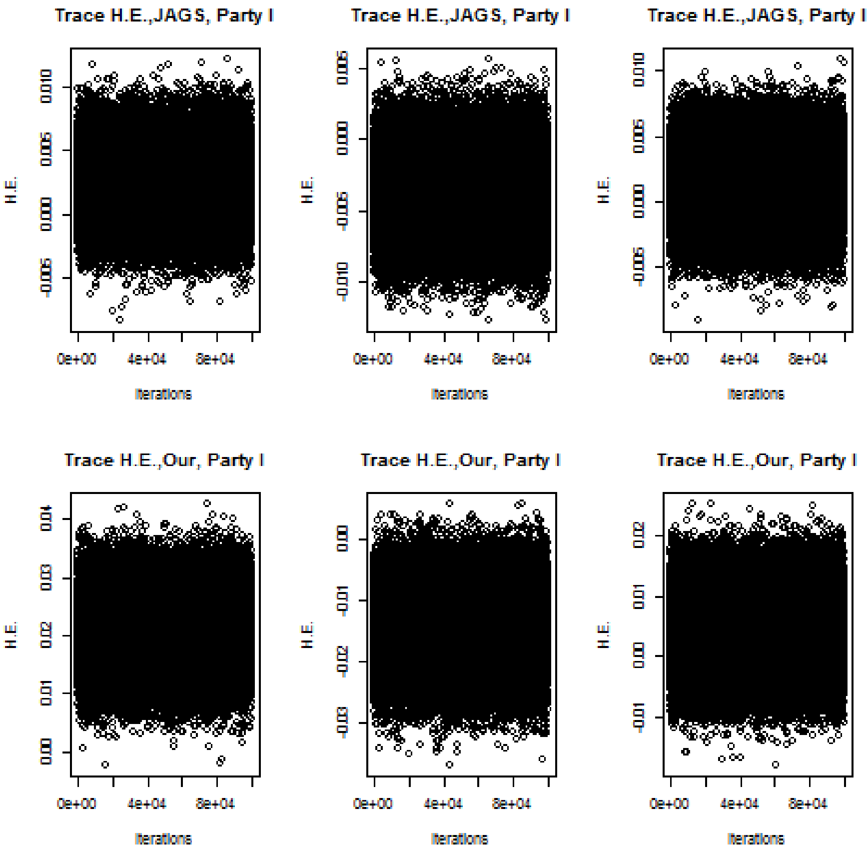


Figure 5: Trace plots for all three house effects - JAGS and Our sampler, first party

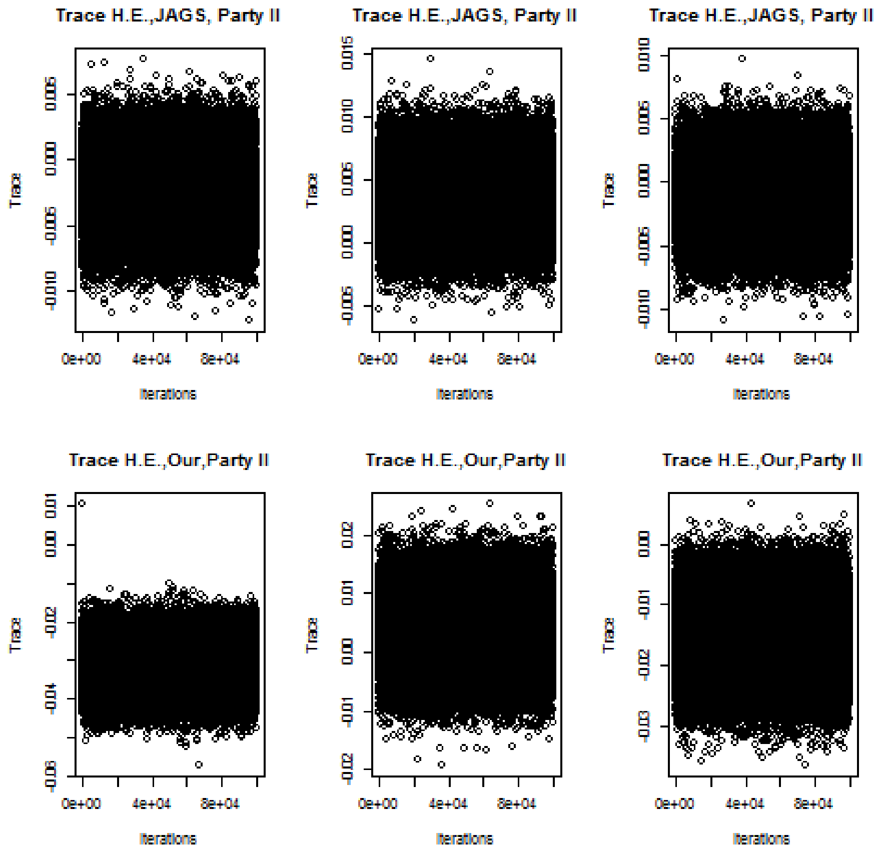


Figure 6: Trace plots for all three house effects - JAGS and Our sampler, second party

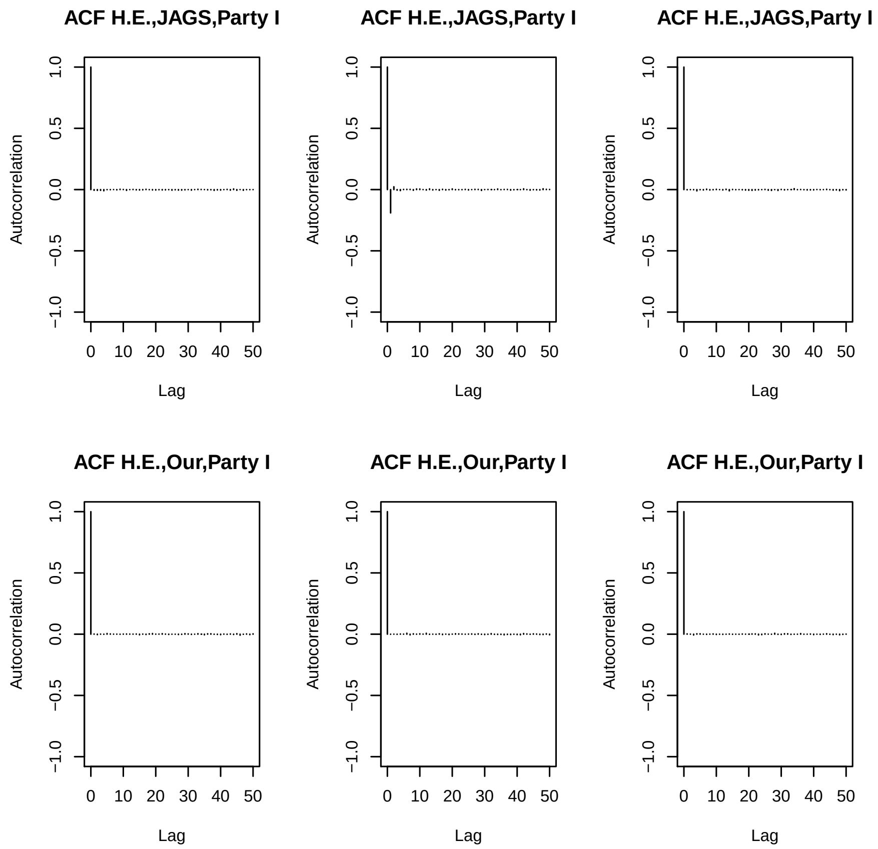


Figure 7: ACF plots for all three house effects - JAGS and Our sampler first party

