

Bayesian hierarchical age-period-cohort models with time-structured effects. An application to religious voting in the US, 1972–2008.*

Daniel Stegmueller[†]

Published in: Electoral Studies, 33, 2014, pp. 52–62

Abstract

To examine dynamics of political processes using repeated cross-section data, effects of age, cohort, and time period have to be disentangled. I propose a Bayesian dynamic hierarchical model with cohort and period effects modeled as random walk through time. It includes smoothly time-varying effects of covariates, allowing researchers to study changing effects of individual characteristics on political behavior. It provides a flexible functional form estimate of age by integrating a semi-parametric approach in the hierarchical model. I employ this approach to examine religious voting in the United States using repeated cross-sectional surveys from 1972 to 2008. I find starkly differing nonlinear trends of de- and re-alignment among different religious denominations.

*I thank my editors, Anja Neundorf and Richard Niemi, as well as my anonymous reviewer for helpful suggestions and criticisms. Furthermore, I thank NanDirk de Graaf, Thomas Gschwend, Simon Jackman, Larry Bartels, Bent Nielsen, and participants at the ESRA 2011 session “Analyzing social change with repeated cross-sections I” for helpful discussions. As always, they are not to be implicated by my shortcomings.

[†]University of Mannheim, dstegmueller@uni-mannheim.de

1 Examining socio-political processes

Students of political behavior are often not only interested in relationships between individual characteristics and political outcomes, but want to describe evolution of these relationships over time. In absence of panel data, repeated cross-sectional surveys are used to examine the dynamics of social and political behavior (e.g., among many, De Graaf et al. 2001; Brooks and Manza 2004; Elff 2007, 2009). When analyzing change, a well-known problem is the tightly knotted relationship between age, time period and cohort membership, as outlined by Neundorf and Niemi (2013) in this issue. Recent proposals (Yang and Land 2006, 2008), which use hierarchical models to disentangle these effects by simultaneously nesting individuals in cohorts and time periods, have become popular (for a recent application see Caren et al. (2011) and Smets and Neundorf (2013) in this issue).

In this paper, I propose to extend these various models to ameliorate some of their potential shortcomings. First, previous contributions have simply assumed that both time and cohort random effects are exchangeable. The assumption that the ‘order’ of periods or cohorts does not matter seems hardly justifiable: both cohort and period effects are (by definition) changes occurring in sequential time. In contrast, I propose a hierarchical model where random effects are time-structured by modeling them as a random walk through time. Second, hitherto proposed hierarchical age-period-cohort (APC) specifications model time period and cohort differences in the dependent variable, but assume time-homogeneity of effects of relevant individual attributes, such as religion or class. However, changes in effects of these variables are often of central interest to researchers (e.g., Manza and Brooks 1997; De Graaf et al. 2001; Brooks and Manza 2004; Elff 2009; Jansen et al. 2011). Borrowing from state-space modeling, I extend my model to estimate smoothly time-varying effects of central covariates.¹ Finally, in most applications researchers specify flexible time and cohort effects, but treat age as linear or quadratic. I detail how to estimate the functional form of age more flexibly by employing a nonparametric strategy based on penalized splines. I specify my model in a Bayesian framework, which allows me to implement this various extensions under a common hierarchical model framework. Furthermore, using the Bayesian inferential paradigm, researchers do not have to rely on the interpretation of random effects as random samples from a (super-)population, which is questionable when working with APC models. Instead, inferences are made based only on actually observed periods and cohorts.²

¹These model specifications are increasingly common in public health and disease mapping research, where researchers face the age-period-cohort problem with grouped data (such as cancer incidence rates). See, e.g., Knorr-Held (2000); Knorr-Held and Rainer (2001); Schmid and Held (2004) and the references therein.

²Add to this the general intellectual attractiveness of Bayesian inference (see, e.g., Jackman 2009: ch.1 or Jaynes 2003). Yang and Land (2008) provide further specific arguments for using a Bayesian approach for APC models.

I demonstrate the practical application of my proposed model specification by analyzing trends in religious voting in the United States. Researchers and pundits alike have put renewed focus on the role of religion in shaping the political landscape of the US (Wald et al. 2005; Wald and Wilcox 2006). Following the perception of increased polarization, specific interest lies in examining changes in traditional party alignments of different denominations (e.g., Brooks and Manza 2004). I use repeated cross-sectional surveys from 1972 to 2008 with detailed measures of religious denominations. I find no effect of cohort membership, but starkly differing nonlinear trends of de- and re-alignment among different religious denominations.

The paper is structured in two central parts. The next section contains a detailed discussion of hierarchical age-period-cohort models and my proposed extensions. I also discuss a strategy for comparing models of differing complexity. Section 3 contains the application of my model to denominational vote trends in the US. I discuss data and imputation strategy and present detailed results. Finally, section 4 concludes the paper with a short discussion.

2 Dynamic hierarchical model

The key to a combined analysis of effects of age, cohort, and time period is to account for the fact that individuals experience the same time period but are also simultaneously members of a specific cohort. To capture this dependence structure, recent contributions (e.g., Yang and Land 2006, 2008) propose the use of cross-classified random effect models (Snijders and Bosker 2012; Browne et al. 2001; Rasbash and Browne 2008).³ One models response y_{itk}^* of individual i in the cell defined by the cross-classification of time period t ($t = 1, \dots, T$) and cohort k ($k = 1, \dots, K$) as a function of age a_i , covariates x_i , and time (p_t) and cohort (c_k) effects:

$$y_{itk} = a_i + p_t + c_k + x_i$$

More specifically, I propose to model responses as resulting from an individual's age a_i , some theoretically relevant variables w_i , a set of controls for socio-economic background characteristics x_i including an overall constant (i.e., for all i , $x_{i1} = 1$), and random effects for time periods ζ_t and cohort membership ξ_k :

$$y_{itk}^* = f(a_i) + \sum_{r=1}^R \beta_r' w_{ir} + \sum_{s=1}^S \gamma_s' x_{is} + \zeta_t + \xi_k + \epsilon_{itk} \quad (1)$$

Here, $f(a)$ is some smooth functional form estimate of age effects (discussed in detail later), β_r and γ_s are effect coefficients for theoretically central and control variables, respectively, and ϵ_{itk} is a white noise error term. In this general formulation, the

³See Smets and Neundorf (2013) in this issue for a more detailed discussion of this model.

model is applicable to outcomes that are either continuous, dichotomous or ordered categorical, by suitably specifying the link between observed y and latent y^* (Greene and Hensher 2010). Since my application uses a binary dependent variable, the following discussion refers to this situation. I use a probit model, obtained by assuming that a latent variable – normally distributed with unit variance – generates observed outcomes via some threshold or utility mechanism (Albert and Chib 1993; Jackman 2000). Thus an individual responds $y_{itk} = 0$ if $y_{itk}^* < 0$ and $y_{itk} = 1$ if $y_{itk}^* \geq 0$.

Model specification is completed by specifying a distribution for cohort and time period random effects. Existing proposals (e.g., Yang 2006) rely on the standard assumption of normally distributed random effects. For example, cohort random effects are assumed to be drawn from a normal distribution with zero mean and estimated variance σ^2 :

$$\xi \sim N(0, \sigma^2).$$

While convenient, I argue that this specification is somewhat ignorant about our knowledge of social processes. It seems implausible *a priori*, that each new cohort of individuals or each new time period, is formed in a social and political vacuum – independent of the past. Rather, most theoretical accounts of socio-political change emphasize continuity between cohorts or time points, as well as elements of change. Following these arguments, I propose to add structure to cohort and time period random effects by including systematic dependencies between neighboring periods and cohorts, while still allowing for the occurrence of sudden change.

