Bayesian Computation for Parametric Models of Heteroscedasticity in the Linear Model *

W. John Boscardin Andrew Gelman
Department of Statistics University of California, Berkeley

May 16, 1994
Revised November 9, 1994

*In *Advances in Econometrics* **11**, A87–10. We thank the National Science Foundation for grants SBR-9223637, DMS-9457824, and a Graduate Research Fellowship.

ABSTRACT

In the linear model with unknown variances, one can often model the heteroscedasticity as $\mathrm{var}(y_i) = \sigma^2 f(w_i, \theta)$, where f is a fixed function, w_i are the "weights" for the problem and θ is an unknown parameter ($f(w_i, \theta) = w_i^{-\theta}$ is a traditional choice).

We show how to do a fully Bayesian computation in this simple linear setting and also for a hierarchical model. The full Bayesian computation has the advantage that we are able to average over our uncertainty in θ instead of using a point estimate. We carry out the computations for a problem involving forecasting U.S. Presidential elections, looking at different choices for f and the effects on both estimation and prediction.

1 Introduction

In both the econometrics and statistics literature, a standard way to model heteroscedasticity in regression is through a parametric model for the unequal variances, as described in many places, e.g. Amemiya (1985), Greene (1990), Judge et al. (1985), Carroll & Ruppert (1988). Modeling heteroscedasticity should improve the efficiency of estimates of regression coefficients, but the most important effect is in prediction, allowing predictive inferences to be more precise for some units and less precise for others.

Previous approaches have tended to focus on obtaining a point estimate for the heteroscedasticity parameter, θ (possibly vector-valued). Unfortunately, in many applications, the parameter governing the heteroscedasticity is not well-identified by the data, so any point estimate may be quite unreliable. Furthermore, there is no consensus on which point estimate to use (Carroll & Ruppert, 1988).

To avoid these problems, we use a fully Bayesian approach, which automatically averages over our uncertainty in the model parameters. One potential pitfall of Bayesian analysis is its sensitivity to the choice of a prior distribution, but, for the example that we consider, our inference and prediction are quite robust to reasonable choices of prior distributions on θ.

In this paper, we will develop a computational framework for parameter estimation and prediction in both non-hierarchical (NH) and hierarchical (HIER) linear models. The computation is quite straightforward for the NH models and harder for the HIER models. We will illustrate our methods with the example of forecasting U.S. Presidential elections.

2 Models

2.1 Heteroscedasticity in the linear regression

We use the following regression model for the n independent units on the k covariates:

$$y_i|\boldsymbol{\beta}, \sigma^2, \theta \sim \mathrm{N}((X\boldsymbol{\beta})_i, \sigma^2 f(w_i, \theta)). \quad (1)$$

For later reference the log-likelihood function is

$$\ell(\boldsymbol{\beta}, \sigma^2, \theta) \;=\; \log p(y|\boldsymbol{\beta}, \sigma^2, \theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_i \log f(w_i, \theta)$$

$$-\frac{1}{2\sigma^2}\sum_i (y_i - (X\boldsymbol{\beta})_i)^2 / f(w_i, \theta). \tag{2}$$

We consider the special case in which the unequal variances are governed by known "weights" w_i, and we wish to define a continuum between the extreme cases of no weighting and having variance proportional to $1/w_i$, i.e., weighted least squares. We consider two different forms for f :

$$\text{HET1:} \quad f(w_i, \theta) = w_i^{-\theta} \tag{3}$$

where $\theta \in [0, 1]$, and

$$\text{HET2:} \quad f(w_i, \theta) = (1-\theta) + \frac{1}{w_i}\theta \tag{4}$$

where $\theta \in [0, 1]$. For either model, the extremes of $\theta = 0$ and 1 correspond to equal variances and variances proportional to $1/w_i$, respectively.

The first model is standard in the literature; see Greene (1990), for example. The second model has a natural interpretation as two independent sources of variation, with one term having variance inversely proportional to w_i and the other term having constant variance (e.g., sampling and modeling errors). We also consider, as a comparison, the equal variance (EV) or homoscedastic model, i.e. $f(w_i, \theta) \equiv 1$.

2.2 Parametrization of the weights

In each case we normalize the n values $f(w_i, \theta)$ to have product equal to 1 as suggested by Box & Cox (1964). This is done to provide a consistent meaning for σ^2 between models with different values of θ and to simplify the expression for the likelihood (see equation 2). In the specification of equation 3, this just amounts to dividing the w_i by their geometric mean, so, without loss of generality, we can take the w_i to have product equal to one. For the form in equation 4, the normalization actually implies a slightly different functional form of HET2:

$$\text{HET2:} \quad f(w_i, \theta) = \frac{(1-\theta) + \frac{1}{w_i}\theta}{\prod_i \left((1-\theta) + \frac{1}{w_i}\theta)\right)^{1/n}}. \tag{5}$$

For HET2, there is no trick that will let us avoid the issue of the normalization.

2.3 Prior distributions on the variance parameters

We consider various non-informative prior distributions to the variance components. For the EV model, the improper uniform prior density on $\log\sigma$ is standard from many perspectives (e.g., Box and Tiao, 1973). For the heteroscedastic models, there does not seem to be any clear choice for a noninformative prior density on θ, but $p(\theta) = 1$ on the unit interval seems reasonable. We cannot use a uniform density on logit(θ) as it would lead to an improper posterior density under either model. Another school of thought suggests calculating the Jeffreys prior density (e.g., Box and Tiao, 1973). The conditional Fisher information for θ is calculated as

$$-\frac{\partial^2 \ell}{\partial\theta^2} = -\frac{1}{2}\sum_i \frac{f'(w_i,\theta)^2}{f(w_i,\theta)^2}. \tag{6}$$

For HET1, this turns out to be a function of the w_i's only, and thus the information is constant, suggesting a uniform prior density on θ. For HET2, the Jeffreys prior density has a particularly unattractive form, and we did not pursue this any further.

In our main analysis here, we use a uniform prior density on $(\log\sigma, \theta)$ for both models. For each model, we assess the sensitivity to the choice of prior distribution by comparing to inferences obtained using the Beta$(\frac{1}{2}, \frac{1}{2})$ density for θ, which is another standard noninformative density for a parameter on $[0, 1]$.

2.4 Models for the regression coefficients

When computing regressions in the nonhierarchical model (NH) using least squares, we are implicitly assigning a uniform prior density on the regression parameters $\boldsymbol{\beta}$ in the model (1).

In the hierarchical model (HIER), we let $\boldsymbol{\beta}$ have an informative prior distribution which depends on unknown parameters. In this paper, we restrict to the case:

$$\boldsymbol{\beta}|\boldsymbol{\tau^2} \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta), \tag{7}$$

where Σ_β is a diagonal matrix whose entries come from $\boldsymbol{\tau^2} = (\tau_1^2, \ldots, \tau_J^2)$. (Lindley and Smith, 1972, discuss more general forms of the hierarchical normal model.) If we want to keep a non-hierarchical structure on some of the β's ("fixed" effects), we can set the corresponding entries of Σ_β^{-1} to zero. Having done this, there are now k_* β's with an informative prior distribution ("random" effects). Of these, k_j have variance $\tau_j^2, j = 1, \ldots, J$ and $k_1 + \cdots + k_J = k_* \leq k$. The hierarchical model is not necessary for our study of heteroscedasticity, but in all our experiences, including the example of election forecasting discussed below, the hierarchical part of the model has been important, both for parameter estimation and prediction. We will elaborate on this importance in the example. We assign a noninformative prior density to the variance components τ_j^2. As is well known in the statistics literature (e.g., Hill, 1965 and Box & Tiao, 1973), we cannot assign a uniform prior density on the parameters $\log \tau_j$ as that would lead to an improper posterior distribution. Instead we consider two alternative noninformative distributions: uniform in τ_j and in τ_j^2.

