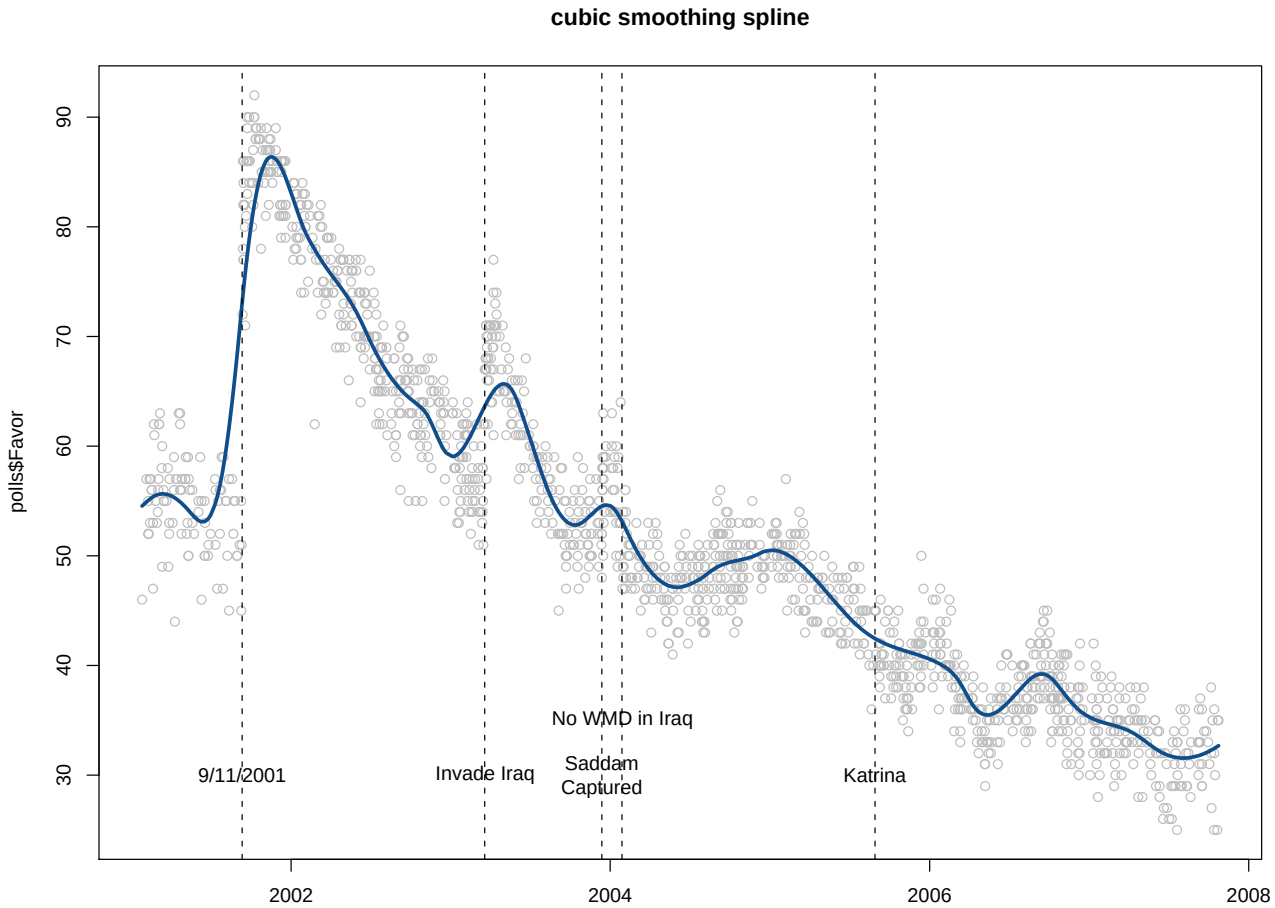
Example: Presidental Polling data

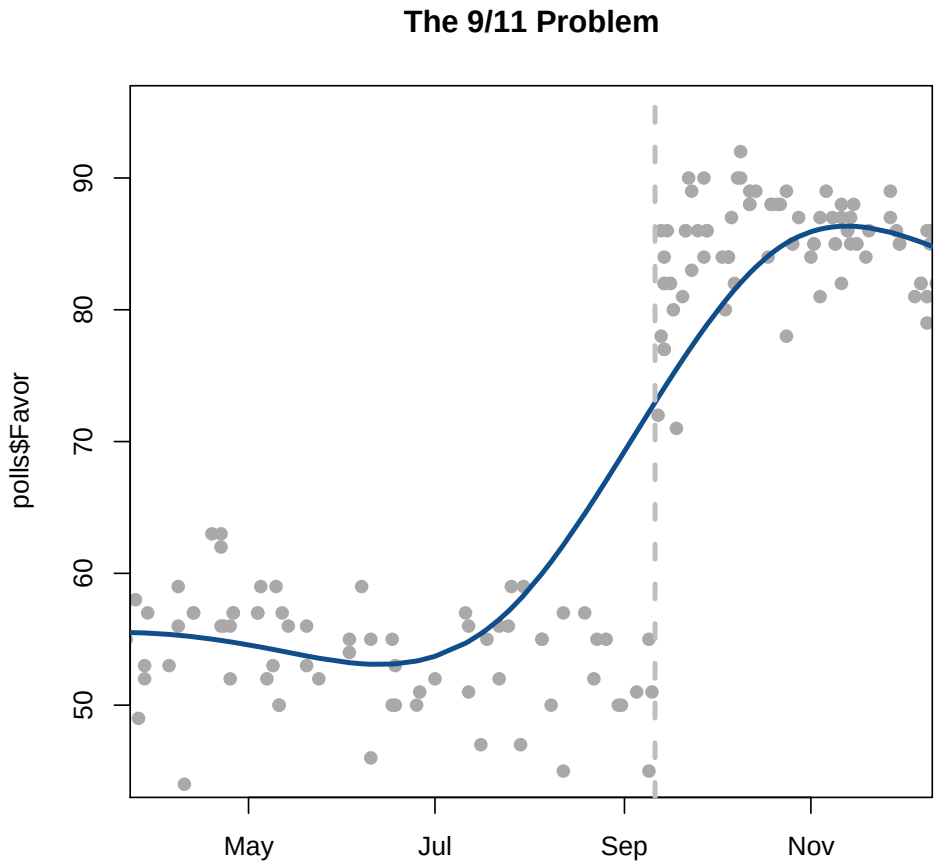
- Results of 1533 polls over President Bush's two terms (January 20, 2001 to present),
- Response (Y) is percent approval
- Domain ($\mathscr{T}$) is calendar date
- Random effects (**b**) for 26 polling houses (e.g., Gallup or Fox News) assumed independent across houses ($B = \sigma^2 I_{26}$), so we can draw inference on the order of a "population of polls"
- These are real random effects: we're curious about the predicted values $\widehat{\mathbf{b}} = E(b|Y)$, since they represent a house specific bias. (We ought to include fixed effects too.)
- This is longitudinal no replicate data, our interest is in finding a good fit to the data not on prediction.
- Thanks to Marc Ratkovic, Dept. of Political Science, for the data/example.