2.1 Time-structured random effects

To yield a more realistic treatment of period and cohort succession, I model random effects as following a non-stationary second order random walk transition process (Besag et al. 1995; Knorr-Held and Rainer 2001):

$$\xi_k = 2\xi_{k-1} - \xi_{k-2} + v_k, \quad v_k \sim N(0, \omega^2). \quad (2)$$

This injects some structure into the cohort random effects, since each new cohort random effect now depends on past values to a certain extent. More precisely, the conditional expected value of a new cohort random effect $E(\xi_k | \xi_{k-1}, \xi_{k-2})$ is a locally linear extrapolation based on past random effects ξ_{k-1} and ξ_{k-2} . As illustrated in Figure 1, the key characteristic of this setup is that *a priori* one expects that neighboring cohorts are similar and changes will occur gradually – that is, ω^2 will be small. However, if the data suggest that a new cohort differs strongly from the previous one, the model will accommodate this and one will see a deviation from the local linear trend.⁴

⁴Note that trends are only *locally* linear, that is, I do not impose a linear trend across all cohorts. This model specification is known in macro-economics as Hodrick-Prescott filter, widely used to

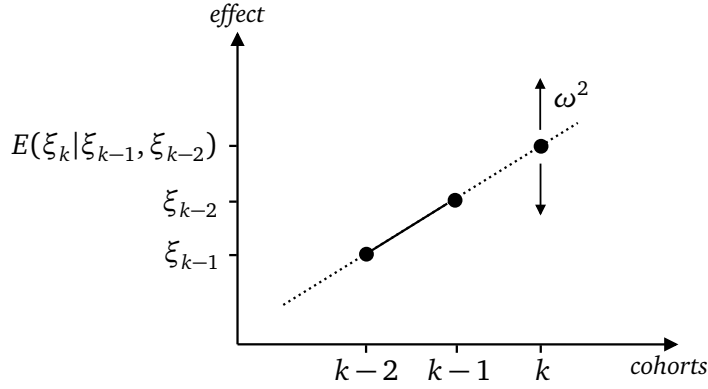


Figure 1: Illustration of locally linear approximation by random walk

To realize this idea in a hierarchical model, I specify the cohort random effects prior as:

$$\xi_k \sim N(2\xi_{k-1} - \xi_{k-2}, \omega^2) \quad (3)$$

Here the mean of the prior distribution in equation (3) specifies the systematic time trend, while the variance allows for deviations from it (for recent political science applications see Jackman 2005 or Martin and Quinn 2002). The extent to which a new cohort random effect value is determined by the time trend is determined by the variance parameter ω^2 . As $\omega^2 \rightarrow 0$ new cohorts will closely follow a linear trend, while large values of ω^2 yield cohort random effects which are rather different from each other. In the Bayesian framework ω^2 is estimated from the data by assigning it an appropriate hyper-prior (details on priors are given in section 2.4.) A random walk has to be initialized from some starting position. In absence of strong a priori information, simply choose a random position on the real line, which in a Bayesian framework amounts to specifying constant priors $\xi_1, \xi_2 \propto c$ and $\zeta_1, \zeta_2 \propto c$ for the first two cohorts and years, respectively. To identify the random effects, one needs some form of restriction, such as a corner constraint. Another option – used here – is a sum-to-zero constraint, i.e.:⁵

$$\sum_{k=1}^K \xi_k = 0. \quad (4)$$

Finally, specification of time period random effects follows the same logic by using a second differences mean structure with estimated evolution variance ψ^2 ,

$$\zeta_t \sim N(2\zeta_{t-1} - \zeta_{t-2}, \psi^2), \quad (5)$$

smooth macro-economic time series (Hodrick and Prescott 1997).

⁵This constraint can be implemented by simply re-centering the cohort random effects on each iteration of the MCMC sampler.

and a sum-to-zero constraint on the vector of random effects.

2.2 Time-trends of covariate effects

So far the proposed model essentially performs a variance decomposition of the dependent variable. However, I argue that many applications are more concerned with following social and political processes through time. In other words, simply stating that cohort or time periods differ is often less interesting than learning how effects of substantive variables vary between cohorts or over time. To enable this type of analysis, I extend the APC model to include varying effects of covariates. This conceptualization of smoothly time-varying effects – as opposed to simple ‘random slopes’ in a standard multilevel model, which again ignore the time-structured nature of data – follows ideas from state space modeling (West and Harrison 1997; Fahrmeir and Tutz 1997: ch.8) as well as the varying coefficient framework in spatial statistics (e.g., Hastie and Tibshirani 1993).

Thus, following the logic of locally linear priors as discussed above, I model changing covariate effects via a second order random walk. Instead of assuming time-constant effects β_r for each relevant covariate r ($r = 1, \dots, R$), I specify them as time period specific (β_{rt}) evolving according to the following stochastic process (cf. West and Harrison 1997):

$$\beta_{rt} = 2\beta_{r,t-1} - \beta_{r,t-2} + u_{rt} \quad \text{with } u_{rt} \sim N(0, v_r^2). \quad (6)$$

In terms of my application this yields denomination effects on vote choice, which evolve according to a locally linear trend. For large estimated values of v_r^2 , no clear trend exists – each new denomination-by-time effect will differ sharply from the previous one. Contrary, if $v^2 \rightarrow 0$ denomination effects (of group r) follow a clear linear trend. The extended model now reads:

$$y_{itk}^* = f(a_i) + \sum_{r=1}^R \beta'_{rt} w_{ir} + \sum_{s=1}^S \gamma'_s x_{is} + \zeta_t + \xi_k + \epsilon_{ijk} \quad (7)$$

$$\beta_{rt} \sim N(2\beta_{r,t-1} - \beta_{r,t-2}, v_r^2) \quad (8)$$

$$\zeta_t \sim N(0, \psi^2) \quad (9)$$

$$\xi_k \sim N(2\xi_{k-1} - \xi_{k-2}, \omega^2) \quad (10)$$

Again, to identify the newly added time-varying effects, one needs to impose a restriction, such as the sum-to-zero constraint in (4). I enforce

$$\forall r : \sum_{t=1}^T \beta_{rt} = 0, \quad (11)$$

which implies that second differences can be identified from the data. In other words,

we can learn about *changing* associations between religious group membership and voting behavior if they depart from a linear trend. If all social change is strictly linear, it is impossible to disentangle effects from the data alone.

Besides adding time-varying effects of theoretically central covariates, equation (9) now contains a simplified distributional assumption for period random effects ζ_t . After including several time trends via (8), I specify remaining period differences as arising from a normal distribution centered at zero with an estimated variance component ψ^2 . The purpose of this specification is to split the overall time period trend into a correlated and an uncorrelated part (e.g., Besag et al. 1991). The part that is correlated over time is due to changing effects of denominational group membership on vote choice. The remaining part captures residual variation not explained by these systematic trends. Cohort random effects ξ_k are still allowed to be correlated through time.

2.3 Nonparametric estimation of age effects

Hierarchical APC models are very flexible in terms of cohort and time period effects. The question remains how to treat age in this framework. This is especially relevant when little is known a priori about its functional form. Simply including linear and quadratic trends of age is a strong, and possible unwarranted, simplification. A more complex strategy, such as including some polynomial transform of age might be equally problematic when local features are unknown (Beck and Jackman 1998).