3 Computation

Our general goal in Bayesian computation is to draw simulations from the posterior distribution of the unknown parameters and then to sample from the predictive distribution of future data by using these simulated parameter draws. The computational procedure is easier for the non-hierarchical case since there are only two unknown variance parameters, and we are able to draw independent realizations of the posterior by factoring it in an appropriate way. In the hierarchical case, we need to use a Gibbs-type sampler, and the computation is more elaborate.

3.1 Non-hierarchical regression

First consider the NHHET models. Conditional on θ, we can use the standard results for Bayesian linear weighted regression (e.g. Box & Tiao, 1973; Zellner, 1971):

$$(\sigma^2|\theta, y) \stackrel{d}{=} \frac{S^2}{\chi^2_{n-k}} \tag{8}$$

$$(\boldsymbol{\beta}|\sigma^2, \theta, y) \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, \sigma^2 V_\beta), \tag{9}$$

where

$$W^{-1} = \mathrm{diag}(f(w_1, \theta), \ldots, f(w_n, \theta)), \tag{10}$$

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (X^T W X)^{-1} X^T W y, & (11)\\
V_\beta &= (X^T W X)^{-1}, & (12)
\end{aligned}
$$

and

$$
S^2 = (y - X\hat{\boldsymbol{\beta}})^T W (y - X\hat{\boldsymbol{\beta}}) \qquad (13)
$$

is the sum of the squared weighted residuals.

We can determine the marginal posterior density of θ using the identity,

$$
\begin{aligned}
p(\theta|y) &= \frac{p(\boldsymbol{\beta}, \sigma^2, \theta|y)}{p(\boldsymbol{\beta}, \sigma^2|\theta, y)}\\
&\propto \frac{p(y|\boldsymbol{\beta}, \sigma^2, \theta)p(\boldsymbol{\beta}, \sigma^2, \theta)}{p(\boldsymbol{\beta}|\sigma^2, \theta, y)p(\sigma^2|\theta, y)}.
\end{aligned}
$$

Substituting the likelihood (equation 2), and equations 8 and 9, we find that

$$
p(\theta|y) \propto \frac{\sigma^{-n}\prod_i f(w_i,\theta)^{-1/2}\exp(-S^2/(2\sigma^2))\sigma^{-2}p(\theta)}{\sigma^{-k}|V_\beta|^{-1/2}\exp(\sigma^{-2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T V_\beta^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}))(S^2)^{(n-k)/2}\sigma^{-(n-k+2)}\exp(-S^2/(2\sigma^2))}.
$$

Since the left side of this expression is a function of θ and y only, the right side of this can not depend on $\boldsymbol{\beta}$. We can therefore set $\boldsymbol{\beta}$ to any value we wish, and we set $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ for both numerical stability and algebraic simplicity. Also, recall that the product of the f's is equal to one as discussed earlier, so we now have:

$$
p(\theta|y) \propto |V_\beta|^{1/2}(S^2)^{-(n-k)/2}p(\theta). \qquad (14)
$$

This is not a recognizable distribution, but we can certainly compute the value of the unnormalized posterior density for any value of θ by doing the weighted least-squares regression and computing the quantity in Equation 14. We do this over a fine mesh of θ values and then use the inverse-cdf method to simulate many draws from this arbitrarily good approximation to $p(\theta|y)$. Given each of our simulated θ values, we can compute W and then use weighted least squares to get $\hat{\boldsymbol{\beta}}$ and S, draw σ^2 from its posterior distribution, compute V_β, and draw a sample of $\boldsymbol{\beta}$ from its posterior distribution. Once we have a set of posterior simulations of $(\boldsymbol{\beta}, \sigma^2, \theta)$, we can simulate from the predictive distribution of y given a new set of covariates, X_{pred}.

The procedure for NHEV is much simpler. By letting $W = I$ and suppressing the conditioning on θ, Equations 8 – 13 give a full factorization of the posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$, and so the computations for this special case are trivial to carry out.

3.2 Hierarchical regression

The hierarchical regression model can be interpreted as a non-hierarchical regression with additional "data" corresponding to the prior distribution (e.g., Dempster, Rubin, and Tsutakawa, 1981):

$$\begin{aligned} y_*|\boldsymbol{\beta}, \Sigma_* &\sim \mathrm{N}(X_*\boldsymbol{\beta}, \Sigma_*) \\ p(\boldsymbol{\beta}|\Sigma_*) &\propto 1, \end{aligned} \tag{15}$$

where y_*, X_* and Σ_* are derived by combining the likelihood for y and the prior distribution on $\boldsymbol{\beta}$:

$$y_* = \begin{pmatrix} y \\ 0 \end{pmatrix}, \quad X_* = \begin{pmatrix} X \\ I_k \end{pmatrix}, \quad \Sigma_* = \begin{pmatrix} \Sigma_y & 0 \\ 0 & \Sigma_\beta \end{pmatrix} = \begin{pmatrix} \sigma^2 W^{-1} & 0 \\ 0 & \Sigma_\beta \end{pmatrix}, \tag{16}$$

(W is defined in Equation 10) with the prior distribution of the variance parameters, $p(\Sigma_*)$, determined by the prior distribution of $(\sigma^2, \boldsymbol{\tau^2}, \theta)$.

Even though we have reduced the problem to a non-hierarchical regression, we can not just proceed as in Section 3.1, in which we were able to analytically integrate out σ^2 to find an expression for $p(\theta|y)$. Here, there are too many variance parameters (σ^2 and $(\tau_1^2, \ldots, \tau_J^2)$), so we need a different method. We use the following general paradigm not just for this problem, but for other problems in our work. First, we attempt to find the mode of $p(\sigma^2, \boldsymbol{\tau^2}, \theta|y)$ by an EM-type algorithm, as in Dempster, Rubin, and Tsutakawa (1981). We then run a Gibbs sampler (e.g., Gelfand and Smith, 1990) to sample from $p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau^2}, \theta|y)$. If a particular conditional is not easy to sample from, then we use a step of Metropolis' algorithm instead of a Gibbs step for that parameter. This Markov chain simulation has the posterior distribution as its equilibrium distribution (see, e.g., Smith and Roberts, 1993). To help monitor convergence, we run multiple parallel sequences from overdispersed starting points simulated from a t_4 distribution centered at the mode of $(\sigma^2, \boldsymbol{\tau^2}, \theta)$ found by the EM procedure (Gelman & Rubin, 1992).

In this particular problem, the Gibbs-Metropolis sampler works as follows; for clarity, we demonstrate the calculations with the uniform prior distribution on $(\log\sigma, \tau_1^2, \ldots, \tau_J^2, \theta)$. At the t-th iteration for a particular sequence, we have $(\boldsymbol{\beta}^t, (\sigma^2)^t, (\boldsymbol{\tau^2})^t, \theta^t)$. We then sample

$$\boldsymbol{\beta}^{t+1}|(\sigma^2)^t, (\boldsymbol{\tau^2})^t, \theta^t, y \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, V_\beta),$$

where $\hat{\boldsymbol{\beta}}$ and V_β come from the weighted least-squares procedure applied to y_*, X_* and weights Σ_*^{-1}. Then, given $\boldsymbol{\beta}^{t+1}$ and θ^t, the conditional posterior distribution of the variance components $(\sigma^2)^{t+1}$ and $(\boldsymbol{\tau^2})^{t+1}$ can be determined from the weighted residuals $r_* = W_*^{1/2}(y_* - X_*\boldsymbol{\beta}^{t+1})$, where

$$W_* = \begin{pmatrix} W & 0 \\ 0 & I \end{pmatrix}. \tag{17}$$

Each variance component is simulated as the sum of the squares of the corresponding elements of r_* divided by a χ^2_ν random variate. For σ^2, $\nu = n - k$, and for each τ_j^2, $\nu = k_j - 2$.