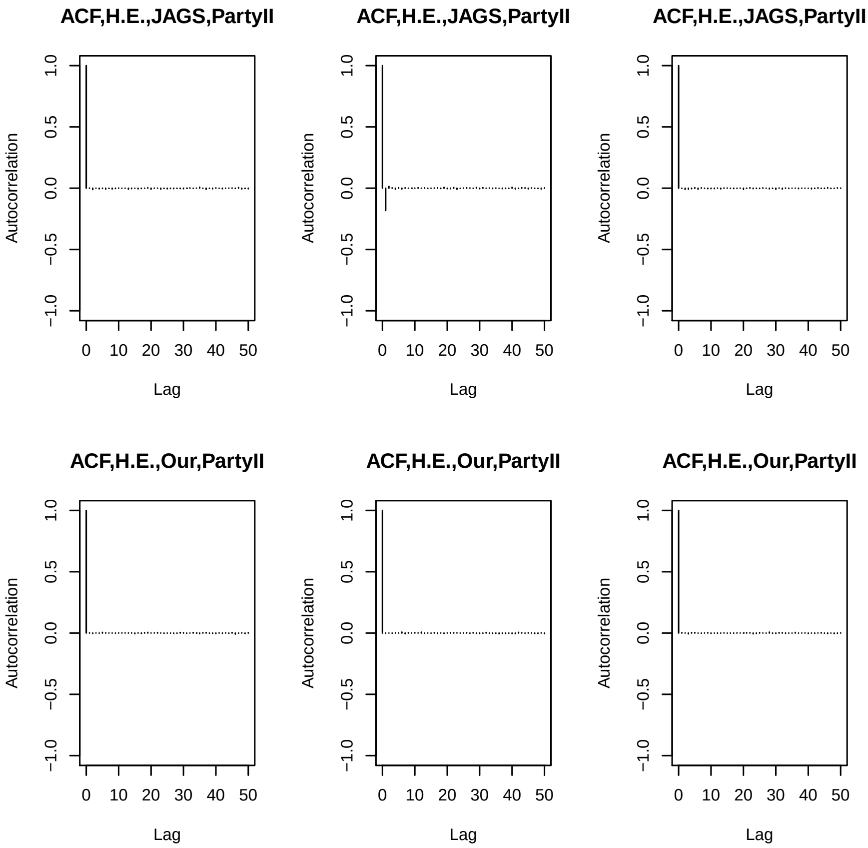


Figure 8: ACF plots for all three house effects - JAGS and Our sampler second party

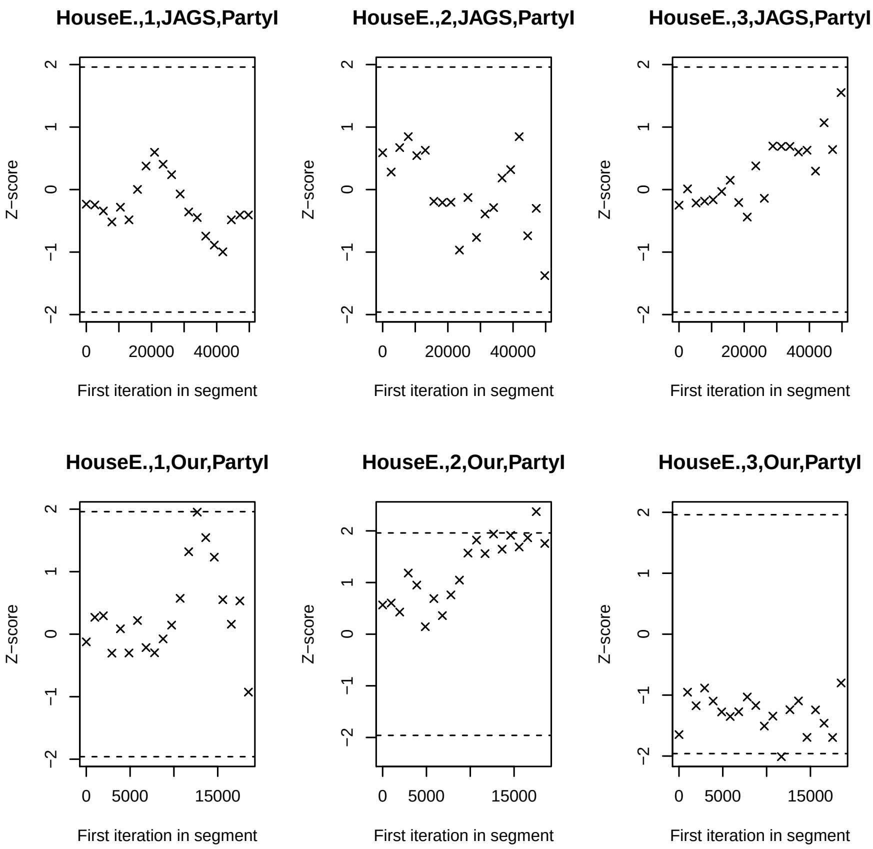


Figure 9: Gewake plots for all three house effects - JAGS and Our sampler first party

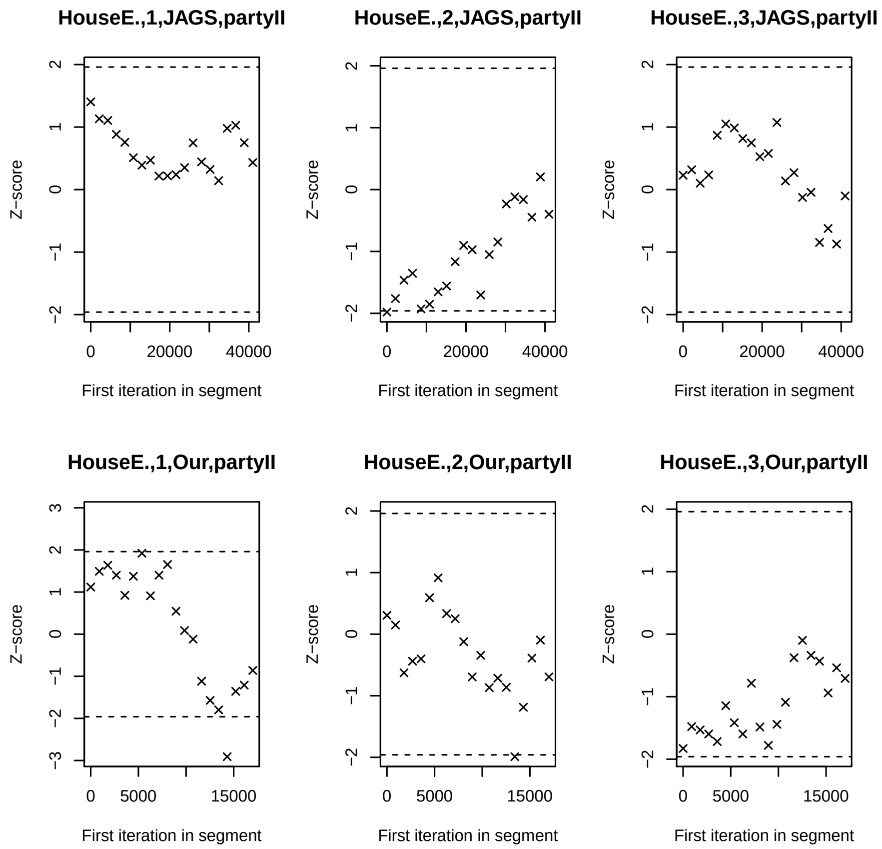


Figure 10: Gewake plots for all three house effects - JAGS and Our sampler second party

