Poststratification Without Population Level Information on the Poststratifying Variable, With Application to Political Polling

Cavan REILLY, Andrew GELMAN, and Jonathan KATZ

We investigate the construction of more precise estimates of a collection of population means using information about a related variable in the context of repeated sample surveys. The method is illustrated using poll results concerning presidential approval rating (our related variable is political party identification). We use poststratification to construct these improved estimates, but because we do not have population level information on the poststratifying variable, we construct a model for the manner in which the poststratifier develops over time. In this manner, we obtain more precise estimates without making possibly untenable assumptions about the dynamics of our variable of interest, the presidential approval rating.

KEY WORDS: Bayesian inference; Poststratification; Sample surveys; State-space models.

1. INTRODUCTION

Poststratification is widely recognized as an effective method for obtaining more accurate estimates of population quantities in the context of survey sampling (Little, 1993). Not only does it correct for nonsampling error, but it can lead to less variable estimates. The basic idea is that if we know our population is composed of distinct groups (strata) that differ with regard to the quantity we are interested in estimating and we know the sizes of these strata in our population, then we can obtain a more accurate estimate of the quantity of interest by correcting for any imbalance in the representation of the strata in our sample. This correction is obtained by using a weighted average (using the known weights from the population) of the averages within strata as our estimate of the population mean. If we calculate the variance of this estimate conditional on the observed number of respondents falling into each of the strata (as is generally recommended; see Holt and Smith 1979), the variance of this estimate will be a linear combination of the variance of the strata means. Hence, the estimate could have zero variance (if group membership exactly determines the quantity of interest), but in practice our gains will depend on how strongly our quantity of interest is related to the variable(s) we use to poststratify. Although poststratification is not always used in academic studies, it is a commonplace tool in commercial public opinion polls (Voss, Gelman, and King 1995).

One of the greatest practical limitations to the use of poststratification is the need to know the proportion of the population in each strata. We have population level information only for certain variables, so it appears that poststratification is useful only if our quantity of interest is related to one of a handful of characteristics for which we have population level information. Here, we overcome this difficulty by constructing a dynamic model for the variable by which we poststratify, thereby estimating the strata weights from our sample. The dynamic model for the poststratifier allows for more efficient estimation of the weights for each time period than would be possible if we analyzed each sample separately. Clearly, if the method for obtaining the samples does not change over time, we cannot hope to correct for sampling bias if we estimate our weights. Hence, here we use poststratification solely to obtain more efficient estimates. Note that we are not required to propose any dynamic model for the quantity of interest, only for the poststratifier. Because we are free to select the poststratifier, we try to choose a variable that is related to the quantity of interest and has dynamic behavior that is relatively well understood (for example, the variable is basically constant over time).

1.1 Structure of the Data and Preliminary Considerations

We analyze data from a (self-weighted) sample survey of U.S. adults, the "WISCON" project, from the Letters and Science Survey Center at the University of Wisconsin at Madison. For each respondent, we have his or her rating of the president on a scale of 1 to 10, the party with which he or she most closely identifies (which we group into one of three categories, Democrat, Republican, or Independent, based on the respondent's answer to two questions about party identification), and the date of the interview. We group each respondent by the week in which he or she was interviewed so as to have a sequence of samples of these quantities (i.e., the approval rating within each party and the size of each party in our sample) from the week starting 1/19/93 until the week starting 8/13/96 (which constitutes most of Clinton's first term). We are ultimately interested in estimating the mean approval rating of the president for each week, μ_t for $t = 1, \ldots, T$, given all of the data up to time T. The weekly samples collect information from about 40–60 respondents (we do not try to estimate the mean approval rating for weeks with too few interviews; hence, we exclude several weeks, leaving a total of 171 weeks of data), and so a natural estimate of μ_t (and a basis for comparison for any other method) is the sample

Cavan Reilly is Assistant Professor, Department of Biostatistics, University of Minnesota, Minneapolis, MN 55454. Andrew Gelman is Professor, Department of Statistics, at Columbia University, New York, NY 10027 (E-mail: *gelman@stat.columbia.edu*). Jonathan Katz is Associate Professor of Political Science, D.H.S.S., California Institute of Technology, Pasadena, CA 91125 (E-mail: *katz@cattech.edu*). C. Reilly and A. Gelman thank the NSF for grant SBR-97-08424 and Young Investigator Award DMS-97-96129, and J. Katz thanks the John M. Olin Foundation.

Journal of the American Statistical Association
March 2001, Vol. 96, No. 453, Applications and Case Studies

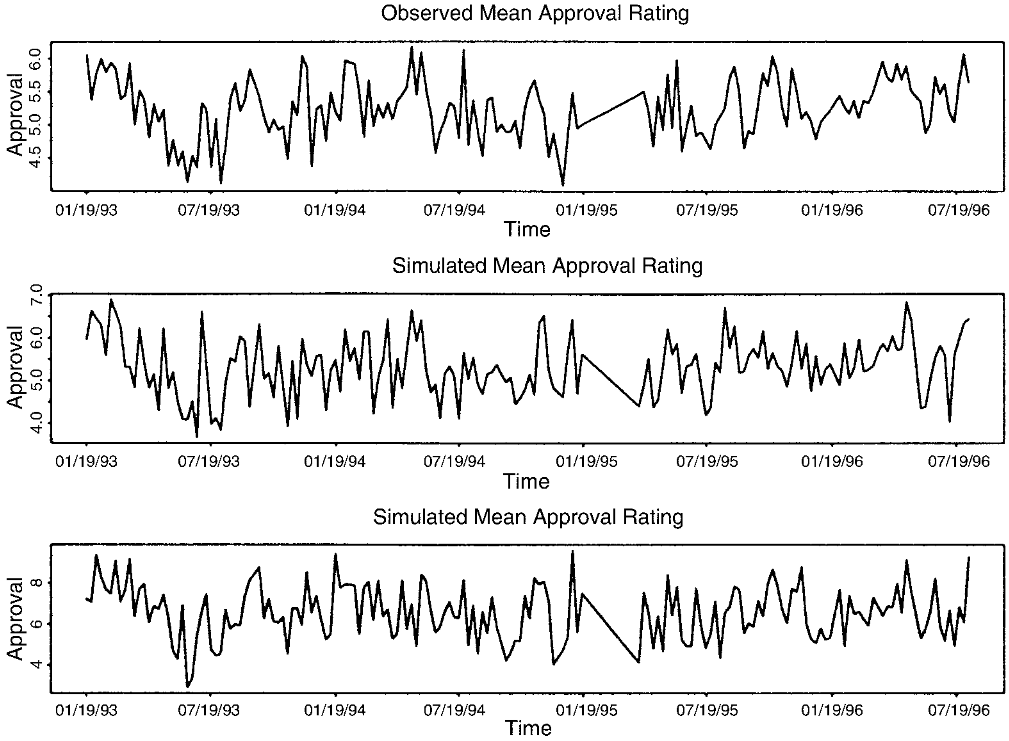

Figure 1. Observed Mean Approval Rating and Two Simulations of the Mean Approval Rating Under the Model.

mean with standard error given by the sample standard deviation divided by the square root of the sample size at time t (moreover, because the sample sizes are large, we can appeal to the central limit theorem to conclude that the distribution of the sample mean is approximately normal).

The top plot in Figure 1 displays the mean approval rating for all of the weeks. We suppose that these sample means are independent over time because they are based on independent random samples. Whereas our dynamic model for the weights is a Bayesian model (as we see subsequently), it is useful to note that using the sample mean (based on samples large enough for the central limit theorem to take effect) with the aforementioned standard error as an estimate of μ_t is equivalent in Bayesian terms to assuming a normal distribution for the sample mean given μ_t and σ_t (where σ_t is the standard deviation of the approval ratings at time t), using a normal prior for μ_t with arbitrarily large variance and using the sample standard deviation as an estimate for the unknown quantity σ_t.