We now need to draw from $p(\theta|\boldsymbol{\beta}^{t+1}, (\sigma^2)^{t+1}, (\boldsymbol{\tau^2})^{t+1}, y)$. Algebraically, we have

$$\begin{aligned} p(\theta|\boldsymbol{\beta}^{t+1}, (\sigma^2)^{t+1}, (\boldsymbol{\tau^2})^{t+1}, y) &\propto p(\boldsymbol{\beta}^{t+1}, (\sigma^2)^{t+1}, (\boldsymbol{\tau^2})^{t+1}, \theta|y) \\ &\propto p(\theta)(\sigma^{-(n+2)})^{t+1} \prod_j (\tau_j^{-k_j})^{t+1} \prod_i f(w_i,\theta)^{-1/2} \\ &\quad \times \exp\left((y_* - X_*\boldsymbol{\beta})^T \Sigma_*^{-1}(y_* - X_*\boldsymbol{\beta})\right). \end{aligned} \tag{18}$$

We use a Metropolis step here because the density has no standard form but is numerically computable for any set of coefficients and variance parameters. The Metropolis algorithm is easiest to describe in the special case of a symmetric random walk chain (see, e.g., Tierney, 1994 for other possibilities). We generate a candidate θ^* by taking a random step away from our current value θ^t according to a jumping density, $J(\theta^*|\theta^t)$. The jumping density is symmetric; that is, $J(\theta^*|\theta^t) = J(\theta^t|\theta^*)$ for any θ^t, θ^*. We then accept this candidate with probability

$$\alpha(\theta^t, \theta^*) = \min\left(\frac{p(\theta^*|\boldsymbol{\beta}^{t+1}, (\sigma^2)^{t+1}, (\boldsymbol{\tau^2})^{t+1}, y)}{p(\theta^t|\boldsymbol{\beta}^{t+1}, (\sigma^2)^{t+1}, (\boldsymbol{\tau^2})^{t+1}, y)}, 1\right) \tag{19}$$

For this problem, we use a symmetric normal jumping kernel, $J(\theta^*|\theta^t) = \mathrm{N}(\theta^*|\theta_t, v^2)$, with v^2 set to 2.38^2 times the estimated variance of θ based on the normal approximation at the EM estimate, as suggested in Gelman, Roberts and Gilks (1994). To maintain symmetry, we reflect the random walk at the boundaries of the space of θ, 0 and 1.

Once we have a set of simulations from $p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau^2}, \theta|y)$, we can use these to make predictions of y for a new set of predictive covariates, X_{pred}. The procedure is not completely straight forward, so we describe it briefly. We use each $\boldsymbol{\tau^2}$ simulation to generate realizations of the k_* "random" β's. We combine these with the "fixed" β's and σ^2 simulation to get y_{pred}. We then repeat this process for the rest of the simulations in out set. The resulting set of y_{pred}'s is a sample from the posterior predictive distribution, $p(y_{pred}|y)$.

4 Numerical Illustration

Judge et al. (1982), Example 9.3.7, give a simulated data set based on the following heteroscedastic linear model:

$$y_i|\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2 \sim N((X\boldsymbol{\beta})_i, \exp(\alpha_1 + \alpha_2 x_{i2})),$$

where the ith row of X is $(1, x_{i2}, x_{i3})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (10, 1, 1)^T$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T = (-3, 0.3)^T$.

We can transform this to the form of HET1 by taking $w_i = \exp(-(x_{i2} - \bar{x}_2))$, $\theta = \alpha_2$, and $\sigma^2 = \exp(\alpha_1 + \alpha_2\bar{x}_2)$. Our posterior inference for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is summarized in Table 1. Given our posterior simulations of θ and σ^2, we just transform back to the $\boldsymbol{\alpha}$ parametrization to get posterior simulations of $\boldsymbol{\alpha}$ ($\alpha_2 = \theta, \alpha_1 = \log\sigma^2 - \theta\bar{x}_2$). The inference is quite similar to the generalized least squares estimates obtained in the reference (by first estimating $\boldsymbol{\alpha}$). Our posterior means are slightly closer than the GLS estimates to the true parameter values in each case.

TABLE 1 ABOUT HERE

5 Example: Forecasting U.S. Presidential Elections

5.1 Background

Most political scientists are now aware that the U.S. Presidential election can be predicted quite well with information available several months before the actual election. In contrast, the Gallup poll taken several months before the election can predict a landslide victory for the eventual loser, as when Dukakis lost to Bush in 1988. The methods used by Rosenstone (1983, 1990), Fair (1978, 1982, 1988), Campbell (1992) and others use standard regression methods and are remarkably successful. We became interested in this problem to see if we could do better by using more refined statistical techniques.

We begin by using a model described by Gelman and King (1993) that is similar to the forecasting model of Campbell (1992). The method is to run a simple linear regression of Democratic share of the two-party vote by state for each election year, beginning with 1948, on several covariates. These covariates can be divided into three groups: national variables, state variables and regional variables. First define the variable INC to be 1 if a Democrat is President and -1 if not. The national variables that we use are the Democratic share in the September Gallup poll (TRIHEAT), the change in GNP in the previous quarter times INC (7ECxINC), a dummy to say if the current president is seeking reelection times INC (PRESINC), and the latest approval rating times INC (APPRxINC). The state variables are the Democratic share in the last 2 Presidential elections (DEV2 and DEV3), home-state advantage for the President and the Vice-President times INC (HOME4 and VP4), state economic growth in the past year times INC (EC11xINC), legislature partisanship (LEGIS), an index of state liberalism (ADAACA), and a measure of the proportion of Catholic voters in each state for the year in which there was a Catholic candidate (CATH60). The regional variables, mostly used to "patch up" problems in "abnormal" years, are a $-1/0$ dummy for the South in '64 ($-$S64B), and similar dummies for the Deep South in '64 ($-$DEEPS64), New England in '64 (NEWENG64), the West in '76 ($-$W76), and the North Central in '72 (NCENT72). There is also a true regional variable (S4) which is 1 for every southern state in the years for which the Democratic candidate had a Southern home state. The variables are signed so that they are all expected to have positive coefficients. Finally we have a column of 1's (CONSTANT), to give a total of $k = 19$ covariates for the NH models. We exclude all cases in which a third-party candidate for President won the plurality of the votes in a state, leaving us with 511 observations.

There are a few issues that need to be addressed in connection with our modeling choices (see Judge et al., 1982, Chapter 19). First, the parameter space is actually discrete and different for each state because we are looking at the Democratic share of a finite number of votes. This is not really a problem, though, because of the enormous population sizes involved (no less than twenty thousand in any particular instance). Second, proportions for count data typically are heteroscedastic, which usually precludes a standard regression analysis, but our models are taking this into account. Finally, in many problems that occur on the unit interval, there is highly non linear behavior near the end points which necessitates a logit or probit transformation. Here, though, 99% of the data points are between 0.24 and 0.77. So while we could model a transformation of the data, this would probably not make much of a difference.

Models NHEV, NHHET1 and NHHET2

Homoscedastic models are standard in empirical studies of elections, and so our first model (NHEV) is that the variance in each state for each election year is a constant σ^2.

Of course, we do not really think that the states are equally variable. In fact, we might expect that the bigger states are less variable than the smaller states. A naive model which says that votes are binomially distributed implies that the variance of the Democratic vote share should be proportional to $\frac{1}{n}$ where n is the number of voters. This motivates using the models of heteroscedasticity discussed in Section 2, and we tried both forms for f with corresponding models NHHET1 and NHHET2, with w_i being the number of voters for the Democrats and Republicans in the corresponding state and year. For predictions of 1992, we set w_i to the voter turnout in the state in 1988, scaled by the increase in the voting-age population of the state from 1988 to 1992.

Models HIEREV, HIERHET1 and HIERHET2

We can immediately notice some potential problems with the NH models. First, most of the regional variables, while they certainly allow the models to fit better (we calculated a multiple R^2 of better than 0.99 for NHEV!), will not help at all for doing predictions. Also, this model ignores the year-by-year structure completely and treats the data as 511 independent observations.