Example: Presidental Polling data



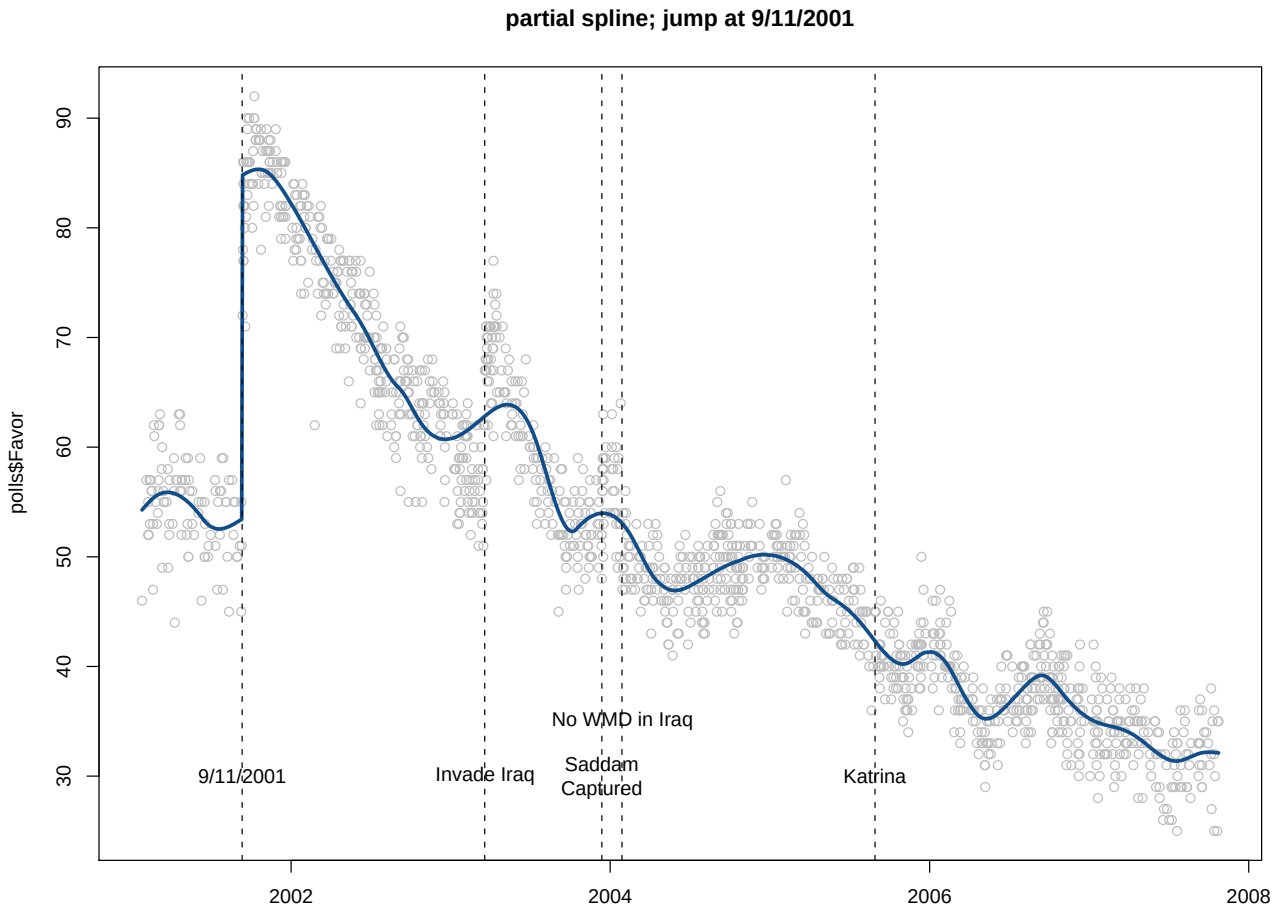
The mixed effects cubic smoothing spline (GCV tuned).