That is, if we let n_{1t} denote the number of Democrats in our sample at time t, n_{2t} denote number of Republicans in our sample at time t, and n_{3t} denote the number of Independents in our sample at time t, and we set $N_t = \sum_j n_{jt}$ for $t = 1, \ldots, T$, then we find the posterior distribution of μ_t by supposing that for $t = 1, \ldots, T$ we have $\overline{y}_t | \mu_t, \sigma_t, N_t \sim \mathrm{N}(\mu_t, \sigma_t^2/N_t)$ and $\mu_t | \sigma_t \sim \mathrm{N}(\mu_0, \sigma_0^2)$, where $\overline{y}_t$ is the mean approval rating for our sample at time t and we take σ_0 to be arbitrarily large. This implies $\mu_t | \overline{y}_t, \sigma_t, N_t \sim \mathrm{N}(\overline{y}_t, \sigma_t^2/N_t)$. We estimate σ_t^2 with the usual unbiased estimate, s_t^2, and so we obtain draws from the posterior distribution of μ_t using the normal distribution in a completely straightforward manner (so we are ignoring the fact that the sample variance is subject to variability). Later, we treat the mean approval rating within each party μ_{jt} for $j = 1, 2, 3$ in the same manner, and we assume that the approval rating is independent across parties. In this sense we have no dynamic model for the approval rating within party. Our intention is to poststratify presidential approval by political party identification and show that by correcting our estimate for imbalances in political party representation, we can obtain a more efficient estimate. Although it is difficult to propose a dynamic model for approval, it is reasonable to suppose that the proportion of a population that holds a given political attitude is almost constant from week to week.

1.2 A Simple Model and Method

As a simple investigation into the efficacy of this method, we use the average over all time periods of the proportion of our sample in each party for the strata weights, and treat these weights as known (Little, 1996). This is equivalent to the dynamic model, which supposes that the proportion in each party is constant over time (and we ignore the uncertainty in the estimation of the weights, an entirely reasonable practice because these averages of sample proportions are sample averages based on $\sum_t N_t = 8{,}462$ observations). If we use these averages for the weights for all the weeks and treat them as known, then we can estimate the efficiency of our poststratification estimate relative to the sample mean for each week by the ratio of the variance of the estimated sample mean s_t^2/N_t to the estimated variance of the poststratification estimate at time t. We find that these estimated efficiencies range from 0.48 to 2.8 with an average of 1.23. The correlation between approval and party identification is about .35 (treating party identification as continuous), so we see that even a weak correlation can be useful. These results are in accord with the findings of others (e.g., Holt and Smith 1979), in particular $\overline{y}_{\text{PS}}$ often has lower variance (and here, on average, has lower variance), but sometimes the sample mean is preferable. Although this simple method indicates that the poststratification estimator can outperform the sample mean (on average here), a model that assumes the strata proportions are constant over time (3.5 years) is not very reasonable (see, for example, MacKuen 1983). A more plausible model is provided in the next section, but we note that in some settings this analysis may be satisfactory.

1.3 Political Polling and the Presidency

Presidential approval has been a central concept in the study of both presidential power and public opinion in political science. With the advent of the "new presidency" in the age of mass media politics, having high levels of approval is seen as an important political resource for presidents (Kernell 1986). Having high levels of approval is thus a central component of presidential power (Neustadt 1990) and influences electoral outcomes and legislative success (Rivers and Rose 1985; Ostrom and Smith 1993; Brody 1991).

Since the early 1970s, a long list of studies have examined various presidential approval series, although the series from the Gallup Organization is most common because it is available starting with the Truman administration. In general, these studies have examined how the percentage of the population that approves of the job of the current president varies with economic conditions and "rally events," such as armed conflict or political scandal (see, inter alia, Kernell 1978; MacKuen 1983; Norpoth and Yantek 1983; Kiewiet and Rivers 1985; Ostrom and Simon 1989; Brace and Hinckley 1991; Brody 1991; Beck 1991, 1992; Clarke and Stewart 1994). More recent work on presidential approval has paid particular attention to the dynamics of presidential approval. The consensus is that approval within the population is highly persistent from month to month, but there has been some debate on how best to model this persistence (see Smith 1992; Williams 1992; Box-Steffensmeier and Smith 1998).

2. MODELS AND POSTERIOR SIMULATION

2.1 Parameterization of a Categorical Poststratification Variable as a Multivariate Outcome

Because we do not have population level information on party identification, to effectively poststratify we first posit a model for the temporal evolution of the party identification series. Rather than directly model the two series n_{1t} and n_{2t} (the number of respondents in each party), we first transform our data so that we model a vector with components that are approximately independent. The approximate independence thereby induced should make our inference less sensitive to our model for the covariance structure utilized in our dynamic model of the proportions. For the political party identification series, we model the proportion of respondents who identify with one of the two major parties, and the proportion of those who identify with the Democrats among those who identify with one of the major parties. So, if we let $n_t = n_{1t} + n_{2t}$ and define the two-vector $y_t = (n_t/N_t, n_{1t}/n_t)$, then, because N_t and n_t are large, it is reasonable to suppose that y_t has a bivariate normal distribution [for the derivations that follow we adopt the convention that $y_t = (0, 0)$ if $n_t = 0$]. If we let θ_{1t} denote the proportion of the population that is in one of the major parties (i.e., Democrat or Republican) and let θ_{2t} denote the proportion of Democrats among those in a major party, then the measurement covariance (i.e., sampling error) of y_t given θ_{1t}, θ_{2t}, and N_t under simple random sampling (ignoring finite population correction factors) can be expressed as

$$V_t^* = \begin{pmatrix} \theta_{1t}(1-\theta_{1t})/N_t & \theta_{1t}\theta_{2t}(1-\theta_{1t})^{N_t} \\ \theta_{1t}\theta_{2t}(1-\theta_{1t})^{N_t} & \frac{\theta_{2t}(1-\theta_{2t})}{N_t}\sum_{j=0}^{\infty}\frac{\mu_j(\theta_{1t},N_t)}{N_t^j} \\ & +\theta_{2t}^2(1-\theta_{1t})^{N_t}(1-(1-\theta_{1t})^{N_t}) \end{pmatrix},$$

where $\mu_j(\theta, N) = \sum_{k=0}^{N-1}\binom{N}{k}k^j(1-\theta)^k\theta^{N-k}$. We obtain this expression by noting that, conditional on N_t, θ_{1t}, and θ_{2t} if we use 1_A to represent the indicator function of the set A, then [if we use the convention that $1_{\{n_{3t}<N_t\}}/(N_t - n_{3t})$ is zero when $N_t = n_{3t}$ in the second line] if y_{jt}, $j = 1, 2$, is the jth element of y_t,

$$\begin{aligned}\operatorname{Var}(y_{2t}) &= \mathrm{E}(\operatorname{Var}[y_{2t}|n_{1t}+n_{2t}]) + \operatorname{Var}(\mathrm{E}[y_{2t}|n_{1t}+n_{2t}]) \\ &= \mathrm{E}\left[1_{\{n_{1t}+n_{2t}>0\}}\frac{\theta_{2t}(1-\theta_{2t})}{N_t-n_{3t}}\right] + \operatorname{Var}(1_{\{n_{1t}+n_{2t}>0\}}\theta_{2t}) \\ &= \frac{\theta_{2t}(1-\theta_{2t})}{N_t}\sum_{j=0}^{\infty}N_t^{-j}\mathrm{E}(1_{\{n_{1t}+n_{2t}>0\}}n_{3t}^j) \\ &\quad + \theta_{2t}^2(1-\theta_{1t})^{N_t}(1-(1-\theta_{1t})^{N_t}),\end{aligned}$$

from which we obtain the element on the second diagonal of V_t^* and the other elements are straightforward. Although we could substitute our sample proportions, y_{jt}, for the unknown population proportions, θ_{jt}, in this expression and thereby obtain an estimate of the measurement covariance matrix (using 20 terms in the infinite sums is more than sufficient to obtain seven digit accuracy, and one or two terms is probably adequate for most practical purposes), we instead use the simple approximation to the desired estimate (which is good to within 1% of the desired estimate of the standard error of y_{2t} and is obviously good for the off-diagonal element because N_t is large and θ_{1t} is at least .7),

$$V_t = \begin{pmatrix} y_{1t}(1-y_{1t})/N_t & 0 \\ 0 & y_{2t}(1-y_{2t})/n_t \end{pmatrix}.$$

We treat these measurement variances as known in our analysis.