These problems are handled with a hierarchical model. We include all the original covariates except for the regional "patch-up" variables: $-$S64B, $-$DEEPS64, NEWENG64, $-$W76, and NCENT72. We leave a flat prior density on the original β's. We then include a covariate for election year which has 11 levels corresponding to the 11 elections from 1948 to 1988, and a covariate for region by year. The regions we used were East, Midwest, West, and South. Thus, there were 44 levels of this factor. We call the 11 β's associated with the year factor the year effects, and the other 44 β's the region $\times$ year effects. The year effects are given independent normal prior distributions with mean zero and variance τ_1^2. Similarly, the region $\times$ year effects are $\mathrm{N}(0, \tau_2^2)$ except for the South $\times$ year effects which are $\mathrm{N}(0, \tau_3^2)$, reflecting the prior belief that the South is a special region that is not exchangeable with the other three, politically. To match some of the earlier definitions, there are $k = 19 - 5 + 11 + 44 = 69$ covariates of which $k* = 55$ have informative prior distributions with $k_1 = 11, k_2 = 33$, and $k_3 = 11$.

This setup is quite arbitrary. We could specify more than four regions, we could give each region a separate variance, and so on. This model represents a simple, first-pass solution, which we hope should address the most serious problems of the NH models. We would want to try more carefully specified models, and check the robustness of our inference to changing the specification.

We fit this hierarchical model with the equal variance assumption (HIEREV) and the two heteroscedastic models (HIERHET1 and HIERHET2) from Section 2.

5.2 Results with non-hierarchical models

Table 2 contains a summary of the posterior simulations for the parameters. Table 3 shows quantiles of the state by state, electoral college and popular vote predictions for the 1992 elections. Figure 1 plots the prediction standard errors as versus the w_i (on a log scale).

TABLE 2 ABOUT HERE
TABLE 3 ABOUT HERE
FIGURE 1 ABOUT HERE

The posterior distribution of the coefficients is comparable across models. The posterior medians of the coefficients all have the expected positive sign, except for APPRxINC which is essentially zero. Both of the HET models find evidence for a small amount of heteroscedasticity. This can be seen in Figures 1b and c and also from the 95% posterior intervals for θ in Table 2. For the predictions, there is nearly exact agreement in the medians as can be seen in Table 2, so we do not show a comparison plot. The standard errors of prediction for NHEV bear no relation to turnout and are centered around 3.8%.

The standard errors for NHHET1 decrease linearly with log turnout whereas those for NHHET2 decrease linearly at first, but then level out following the parametric model. Predictive standard errors range from 3.2% to 4.2% for NHHET1 and from 3.4% to 4.7% for NHHET2.

To confirm our suspicion that the year by year structure of the data is not being dealt with effectively by the NH models, we compute the average residuals by election year. This gives us a list of 11 numbers for each of the three models. For NHEV, we expect these 11 numbers $\bar{r}_t$ to be approximately independently normally distributed with variances σ^2/n_t where n_t is the number of observations we used in election year t, i.e. that $n_t^{1/2}\bar{r}_t \approx N(0, \sigma^2)$. The posterior 95% interval for σ from Table 2 is roughly $(3.5\%, 4.0\%)$, but empirically the eleven values of $n_t^{1/2}\bar{r}_t$ have a standard deviation of nearly 8%; i.e., the year by year swings are about twice as variable as the model would predict. Similarly, for NHHET1 and NHHET2, the model predicts that $n_t(\sum_{i=1}^{n_t} f(w_i, \theta))^{-1/2}\bar{r}_t \approx \mathrm{N}(0, \sigma^2)$. The results are roughly the same: despite a 95% posterior interval for σ of $(3.5\%, 4.0\%)$ in each of the models, the empirical variation of the eleven quantities is nearly 8% in each case. We perform a more formal check of the model in Section 5.4.

5.3 Results with hierarchical models

For each of the hierarchical models, we computed the mode of the variance components using the EM algorithm and then ran 10 parallel simulation sequences of the Gibbs-type sampler, each of length 500. As described in Gelman and Rubin (1992), we can use the multiple sequences to help decide whether the sampler has run for a long enough time. We calculated the potential scale reductions for all the parameters, $\hat{R}$, based on discarding the first half of each sequence (to allow "burn-in" to occur), and they were all less than 1.1. Since running the sampler for an infinitely long time would only bring these numbers down to 1.0, we are satisfied with the convergence of the sampler. Table 4 summarizes the posterior distribution of the parameters, Table 5 the quantiles of the 1992 predictions, and Figure 2 the prediction standard errors.

The coefficient estimates appear to be relatively similar. They are positively signed as expected. If anything, the HET estimates are more efficient (have narrower 95% intervals) than the EV estimates in most of the cases. The hierarchical standard deviations τ_1, τ_2, τ_3 are not determined with very much precision. This points out one great advantage of doing Bayesian computations; if we had simply made point estimates of these variance components, we would have been ignoring a wide range of possible values for all the parameters. Also, there is more evidence for heteroscedasticity than in the NH models. Intuitively, this makes sense; because we are accounting for the year and the region by year variability, it should be easier to detect unequal variances in the state by state errors.

TABLE 4 ABOUT HERE
TABLE 5 ABOUT HERE
FIGURE 2 ABOUT HERE

Again, the median predictions are virtually identical for both the HIER and NH models, and we do not show a plot. The prediction SE's split nicely into southern and non-southern states, with the southern states about one percent higher. Within each of the groups, we see the same behavior as a function of turnout as we saw in the NH models. The prediction standard errors are much larger than the ones for the NH models (they range from 6% to 8.5%), perhaps too large, because of our simplistic model of the regions of the country. To develop a model with narrower prediction intervals requires including more information about the elections.

To check that idea, we fit HIERHET1 again, but this time including the five regional covariates (HIERHETALL). Although this is not a well-defined model for prediction, as explained earlier, we certainly get better results by doing this. Now the posterior distributions of the τ's have both smaller medians and lower variance. Also, the prediction standard errors are more reasonable, ranging between 5% and 6%. Tables 6 and 7 and Figure 3 display some of these results.

TABLE 6 ABOUT HERE
TABLE 7 ABOUT HERE
FIGURE 3 ABOUT HERE

5.4 Model Checking

For each of our simulated parameter vectors $\Phi = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau^2}, \theta)$, we can generate a replicated data set, y_{rep}. Then to check whether the model fits well in some way, we can calculate an appropriate test variable, T, which is a function of both data and parameters. For example, if we are concerned about the yearly swings as discussed at the end of Section 5.2, we could calculate the mean square national prediction error:

$$T(y,\Phi) = \frac{1}{11} \sum_{t\in(1948,1952,\ldots,1988)} \left[\frac{1}{\sum_i w_{it}} \sum_{i\in(AL,\ldots,WY)} w_{it} \left(y_{it} - (X\boldsymbol{\beta})_{it}\right) \right]^2$$

(recall that w_{it} is the turnout in state i and election year t). Then, we can look at a scatter plot of $T(y_{rep}, \Phi)$ vs $T(y, \Phi)$ for our entire set of simulations of Φ. If the model fits well, the points should cluster around the 45 degree line. We can regard the percentage below the line as a p-value (Gelman, Meng and Stern, 1994 and Rubin, 1984) for formally testing the fit of the model.

We carried out this computation for the three NH models and three HIER models that were originally introduced in this paper. The scatter plots appear in Figure 4. The p-value is essentially 0 for the three NH models, and is acceptable for the HIER models, telling us that the year to year variability has been captured well by the HIER models.

FIGURE 4 ABOUT HERE

5.5 Sensitivity analysis

Given our set of simulations from the posterior distribution of the parameters for an original model, it is not difficult to get approximate posterior simulations from a perturbed model (i.e. with different likelihood or prior assumptions). To do this, we use the importance resampling (SIR) algorithm (Rubin, 1987; see also Gelman, 1992, and Smith and Roberts, 1993). For each of the original simulations we calculate its importance weight; i.e. the ratio of the (unnormalized) perturbed and original posterior densities. We then choose a subsample without replacement from the original set of simulations with sampling probabilities proportional to the importance weights. This subsample is approximately from the perturbed posterior distribution.