An alternative is to leave the functional form of age unspecified and estimate it from the data. Estimation based on splines provides an established and easily implementable approach (Keele 2008). I model the unknown functional form of age by using B-splines with a quadratic basis (see de Boor 1978 for an introduction) on L equally spaced knots:⁶ $\min(a) = \kappa_1 < \kappa_2 < \dots < \kappa_{L-1} < \kappa_L = \max(a)$. Thus, age is modeled via L B-spline basis functions $B_l(a_i)$ and their associated regression coefficients (η_1, \dots, η_L)

$$f(a_i) = \sum_{l=1}^L \eta_l B_l(a_i) \quad (12)$$

Specifying the number of knots L is a nontrivial enterprise. To avoid multiple trial and error procedures, Eilers and Marx (1996, 2010) propose to use penalized splines, or P-splines. The idea here is to specify the spline using a rather high number of knots (often between 20 and 40) over the domain of the variable, and to avoid overfitting by introducing a term which penalizes for too abrupt function jumps.

⁶When calculating the B-spline basis we actually need additional knots outside the range of the variable. Thus the extended knot set is $\kappa_{-1}, \kappa_0, \kappa_1, \dots, \kappa_{L+1}, \kappa_{L+2}$. Knots are placed at equidistant intervals: Let $m = (\max(a) - \min(a))/(l - 1)$, then knots are placed at $\kappa_l = \min(a) + (l - 1)m$ with $l = -1, \dots, L + 2$.

More precisely, they introduce a penalty term λ for adjacent B-Spline coefficients based on first differences (Eilers and Marx 1996):

$$\lambda \sum_{l=2}^L (\Delta \eta_l)^2 \quad (13)$$

where Δ is the forward difference operator. Adding this term to (12) leads to a penalized estimator which produces a smooth function as λ gets larger. This penalized spline approach reduces the dependence of functional form estimates on parameters like the number of knots, and thus leads to robust substantive inferences. Currie and Durban (2002) show that this specification can be reformulated as a hierarchical model; consequently, by choosing an appropriate hyperprior the degree of smoothing (λ) can be estimated from the data (see Ruppert et al. 2003 for an excellent introduction). Details on all my prior specifications are given in the next section. As before, for identification age is restricted to have a zero mean effect (the overall mean is captured by the constant in x_i) by imposing a sum to zero constraint.

2.4 Priors

For fully Bayesian inference I need to specify appropriate (hyper-)priors for all parameters. In my proposed model, the degree of smoothing of cohort and time period random effects is estimated from the data. Thus, priors for evolution variances of cohort and time period random walks are drawn from conjugate inverse Gamma distributions (Gilks et al. 1996):

$$\begin{aligned} \omega^{-2} &\sim \Gamma(a_0, b_0) \\ \psi^{-2} &\sim \Gamma(a_0, b_0). \end{aligned}$$

Similarly, evolution variance priors for varying covariate effects are for all r :

$$v_r^{-2} \sim \Gamma(a_0, b_0).$$

To yield diffuse or non-informative variance priors, common parametrizations of the Gamma distribution used in practical applications are shape and scale parameters close to zero, such as $a_0 = b_0 = 0.001$. An alternative is to set $a_0 = 1$ and b_0 set to a small value such as 0.005 (Besag et al. 1991). I use the former for all reported results.

A spline smoothness prior needs to ensure that the large number of spline coefficients does not lead to an estimated functional form of age that is too rough. Following the idea of a penalty term based on first differences of adjacent coefficients (Eilers and Marx 1996), Brezger and Lang (2006) propose a joint penalty prior for the vector of

all spline coefficients $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$

$$\boldsymbol{\eta} \propto \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\eta}'\mathbf{P}\boldsymbol{\eta}\right) \quad (14)$$

where \mathbf{P} is an appropriate penalty matrix indicating neighboring coefficients. More details are given in appendix A1. The variance τ^2 acts as an inverse smoothing parameter, in other words, the penalty term λ given in equation (13) is estimated by $\lambda = 1/\tau^2$. Thus, smaller variances result in more smoothed function estimates. In this mixed model setup the degree of smoothing is estimated from the data by assigning it a conjugate inverse gamma prior: $\tau^{-2} \sim \Gamma(a_0, b_0)$. Following the same logic as in my discussion above, I use values leading to a rather diffuse prior, i.e., $a_0 = b_0 = 0.001$. Finally, using flat priors for β_r and γ_s yields regression-type estimates for all remaining covariate effects.

To check for the sensitivity of my results, I follow Geisser (1993) and perturb my model with both small and large prior variances, in order to detect their effect on the analysis. I also consider a prior which is not defined for the variance (or precision) but for the standard deviation, as suggested by Gelman (2006). Gill (2008: ch 6.2) provides further discussion of robustness check strategies in Bayesian analysis. Generally, results are rather insensitive to different prior choices. This is especially the case for age estimates. Varying priors for period and cohort random effects leads to estimated trends that are more or less ‘smooth’, but substantive results do not change.

2.5 Model comparisons

With models of this level of complexity a means of comparing specifications is helpful.⁷ Different models can be compared using model deviance. However, deviance decreases with added model complexity. To take model complexity into account, Spiegelhalter et al. (2002) propose the Deviance Information Criterion (DIC), which can be seen as a Bayesian analogue to the widely used Akaike information criterion. They define DIC as deviance plus complexity penalty:

$$DIC = \overline{D(\boldsymbol{\theta})} + p_D$$

where $\overline{D(\boldsymbol{\theta})}$ is the mean deviance, $\boldsymbol{\theta}$ represents the vector of all model parameters, and p_D is a penalty term for model complexity. Thus, DIC penalizes improvements in

⁷The most elegant solution – Bayes factors – are hardly computationally feasible for practical applications, due to its computational intractability and its insistence on non-diffuse priors. Other proposals exist that are easier to implement, such as conditional predictive ordinates (Geisser 1993; Gelfand et al. 1992). I do not discuss these here as the close relationship to the AIC makes DIC the most straightforward choice for practical applications. See the discussion following Spiegelhalter et al. (2002) for a number of disadvantages of the DIC.

deviance for the fact that a more complex model has been used. The penalty term is given by an estimate of model complexity, the *effective* number of model parameters.⁸ This strategy is preferable to simply ‘counting parameters’, since in models where random effects priors are highly structured, each random effect is likely to represent far less than one parameter.⁹ For more elaborate discussions see Spiegelhalter et al. (2002) and Gill (2008: 260).

3 Application: Religious voting in the US, 1972–2008

To illustrate the application of this setup, I analyze trends in religious voting in the United States between 1972 and 2008. Previous research on religious trends usually does not include cohort differences, and a more complete APC specification will show if this omission was reasonable. However, my application goes beyond a simple decomposition of age, period, and cohort effects and examines trends in voting behavior of different groups (following the seminal work by Manza and Brooks (1997) and Brooks and Manza (2004)). Changing alignments of major religious groups have been of central interest in recent research, as they are supposed to be a major source of (macro-)political change (e.g., Rozell and Wilcox 1996; Wuthnow and Evans 2002; Green et al. 2003). For a more detailed discussion of arguments about, and previous research on, denominational voting differences, see Brooks et al. (2003) and Brooks and Manza (2004).