2.2 Dynamic Model for the Poststratifying Variable

Given V_t and N_t for $t = 1, \ldots, T$, and the initial conditions m_0 and C_0, we propose a state-space model for $t = 1, \ldots, T$,

$$\begin{aligned} y_t &= \theta_t + \nu_t, \quad \text{where} \quad \nu_t \sim \mathrm{N}(0, V_t), \\ \theta_t &= \theta_{t-1} + \omega_t, \quad \text{where} \quad \omega_t \sim \mathrm{N}(0, W), \\ \theta_0 &\sim \mathrm{N}(m_0, C_0), \end{aligned}$$

where $\{\nu_t\}$ and $\{\omega_t\}$ are mutually orthogonal sequences of independent disturbances. We treat the matrix W as a random variable and estimate it from the data. This model is motivated by the fact that political attitudes in the contemporary United States do not change much over the course of a single week. For known W, this is a special case of a model for which the Kalman filter can be used to obtain the posterior moments of the state vectors θ_t for $t = 0, \ldots, T$ (see, e.g., West and Harrison 1997).

2.3 Analytic Expressions for Posterior Inference

To obtain samples from the posterior distribution of the weights for our poststratification estimate, we first obtain samples from the posterior distribution of the state process in our dynamic model given all of the data up to time T, but because we do not know W, we suppose this is a (matrix-valued) random variable and conduct Bayesian inference for this matrix. Our goal is first to simulate W from its marginal posterior distribution, and then to simulate the state vectors, θ_t, given W; that is, we use the fact $p(\theta, W|y) = p(\theta|W, y)p(W|y)$, where $\theta = (\theta_0, \theta_1, \ldots, \theta_T)$ and $y = (y_1, \ldots, y_T)$. These results can be given a non-Bayesian interpretation as predictive inference for θ conditional on a marginal likelihood estimate of W.

We find the posterior distribution of the state vectors given the state covariance matrix W by using standard formulas from the Kalman filter. Now, under our model, we have (by the Kalman filter)

$$\theta_t | y_1, \ldots, y_t, W \sim \mathrm{N}(m_t, C_t)$$

with

$$\begin{aligned} m_t = {} & V_t(C_{t-1} + W + V_t)^{-1} m_{t-1} \\ & + (C_{t-1} + W)(C_{t-1} + W + V_t)^{-1} y_t \end{aligned}$$

and

$$\begin{aligned} C_t = {} & C_{t-1} + W - (C_{t-1} + W) \\ & \times (C_{t-1} + W + V_t)^{-1}(C_{t-1} + W) \end{aligned}$$

for $t = 1, \ldots, T$. Hence it is elementary to show

$$p(\theta|W, y) = \mathrm{N}(\theta_T | m_T, C_T) \prod_{t=1}^{T} \mathrm{N}(\theta_{t-1} | h_{t-1}, H_{t-1}),$$

where

$$h_t = W(C_t + W)^{-1} m_t + C_t(C_t + W)^{-1}\theta_{t+1}$$

and

$$H_t = C_t - C_t(C_t + W)^{-1} C_t'$$

for $t = 0, \ldots, T-1$.

We can obtain the marginal posterior density of the state covariance matrix by writing the likelihood for y as a function of W, that is, $y_t = \sum_{s=1}^{t} \omega_s + \theta_0 + \nu_t$, and so

$$p(W|y) = p(W) \prod_{t=1}^{T} \mathrm{N}(y_t | m_0, tW + C_0 + V_t).$$

In this manner we obtain the posterior distribution of the state covariance matrix once we determine an appropriate prior. We take $p(W) \propto 1$ [so that our posterior mode coincides with the maximum likelihood estimator (MLE) of W, treating θ as a nuisance].

2.4 Other Modeling Issues

In light of the previous development, simulation is relatively straightforward, but we must attend to some details. For example, a minor complication is the fact that we have no (or insufficient) data for some weeks and so our time series has unequal time increments (so in the previous development, W should have been a function of t). The simple remedy is to realize that because we assumed that θ_t follows a random walk, if it has been k weeks since we last obtained survey results, and the state covariance matrix is W (i.e., the covariance matrix of an increment of the state process based on one week of data is W), then the covariance matrix of the state process over an increment of k weeks is kW. For our dataset and the way in which we use the Kalman filter, this correction has no discernible impact on our results. We also must specify initial values m_0 and C_0 for the Kalman filter. Based on rough guesses, we set $m_0 = (.8, .5)$, and to convey our lack of accurate information on these quantities, we make C_0 a diagonal matrix with elements $.2^2$. With 171 weeks of data, the specification of the initial values has little impact on our estimation of the state process θ_t for $t = 1, \ldots, T$ and has no practical impact on our poststratification estimator (this was verified experimentally by altering m_0 and C_0).

2.5 Computation

2.5.1 Posterior Simulation of the Poststratification Proportions. We use the Metropolis algorithm to obtain draws from $p(W|y)$. Then we draw θ from the appropriate sequence of normal distributions. Our methodology follows that outlined in Gelman, Carlin, Stern, and Rubin (1995): Our candidate distribution is a multivariate normal with variance based on the curvature of the posterior at the mode (and we scale this matrix so that the proportion of accepted jumps is in the 40% range), and we use multiple sequences that begin from overdispersed starting points (which were selected by drawing deviates from a properly centered and scaled Student's t distribution with 4 degrees of freedom). We used four sequences of 10,000 iterations, and the resulting values of the convergence diagnostic statistic, $\sqrt{\hat{R}}$, were all less than 1.1.

Given W, it is completely straightforward to simulate θ. Note that we do not require iterative simulation to simulate θ. We simply use draws from the bivariate normal distribution with mean and covariance matrix given by h_t and H_t, because the joint distribution of the state vectors was found previously; that is, we use the forward filtering, backward sampling algorithm of Carter and Kohn (1994) and Frühwirth-Shnatter (1994).

2.5.2 Simulation of the Mean Within Each Poststratification Category and the Poststratified Estimate of the Population Mean. We estimate the mean within each party in the same manner that we estimated the mean approval without regard to party identification; hence, it is trivial to obtain simulations of μ_{jt}. To obtain draws from the posterior distribution of the poststratification estimate, we assume that the approval rating within each party is conditionally independent of the proportion of the population in each of the parties given the sample means within parties and the number of respondents in each party. Therefore, we simulate a draw from the posterior distribution of μ_t^{PS} by simply combining simulations from both parts of the foregoing model in the obvious fashion, namely, if we let $\pi_{1t} = \theta_{1t}\theta_{2t}$, $\pi_{2t} = \theta_{1t}(1-\theta_{2t})$, and $\pi_{3t} = 1-\theta_{1t}$, then we obtain $\mu_t^{\text{PS}} = \sum_{j=1}^{3} \mu_{jt}\pi_{jt}$.