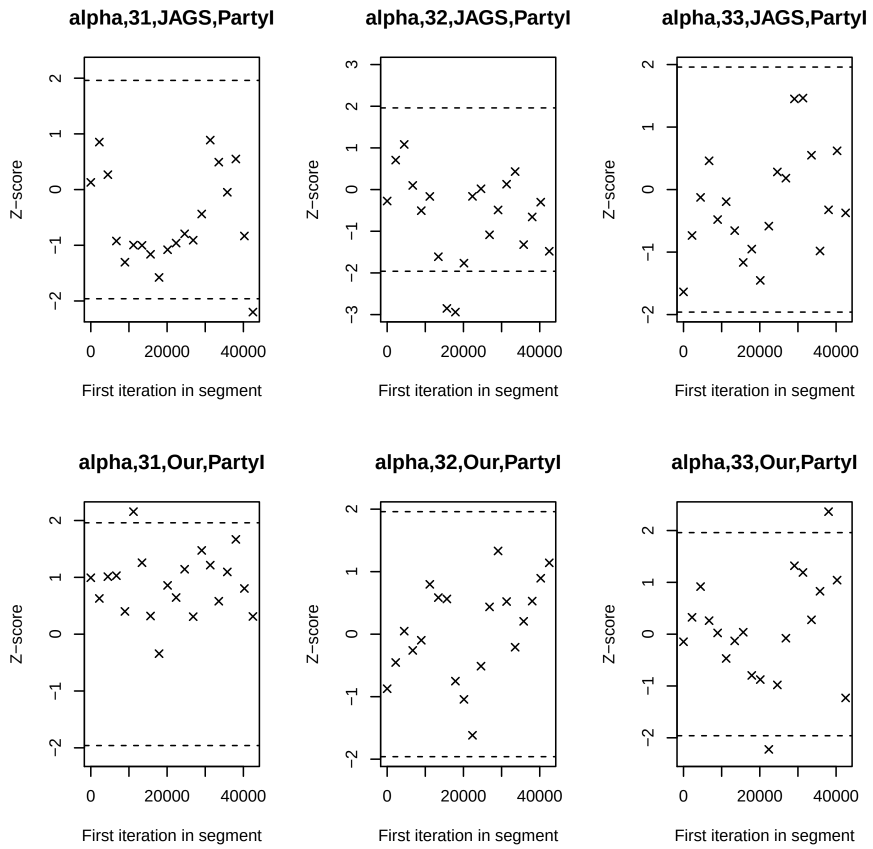


Figure 11: Gewake plots for the last 3 alphas - JAGS and Our sampler, first party

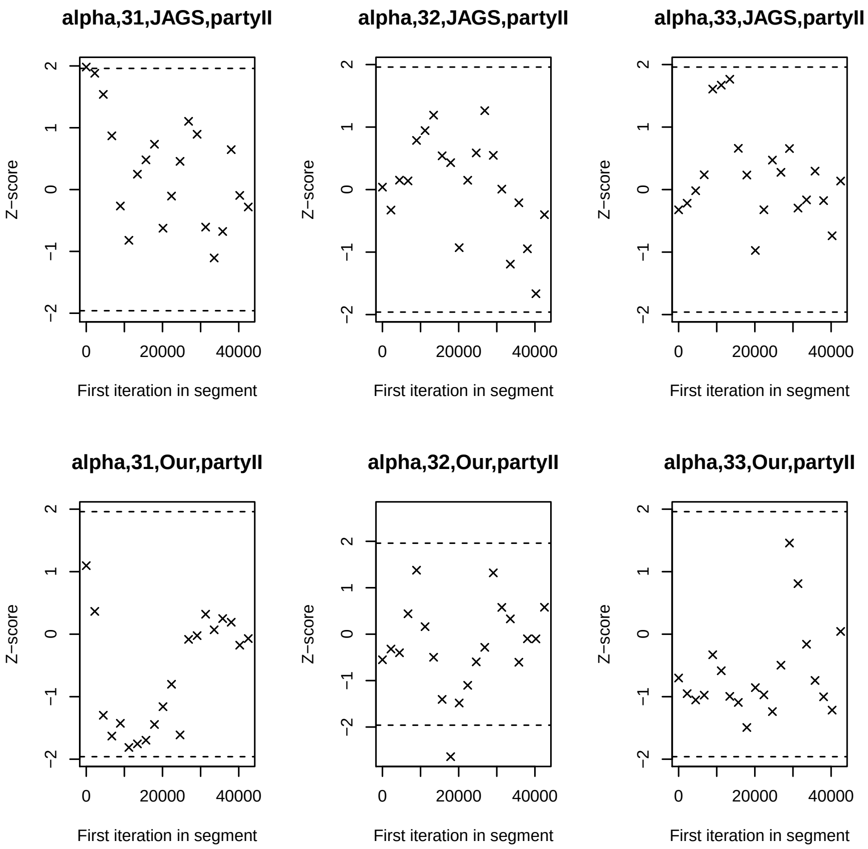


Figure 12: Gewake plots for the last 3 alphas - JAGS and Our sampler, second party

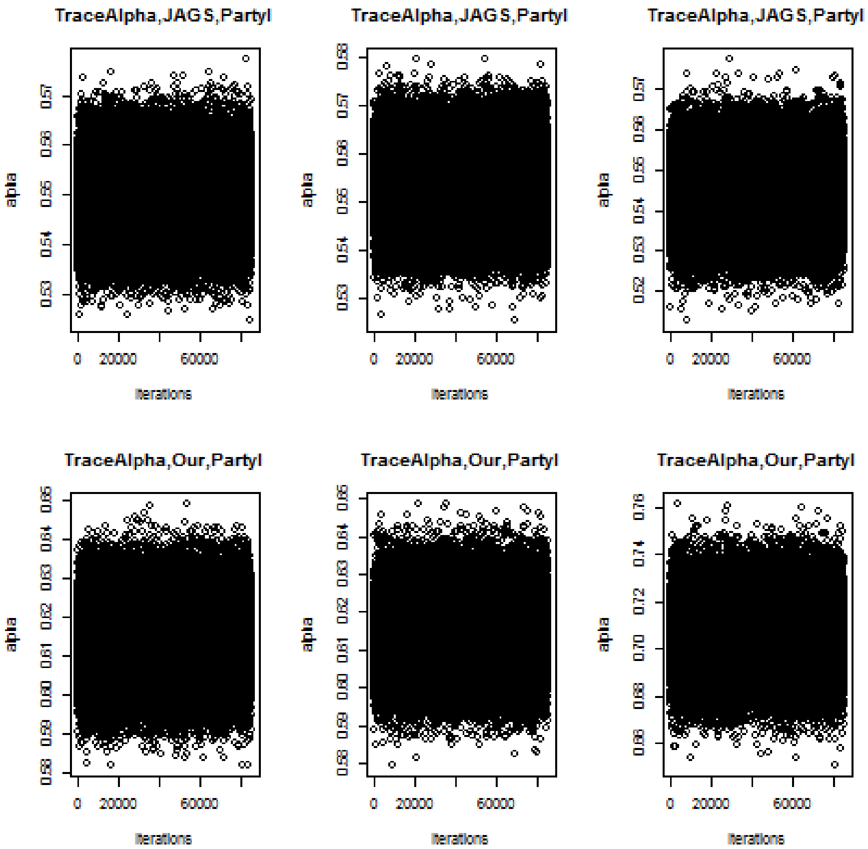


Figure 13: Trimmed trace plots for the last 3 alphas - JAGS and Our sampler, first party

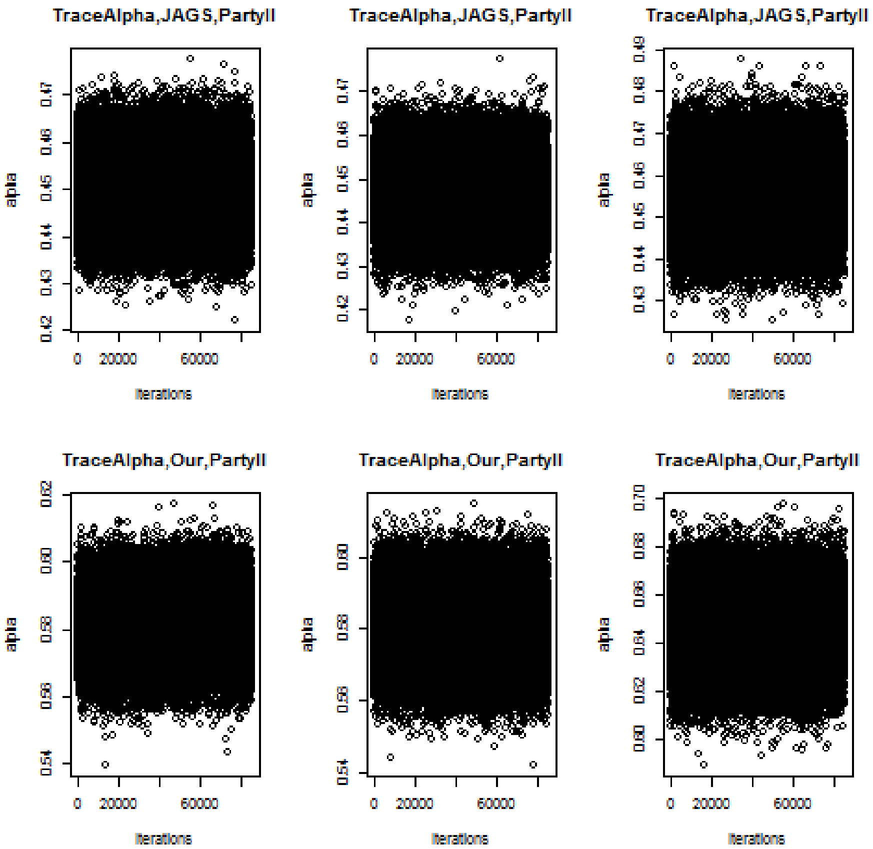


Figure 14: Trimmed trace plots for the last 3 alphas - JAGS and Our sampler, second party