3.1 Data

Following previous research on religion and voting (e.g., Brooks and Manza 2004), I analyze data from the American National Election Studies for the ten presidential elections and seven midterm elections from 1972 through 2008. Possible level differences for midterm elections are checked by including an indicator variable. The pooled data-set provides information for 17,896 individuals. Their vote choice is represented by an indicator variable equal to one if they voted for the Democratic candidate and zero for the choice of Republican candidates. Individuals who explicitly state that they abstained from voting or supported third candidates are excluded.¹⁰

⁸It is calculated as $pD = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$, i.e., the mean Deviance minus the Deviance evaluated at the means of parameters. The Deviance is defined as minus two times the log-likelihood, i.e., $D(\boldsymbol{\theta}) = -2\log(p(y|\boldsymbol{\theta}))$.

⁹Thus, p_D can be understood as the sum of all model parameters each weighted by (Gill 2008: 261): $w_p = 1$ if a parameter is not constrained by any prior information, $w_p = 0$ if all information comes from the prior, and $w_p \in [0, 1]$ for parameters which depend on priors to a varying extent, such as random walks.

¹⁰I do not use data prior to 1972 since they do not contain detailed enough denominational information. Midterm elections in 2003 and 2006 are not covered by the ANES. Furthermore, including third candidates such that the dependent variable indicates choice of Democratic vs.

The three central time variables are age, time period (or survey year) and cohort. Age is measured in years and only individuals older than eighteen are included. Seventeen time periods represent bi-annual elections from 1972 to 2008. Finally, I constructed 18 five-year birth cohorts ranging from 1895 to 1980. 115 individuals born before 1895 and 104 individuals born after 1985 are included in the first, respectively last, cohort since they are too small in size to form their own cohort.

My central theoretical variable, major American religious groups, is captured by a six category scheme comprised of Protestant, Catholic and Jewish faiths. Protestants are subdivided into three denominational families, viz., Mainline, Evangelical and Black Protestants. The latter are members of historically Black protestant churches, which recent scholarship argues form a distinct religious group “manifestly different from that of other white Protestant traditions” combining economic liberalism with social conservatism (Steensland et al. 2000: 294; and the references therein). The final, sixth category is made up of those who explicitly state that they have no denomination. This group also includes a small number of individuals who belong to ‘other’ denominations, such as Muslim and Hindu, which are too few in numbers to include on their own. This classification is derived from detailed information (with roughly 70 detailed categories) on individuals’ religious preference contained in the ANES, following the classification scheme of Steensland et al. (2000) and Brooks and Manza (2004). Differences in the frequency of church attendance are included by a pseudo-continuous variable measuring how often a respondent attends church.

Table 1 shows means and standard deviations of all variables included in my analysis before and after imputation (discussed below). Income is measured in real dollars deflated using the consumer price index with base year 2000. Education is measured by the number of years spent in full-time education. An individual’s party identification is included via the well-known three category measure, differentiating Democrats, Republicans and Independents. Stable regional differences in voting behavior are captured by a categorical variable for a respondent’s region of residence. Finally, further individual background characteristics are captured by dummy variables indicating if a respondent is female, African-American, and currently married.

3.2 Imputation of missing data

As Table 1 shows, several variables have imputed missing values. Missing data is an ubiquitous feature of social surveys and a problem likely to be encountered by every practical application of the models presented in this special issue. Thus, I provide details of my imputation strategy, which takes into account the repeated cross-section nature of the pooled dataset.

‘other’ candidates does not change results.

Table 1: Means and standard deviations (Pre- and Post-Imputation) of variables

	Pre-imputation		Post-imputation		Imputed [%]
	means	sd	mean	sd	
Vote (Democrats)	0.53	0.50	0.53	0.50	2.7
Age (years)	48.15	16.97	48.15	16.97	0.3
Education (years)	13.12	2.86	13.12	2.86	0.2
Income (10,000 \$) ^a	4.77	3.33	4.74	3.34	7.6
Female	0.55	0.50	0.55	0.50	
African-American	0.12	0.32	0.12	0.32	0.5
Married	0.62	0.49	0.62	0.49	0.2
Denomination ^b					0.8
Mainline Protestant	0.33	0.47	0.33	0.47	
Evangelical Protestant	0.22	0.41	0.22	0.41	
Black Protestant	0.06	0.24	0.06	0.24	
Catholic	0.25	0.44	0.25	0.43	
Jewish	0.03	0.16	0.03	0.16	
Church attendance	3.11	1.55	3.11	1.55	0.5
Party identification ^c					0.3
Independent	0.33	0.47	0.33	0.47	
Democrat	0.39	0.49	0.39	0.49	
Region ^d					
North	0.19	0.39	0.19	0.39	
South	0.32	0.47	0.32	0.47	
West	0.21	0.40	0.21	0.40	
N individuals			17896		
N periods			17		
N cohorts			18		

Note: Post-imputation statistics calculated from 5 imputed data sets. Full data set contains 17,896 individuals, 18 cohorts, 17 periods.

^a Deflated using CPI, base year: 2000

^b Reference category is no or other denomination

^c Reference category is Republican party ID

^d Reference category is East

Following the strategy outlined by Gelman et al. (1998), I employ a hierarchical multivariate normal model. More precisely, let individuals i , ($i = 1, \dots, n_s$) partake in several surveys s , ($s = 1, \dots, S$) which consists of Q questions ($q = 1, \dots, Q$). An individual's vector of survey responses is assumed to follow a multivariate normal distribution,

$$y_{is} \sim N_Q(\mu_s, \Sigma)$$

where μ_s is a vector of means and Σ is a $Q \times Q$ variance covariance matrix. The

survey means are (conditionally) exchangeable and follow a normal distribution

$$\mu_s \sim N_Q(\theta, \Phi)$$

with its location determined by a vector of means, θ , and its variance governed by a $Q \times Q$ diagonal matrix of variances, $\Phi = \text{diag}(\phi_1^2, \dots, \phi_Q^2)$. Informally, this model can be seen as an extension of the widely used multivariate normal imputation model (e.g., King et al. 2001) to a multilevel setting, where the ‘second level’ hierarchy is given by the ANES surveys.

To set up this model in a Bayesian framework, one needs to choose priors for all elements. The prior for the individual level variance-covariance matrix Σ is an inverse Wishart distribution:

$$\Sigma \sim W^{-1}(\Sigma_0, \nu_0).$$

Following Gelman et al. (1998), I use a relatively non-informative specification, where the scale matrix Σ_0 is specified as identity matrix, I_Q , and the degrees of freedom parameter ν_0 equals the number of survey questions included. Elements of the survey level variance matrix Φ follow an inverse gamma distribution,

$$\phi_q^{-2} \sim \Gamma(0.001, 0.001), \quad q = 1, \dots, Q,$$

with shape and scale set to 0.001 to yield a diffuse prior. Imputed values are not sensitive to different parametrization of the gamma distribution. Finally, priors for means μ and θ are normally distributed, centered at zero and with a variance of 100.

I obtain five imputations by running two chains of the Gibbs sampler (Tanner and Wong 1987; Gelfand and Smith 1990) for 5000 iterations (after which they have settled in their equilibrium distribution) and by drawing from the posterior distribution every 500 iterations. This creates five data sets with missing values imputed from the resulting posterior predictive distribution of each set of draws. All results presented in the following are based on these five imputed data sets (cf. Rubin 1987).

3.3 Results

Table 2 shows Bayesian fit statistics for a series of models of increasing complexity.¹¹ I start with a simple decomposition of the Democratic vote into effects of age, cohort

¹¹All models are fitted using Markov Chain Monte Carlo sampling following the strategy outlined in Knorr-Held and Rainer (2001) and Rue and Held (2005). For further details and code of the models see the supplementary material online. The sampler is run for 100,000 iterations (thinned by a factor of 20) after a burn-in of 50,000. Mixing is very good quite early in the evolution of the chain. Standard convergence diagnostics (Cowles and Carlin 1996) suggest that the chain has reached its steady state.