2.5.3 Comments on Computations. This method for obtaining draws from the posterior distribution of θ by averaging over our uncertainty in the estimation of the state covariance matrix can be generalized to deal with any unknown parameters in the usual Gaussian linear Kalman filter, such as unknown autoregressive coefficients in state-space autoregressions or unknown variance components in dynamic regression models. For example, we tried fitting first order state-space autoregressions with unknown state variances and unknown autoregressive coefficients to the two series y_{1t} and y_{2t} separately using this methodology (with only two parameters, we were able to obtain simulations for the autoregressive coefficient and the state-space variance by discretizing the bivariate posterior distribution and using the inverse cdf method; see, for example, Gelman et al. 1995). Whereas the autoregressive coefficients were definitely very close to 1 (as we expect with such low values of the state variances), we ignored the complication that the autoregressive coefficient matrix might be different from the identity matrix in our model for y_t (because this would augment the dimension of the state space of our Markov chain by 3 in the implementation of the Metropolis algorithm). In any event, we see how simple our approach to unknown model parameters can be. Indeed, no iterative simulation is required at all for these low dimensional problems. The advantage of this technique for averaging over our uncertainty in the model parameters compared to simply using the Gibbs sampler to simulate the state process given the model parameters and then simulate the model parameters given the state process (as is frequently done; see, e.g., West and Harrison 1997) is that no iterative simulation is required for the state vectors in our method. This is a great simplification, because adjacent state vectors are highly correlated in their joint posterior distribution; hence convergence of the chain can be difficult to obtain if we must use an iterative simulation method to simulate the state vectors. This posterior correlation is especially troubling for typical filtering applications, because hundreds (or even thousands) of state vectors may be involved. For our application, this means that we need to obtain draws only from the equilibrium distribution of a 3-dimensional Markov chain, rather than a 345-dimensional Markov chain.

In the sample survey literature, difficulty using the MLE of the state variance has been reported when the series is short (see, e.g., Pfeffermann 1991). In such cases, averaging over the uncertainty in the estimation of the state variance in the preceding manner should eliminate these problems. In particular, with short series, the MLE of the state variance occasionally will be zero (even if the data were produced by a mechanism with a nonzero state variance), but because this point estimate is subject to uncertainty, if we average over the uncertainty of the estimated state variance, we will find that the Kalman filter can still lead to more accurate inference without implying that the level of the process is constant. Moreover, if we have information about the state covariance matrix (or any parameters in the more general linear Gaussian model), we can incorporate this information through a prior on W (rather than taking the flat prior $p(W) \propto 1$ as we have here). With short series, the judicious use of such prior information can lead to more reliable inference, because the posteriors of the model parameters may be quite diffuse if we use flat priors.

2.6 An Alternative Model for the Time Series of Poststratification Proportions

The model described in the previous sections was not the first model we fit to these data. The first model we fit follows the approach to multinomial time series developed in Cargnoni, Müller, and West (1997). We did not end up using this model because we found that it did not fit our data (see Section 3.2.2); however, we present it here for completeness and because it might be useful in other settings. Using the same notation as before, if we let $\pi_t = (\pi_{1t}, \pi_{2t}, \pi_{3t})$, then we first assume that for $t = 1, \ldots, T$,

$$n_{1t}, n_{2t}, n_{3t} | N_t, \pi_t \sim \text{Mult}(N_t, \pi_t).$$

Next, let $\eta_{jt} = \text{logit}(\theta_{jt})$ for $j = 1, 2$. These transformations separate partisan changes from changes in affiliation within the two largest parties and change scale in such a way that additive models are more reasonable (they also yield diagonal measurement covariance matrices, as we saw before). Now we define the vector $\eta_t = (\eta_{1t}, \eta_{2t})$ and we suppose that for $t = 1, \ldots, T$,

$$\eta_t = \xi_t + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{N}(0, V),$$
$$\xi_t = \xi_{t-1} + \delta_t, \quad \text{where} \quad \delta_t \sim \text{N}(0, W^*),$$

where $\{\epsilon_t\}$ and $\{\delta_t\}$ are mutually orthogonal sequences of independent disturbances. We finish our specification of the dynamics of π_t by supposing $\xi_0 | m_0^*, C_0^* \sim \text{N}(m_0^*, C_0^*)$. In addition, we suppose that V and W^* are random variables (matrices) and we specify inverse Wishart priors with scale equal to the identity matrix and 2 degrees of freedom in the hope of obtaining a prior that has little impact on our inference (a hope that is realized, as we see by experimentation). This model implies that the dynamics of the vector η_t are basically equivalent to a vector process that follows an autoregressive integrated moving average (ARIMA) (0, 1, 1) model. The values of the moving average parameters in the equivalent ARIMA(0, 1, 1) model are determined by V and W^* (for more on this equivalence, see West and Harrison 1997). Although we may be tempted to set $V = 0$ in the hope of obtaining an efficient algorithm for simulating draws from a model that specifies that the transformed proportions follow a vector random walk, this will not work because, if we use the sampling algorithm of Cargnoni et al. (1997), we iteratively sample from two conditional distributions that degenerate into point masses as V approaches zero (thus no mixing takes place for the parameters of interest). We then can draw samples from the posterior distribution of all parameters in our model for the party identification series using the Metropolis–Hastings algorithm as explained in Cargnoni et al. (1997). To assess convergence, we used four independent sequences that began from overdispersed starting points (and the general methodology presented in Gelman et al. 1995). To obtain overdispersed starting points for our example, we conducted a preliminary run of 1,000 iterations and then used two times the medians for the variance parameters as our starting values for these parameters, whereas for the η_{jt}'s, we used the medians of the values obtained from this trial run as our starting values. By specifying unrealistically large values for the variance parameters, we got the sampler to spread out the values of ξ_t and η_t in the first iteration in a way that would be very difficult to do "by hand," because there are over 370 initial values that we must supply. The convergence of the chains was rapid: after a burn-in of 2,000 iterations, the next 1,000 were saved and all of the values of the $\sqrt{\hat{R}}$ statistic were less than 1.02.

2.7 Model Criticism

Our models do not attempt to represent every conceivable facet of the phenomenon under investigation, and so it is essential to understand the shortcomings of our models. A simple, yet sensitive, method for detecting model weaknesses is to use the model to simulate another dataset and then compare the simulated data to the observed data (posterior predictive checks; see, e.g., Gelman et al. 1995). The first step is to examine several of the simulated datasets graphically. After this, one can design test statistics and compare the distribution of these test statistics under the posterior predictive distribution to their distribution under the posterior distribution (if a test statistic does not depend on any of the model parameters, it is constant under the posterior distribution). In the time series modeling context, several natural test statistics can be proposed on general grounds. First, if our series is x_t for $t = 1, \ldots, T$, then the average absolute value of the change in the level of the series $T_1(x_1, \ldots, x_T) = (1/(T-1)) \sum_{t=2}^{T} |x_t - x_{t-1}|$ is a simple measure of the volatility of the series (if our fitting method smoothes the data too much, then T_1 will be too large under the posterior predictive distribution). If ϕ_t is the forecast of x_t conditional on the observed data, another natural diagnostic is the average of the absolute value of the prediction error, $T_2(x_1, \ldots, x_T, \phi_1, \ldots, \phi_T) = (1/T) \sum_{t=1}^{T} |x_t - \phi_t|$. If the fitting method smooths too much, the prediction errors will be too large on average. Although obtaining analytic expressions for these quantities is a daunting task, it is simple to draw simulations of these quantities from the appropriate distributions.