We began by changing the prior distribution assumption on θ from uniform on the unit interval to Beta$(1/2, 1/2)$. A summary of the changes appears in Tables 8 and 10 Since we already know that the models favor θ values of less than $1/2$, changing to this new prior model will tend to favor even smaller values of θ. Indeed for all five models, the posterior quantiles of θ are shifted down slightly, and the predictions became slightly more like the EV results (Tables 9 and 11).

TABLE 8 ABOUT HERE
TABLE 9 ABOUT HERE
TABLE 10 ABOUT HERE
TABLE 11 ABOUT HERE

Next we changed the prior distribution on the τ_j from uniform on τ_j^2 to uniform on τ_j. This will lead us to favor lower values for these variance parameters as can be seen in Table 10. The new predictions are in Table 11.

The most important thing to notice is that the predictions are quite robust to either change in the prior density. A more complete robustness analysis would perturb either the normal likelihood or normal prior density on the β's. We examined the residuals from the regression and found no outliers, suggesting that the normal likelihood model is acceptable.

6 Discussion

Heteroscedasticity was an easy thing to add in the Bayesian context, with little added computational effort, once we decided on the weights and specified the priors. Putting in the hierarchical model is the most important thing, but the heteroscedastic model adds some model fit and makes the predictions more trustworthy. The Bayesian model allows us to fit a parametric model for unequal variances even in situations in which the parameter is not well identified, thus allowing us to use information encoded in the weights w_i in a more general and flexible setting than weighted least squares.

NOTE: The code and data used to do the computations in this paper are available from the authors upon request.

7 References

Amemiya, T. (1985) *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B* **26**, 211–252.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics.

Campbell, J. E. (1992). Forecasting the Presidential vote in the states. *American Journal of Political Science* **36**.

Carroll, R. J. and Ruppert, D. (1988). *Transformations and Weighting in Regression*. New York: Chapman and Hall.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**, 1–38.

Fair, R. C. (1978). The effect of economic events on votes for President. *Review of Economics and Statistics* **60**, 159–173.

Fair, R. C. (1982). The effect of economic events on votes for President: 1980 Update. *Review of Economics and Statistics* **64**, 322–325.

Fair, R. C. (1988). The effect of economic events on votes for President: 1984 Update. *Political Behavior* **10**, 168-179.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A. (1992). Iterative and non-iterative simulation algorithms. *Computing Science and Statistics* **24**, 433–438.

Gelman, A., Meng, X. L., and Stern, H. S. (1994). Bayesian model checking using tail area probabilities. Under revision for *Statistica Sinica*

Gelman, A., Roberts, G. O., and Gilks, W. R. (1994). Efficient Metropolis jumping rules. Presented at Bayesian Statistics 5, Valencia, Spain.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.

Gelman, A., and King, G. (1993). Why are American Presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science* **23**, 409–451.

Greene, W. H. (1990). *Econometric Analysis.* New York: Macmillan.

Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association* **60**, 806–825.

Judge, G., Hill, C., Griffiths, H., Lutkepohl, H., and Lee, T. (1982). *Introduction to the Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.

Judge, G., Hill, C., Griffiths, H., Lutkepohl, H., and Lee, T. (1985). *The Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society B* **34**, 1–41.

Rosenstone, S. J. (1990). Predicting elections. Unpublished manuscript. Department of political science, University of Michigan.

Rosenstone, S. J. (1983). *Forecasting Presidential Elections.* New Haven: Yale University Press.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. §5. *Annals of Statistics* **12**, 1151–1172.

Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Comment on "The calculation of posterior distributions by data augmentation," by M. A. Tanner and W. H. Wong. *Journal of the American Statistical Association* **82**, 543–546.

Smith, A. F. M. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B* **55**, 3–23.

Tierney, L. (1994). Markov Chains for exploring posterior distributions. Tech report 560. Department of Statistics. University of Minnesota.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.

	BAYES		GLS		True Parameter
	$E(\cdot\|y)$	$SD(\cdot\|y)$	Estimate	SE	**Value**
β_1	1.689	6.836	1.010	7.090	10
β_2	1.641	0.415	1.657	0.417	1
β_3	0.902	0.365	0.896	0.375	1
α_1	-2.338	2.202	-4.376		-3
α_2	0.275	0.104	0.366		0.3

Table 1: Comparison of Posterior Distribution with GLS Estimates for the simulated data example of Judge et al. (1982).

	NHEV			NHHET1			NHHET2		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
Constant	22.08	24.67	27.24	22.30	24.99	27.49	22.25	24.88	27.56
TRIHEAT	0.42	0.48	0.53	0.42	0.47	0.53	0.42	0.47	0.53
7ECxINC	2.02	2.36	2.69	2.03	2.37	2.69	2.05	2.38	2.72
PRESINC	1.34	2.20	3.05	1.23	2.09	2.98	1.20	2.11	2.97
APPRxINC	-0.02	-0.01	0.00	-0.02	-0.01	0.00	-0.02	-0.01	0.00
DEV3	0.27	0.35	0.41	0.28	0.34	0.41	0.26	0.33	0.40
DEV2	0.20	0.26	0.32	0.20	0.25	0.31	0.20	0.25	0.31
HOME4	2.30	3.92	5.72	2.02	3.57	5.13	2.20	3.78	5.37
VP4	0.39	1.88	3.54	0.41	1.82	3.35	0.36	1.90	3.48
LEGIS	0.03	0.05	0.07	0.03	0.05	0.07	0.03	0.05	0.07
EC11xINC	0.01	0.10	0.18	0.01	0.10	0.19	0.01	0.10	0.19
ADAACA	0.02	0.03	0.04	0.02	0.03	0.05	0.02	0.04	0.05
CATH60	0.05	0.14	0.22	0.06	0.14	0.23	0.05	0.14	0.22
S4	5.63	7.58	9.29	5.73	7.48	9.28	5.68	7.51	9.29
-S64B	4.65	7.81	11.10	4.44	7.50	10.82	4.34	7.57	10.73
-DEMS64	12.70	18.11	23.23	13.22	18.28	23.33	12.85	18.03	22.96
NEWENG64	4.30	6.73	9.16	4.18	6.58	9.02	4.09	6.54	8.87
-W76	4.95	6.74	8.63	4.76	6.66	8.54	4.77	6.76	8.55
NCENT72	4.01	6.31	8.82	3.71	6.26	8.81	3.90	6.28	8.81
SIG	3.48	3.70	3.95	3.47	3.68	3.93	3.46	3.68	3.94
THETA				0.02	0.13	0.24	0.02	0.10	0.22

Table 2: 2.5%, median and 97.5% quantiles of the posterior distributions of the parameters for the NH models. All values are obtained by posterior simulation.