Table 2: Posterior deviance, effective number of parameters, and deviance information criterion for fitted models.

		Deviance	pD	DIC
M0	APC decomposition	24132	32.9	24165
M1	M0 + socio-economic variables	18022	37.0	18059
M2	M1 + religion variables	17678	44.8	17723
M3	M2 + denomination time trends	17544	71.9	17615

Note: Combined results from five imputations.

membership, and time period in Model M0 and add control variables and religious variables in Models M1 and M2. The impressive decrease in deviance from M0 to M1 shows the importance of taking into account the changing composition of the electorate – e.g., due to educational expansion and rising incomes. M2 stresses that, even after controlling for individuals’ basic social and economic characteristics, religion plays an important part in explaining Democratic vote choice. M3 is the most complex specification as it includes time trends of denomination effects on vote choice (as specified in section 2.2). This complexity is reflected by the jump of effective number of parameters pD relative to M2. However, the clear decrease in Deviance as well as the DIC difference of over 100 makes clear that this specification represents the data much better. Thus, I focus most of my following discussion on results obtained from M3. In practical applications, one would likely explore simpler specifications suggested by this flexible model, say by restricting age effects to simpler functional forms.

Table 3 shows estimates of covariate effects for both M2, which includes time-constant denomination effects, and M3, which models denomination effect time trends. Displayed are posterior means and highest posterior density (HPD) regions, which can be understood as Bayesian analogue to the frequentist confidence interval.¹² Estimated effects differ little or not at all between both specifications, which is reassuring in light of the added complexity of M3. Not surprisingly, strong effects of individual background characteristics are apparent. Individuals who obtain higher incomes, are married, and live in the South are less likely to vote for the Democrats.¹³ On the other hand, women and African-Americans show a clear tendency to cast a Democratic vote, which is also true for individuals who identify themselves as Democrats and (to a lesser extent) as Independents. Individuals who attend church more frequently are

¹²More precisely a region R is a $100(1 - \alpha)$ percent HPD region (not necessarily contiguous) for parameter θ if (1) $P(\theta \in R) = 1 - \alpha$ and (2) $P(\theta_1) \geq P(\theta_2)$ for all $\theta_1 \in R$ and $\theta_2 \notin R$, i.e., it yields an interval estimate with the added requirement that each value in the interval is larger than those outside of it. For a discussion of Frequentist versus Bayesian intervals see Jaynes (1976).

¹³For a much more sophisticated discussion of income and regional differences see Gelman et al. (2008). In a future extension of this paper, I plan to allow for spatially varying effects.

Table 3: Posterior summary of covariate effects on the Democratic vote. Posterior means and 90% highest posterior density regions

	M2			M3		
	Mean	HPD region		Mean	HPD region	
Education	-0.032	-0.076	0.008	-0.030	-0.073	0.014
Income	-0.237	-0.280	-0.193	-0.240	-0.284	-0.197
Female	0.118	0.084	0.156	0.119	0.082	0.155
Black	1.296	1.191	1.405	1.305	1.200	1.412
Married	-0.076	-0.118	-0.037	-0.075	-0.117	-0.032
Church attendance	-0.235	-0.276	-0.197	-0.237	-0.276	-0.196
Denomination						
Mainline Protestant	-0.203	-0.266	-0.141			
Evangelical Protestant	-0.342	-0.413	-0.277			
Black Protestant	-0.283	-0.441	-0.145		see Fig. 4	
Catholic	-0.024	-0.089	0.042			
Jewish	0.518	0.387	0.647			
Party identification						
Independent	0.877	0.830	0.923	0.887	0.839	0.934
Democrat	1.662	1.614	1.709	1.683	1.635	1.731
Region						
North	-0.042	-0.097	0.009	-0.036	-0.091	0.018
South	-0.079	-0.124	-0.029	-0.087	-0.135	-0.038
West	-0.007	-0.057	0.048	-0.004	-0.058	0.050
Midterm election	0.240	0.085	0.403	0.253	0.051	0.461
(Intercept)	-0.869	-0.980	-0.769	-0.877	-1.032	-0.725

Note: Based on 5,000 MCMC samples. Combined results from five imputations. Evolution variances of random walks and variance of period random effects not shown.

less likely to vote for Democratic candidates.¹⁴ My central variables of interest show expected patterns: Protestants are far less likely to vote Democratic, a tendency most marked among Evangelical Protestants, followed by Black Protestants. Those of Jewish faiths are of course more likely to vote for Democratic candidates. Finally, the estimated Catholic effect is substantively small and statistically indistinguishable from zero. Note that my results for Catholics and Jewish voters differ from findings in Brooks and Manza (2004), who use data from 1972–2000. This sensitivity to the analysis time-span stresses the importance of modeling changing denominational vote trends. But before analyzing time-varying variants of these effects, I inspect the

¹⁴The model does of course allow more detailed investigations of time trends in covariates described here. For example, a specification of a smoothly time-varying effect of church attendance shows a slight curvilinear pattern with a decreasing effect starting from 1988 and a slight increase since 2002. However, the focus of this application is one *denominational* effects – and these results remain unchanged.

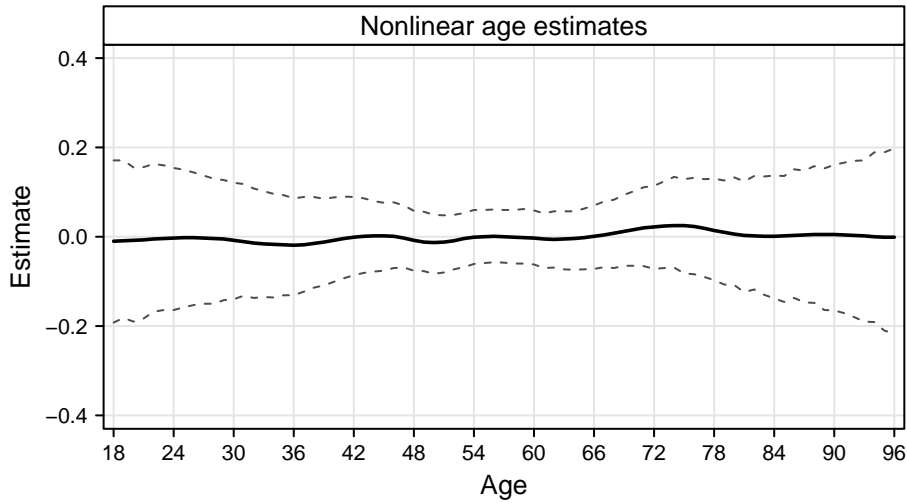


Figure 2: Penalized spline estimate of age effect $f(a_i)$ on propensity of Democratic vote. Posterior means and 90% highest posterior density regions.

estimated functional form of age and systematic cohort differences.

To gauge the role of individual aging, looking at dozens of estimated spline coefficients will not be very helpful. Therefore, I plot the estimated functional form of age on the propensity to vote for the Democrats in Figure 2. It shows posterior means with highest posterior density regions, which together indicate that age estimates are not systematically different from zero. This is reflected in the estimated magnitude of the penalty parameter λ , which is estimated by its inverse as $E(1/\lambda) = 0.002$. This shows that there is virtually no variability between spline coefficients and thus no deviation from a constant trend. In fact a model that replaces the flexible spline function of age with a simple linear specification or even a zero coefficient restriction describes the data just as well.¹⁵ Thus, these results provide strong evidence for the absence of any ‘age effect’. In other words, there is nothing intrinsic in individuals’ aging – as opposed to individual socio-economic characteristics and changing political and social contexts – that would influence voting behavior.