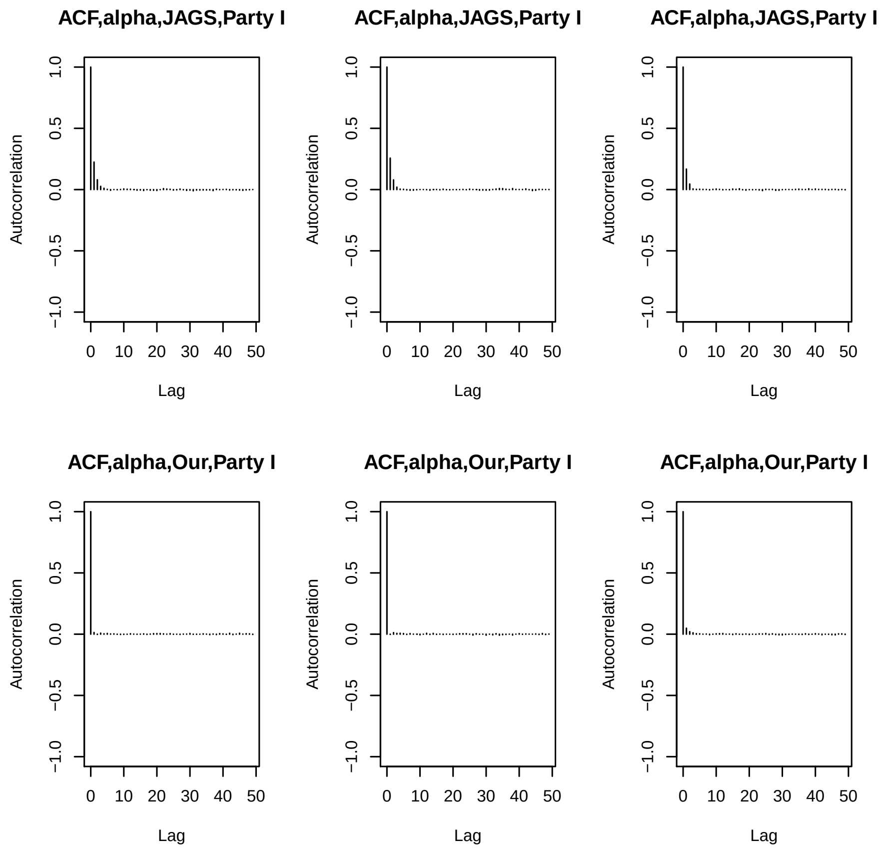


Figure 15: ACF plots for the last 3 alphas - JAGS and Our sampler, first party

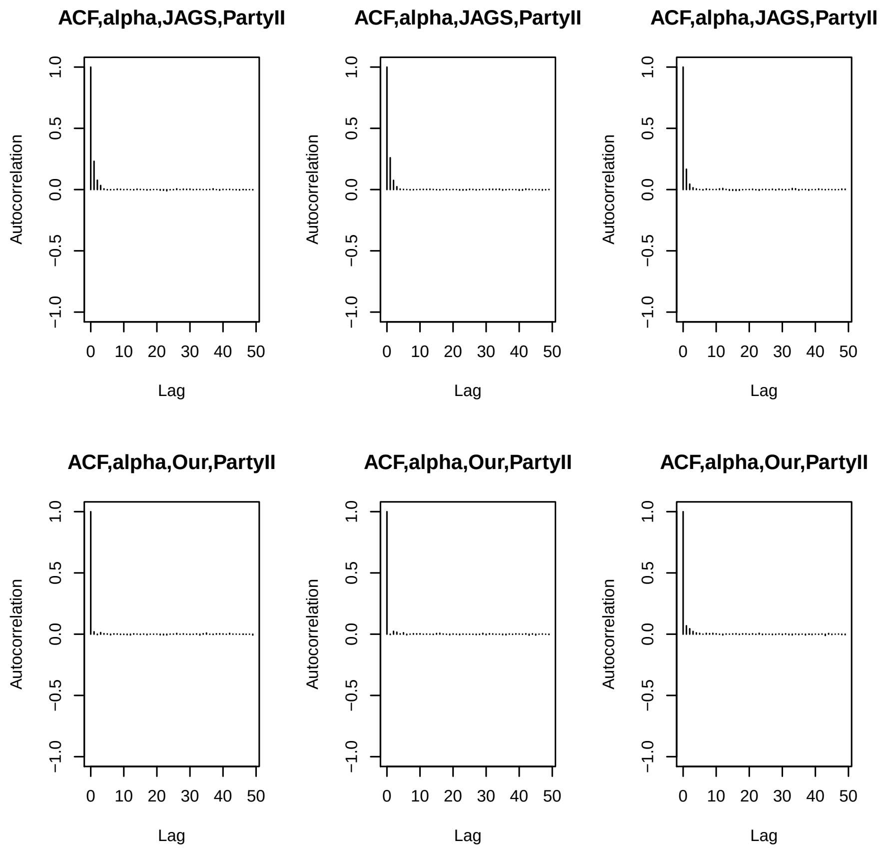


Figure 16: ACF plots for the last 3 alphas - JAGS and Our sampler, second party

is that we use corresponding value from the previous iteration within Gibbs as a proxy for the forward value of the alphas. However, Gewake plots for House effects showed that there is a big burning amount of iterations for some of the house effects, especially for those estimated from Our Sampler.

So we burned:

house effects from: JAGS I party 550 iterations

house effects from: JAGS II party 18000 iterations

house effects from: Our modified sampler I party 63000 iterations

house effects from: Our modified sampler II party 66000 iterations

We can see that numebr of burning iterations for house effects estimated from Our modified sampler are pretty large, and even after that huge burning, some of the three house effects are still non stationary.

Gewake plots for our state parameters are burned up to the first 15500 iterations, for both samplers, even that afterwards they exhibit non-convergence. However, house effects and alphas from both samplers look promising that if we had more samples they would converge.

Distributions of parameters

Histograms of trimmed House Effects

Histograms of house effects are organized in four rows (first two are from JAGS sampler and second two from Our modified sampler for each of the two parties) and three columns (one per each house effect).

From the histograms of the house effects we can see that JAGS does not detect bias correctly for each of the three houses and for both parties. Namely, our plugged bias for the first party is 0.02,-0.04 for the first and second house, respectively. So we expect, our samplers to determine these house effects correctly.

Both samplers return somewhat sensible house effects, but not the exact as we expected. It is worth mentioning here that JAGS managed to find correct bias in the cases when we had smaller marginal error in our simulated data.

As regards the bias for the second party, for instance, in our simulated data we plugged in positive bias for the first house for the first party. Then the bias for the second party should be equal to the negative bias of the first party for the corresponding house. This is due our model assumption that observations from the two parties should add up to one (each party's observation represent the percentage share) and consequently, the house effects should add up to zero. For the second house effect we plugged in negative bias for the first party, and so the second party should exhibit bias negative bias of the first party.

Both samplers determine the direction of the bias (positive or negative) for both parties correctly, but results for both parties are just somewhat close to 0.02,-0.04 for the first party, for the first two houses respectively. Having the expected results in mind, it is clear that for the first house effect, first party, JAGS underestimates bias whereas Our modified sampler gives results close to 0.02 as expected. However, for the second house, both samplers, JAGS and Our modified sampler, overestimate the pluged in bias, with Our modified sampler being closer to the expected result (Our modified sampler estimated results is -0.02 against the expected -0.04)

Descriptive statistics tables for the results estimated from both samplers

In this subsection we present summary statistics for the results, that are obtained from JAGS and Our modified sampler, for both parties for each house effect. These results are shown in the Table3,Table4, Table5 and Table6.