3. RESULTS FOR OUR EXAMPLE

3.1 Fitting the Normal Theory Model

Figure 2 shows the marginal posterior distribution of the components of W and the correlation between the elements of the state vectors, based on 40,000 simulation draws from the Metropolis algorithm. Figure 3 displays 95% probability intervals for the proportion in each party obtained by the model (these intervals are laid over the sample proportions), and Figure 4 shows posterior predictive draws of the sample proportions. In Figure 5, we show the 95% confidence intervals for the average approval rating within each party, and Figure 6 displays the 95% probability intervals given by our poststratification estimate and 95% confidence intervals based on the sample mean (whose construction was given in the Introduction, but, of course, no simulation was used here). From the last graph, we see that our poststratification estimator is more precise than the sample mean.

To more fully understand how the poststratification estimator works, it is instructive to see if our estimator really does respond to imbalances in the representation of the parties within our samples. To examine this consider Figure 7. From these graphs, we easily see that if the proportion of Democrats relative to the proportion of Republicans in our sample is too large (relative to the estimate based on our dynamic model), then our poststratification estimator will have a tendency to make the estimated approval rating smaller than the raw estimate (based on the sample mean). The same correction is made if there are too many Democrats in our sample (but the relative proportion of Democrats to Republicans is seen to be more important in determining the correction), and the opposite correction is made if there are too many Republicans. This is exactly the sort of behavior we expect, because Clinton is a Democrat. From Figure 8, we see that the poststratification estimate performs best for moderate sized samples (again, each dot represents one week of data in all of the plots). We also see that the largest corrections are for the smaller samples (as we would expect) and that the size of the correction does not have much to do with the estimated efficiency. Last, the fact that our state-space model for the party identification series is actually a hierarchical model for the increments of the state-space process is manifested in the shrinkage of the increments of our poststratification estimate (as witnessed in the lower right hand corner of Fig. 8).

3.2 Model Checking

3.2.1 Checking the Fit of Our Basic Model. The normal theory Kalman filter model presented herein seems acceptable for our purposes. Figure 4, shows a draw from the posterior predictive distribution for the number of respondents falling into each of the parties, and the lower two panels of Figure 1 displays two draws from the posterior predictive distribution for the average approval rating for each week. We obtain a draw from the posterior predictive distribution of the average approval rating by using a weighted mean of draws from approval within party, with weights given by the simulated sample proportions in each party under the posterior predictive distribution for these proportions. We find the observed value of T_1, where

$$T_1(n_{1,1}, \ldots, n_{1,T}) = \frac{1}{T-1} \sum_{t=2}^{T} \left| \frac{n_{1,t}}{n_t} - \frac{n_{1,t-1}}{n_{t-1}} \right|$$

is .089, and the 95% probability interval for T_2, where

$$T_2(n_{1,2}, \ldots, n_{1,T}, \theta_{2,1}, \ldots, \theta_{2,T-1}) = \frac{1}{T-1} \sum_{t=2}^{T} \left| \frac{n_{1,t}}{n_t} - \theta_{2,t-1} \right|$$

under the posterior distribution is (.065, .072). We find that 95% probability intervals for these two quantities based on 1,000 simulation draws from their posterior predictive distributions under the normal theory model are (.076, .099)

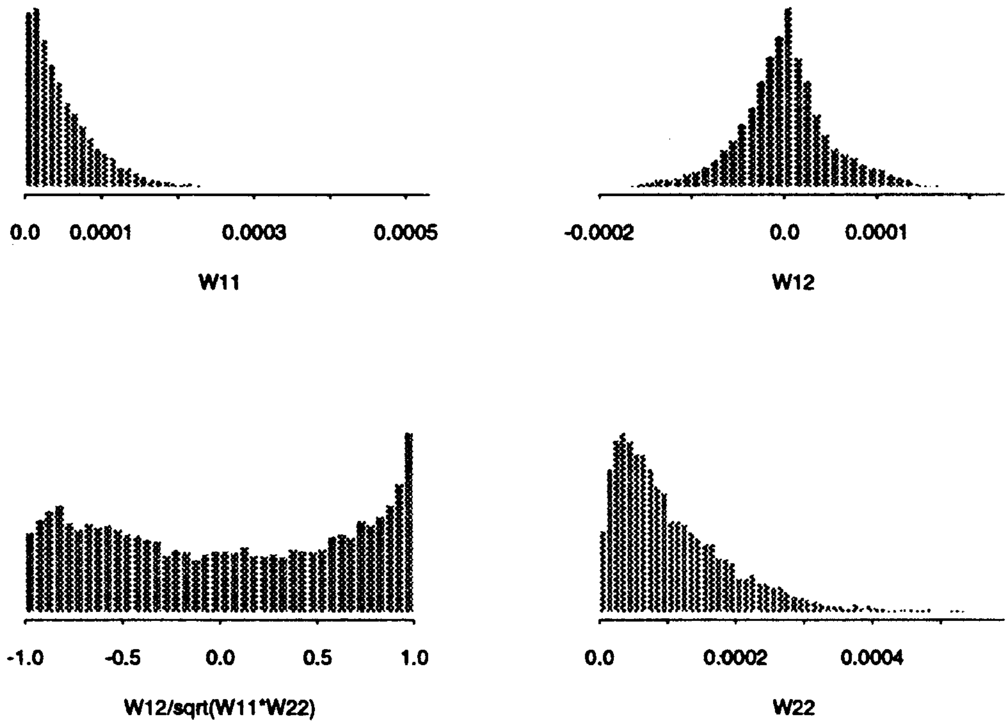

Figure 2. The Marginal Posterior Distribution of Each Element of the State Covariance Matrix Assuming a Flat Prior. The lower left plot shows the marginal posterior distribution of the correlation of the states.

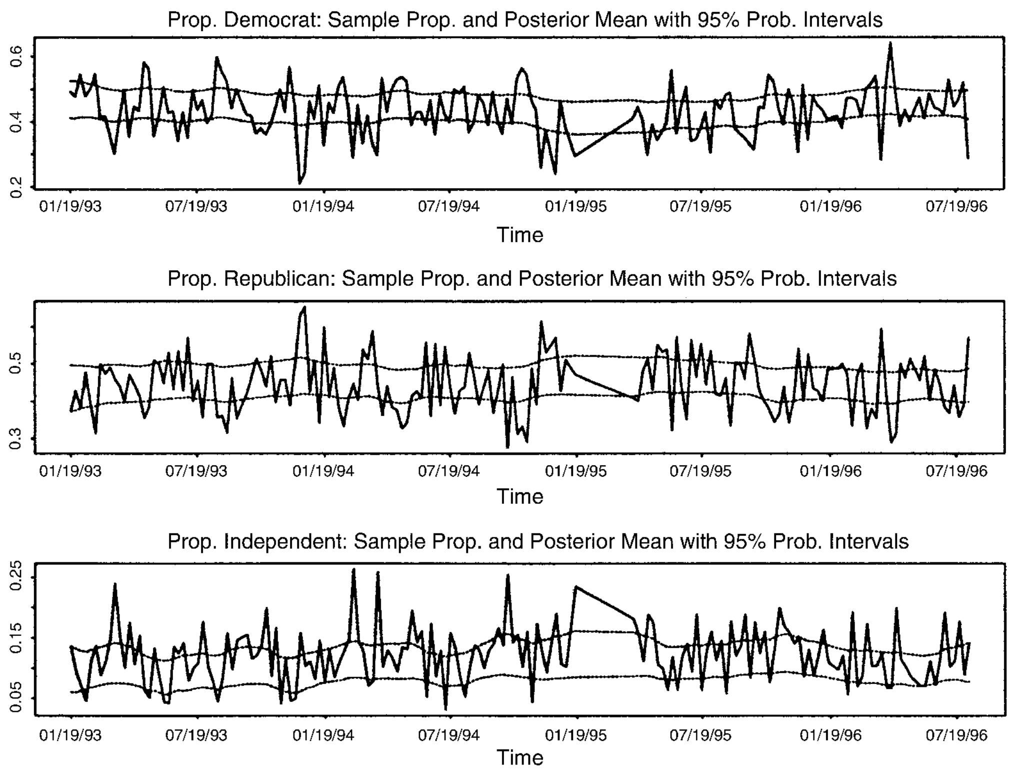

Figure 3. The Proportion in Each Party for All Weeks With 95% Probability Intervals Given by the Model.