		NHEV			NHHET1			NHHET2		
	Actual	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
AL	46.1	46.7	53.7	62.2	46.8	54.3	61.2	46.6	54.1	60.9
AK	43.8	37.4	44.5	51.7	37.1	45.3	53.2	36.9	45.3	54.5
AZ	48.7	36.3	43.2	51.1	36.9	43.9	51.7	36.7	44.0	51.2
AR	60.0	51.3	60.0	67.9	52.3	60.3	68.0	52.9	60.4	67.4
CA	59.5	44.9	52.1	59.8	46.1	52.5	59.0	46.0	52.4	59.1
CO	52.6	39.8	47.2	54.5	40.3	47.7	55.0	40.5	47.7	54.8
CN	53.8	43.9	51.2	58.8	44.3	51.7	58.7	44.7	51.7	58.8
DE	55.0	40.3	47.7	55.1	39.6	48.1	55.5	39.3	48.2	55.7
FL	48.8	44.4	51.9	59.7	45.4	52.6	59.6	45.5	52.4	59.6
GA	50.6	46.0	53.9	61.4	47.4	54.5	61.7	47.6	54.6	61.6
HA	57.0	49.0	56.6	64.2	49.3	57.1	64.7	49.1	56.9	65.2
ID	40.3	32.6	40.1	47.2	33.1	40.9	49.1	33.3	40.7	48.7
IL	57.8	44.1	51.2	58.8	44.8	51.7	58.5	44.7	51.7	58.9
IN	46.2	38.2	45.1	53.1	38.5	45.7	52.7	38.3	45.6	52.8
IO	53.7	47.0	54.4	61.6	47.3	54.8	62.1	47.0	54.5	61.5
KS	46.6	37.8	44.6	51.9	37.2	44.8	52.3	37.4	44.9	51.9
KY	51.7	41.4	48.7	55.9	41.5	48.9	56.2	41.8	49.1	56.6
LA	52.3	47.9	55.9	63.3	49.5	56.4	63.5	49.1	56.2	63.9
ME	55.7	43.1	50.2	58.0	43.5	50.9	58.7	43.6	51.0	59.2
MD	58.1	46.5	54.2	62.0	47.3	54.6	62.1	47.7	54.6	61.5
MA	62.3	48.0	55.6	62.7	48.9	55.9	62.8	48.7	56.0	63.0
MI	54.3	42.4	49.6	56.6	43.3	50.0	56.7	43.6	50.1	57.0
MN	57.9	46.2	53.4	61.0	46.6	54.0	61.3	47.0	54.1	61.0
MS	45.1	45.2	52.7	61.2	45.6	53.0	60.6	45.9	52.9	60.4
MO	56.4	43.3	50.5	57.5	43.4	51.0	57.8	43.8	50.8	57.6
MT	51.4	43.0	50.7	58.1	43.3	51.1	58.9	43.2	51.0	59.1
NB	39.0	32.5	40.0	47.2	32.8	40.6	48.5	33.2	40.5	48.2
NV	52.1	36.8	43.9	51.8	36.5	44.3	52.4	36.7	44.2	51.9
NH	50.6	37.0	43.8	51.6	36.2	44.1	51.9	37.0	44.6	52.2
NJ	51.2	42.3	49.6	57.6	43.1	50.1	56.7	43.6	50.2	57.2
NM	54.8	42.3	49.6	56.9	42.0	49.9	57.7	42.0	49.8	57.6
NY	59.5	46.3	53.7	60.7	47.6	54.4	61.3	47.4	54.3	60.8
NC	49.4	45.9	53.1	61.0	46.2	53.7	61.1	46.3	53.4	61.0
ND	42.1	38.3	45.6	52.7	38.0	46.1	53.7	38.3	46.0	54.6
OH	50.6	40.5	48.3	56.0	42.3	48.9	55.7	42.2	49.0	55.5
OK	44.2	37.7	45.0	53.0	38.4	45.4	53.0	38.3	45.5	52.8
OR	57.3	46.2	53.1	60.6	46.3	53.6	61.2	46.3	53.5	61.2
PA	55.6	45.0	52.3	59.7	46.0	52.5	59.2	45.7	52.5	59.4
RI	62.3	51.7	59.0	66.2	51.0	59.5	66.9	51.2	59.3	67.2
SC	45.5	44.5	51.6	58.7	44.5	52.2	59.8	44.6	52.0	59.5
SD	47.4	40.5	47.3	54.5	40.0	47.9	56.5	39.8	47.8	56.1
TN	52.2	49.2	56.8	64.4	49.5	57.2	65.0	49.8	57.1	64.6
TX	48.1	43.6	51.4	59.8	45.8	52.6	60.1	44.9	52.6	60.1
UT	36.1	31.9	39.1	46.2	31.8	39.7	47.3	32.4	39.8	47.1
VT	59.7	42.7	50.5	57.9	43.5	51.1	59.5	42.9	50.9	59.7
VA	47.7	44.4	52.1	59.6	45.1	52.2	59.9	44.8	52.3	59.2
WA	58.7	45.9	53.3	60.5	46.9	53.8	60.6	46.9	53.8	60.5
WV	57.6	47.4	54.2	61.4	47.2	54.9	62.4	47.2	54.6	61.9
WS	52.6	45.4	53.1	60.4	46.4	53.8	60.9	46.3	53.6	60.6
WY	45.9	34.8	42.0	49.5	34.2	42.5	50.9	33.2	42.7	51.1
ELECCOLL	370.0	227.0	336.0	423.0	252.0	362.0	441.0	255.0	360.0	435.0
POPVOTE	53.6	49.8	51.7	53.5	50.2	51.9	53.8	50.0	51.9	53.7

Table 3: 2.5%, median and 97.5% quantiles of the predictions for 1992 for NH models

	HIEREV			HIERHET1			HIERHET2		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
Constant	8.63	26.63	43.47	11.02	27.28	42.88	10.74	27.01	42.74
TRIHEAT	0.09	0.44	0.82	0.09	0.42	0.77	0.10	0.43	0.78
7ECxINC	0.40	2.60	4.71	0.56	2.59	4.81	0.61	2.62	4.81
PRESINC	-3.68	2.04	7.76	-3.90	1.96	7.68	-3.74	2.04	7.65
APPRxINC	-0.09	-0.02	0.06	-0.09	-0.01	0.06	-0.09	-0.01	0.06
DEV3	0.26	0.34	0.42	0.25	0.33	0.41	0.23	0.31	0.39
DEV2	0.13	0.20	0.27	0.12	0.19	0.26	0.12	0.19	0.26
HOME4	2.03	3.70	5.37	1.69	3.15	4.60	2.02	3.52	5.09
VP4	0.11	1.85	3.43	0.29	1.77	3.30	0.31	1.86	3.47
LEGIS	0.02	0.05	0.07	0.02	0.05	0.07	0.03	0.05	0.08
EC11xINC	-0.01	0.08	0.17	-0.01	0.08	0.18	-0.02	0.07	0.17
ADAACA	0.03	0.05	0.06	0.04	0.05	0.06	0.04	0.05	0.07
CATH60	0.11	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.29
S4	3.32	6.72	10.26	2.98	6.50	9.95	3.12	6.58	9.99
SIG	3.37	3.60	3.85	3.34	3.56	3.81	3.35	3.57	3.82
TAU1	0.61	2.58	7.43	0.37	2.54	7.04	0.36	2.55	6.99
TAU2	1.64	2.46	3.63	1.59	2.34	3.56	1.62	2.38	3.61
TAU3	2.86	4.85	9.22	2.71	4.73	8.65	2.73	4.76	8.73
THETA				0.09	0.22	0.36	0.03	0.12	0.28

Table 4: 2.5%, median, and 95% quantiles of the posterior distributions of the parameters for the HIER models (random effects omitted)