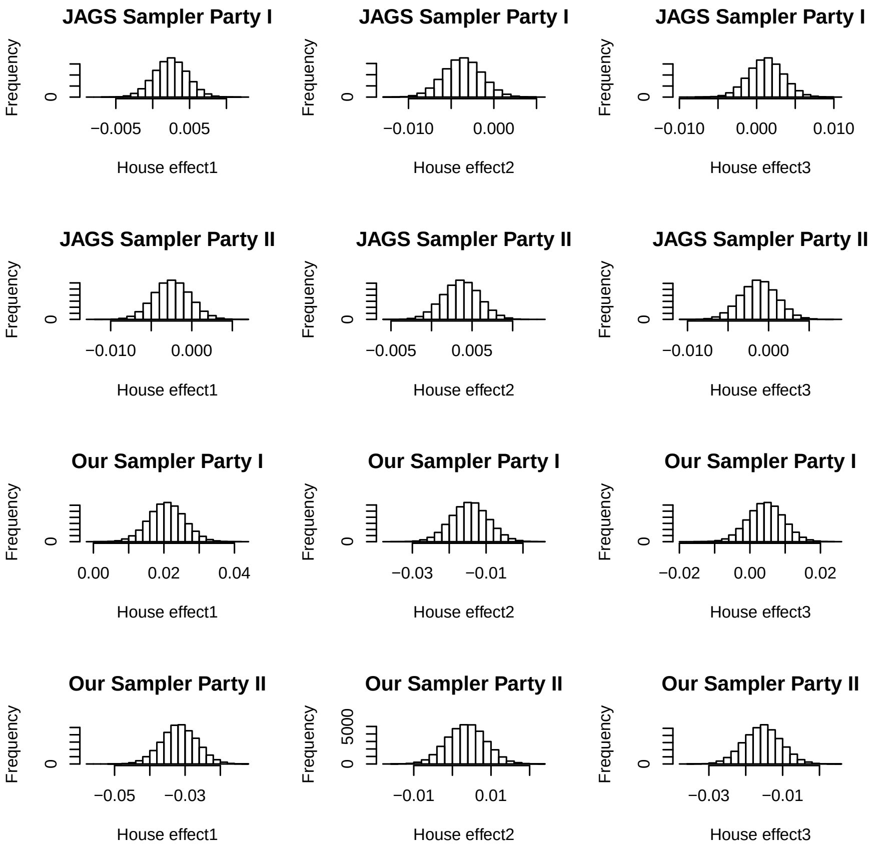


Figure 17: Histograms of all three house effects, from both samplers for two parties, trimmed chains

We can see from these tables more closely the mean, median, standard deviation and 95 % Highest posterior density region.

These results confirm the conclusions that we draw from the histograms of house effects. Namely, for the first party, first house, Our modified sampler determines correct bias that we plugged in our data, but it does not find accurately the bias of the second party for the first house, which should be negative of the bias for the first party. However, at least the direction (negative) of the second party is correct. This is not the case with JAGS sampler. In fact, for these particular simulated data, JAGS underestimates the bias for the first party, first house. But the second party from the corresponding house, the relationship between bias from the first and second party is satisfied: they add up to zero; and this is true for all the three house effects. So JAGS does at better than Our modified sampler with respect to the assumption that bias for two parties, within corresponding house add up to zero. Standard deviation for all the estimates is really small (0.002 for JAGS estimates and 0.005 for Our modified sampler estimates).

	House	mean	median	st. dev.	95%HPD R. Lower	95%HPD R. Upper
1	House Effect 1	0.002	0.002	0.002	-0.002	0.007
2	House Effect 2	-0.004	-0.004	0.002	-0.008	0.001
3	House Effect 3	0.001	0.001	0.002	-0.003	0.006

Table 3: Descriptive statistics tables for results estimated from JAGS sampler, first party

	House	mean	median	st. dev.	95%HPD R. Lower	95%HPD R. Upper
1	House Effect 1	-0.002	-0.002	0.002	-0.007	0.002
2	House Effect 2	0.004	0.004	0.002	-0.001	0.008
3	House Effect 3	-0.001	-0.001	0.002	-0.006	0.003

Table 4: Descriptive statistics tables for results estimated from JAGS sampler, second party

	House	mean	median	st. dev.	95%HPD R. Lower	95%HPD R. Upper
1	House Effect 1	0.021	0.021	0.005	0.011	0.031
2	House Effect 2	-0.014	-0.014	0.005	-0.024	-0.005
3	House Effect 3	0.005	0.005	0.005	-0.005	0.014

Table 5: Descriptive statistics tables for results estimated from Our modified sampler, first party

Histograms of Omega - variance of the state parameters

From the histogram of omegas, the variance of alphas, we can see that neither of the samplers produce omega samples from a supposed posterior for omega. Namely, it should be a Chi-square distribution. We believe that the reason in JAGS case is because of the small sample size, but in Our modified sampler case we believe that omegas are affected by alphas, because when we troubleshooted Our sampler, we fixed alphas and house effects to sample from priors, and obtained nice omega Chi-squared samples. But when we tried to test alphas, with omega and house effects fixed to priors, alphas were not sampled from the supposed normal distribution. However, omegas from both samplers are pushed towards the upper boundary (0.01).

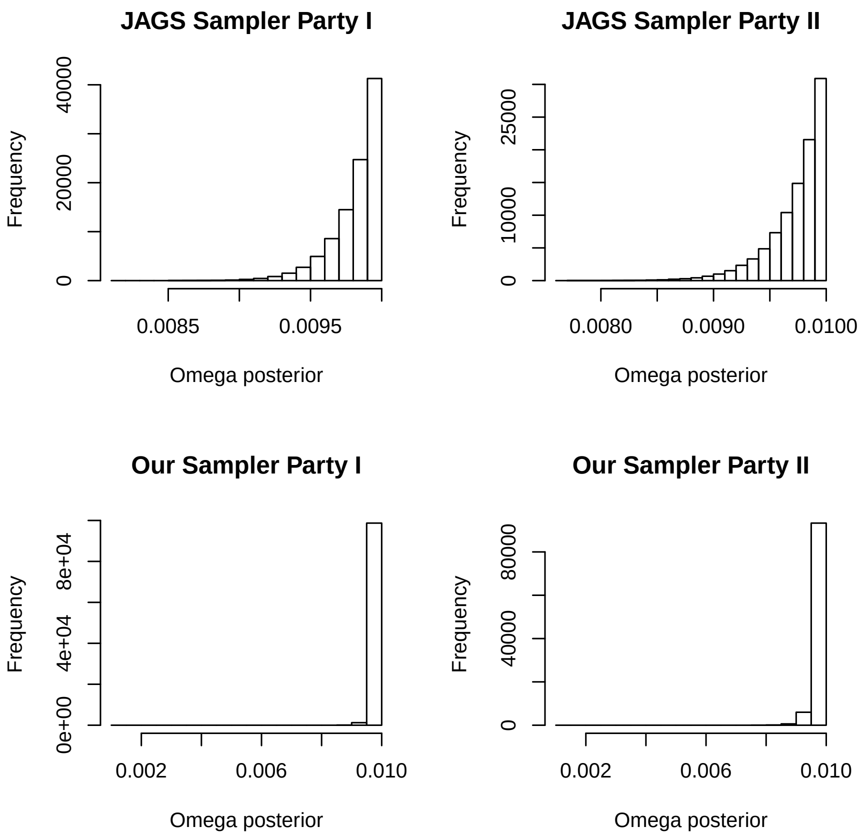


Figure 18: Histograms of variance of alphas from both samplers for the two parties

	House	mean	median	st. dev.	95%HPD R. Lower	95%HPD R. Upper
1	House Effect 1	-0.032	-0.032	0.005	-0.041	-0.022
2	House Effect 2	0.004	0.004	0.005	-0.006	0.013
3	House Effect 3	-0.015	-0.015	0.005	-0.025	-0.006

Table 6: Descriptive statistics tables for results estimated from Our modified sampler, second party

Discussion on Computation

As we saw in previous sections, we did not achieved greater efficiency nor accuracy with Our sampler in comparison with JAGS sampler. Diagnostic shows Our sampler weaknesses. In this sections we will discuss the problems that we encountered along the way and what have we learned from this study.