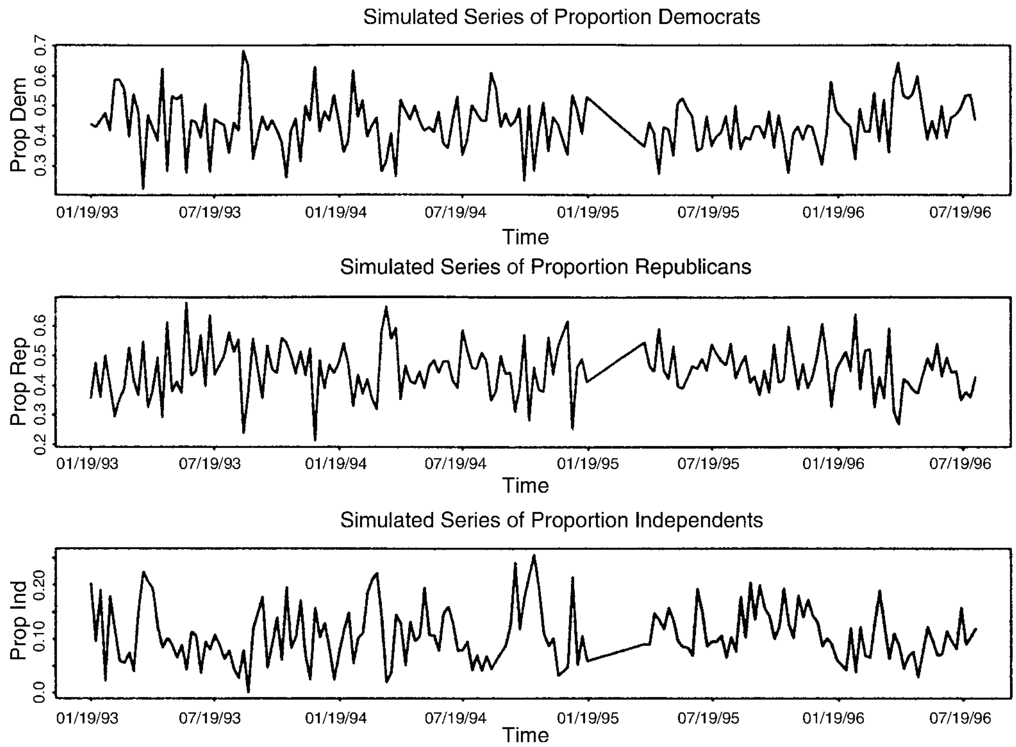

Figure 4. Simulated Sample Proportions for Each Week Under the Model. Compare to Figure 3.

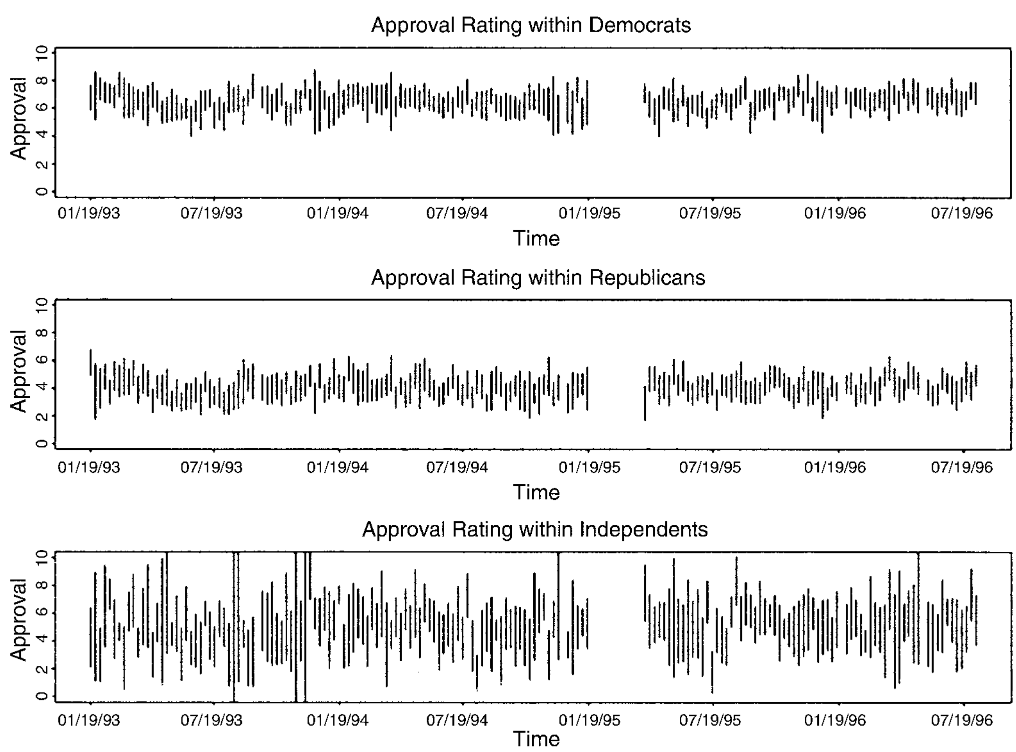

Figure 5. 95% Confidence Intervals for the Mean Approval Rating Within Party.

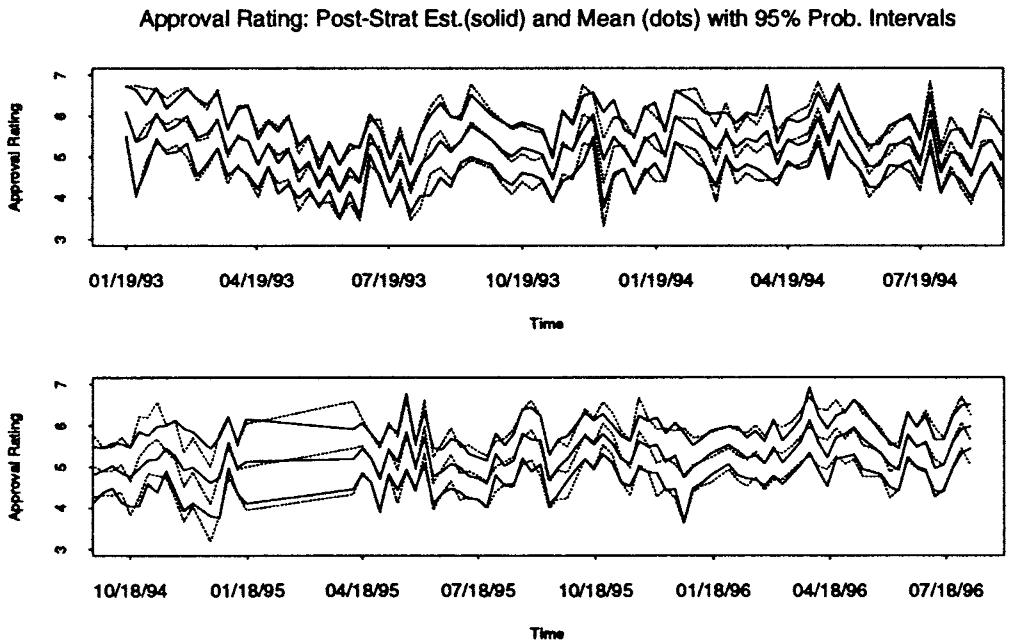

Figure 6. 95% Probability Intervals for the Mean Approval Rating for Each Week Based on the Poststratification Estimate (solid line) and Based on the Sample Mean (dotted line). The series is broken up to fit on one page.