Are there systematic time trends in cohorts’ propensity to vote for the Democrats? To answer that question, Figure 3 plots posterior means of cohort random effects ξ_k and their associated HPD regions. Estimates of ξ_k can be interpreted as differences in trends between cohorts after taking into account compositional differences of the electorate, differences between time periods, and changing trends in the voting

¹⁵I re-estimated M3 with these restrictions. A model with a linear age effect has a deviance of 17544 and a DIC of 17614. A model with a zero restriction for age yields a deviance of 17544 and a DIC of 17612. Thus both show little or no difference in fit to the more complex specification. Estimated denomination time trends (Figure 4) as well as covariate effects (Table 3) stay the same in both specifications.

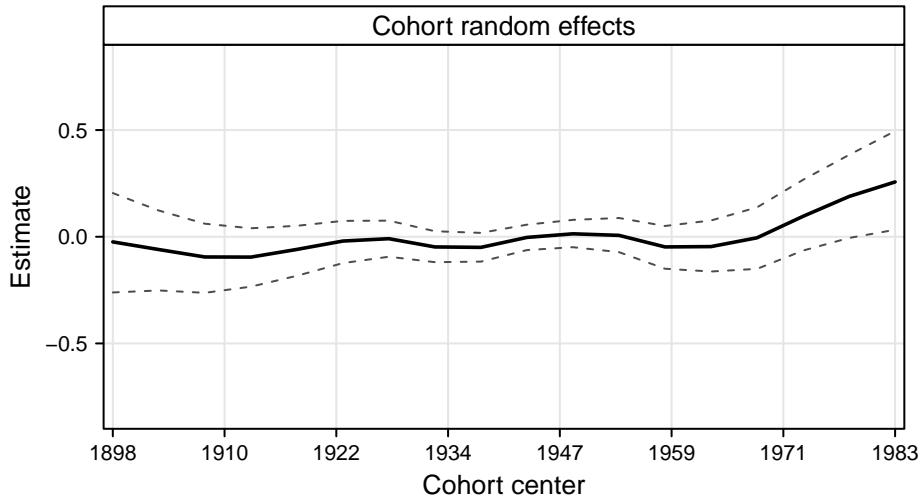


Figure 3: Cohort random effects ξ_k . Posterior means and 90% HPD regions.

behavior of religious groups. Overall, Figure 3 suggest that cohorts are of little relevance in explaining long-term trends in voting behavior in the United States. This is reflected in an estimated evolution variance for the second order random walk used to model the cohort random effects of $\omega^2 = 0.005$. Its small value indicates that successive cohort random effects vary little, and instead follow an overall systematic trend, which here is essentially constant. Only recent cohorts born in the seventies and early eighties show some tendencies to favor the Democrats, even after including socio-economic changes such as incomes, education, and denominations. However, given that these cohorts are at the boundary of our data set, where estimates are considerable less precise, these finding are very premature.

If cohorts matter little (or not at all) and neither does aging, what does? Estimates from M2 in Table 3 already showed large time-constant denomination effects. The more flexible specification in M3 allows denomination effects to vary over time. More specifically, my proposed model specifies denomination-specific effects as following a (locally) linear trend, unless changes are suggested by the data. These smoothed time-varying denomination effects on the Democratic vote are displayed in Figure 4. Before discussing trends in denomination effects on vote choice, I compare their average levels to estimates obtained from the simpler, time-constant model specification (M2, Table 3). It is generally advisable to inspect time-constant estimates and compare them to more flexible specifications to ensure that sign and magnitude of time-varying estimates are reasonable. This is the case here. However, one can immediately discern interesting changes over time, which are obscured by the simpler specification. Note that many of the found patterns are also likely to be missed when exploring simpler denomination times linear-time interaction models. These are included in this model setup as a limiting case, and can be substituted back into the model if findings suggest the appropriateness of simpler trends (e.g.,

the almost linear trend found for Catholics below).

Mainline Protestants have moved (even) closer to the center, rendering their voting behavior almost indistinguishable from the non-religious. This trend towards more ‘centrism’ is markedly reversed for Evangelical Protestants, whose opposition to voting for the Democrats increases at an almost constant rate since the early eighties, bringing them close to previous levels of support for Nixon in his landslide victory of 1972. Most marked is the recent turn of Black Protestants towards Democratic candidates. This trend is of course most marked for the 2008 election of Barack Obama. But it is already evident in the re-election of George W. Bush in 2004.¹⁶ This demonstrates that when the data overwhelm the prior (which prefers smooth trends) the model allows for large shifts in estimated effects.

Jewish voters, generally regarded as core constituency for Democratic candidates, seem to become even closer allies. Starting from the mid eighties their propensity to vote Democratic has increased markedly. In contrast, time-varying estimates of Catholic voting behavior confirm the finding from M2 that they do not differ from the secular population. A model using a constant effect for Catholics fits the data equally well (with a DIC of 17620). This emphasizes an important characteristic of the proposed model: simple linear or constant time trends are included as limiting cases and will be estimated if the data suggest little or no variation over time. This makes superfluous exhaustive (and problematic) ‘specification searches’ with multiple tests of a large number of models with linear, quadratic or even higher covariate-by-time interactions.

After accounting for trends in denomination effects, variation between time periods that is not explained by denominational trends or cohort differences is captured by time period random effects ζ_t , which are displayed in Figure 5. It shows that many random effects are indistinguishable from zero or close to it, suggesting that most characteristic features of these elections have been captured by systematic parts of the model. Obvious exceptions are the swing in the midterm election following the Watergate scandal, Reagan’s landslide victory in 1980 and re-election in 1984, and Clinton’s re-election in 1996. Those random effects capture swings in voter behavior likely to be influenced by current events and conditions – beyond what can be predicted from individuals’ basic economic and social characteristics. These results could be improved upon by extending the model to include macro-level changes in inflation or national or regional unemployment rates – but this beyond the scope of this illustration.

¹⁶This trend reversal is also found by a model with normally distributed random effects. In other words, this finding is not due to the random walk prior ‘pulling’ up the estimate towards the 2008 election. A model specification where the fixed effect of African-American is allowed to differ in 2008 – capturing the fact that this group was more likely to vote for Obama – yields essentially the same Black Protestant trend estimate with increased statistical uncertainty.