Our modified sampler weaknesses

Alphas are non stationary, but this is much more improved when compared to stationarity of the alphas sampled using Our original sampler. Fails to determine the bias accurately, which is also much more improved when compared to Our original sampler. For instance, It overestimates the bias for the first house effect, for the first party. Omega, variance of the state parameters are pushed towards the upper boundary of the omega, its posterior does not look like chi-squared.

JAGS sampler weaknesses

Alphas are non stationary, but they are close to convergence. Fails to determine the bias accurately which means, for instance, it underestimates the bias for the first house effect, for the first party. Omega, variance of the state parameters are pushed towards the upper boundary of the omega, its posterior does not look like chi-squared.

What problems we encountered?

The main challenge was to develop our own sampler for this particular problem. We implemented the sampler using already derived posterior distributions. But following a paper with derived posterior distributions is not enough. We had to figure out whether we can sample from given distributions directly, or we should use some additional sampler such Metropolis Hastings. After we tried Metropolis Hastings for sampling house effects, we learned that this samples house effects rather randomly. So we decided that it is better to sample directly from the posterior distribution of the house effects. Also, the Jackman's paper that we used has a lot of unexplained details, that made detecting the problems related to the model out of scope of this project.

For instance, the paper mentions truncated Normal posterior distribution for alphas in the first time point as well as truncated Gamma distribution for the posterior of the omega, the variance of state parameters(alphas). We do not know how JAGS handles this types of distributions, but we tried Metropolis Hastings for sampling omega as well as Truncate function from the "distr" R package. Both approaches sampled reasonable results for omega, so for tractability we decided to use "distr" R package.

The other big issue that was causing a lot of troubles was the forward value of alphas on which posterior alphas depend heavily. For this problem we tried several different approaches such as corresponding value from the previous iteration for each iteration within Gibbs (Our original sampler version), Kalman filter approach and anchoring forward value of alphas to the election results. As we already explained, in Our original sampler version, alphas grow beyond one (which is unreasonable, because alphas are percentage of voters that favor certain party at given time point). The Kalman filter version samples sensible values for alphas, but house effects are too big or too small. Version of Our ampler that we presented here, that uses assumption that we know the election results, is the only version that samples somewhat reasonable results.

Can we trust answers from Our sampler?

Since for sample size of 10^5, we found that Our modified sampler overestimates and JAGS sampler underestimates the bias for the first house, for the first party, we cannot trust the answers. We also cannot trust our sampler because we set up Our modified sampler to work under the assumption that we know the results in advance. It means that we set up alpha chains to sample the forward value from our prior believe, so we do not let the data to reveal what is hidden in them self. The main issue here, we believe is the analytic problems of the underlying model that we used. But that issue is out of scope this project.

Insights

JAGS with Canadian election data

Canadian Federal Election polling data

To give an idea of what results might look like for our method, we used JAGS to analyze some actual polling data. We used data from the Canadian Federal election in 2011. We considered only a subset of the available data, considering it to be a reasonable assumption that the most important polling data would be from the date that the election was called to the election date itself. The actual polling data under consideration is given in the following figure.

JAGS results

We ran JAGS on the Canadian polling data with 1,000,000 iterations. To ensure reasonable results, we used very strong prior assumptions that the state parameters are initialized at values near the election results. These initial values are $\alpha_1 = 0.05$ for the Bloc Québécois, $\alpha_1 = 0.4$ for the Conservatives, $\alpha_1 = 0.04$ for the Green Party, $\alpha_1 = 0.19$ for the Liberals, and $\alpha_1 = 0.3$ for the NDP. The model then assumes the following priors: $\alpha_1 \sim Unif(0, 0.5)$, $\omega \sim Unif(0, 0.1)$, $\delta_j^* \sim N(0, 1000)$ where $\delta_j = \delta_j^* - \bar{\delta}^*$ and hence $\Sigma_j \delta_j = 0$. These priors are meant to provide results that are consistent with the election results and are not necessarily well chosen. The assumption that $\Sigma_j \delta_j = 0$ implies that each polling house has some bias but the ‘industry-wide’ bias is zero. This is a poor assumption, especially for this dataset. In the period of data in question, support for the Conservative party was consistently low for the campaign, with the Conservatives getting more votes than any of the top firms indicated in their polls. The only firm that overestimated the Conservatives is the ultra-right wing polling company Compas. Similarly, polling data for the Liberal had a positive bias, perhaps as a result of expectations that their party would remain one of the top two parties. The house effects results for the JAGS analysis are given in the proceeding figure. The posterior means of the house effects are given in the proceeding table. From these results, we see that under our model assumptions, all the polling firms have a large bias for at least one of the parties.

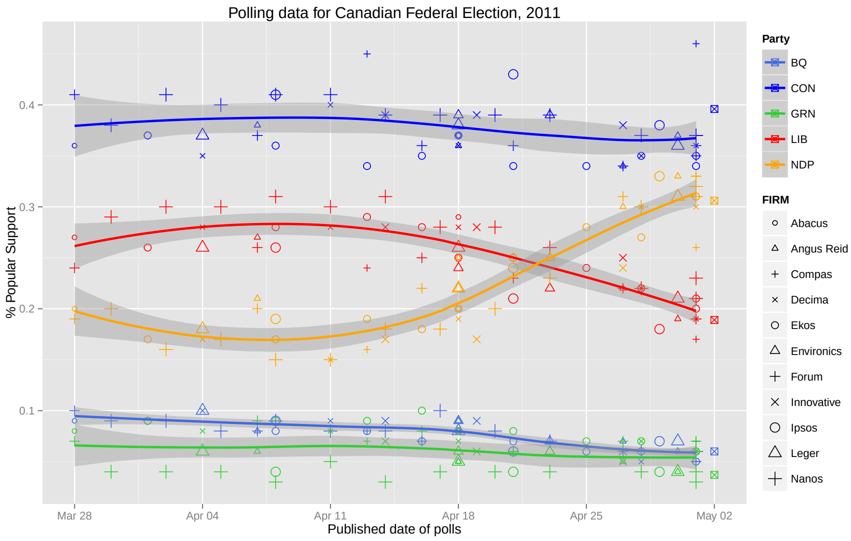


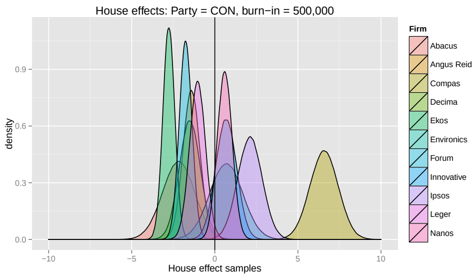
Figure 19: The polling data from the date that the election was called to election day, May 2, 2011. The points on May 2, 2011 represent the actual election results. The y-axis denotes the proportion of popular support for each party. The different points and shapes denote the different polling firms. The regression line and confidence band were not used in the actual analysis. It is a LOESS curve fitted to the polling data to get an idea of the trends in popular support over the campaign period.

	BQ	CON	GRN	LIB	NDP
Angus Reid	0.1796	-1.4058	-0.7686	-0.4438	2.6519
Decima	-0.0280	-1.5059	1.3180	0.7374	-1.5831
Ekos	-0.5126	-2.7671	1.7673	0.6326	-0.3135
Forum	-0.9667	-1.7594	1.2323	-1.3233	2.6143
Nanos	-0.1615	0.5970	-2.3069	2.4460	-0.6504

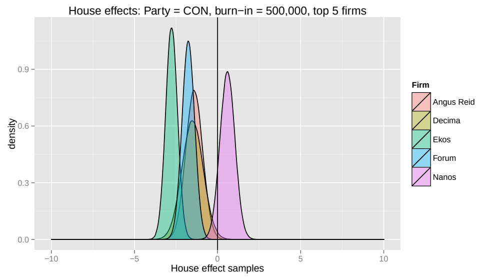
Table 7: Posterior means of the house effects from JAGS using 1,000,000 iterations and burn in of 500,000.