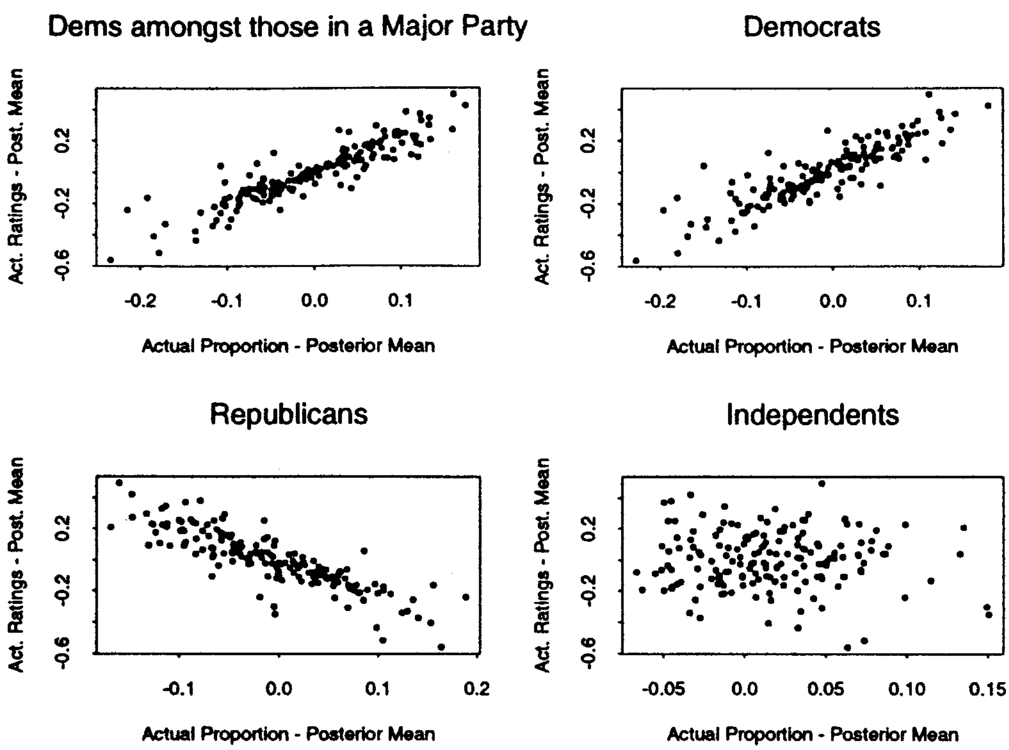

Figure 7. The Difference in the Observed Proportions and the Posterior Means by the Difference in the Observed Ratings and the Posterior Means for Subsets of the Samples. The poststratification estimate corrects for unequal representation of the parties in our samples. Each dot represents one week.

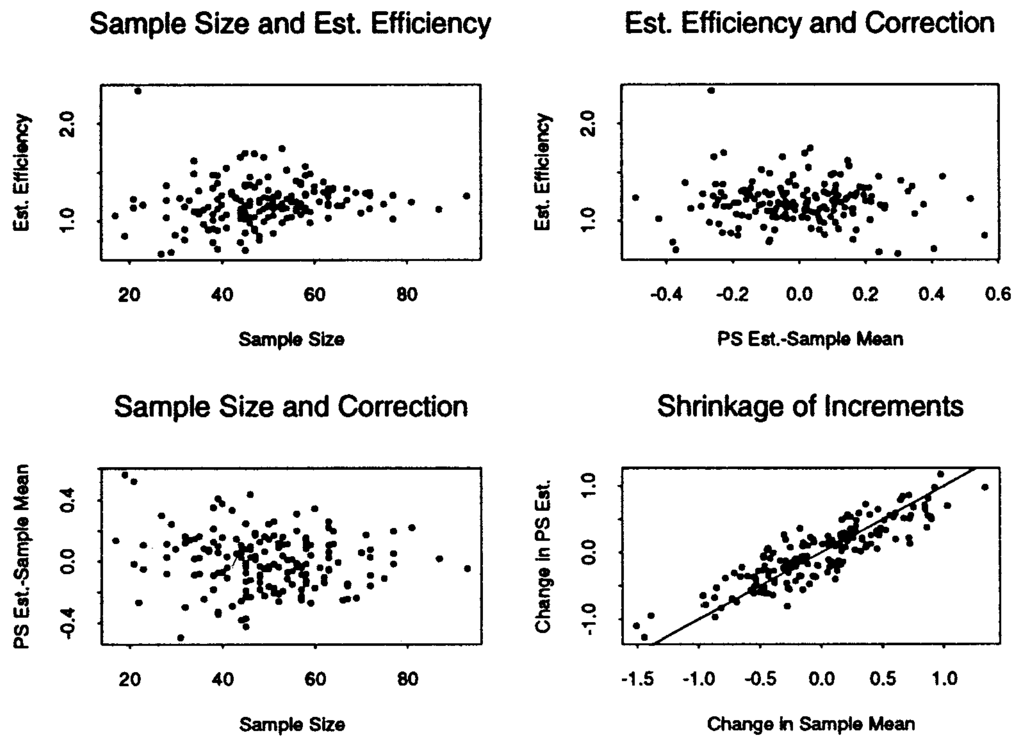

Figure 8. The Poststratification Estimate Performs Best for Moderate Sized Samples. The line in the plot that illustrates the shrinkage of the increments is a y = x line. Each dot represents one week.

and (.056, .069), and we find a 95% probability interval for the difference $T_2(n^k_{1,2}, \ldots, n^k_{1,T}, \theta^k_{2,1}, \ldots, \theta^k_{2,T-1}) - T_2(n_{1,2}, \ldots, n_{1,T}, \theta^k_{2,1}, \ldots, \theta^k_{2,T-1})$ (where $n^k_{1,t}$ is the draw from the posterior predictive distribution corresponding to θ^k_t from the posterior distribution for $t = 1, \ldots, T$ and $k = 1, \ldots, 1000$), is $(-.014, .002)$. These posterior predictive checks indicate our normal theory model fits these aspects of the data.

3.2.2 Checking the Fit of Our Alternative Model. Once we examine our simulations for the proportion in a major party and the proportion of those in a major party who are Democrats based on the multinomial model, it appears that the posterior medians of these variables are too variable. The normal theory model for the party identification series is actually based only on a subset of the data we used to fit the multinomial model (our multinomial model was fit to data that included a portion of Bush's presidency), and so our observed value of T_1 is not the same as before (and we do not expect T_2 under the posterior distribution to be the same as before). Based on 1,000 posterior predictive samples, we found that a 95% probability interval for T_1 under the multinomial model is (.115, .147), whereas our observed value is .097. For our other test statistic, T_2, we find a 95% probability interval based on the posterior predictive distribution is (.102, .129), whereas a 95% probability interval for T_2 based on the posterior distribution is (.089, .107). We also find that a 95% probability interval for the difference, $T_2(n^k_{1,2}, \ldots, n^k_{1,T}, \theta^k_{2,1}, \ldots, \theta^k_{2,T-1}) - T_2(n_{1,2}, \ldots, n_{1,T}, \theta^k_{2,1}, \ldots, \theta^k_{2,T-1})$ is (.005, .028). These shortcomings indicate that the model is overfitting (i.e., this model does not smooth the series of proportions enough). It is difficult to construct a simpler model for the party identification series within the context of the model proposed by Cargnoni et al.(1997), and so we chose to use the model based on the normal theory Kalman filter for the sample proportions.