Example: Presidental Polling data



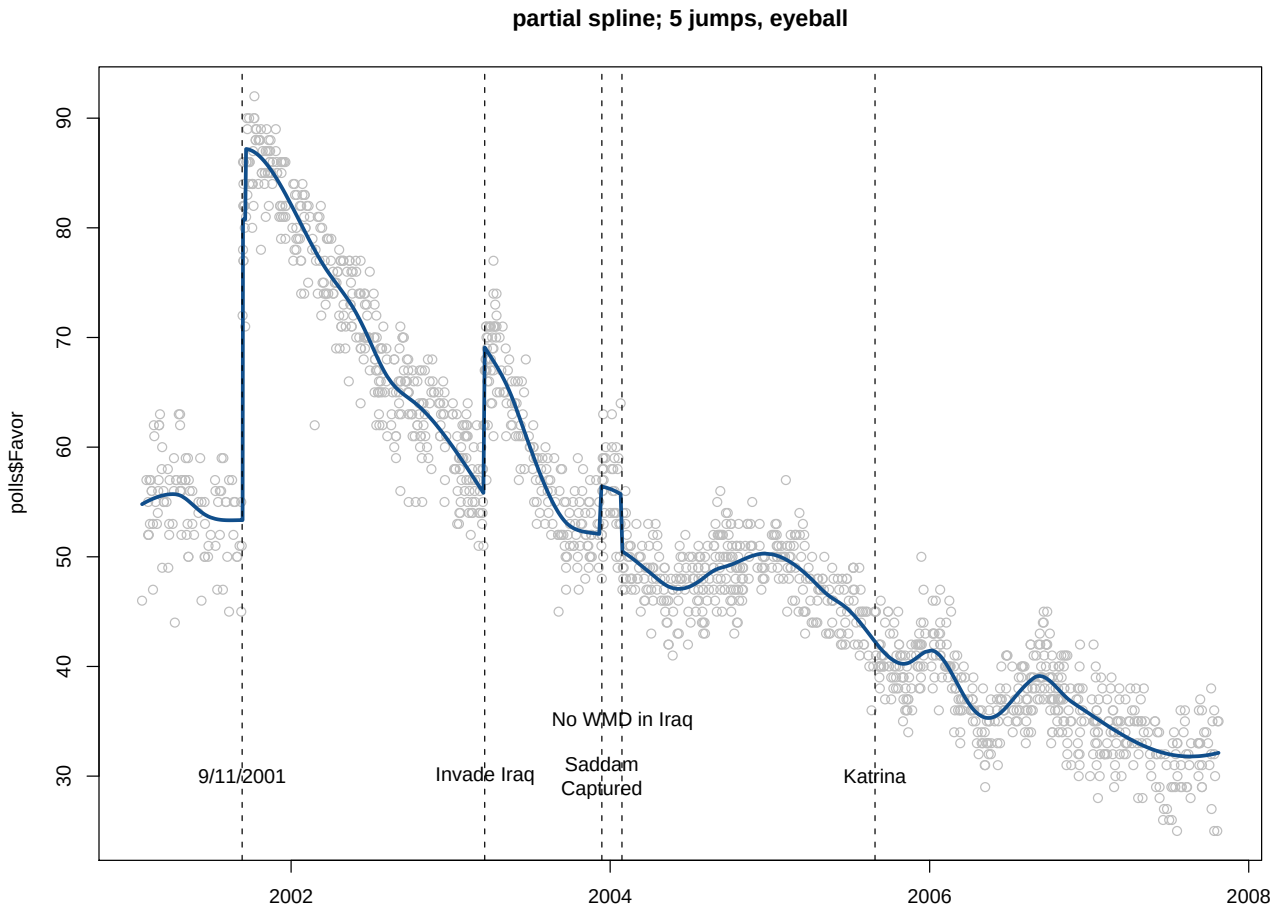
The smoothing spline is “too smooth", it shows the influence of 9/11 months before it happens.

Example: Presidental Polling data



The partial spline handles breaks in the “underlying function," when they’re known. Can we find others?

Example: Presidental Polling data



A first attempt: a forward selection algorithm picks the breaks, tuned by eye.

Follow-up Work

- Partial spline parts (Pa) unpenalized:

$$||(Y - Pa - Zb - Rc - Wd)||_2^2 + n\lambda(\text{c smooth}) + n(\text{b random})$$

- Variable selection problem (recall Xiwen's segmentation problem)

$$||(Y - Pa - Zb - Rc - Wd)||_2^2 + n\lambda_1(\text{a sparse}) + n(\text{b random}) + n\lambda_2(\text{c smooth})$$

- How do pick tuning parameters (sparse, smooth) and fit random effects simultaneously? We need a criterion. Stay tuned.