The figure shows the posterior samples of the house effects for all houses for the Conservative party, as well as the house effects for the Conservatives, the Liberals and the NDP for the top five firms. The posteriors come from 1,000,000 iterations of JAGS with burn-in chosen as 500,000. The top five firms were chosen as the five firms that conducted the most polls over the campaign period. These firms are Angus Reid, Decima, Ekos, Forum and Nanos. The posteriors Figure (b) shows that very few of the house effect posteriors for the Conservatives have much density around 0, with all of them clearly skewing positive or negative. This result is similar for the NDP, shown in (d). For the Liberals in (c), Decima and Angus Reid do not have significant bias, while the other three do.

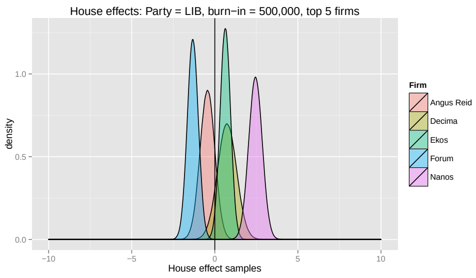
JAGS diagnostics



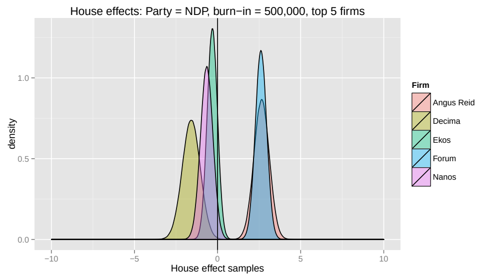
(a) House effects, all firms, party = CON



(b) House effects, top five firms, party = CON



(c) House effects, top five firms, party = LIB



(d) House effects, top five firms, party = NDP

Figure 20: Percentage house effects for the top three parties. The range of the x-axis for each plot is from -10 to 10, indicating house effects in the range of -10% to 10%. The top five firms were chosen as the five firms that conducted the most polls during the election campaign period.

We performed the Raftery-Lewis and Geweke diagnostics on the JAGS output. The diagnostic output looks favourable with no diagnostic showing problems. Under the model assumptions, JAGS allows for good mixing and parameters can converge to their posteriors. Unfortunately, the model assumptions are not reasonable.

Diagnostics on our sampler

When implementing our own sampler, we encountered many difficulties. Many of these difficulties stemmed from our underlying model. Some of the major difficultes arise from the analytic form of the posterior densities for the state parameters $\alpha_1, \ldots, \alpha_T$ and the house effects δ_j for all firms.

Posterior mean of $\boldsymbol{\alpha}$

Since we are using a pooling framework, we are using the 'multiple surveys on one day' set up described by Jackman. In this set up, the posterior density for α_s at time point s is a normal distribution with mean

$$\left[\Big(\sum_{i \in \mathcal{P}_s} \frac{y_i - \delta_{j_i}}{\sigma_i^2}\Big) + \frac{\alpha_{s-1} + \alpha_{s+1}}{\omega^2}\right]\left[\frac{1}{\sum_{i \in \mathcal{P}_s} \sigma_i^2} + \frac{2}{\omega^2}\right]^{-1}.$$

The second term in the mean, with numerator $\alpha_{s-1} + \alpha_{s+1}$, is implemented in Gibbs sampling by sampling the backward value from the previous state current chain and the forward value from

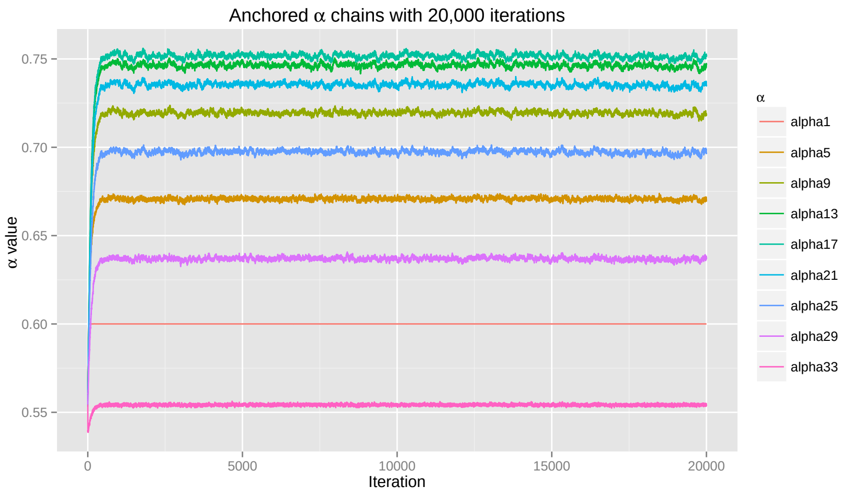


Figure 21: Select α chains from anchored sampler results.

the next state in the previous chain. Although this seems reasonable, it relies heavily on the assumption that these α chains will be well behaved. One way in which we saw them behaving poorly is when the α difference term increases at each step in Gibbs sampling. When this happens, the α difference term dominates the posterior mean term and future values of α only increase more. One way of mitigating this issue is by making ω, the prio variance of the random walk prior on the αs, large enough that this term has less influence on the posterior mean. This is problematic however, because Jackman has truncated the ω term to be between 0 and 0.1, to indicate that a swing in the state parameter of more than 20% from one period to the next is unlikely. Although this isn't addressed in Jackman, he does mention that for identifiability the model must be constrained. He chooses to constrain the α chains by the election results, a tactic that we also tested.

Anchoring α to election results

As a first attempt at addressing the problem of increasing αs, we fixed the δ and ω values at their true values based on the simulated data and anchored the α chains to an 'election result'. This was accomplished by adding an observation to the end of the α chains that was an election result. Since the data was simulated for the true α values to be between 0.4 and 0.6, we fixed the final α at 0.53. We used a prior variance for δ_j of $d = 0.08$. We then proceeded to run our sampler for 20,000 iterations. The results from this diagnostic are in the following figure. The figure shows select chains from this test. The chains for different periods are colour coded. First, it must be mentioned that the model assumes that $\alpha_1 \sim Unif(0.4, 0.6)$. Consistent with out increasing results, we see that the α_1 chain moves up to the boundary and does not come back down. From this plot, we see that the chains up to period 17 increase and remain fairly stable. Chains 21, 25, 29 and 33 are pulled down by the boundary condition, but these chains do not exhibit ideal behaviour. We also ran this same test with 500,000 iterations, with little improvement. In addition to getting unreasonable α chains, we were not able to sample δ or ω.

Anchoring α to election results, FFBS

A solution to the issue of increasing α is to introduce filtering. Jackman does this in Section 9.4.2 in (Jackman, 2009) by introducing a 'filter-forward-backward-sample' approach. This approach uses a Kalman filter to obtain the mean and variance of α_t given the data, $\delta_j \; \forall \; j$, $\sigma_i \; \forall \; i$, and ω. The conditional mean and variance are denoted m_t and C_t, respecitively. They are found recursively using the following equations.

$$m_t = \Big[\Big(\sum_{j\in\mathcal{P}_\sqcup} \frac{y_{jt}-\delta_j}{\sigma^2_{y_j}} + \frac{m_{t-1}}{C_{t-1}+\omega^2}\Big)\Big]\Big[\Big(\sum_{j\in\mathcal{P}_\sqcup} \frac{1}{\sigma^2_{j_y}} + \frac{1}{C_{t-1}+\omega^2}\Big)\Big]^{-1} \quad (1)$$

$$C_t = \Big[\Big(\sum_{j\in\mathcal{P}_\sqcup} \frac{1}{\sigma^2_{j_y}} + \frac{1}{C_{t-1}+\omega^2}\Big)\Big]^{-1} \quad (2)$$

These conditional moments were obtained as described, however, they too were not stationary and were consistently increasing. Hoping to continue in the spirit of this method, the R package `dlm` was used to implement a Kalman filter and find these estimates. Doing this, stationary estimates were found, and we were able to implement backward samling of the α's based on the results from this filter. This was our most successful attempt at obtaining results consistent with the true parameters, however it too had issues. We ran this method for 50,000 iterations. The trace plot from this method is found below. The large pink strip is from the initial starting point being initiliazed anywhere in the range of 0.45 and 0.55. The rest of the chains are well behaved and are closely spaced together, indicating no large jumps from one period to the next.