4. CONCLUSIONS

The resulting estimates are more precise than the weekly sample means (the estimated efficiencies ranging from .66 to 2.3 with a mean of 1.19). If we consider the cost of obtaining survey data (because many questions are asked of each respondent, it can take 30 minutes to complete an interview), this is a great savings (with 8,462 observations, it is like getting over 1,600 more observations for free). If one has a long series for the quantity of interest, it may be feasible to identify an appropriate time series model for the quantity of interest. In such a case, one could base estimates on this model and obtain substantially more precise estimates (for example, one may be able to conclude that a random walk plus error model describes the movement of the series of interest over time). One advantage of this poststratification estimate is that we are not required to propose a dynamic model for the quantity of interest: we need only a dynamic model for some quantity that is related to our quantity of interest. This is a great help here because specification of a dynamic model for a volatile variable (like approval rating) is controversial, whereas the slowly changing nature of political attitudes implies that models that allow for almost constant levels are suitable for separating measurement error from shifts in attitudes. Also, the results from our model for the party identification series can be used to construct poststratified estimates for other variables. In this manner, a short series can be poststratified by using simulations based on a more extensive dataset, and thereby more precise estimates are obtained.

The failure of the multinomial model led us to consider other sorts of state-space models for discrete variables. The fact that state-space models are hierarchical models for the increments of the state process suggests that one can treat discrete variable filtering problems (by filtering, we also refer to the associated problems of smoothing and prediction) exactly like random effects generalized linear models (on which there is an extensive literature, ranging from analytic approximations to several methods of posterior simulation); see the comments by Meyer in West, Harrison and Mignon (1985). Because adjacent states will have high posterior correlation, it seems sensible to parameterize the state process in terms of the increments of the state process rather than the levels of the process (this should yield a sampling algorithm that converges faster than one that samples the levels of the state process). This reparameterization is quite natural when the filtering problem is treated as a random effects generalized linear model.

There are also many approximations for filtering and smoothing in the time series literature (see, for example, West et al. 1985). These approximations provide reasonable initial values for iterative methods or, of course, can be used as estimates themselves. If we are going to use approximate smoothing methods, a convenient way to obtain an approximation to the marginal likelihood of any model parameters, ϕ (e.g., state variances or autoregressive coefficients), is to use a formula common in the random effects literature (see, for example, Rubin 1981 or Besag 1989), namely

$$p(\phi|y) \propto \frac{p(y|\theta, \phi)p(\theta, \phi)}{p(\theta|\phi, y)}.$$

However, if the state space is Markovian,

$$p(\theta, \phi) = p(\phi)p(\theta_0)\prod_{t=1}^{T} p(\theta_t|\theta_{t-1}, \ldots, \theta_0, \phi);$$

thus, it is typically straightforward to write the numerator in the marginal likelihood. For the denominator, we can use a multivariate normal with moments given by our approximate method. We also note that this expression is the easiest way to obtain the posterior distribution of the model parameters in the context of the extended Kalman filter.

In conclusion, we find that the poststratification estimator gives more precise results than the sample mean and it does this by correcting our estimate for imbalances in the representation of the political parties in our sample. Moreover, these gains are achieved without recourse to an explicit dynamic model for the quantity of interest.

[Received June 1999. Revised July 2000]

REFERENCES

Beck, N. (1991), "Comparing Dynamic Specifications: The Case of Presidential Approval," *Political Analysis*, 3, 51–88.

Besag, J. (1989), "A Candidate's Formula: A Curious Reult in Bayesian Prediction," *Biometrika*, 76, 183.

Box-Steffensmeier, J. M., and Smith, R. M. (1998). "The Dynamics of Aggregate Partisanship" *American Political Science Review*, 90, 567–580.

Brace, P., and Hinckley, B. (1991), "The Structure of Presidential Approval: Constraints Within and Across Presidencies," *The Journal of Politics*, 53(4), 993–1017.

Brody, R. (1991), *Assessing the President*, Stanford: Stanford University Press.

Cargnoni, C., Müller, P., and West, M. (1997), "Bayesian Forecasting of Multinomial Time Series Through Conditionally Gaussian Dynamic Models," *Journal of the American Statistical Association*, 92, 640–647.

Carter, C., and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Clarke, H., and Stewart, M. (1994), "Prospections, Retrospections, and Rationality: The "Bankers" Model of Presidential Approval Reconsidered." *American Journal of Political Science*, 38(4), 1104–1123.

Frühwirth-Shnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Holt, D., and Smith, T. M. F. (1979), "Post Stratification," *Journal of the Royal Statistical Society*, Ser. A, 142, 33–46.

Kernell, S. (1978), "Explaining Presidential Popularity," *American Political Science Review*, 72, 506–522.

Kernell, S. (1986), *Going Public*, Washington, DC: CQ Press.

Kiewiet, D., and Rivers, D. (1985), "The Economic Basis of Reagan's Approval," in *Contemporary Political Economy*, eds. D. A. Hibbs and J. Fasbender, Amsterdam: Elsevier, pp. 49–71.

Little, R. J. A. (1993), "Post-Stratification: A Modeler's Perspective," *Journal of the American Statistical Association*, 88, 1001–1012.

Little, T. (1996), "Models for Non-response Adjustment in Sample Surveys," unpublished Ph.D. thesis, University of California, Berkeley.

MacKuen, M. (1983), "Political Drama, Economic Conditions, and the Dynamics of Presidential Popularity," *American Journal of Political Science*, 27, 165–192.

Neustadt, R. (1990), *Presidential Power*, New York: Free Press.

Norpoth, H., and Yantek, T. (1993). "Macroeconomic Conditions and Fluctuations of Presidential Popularity: The Question of Lagged Effects," *American Journal of Political Science*, 33, 516–536.

Ostrom, C., Jr., and Simon, D. (1989), "The Man in the Teflon Suit," *Public Opinion Quarterly*, 53, 353–387.

Ostrom, C., Jr., and Smith, R. M. (1993), "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval," *Political Analysis*, 4, 127–184.

Pfeffermann, D. (1991), "Estimation of Seasonal Adjustments of Population Means Using Data from Repeated Surveys," *Journal of Business and Economic Statistics*, 9, 163–177.

Rivers, D., and Rose, N. (1985), "Passing the President's Program: Public Opinion and Presidential Influence in Congress," *American Journal of Political Science*, 29, 183–196.

Rubin, D. B. (1981), "Estimation in Parallel Randomized Experiments," *Journal of Educational Statistics*, 6, 377–401.

Smith, R. (1992), "Error Correction, Attractors, and Cointegration: Substantive and Methodological Issues," *Political Analysis*, 4, 249–254.

Voss, S., Gelman, A., and King, G. (1995), "Pre-Election Survey Methodology: Details From Nine Polling Organizations, 1988 and 1992," *Public Opinion Quarterly*, 59, 98–132.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.

West, M., Harrison, P. J., and Migon, H. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting" (with discussion), *Journal of the American Statistical Association*, 80, 73–97.

Williams, J. (1992), "What Goes Around Comes Around: Unit Root Tests and Cointegration," *Political Analysis*, 4, 229–236.