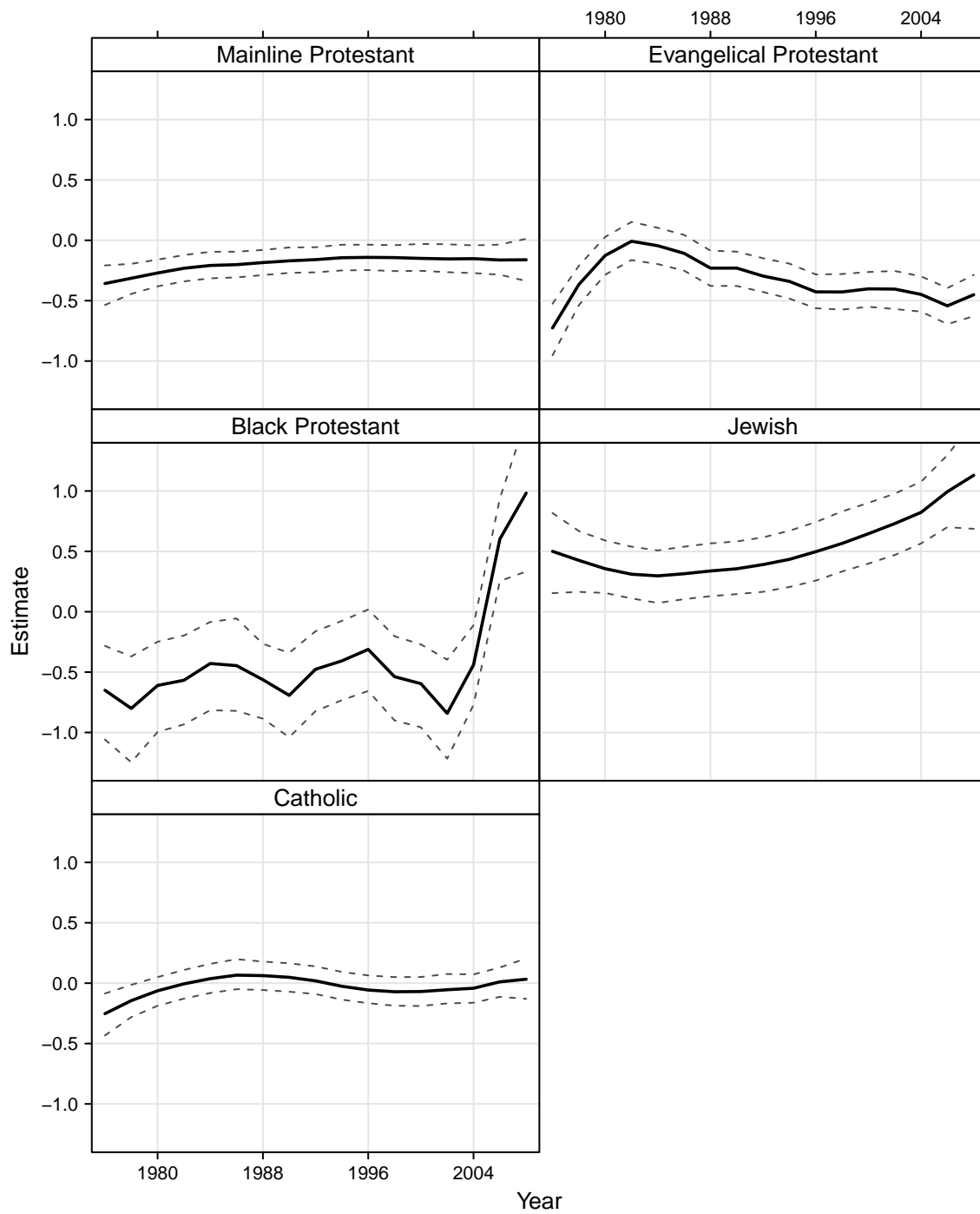


Figure 4: Trends in estimates of denomination effects on the Democratic vote, 1972–2008. Reference group are secular respondents. Posterior means and 90% HPD regions.

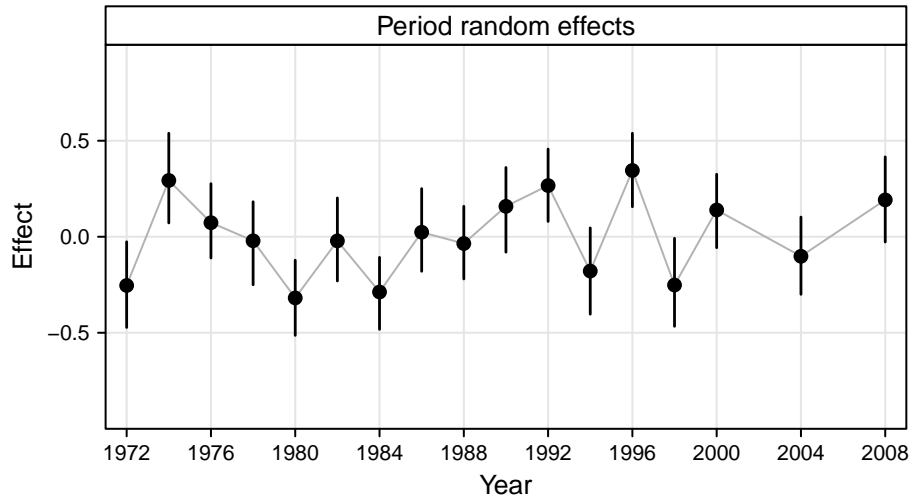


Figure 5: Time period random effects, ζ_t . Posterior means and 90% HPD regions.

4 Conclusion

In this paper, I outlined a general model setup for analyzing age-period-cohort models. My key contribution concerns three issues. First, I link the model closer to theoretical ideas about the nature of social change, which is – in the absence of strong shocks – gradual and evolutionary. By using time-structured priors for cohort and time-period random effects, I explicitly model dependence of new cohorts and periods on the past. Shocks or important events can still influence random effect estimates provided that there is enough evidence for them in the data.

Second, I detail how to estimate changing effects of individual characteristics on the outcome of interest. For many applications, decomposing a dependent variable into age, cohort and time period effects is only a first step. Describing how effects of a small number of theoretically relevant variables change over time can yield new insights into social and political change, and may explain large parts of previously unmodeled period or cohort effects. Again, I propose to model changing effects as essentially gradual, unless they are upset by some strong event. The slow convergence of Mainline Protestants towards the center of the political spectrum on the one hand, and the reversal in Black Protestants' voting alignments on the other, illustrate that the model can capture gradual change as well as large-scale shifts in behavior.

Third, I extend the model to deal with situations where little previous knowledge about the shape of a possible age effects exists and researchers may worry that misspecification of its functional form may lead to distorted results for periods and cohorts. Age – or any other covariate for which strong nonlinearities are expected – can be estimated via penalized spline regressions. Embedding this nonparametric model in the already existing model allows for straightforward and robust inference.

Not every part of my model setup described here will be necessary, or even sensible, in all settings. However, its different building blocks enable flexible construction of models tailored to all kinds of research questions where focus lies on change rather than static relationships. Thus, even if one's main interest is not in a 'variance decomposition' of age-period-cohort effects, the framework outlined here hopefully contributes to the study of the dynamics of individual behavior.

Appendix

A1. Smoothness prior

To derive the multivariate penalty prior used in (14) start from a conditional random walk prior of first order (i.e., the stochastic equivalent of first differences:)

$$\eta_l | \eta_{l-1}, \dots, \eta_1 \sim N(\eta_{l-1}, \tau^2)$$

It's multivariate variant is:

$$p(\eta | \tau^2) = \prod_{l=1}^L p(\eta_l | \eta_{l-1}, \dots, \eta_1)$$

which can be refactored into

$$p(\eta_1) \prod_{l=2}^L p(\eta_l | \eta_{l-1})$$

using the law of total probability. Assuming a constant prior for η_1 and using the first order Markov property, we get after some rearranging:

$$\begin{aligned} p(\eta | \tau^2) &\propto \prod_{l=2}^L \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2\tau^2}(\eta_l - \eta_{l-1})^2\right] \\ &= \frac{1}{(2\pi\tau^2)^{(L-1)/2}} \exp\left[-\frac{1}{2\tau^2} \sum_{l=2}^L (\eta_l - \eta_{l-1})^2\right] \\ &= \frac{1}{(2\pi\tau^2)^{(L-1)/2}} \exp\left[-\frac{1}{2\tau^2} \boldsymbol{\eta}' \mathbf{P} \boldsymbol{\eta}\right] \end{aligned}$$

with \mathbf{P} as appropriate penalty matrix. \mathbf{P} is constructed by using the first difference matrix \mathbf{D} with $\mathbf{P} = \mathbf{D}'\mathbf{D}$. Higher order penalties can be using appropriate difference matrices.