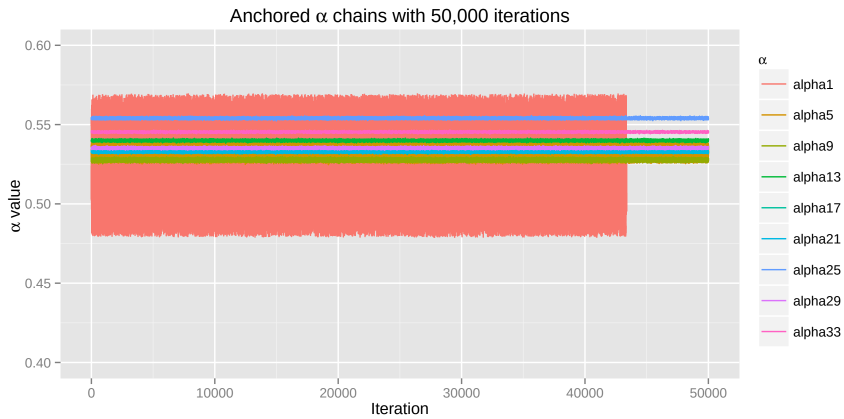
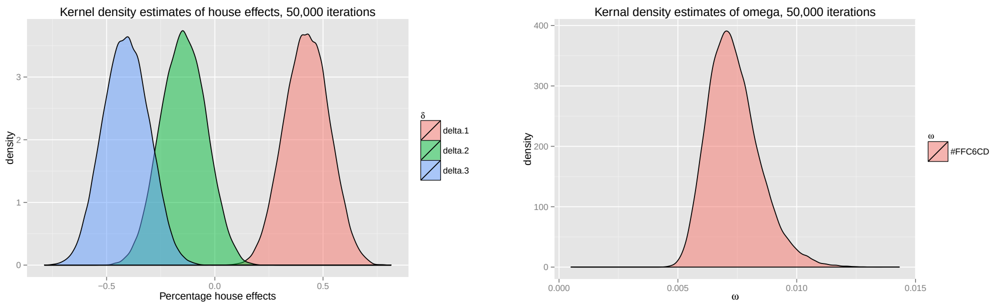


Figure 22: Select anchored α chains from Kalman filter method.

Using this method, we finally come across some house effects and ω values that are within reasonable range! Kernel density plots of these two methods are given in the figures below. For the house effects plot, the densities are plotted with their parameter value on the x-axis, not their percentage value on the x-axis. The true values of the house effects are 0.02, -0.04 and 0.005 for δ_1, δ_2 and δ_3. The density plots show the densities are fairly close, with the posterior for δ_3 being closer to the true value of $delta_2$ and vice versa. In addition, the posterior of δ_2 skews negative when its true value is positive but very close to zero. The true value of ω was 0.001, whereas our ω posterior is centered around 0.08.



(a) Kernel densities of house effects, δ_j (b) Kernel density of ω

Although the Kalman filter method is actually able to produce estimates, it does not do a great job of estimating the house effects accurately. Still, it is much better than our previous attempts.

Final conclusions

What have we learned?

The model explained in Jackman's paper does not impose sufficient constraints for this problem. There are some details that are vague in Jackman's paper, sometimes ad hoc methods are used such as truncation as explained in the section on problems that we encountered. We beleive that there are more constraints that are required for this particular problem. Moreover, the model is very sensitive to our well behaved simulated data. As we explained in previous sections, this is not the problem for Our sampler only, but also for the JAGS sampler.

The way that we modified our sampler, by anchoring forward values of alphas to the election results, we feel like it is not exactly what Gibbs sampler is supposed to do. Instead, we believe that forward value for the alphas should be sampled either from updated Kalman filter or from the corresponding value from the previous iteration for each iteration within Gibbs. As we explained previously, the Kalman filter approach samples reasonable alphas, but the house effects are not reasonable at all. The other approach, that we treated as Our original sampler (because it is developed according to the model explanations in the Jackman's model), where forward alpha values are taken from the corresponding values from the previous iteration within Gibbs sampler, give very unreasonable results for alphas (in this case alphas grow beyond 1 without control). We also saw that on our well behaved simulated data, JAGS sampler sometimes fails to detect the correct bias. This is easy to check, because we know how large the bias was we plugged into our simulated data, so we know what the results should look like. Based on the conducted diagnostic checks and analysis of estimated parameters in our case, we can conclude that we cannot trust nor Our modified results, nor the JAGS results.

Limitations

One of the purposes of developing this model is to use the knowledge of the house effects to aid in building a predictive model for future election outcomes. One of the major downfalls of the current model is that it requires having the election result outcome (or to assume there is no industry wide bias and hence have the house effects sum to 1), therefore the only way to build a predective model would be to use a previous election to determine house effects and use those to predict the current election. However, it is unlikely that house effects are going to stay the same over a number of election cycles. For instance, Pickup and Johnston (2007) found that support for the Conservative party was over-estimated by the polling organisations on an industry wide basis in the 2004 and 2006 elections, however it was the opposite in the 2008 and 2011 elections, the Conservatives were under-estimated (Rosenthal, 2011). This poses a significant problem, and so it may be that the best method for predicting is to weight heaviest the polling companies that historically show the least amount of bias from election to election, instead of trying to outright fix their bias which could change from over-estimating to under-estimating (or vice-versa) over time.

Future Work

It is very difficult and time consuming to develop an efficient sampler for this particular problem. The original study of this problem (Jackman, 2005) that we used as our guide during development of Our sampler says that state parameters hardly achieve stationarity and they had to run the JAGS sampler for $25 * 10^6$ iterations. Our sampler could not produce 10^6 iterations even for one party in 15 hours! This is where C or C++ could take place and improve the efficiency. But in the given time frame for this project we were not able to try C++, so this would be our idea for future development of this sampler.

During debugging when we were simply trying to get our α parameters to behave well, using a very large d value (0.2) in the prior for δ_j seemed to give us well behaved α values. A value for d this large signifies a very vague prior on the house effects so it could be that we did not have enough data in our study to overtake the priors we were originally using. In Jackman's original paper he uses a period (T) of 114 days, whereas we only have 33 weeks and 5 weeks, for our simulated data and real data cases, respectively. Due to time constraints we could not fully test this, but it would be beneficial to explore this in the future.

Available materials

Since we performed a significant amount of troubleshouting trying to develop our own sampler, and we did not manage to develop a sampler that would give relaible results, we are submitting two versions. Namely, all necessary files for Our original sampler (developed according to the Jackman's specifications) to run are compiled in the folder "original model". All necessary files for Our modified sampler (uses modified prior of the house effects and taking future value of alphas from the prior) that is used for conducting diagnostics and results is compiled in the folder "anchored alphas - diagnostic". We also submit a Metropolis Hastings version of our sampler that we presented in class "MH version - presented in class", where we used metroplois hasting algorithm for sampling House Effects. As well we are submitting code used to do the analysis of real data with JAGS filename RealDataAnalysis.R. Finally, we have saved results for 100 000 iterations that can be made available upon request if you would like to test the diagnostic part (the file is too large for emailing).

References

Jackman, S. 2005. 'Pooling the Polls Over an Election Campaign' Australian Journal of Political Science, V40(4): 499 - 517

Jackman, S. 2005b. Appendix to 'Pooling the Polls Over an Election Campaign'. Accessed online, heading 27. http://jackman.stanford.edu/papers/

Jackman, S. 2009. Bayesian Analysis for the Social Sciences, Wiley, 471-485.

Pickup, M. and Johnston, R. 2007. 'Campaign tial heats as electoral information: Evidence from the 2004 and 2006 Canadian federal elections'. Electoral Studies, 26, 460-476.

Rosenthal, J. 2011. 'Was the Conservative Majority Predictable?',
Accessed online http://www.probability.ca/jeff/research.html