		HIEREV			HIERHET1			HIERHET2		
	Actual	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
AL	46.1	36.0	52.5	68.9	36.8	52.4	68.5	35.2	52.1	69.3
AK	43.8	32.9	45.9	58.8	32.5	45.8	60.2	32.2	46.0	60.0
AZ	48.7	30.5	43.8	56.4	30.7	43.5	56.2	30.1	43.6	56.6
AR	60.0	42.4	59.0	76.1	41.6	58.0	74.6	41.7	58.5	76.8
CA	59.5	38.5	52.3	65.5	40.0	52.1	64.1	38.2	52.1	64.6
CO	52.6	34.5	47.7	60.4	34.6	47.3	60.8	33.5	47.4	60.6
CN	53.8	39.5	52.1	65.6	40.0	52.2	65.3	37.8	52.0	66.1
DE	55.0	34.5	47.6	60.1	34.5	47.1	61.3	32.2	47.5	61.4
FL	48.8	34.3	51.2	67.7	35.3	51.1	67.6	33.7	51.2	69.0
GA	50.6	36.1	53.0	69.5	37.1	52.9	68.3	35.3	52.3	70.7
HA	57.0	44.0	57.6	70.8	43.5	57.4	71.6	42.8	57.4	71.7
ID	40.3	27.7	40.8	54.2	27.3	40.9	54.7	26.5	41.0	54.5
IL	57.8	38.2	51.5	65.1	39.3	51.4	63.9	37.1	51.1	63.9
IN	46.2	32.7	45.4	58.7	33.2	45.5	58.6	32.2	45.4	58.5
IO	53.7	40.4	54.2	67.5	41.2	53.7	66.8	40.2	53.5	67.5
KS	46.6	31.0	44.5	57.8	32.0	43.9	57.2	30.3	44.1	57.2
KY	51.7	33.2	48.8	65.7	33.6	48.8	65.4	32.0	48.4	66.3
LA	52.3	38.7	54.7	71.5	37.9	54.6	70.8	37.9	54.6	73.1
ME	55.7	38.3	51.4	63.6	38.6	51.5	64.7	37.9	51.8	65.4
MD	58.1	41.5	54.3	67.0	41.6	54.7	66.8	40.3	54.2	67.5
MA	62.3	43.2	56.0	68.4	43.6	56.1	68.7	42.6	56.4	70.0
MI	54.3	37.2	50.2	62.9	37.9	50.1	62.6	36.7	50.1	63.3
MN	57.9	40.7	54.2	67.4	41.7	53.8	66.2	41.3	53.8	67.3
MS	45.1	34.6	51.1	67.6	34.5	50.8	68.5	34.1	50.4	69.5
MO	56.4	37.8	50.8	63.7	38.5	50.4	63.1	37.0	50.7	64.5
MT	51.4	38.1	51.5	64.1	37.8	51.5	64.9	36.9	51.4	64.9
NB	39.0	26.9	39.8	53.3	26.9	39.8	53.7	26.2	39.5	53.9
NV	52.1	31.3	44.1	56.7	30.0	44.3	57.9	29.9	44.0	58.9
NH	50.6	30.9	44.4	57.0	32.3	44.7	56.9	31.0	44.4	58.9
NJ	51.2	37.5	50.6	63.3	38.4	50.7	62.5	36.8	50.7	64.6
NM	54.8	35.9	49.8	62.7	36.5	49.2	62.8	34.8	49.2	62.2
NY	59.5	40.8	54.4	67.1	42.6	54.4	66.3	41.1	54.1	66.7
NC	49.4	35.3	52.2	68.4	34.9	51.9	67.9	34.5	51.4	69.4
ND	42.1	32.9	45.8	59.2	33.1	46.1	59.5	32.0	45.8	60.7
OH	50.6	35.4	48.4	61.0	35.9	48.2	60.3	35.3	48.1	61.6
OK	44.2	29.6	45.3	62.2	30.1	45.7	62.4	28.2	45.2	62.8
OR	57.3	40.2	53.6	67.2	40.6	53.5	66.3	39.3	53.7	66.5
PA	55.6	39.5	52.4	65.5	40.0	52.1	65.5	38.4	51.9	65.0
RI	62.3	46.5	59.7	72.4	46.3	59.5	72.8	45.3	59.5	73.2
SC	45.5	33.3	50.8	67.7	34.9	50.4	66.4	32.9	49.9	67.4
SD	47.4	34.5	48.0	61.0	34.7	47.9	61.8	32.7	47.6	61.4
TN	52.2	38.5	55.8	71.9	40.0	55.4	72.6	38.3	55.2	72.5
TX	48.1	34.5	51.4	67.3	36.0	51.6	67.8	33.5	50.9	69.1
UT	36.1	26.9	40.1	53.2	26.8	40.1	52.4	26.2	40.4	52.6
VT	59.7	37.7	51.1	63.6	37.5	50.8	64.9	36.7	51.1	65.4
VA	47.7	34.2	50.5	67.0	35.2	50.1	67.5	33.2	50.1	68.5
WA	58.7	41.0	53.8	66.6	41.0	53.5	66.3	40.8	53.7	67.1
WV	57.6	40.8	54.7	66.5	41.7	54.6	67.8	40.8	54.4	67.9
WS	52.6	40.1	53.5	66.6	40.7	53.3	66.3	39.3	53.4	66.6
WY	45.9	29.4	42.5	55.9	29.0	42.2	56.3	27.0	42.4	57.4
ELECCOLL	370.0	16.0	324.0	526.0	20.0	326.0	526.0	13.0	314.0	525.0
POPVOTE	53.6	40.8	51.5	62.2	41.7	51.4	62.3	39.8	51.3	62.9

Table 5: 2.5%, median and 97.5% quantiles of the predictions for 1992 for HIER models

	HIERHET1			HIERHETALL		
	2.5%	50%	97.5%	2.5%	50%	97.5%
SIG	3.34	3.56	3.81	3.02	3.21	3.42
TAU1	0.37	2.54	7.04	0.69	2.10	5.48
TAU2	1.59	2.34	3.56	1.23	1.83	2.78
TAU3	2.71	4.73	8.65	0.90	2.20	4.76
THETA	0.09	0.22	0.36	0.03	0.16	0.28

Table 6: Posterior quantiles of variance components for HIERHET1 and HIERHETALL

		HIERHET1			**HIERHETALL**		
	Actual	2.5%	50%	97.5%	2.5%	50%	97.5%
CA	59.5	40.0	52.1	64.1	42.7	52.4	63.6
NJ	51.2	38.4	50.7	62.5	39.7	50.6	61.4
RI	62.3	46.3	59.5	72.8	48.9	59.8	71.8
VA	47.7	35.2	50.1	67.5	39.4	51.7	63.6
ELECCOLL	370.0	20.0	326.0	526.0	39.0	354.0	529.0
POPVOTE	53.6	41.7	51.4	62.3	43.4	51.9	61.9

Table 7: Prediction summary for HIERHET1 and HIERHETALL

	Original			Beta(.5, .5)		
	2.5%	50%	97.5%	2.5%	50%	97.5%
NHHET1						
SIG	3.47	3.68	3.93	3.46	3.68	3.94
THETA	0.02	0.13	0.24	0.00	0.11	0.23
NHHET2						
SIG	3.46	3.68	3.94	3.46	3.69	3.93
THETA	0.02	0.10	0.22	0.01	0.09	0.21

Table 8: Variance component sensitivity analysis comparing original and new prior densities on theta for NH models

		Original			Beta(.5, .5)		
	Actual	2.5%	50%	97.5%	2.5%	50%	97.5%
NHHET1							
CA	59.5	46.1	52.5	59.0	46.2	52.7	59.6
NJ	51.2	43.1	50.1	56.7	43.1	50.2	56.9
RI	62.3	51.0	59.5	66.9	52.1	59.1	67.5
VA	47.7	45.1	52.2	59.9	45.1	52.5	59.5
ELECCOLL	370.0	252.0	362.0	441.0	263.0	368.0	437.0
POPVOTE	53.6	50.2	51.9	53.8	50.2	52.0	53.6
NHHET2							
CA	59.5	46.0	52.4	59.1	46.1	52.3	59.1
NJ	51.2	43.6	50.2	57.2	43.6	50.1	56.6
RI	62.3	51.2	59.3	67.2	51.9	59.4	66.5
VA	47.7	44.8	52.3	59.2	44.7	52.2	59.1
ELECCOLL	370.0	255.0	360.0	435.0	243.0	360.0	440.0
POPVOTE	53.6	50.0	51.9	53.7	49.9	51.9	54.0

Table 9: Prediction sensitivity analysis comparing original and new prior densities on theta for NH models