References

- Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669–679.
- Beck, N., Jackman, S., 1998. Beyond linearity by default: generalized additive models. *American Journal of Political Science* 42, 596–627.
- Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. *Statistical Science* 10, 3–41.
- Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, 1–59.
- de Boor, C., 1978. *A Practical Guide to Splines*. Springer, New York.
- Brezger, A., Lang, S., 2006. Generalized structured additive regression based on bayesian P-splines. *Computational Statistics and Data Analysis* 50, 967–991.
- Brooks, C., Manza, J., 2004. A great divide? religion and political change in U.S. national elections, 1972–2000. *Sociological Quarterly* 34, 421–450.
- Brooks, C., Manza, J., Bolzendahl, C., 2003. Voting behavior and political sociology. theories, debates, and future directions. *Research in Political Sociology* 12, 137–173.
- Browne, W.J., Goldstein, H., Rasbash, J., 2001. Multiple membership multiple classification (MMMC) models. *Statistical Modelling* 1, 103–124.
- Caren, N., Ghoshal, R.A., Ribas, V., 2011. A social movement generation: cohort and period trends in protest attendance and petition signing. *American Sociological Review* 76, 125–151.
- Cowles, M.K., Carlin, B.P., 1996. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91, 883–904.
- Currie, I.D., Durban, M., 2002. Flexible smoothing with P-splines: a unified approach. *Statistical Modelling* 2, 333–349.
- De Graaf, N.D., Heath, A., Need, A., 2001. Declining cleavages and political choices: the interplay of social and political factors in the Netherlands. *Electoral Studies* 20, 1–15.
- Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11, 89–121.
- Eilers, P.H.C., Marx, B.D., 2010. *Splines, knots, and penalties*. Wiley Interdisciplinary Reviews: Computational Statistics 2, 637–653.

- Elff, M., 2007. Social structure and electoral behavior in comparative perspective: the decline of social cleavages in Western Europe revisited. *Perspectives on Politics* 5, 277–294.
- Elff, M., 2009. Social divisions, party positions and electoral behavior. *Electoral Studies* 28, 297–308.
- Fahrmeir, L., Tutz, G., 1997. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Second edition. Springer.
- Geisser, S., 1993. *Predictive inference: an introduction*. Chapman & Hall, New York.
- Gelfand, A.E., Dey, D.K., Chang, H., 1992. Model determination using predictive distributions with implementation via sampling-based methods, in: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics 4*. Oxford University Press, Oxford, pp. 147–167.
- Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1, 515–534.
- Gelman, A., King, G., Liu, C., 1998. Not asked and not answered: multiple imputation for multiple surveys. *Journal of the American Statistical Association* 93, 846–857.
- Gelman, A., Park, D., Shor, B., Bafumi, J., Cortina, J., 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press, Princeton.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Gill, J., 2008. *Bayesian Methods. A Social and Behavioral Sciences Approach*. Chapman & Hall, Boca Raton.
- Green, J.C., Rozell, M.J., Wilcox, C. (Eds.), 2003. *The Christian Right in American Politics: Marching to the Millennium*. Georgetown University Press.
- Greene, W., Hensher, D., 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.
- Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. *Journal of the Royal Statistical Society B* 55, 757–796.
- Hodrick, R., Prescott, E.C., 1997. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking* 29, 1–16.

- Jackman, S., 2000. Estimation and inference are missing data problems: unifying social science statistics via bayesian simulation. *Political Analysis* 8, 307–332.
- Jackman, S., 2005. Pooling the polls over an election campaign. *Australian Journal of Political Science* 40, 499–517.
- Jackman, S.D., 2009. *Bayesian Analysis for the Social Sciences*. Wiley, New York.
- Jansen, G., De Graaf, N.D., Need, A., 2011. Class voting, social changes and political changes in the Netherlands 1971–2006. *Electoral Studies* 30, 510–524.
- Jaynes, E.T., 1976. Confidence intervals vs. bayesian intervals, in: Harper, W.L., Hooker, C.A. (Eds.), *Foundations of Probability theory, statistical inference, and statistical theories of science*. D. Reidel, Dordrecht, pp. 175–257.
- Jaynes, E.T., 2003. *Probability Theory. The Logic of Science*. Cambridge University Press, Cambridge.
- Keele, L., 2008. *Semiparametric Regression for the Social Sciences*. Wiley, Chichester.
- King, G., Honacker, J., Anne, J., Kenneth, S., 2001. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review* 95, 49–69.
- Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 19, 2555–2567.
- Knorr-Held, L., Rainer, E., 2001. Projections of lung cancer mortality in West Germany: a case study in bayesian prediction. *Biostatistics* 2, 109–129.
- Manza, J., Brooks, C., 1997. The religious factor in U.S. presidential elections, 1960–1992. *American Journal of Sociology* 103, 38–81.
- Martin, A.D., Quinn, K.M., 2002. Dynamic ideal point estimation via Markov Chain Monte Carlo for the U.S. supreme court, 1953–1999. *Political Analysis* 10, 134–153.
- Neundorf, A., Niemi, R., 2013. Beyond political socialization: new approaches to age, period, cohort analysis. *Electoral Studies* xx, xx–xx.
- Rasbash, J., Browne, W.J., 2008. Non-hierarchical multilevel models, in: de Leeuw, J., Meijer, E. (Eds.), *Handbook of Multilevel Analysis*. Springer, New York, pp. 301–334.
- Rozell, M., Wilcox, C., 1996. *God at the Grass Roots. The Christian Right in the 1994 elections*. Rowman & Littlefield, Lanham.
- Rubin, D., 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley, Hoboken.

- Rue, H., Held, L., 2005. *Gaussian Markov Random Fields. Theory and Applications.* Chapman & Hall.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression.* Cambridge University Press, Cambridge.
- Schmid, V., Held, L., 2004. Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* 60, 1034–1042.
- Smets, K., Neundorff, A., 2013. The hierarchies of age-period-cohort effects: political context and the development of generational turnout patterns. *Electoral Studies* xx, xx–xx.
- Snijders, T.A.B., Bosker, R.J., 2012. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling.* 2nd edition. Sage, London.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society Series B* 64, 583–639.
- Steensland, B., Park, J.Z., Regnerus, M.D., Robinson, L.D., Wilcox, B.W., Woodberry, R.D., 2000. The measure of American religion: toward improving the state of the art. *Social Forces* 79, 291–318.
- Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, 528–540.
- Wald, K.D., Silverman, A.L., Fridy, K.S., 2005. Making sense of religion in political life. *Annual Review of Political Science* 8, 121–143.
- Wald, K.D., Wilcox, C., 2006. Getting religion: has political science rediscovered the faith factor? *American Political Science Review* 100, 523–529.
- West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models.* Springer, New York.
- Wuthnow, R., Evans, J.H. (Eds.), 2002. *The Quiet Hand of God. Faith-Based Activism and the Public Role of Mainline Protestantism.* University of California Press, Berkeley.
- Yang, Y., 2006. Bayesian inference for hierarchical age-period-cohort models of repeated cross-section survey data. *Sociological Methodology* 36, 39–74.
- Yang, Y., Land, K.C., 2006. A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology* 36, 75–97.
- Yang, Y., Land, K.C., 2008. Age-period-cohort analysis of repeated cross section surveys. fixed or random effects? *Sociological Methods & Research* 36, 297–326.