	Original			$\theta \sim$ Beta(.5, .5)			$\tau_j \propto 1$		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
HIERHET1									
SIG	3.35	3.56	3.80	3.35	3.56	3.81	3.36	3.56	3.80
TAU1	0.84	2.49	7.63	0.92	2.50	6.96	0.66	2.11	5.41
TAU2	1.60	2.35	3.55	1.59	2.33	3.58	1.60	2.38	3.40
TAU3	2.84	4.76	9.15	2.90	4.74	9.46	2.62	4.44	8.19
THETA	0.09	0.22	0.35	0.08	0.21	0.35	0.09	0.22	0.35
HIERHET2									
SIG	3.36	3.58	3.80	3.37	3.58	3.79	3.36	3.58	3.80
TAU1	0.79	2.61	9.88	0.75	2.61	9.70	0.61	2.09	5.73
TAU2	1.61	2.36	3.62	1.63	2.40	3.69	1.64	2.40	3.54
TAU3	2.71	4.69	9.40	2.65	4.60	8.80	2.56	4.27	8.01
THETA	0.03	0.12	0.25	0.02	0.10	0.24	0.03	0.12	0.25

Table 10: Variance component sensitivity analysis comparing original and new prior densities for HIER models

		Original			**Beta**(.5, .5)			$\tau_j \propto 1$		
	Actual	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
HIERHET1										
CA	59.5	40.0	52.1	64.1	41.0	52.1	64.6	42.9	52.0	62.5
NJ	51.2	38.4	50.7	62.5	38.3	50.4	64.0	40.0	50.3	60.8
RI	62.3	46.3	59.5	72.8	47.1	59.6	72.9	47.5	59.3	70.9
VA	47.7	35.2	50.1	67.5	33.4	50.1	66.4	35.4	50.6	65.0
ELECCOLL	370.0	20.0	326.0	526.0	31.0	315.0	527.0	57.0	325.0	510.0
POPVOTE	53.6	41.7	51.4	62.3	41.8	51.3	62.3	43.6	51.2	59.5
HIERHET2										
CA	59.5	38.2	52.1	64.6	41.0	51.4	62.8	41.0	51.4	62.8
NJ	51.2	36.8	50.7	64.6	38.1	50.5	60.6	38.1	50.5	60.6
RI	62.3	45.3	59.5	73.2	47.5	59.6	70.6	47.5	59.6	70.6
VA	47.7	33.2	50.1	68.5	36.6	50.0	65.1	36.6	50.0	65.1
ELECCOLL	370.0	13.0	314.0	525.0	11.0	316.0	532.0	24.0	314.0	511.0
POPVOTE	53.6	39.8	51.3	62.9	39.7	51.2	63.9	42.4	51.0	59.9

Table 11: Prediction sensitivity analysis comparing original and new priors on theta for HIER models

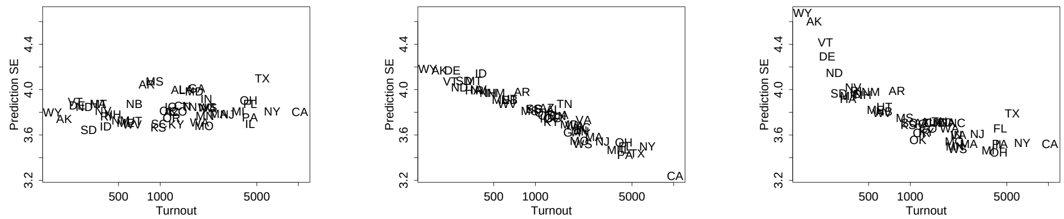

Figure 1: Standard error of predictions versus turnout (in 1000's on a log scale) for a) NHEV, b) NHHET1 and c) NHHET2

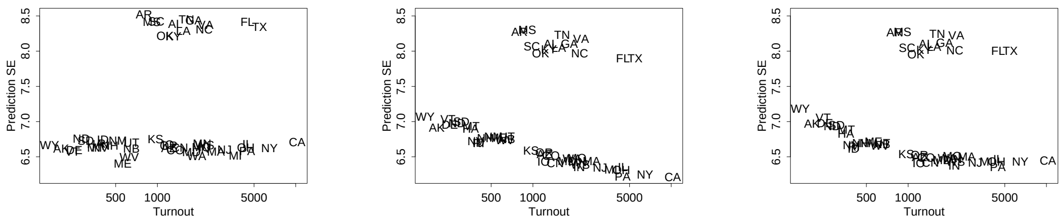

Figure 2: Standard error of predictions versus turnout (in 1000's on a log scale) for a) HIEREV, b) HIERHET1 and c) HIERHET2

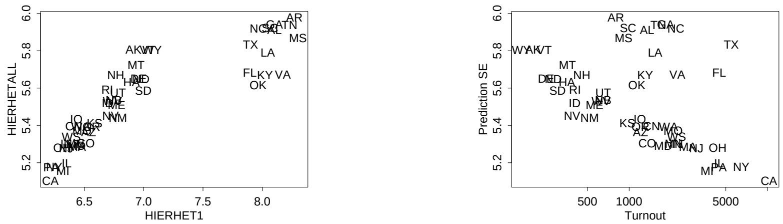

Figure 3: Standard error of predictions for HIERHETALL vs a) SE's for HIERHET1 and b) turnout

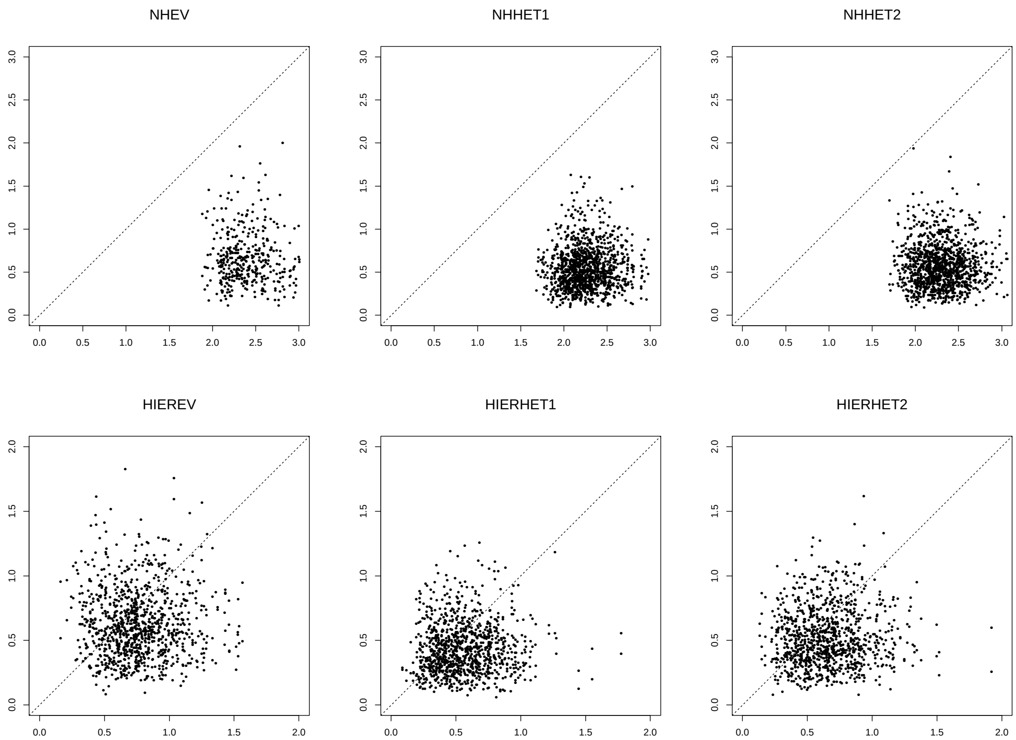

Figure 4: Model checking: scatter plots of $T(y_{rep}, \Phi)$ vs $T(y, \Phi)$ under different models. $T(y, \Phi)$ is the mean square error in the national vote compared to the model. The top three plots indicate that $T(y, \theta)$ is consistently higher than $T(y_{rep}, \Phi)$ for the NH models, showing that these models do not accurately capture the variation in the national vote. The HIER models fit much better in this respect.