American Religion Polling Aggregator

Technical Methodology Report (Version 1.2)

Alex Bass

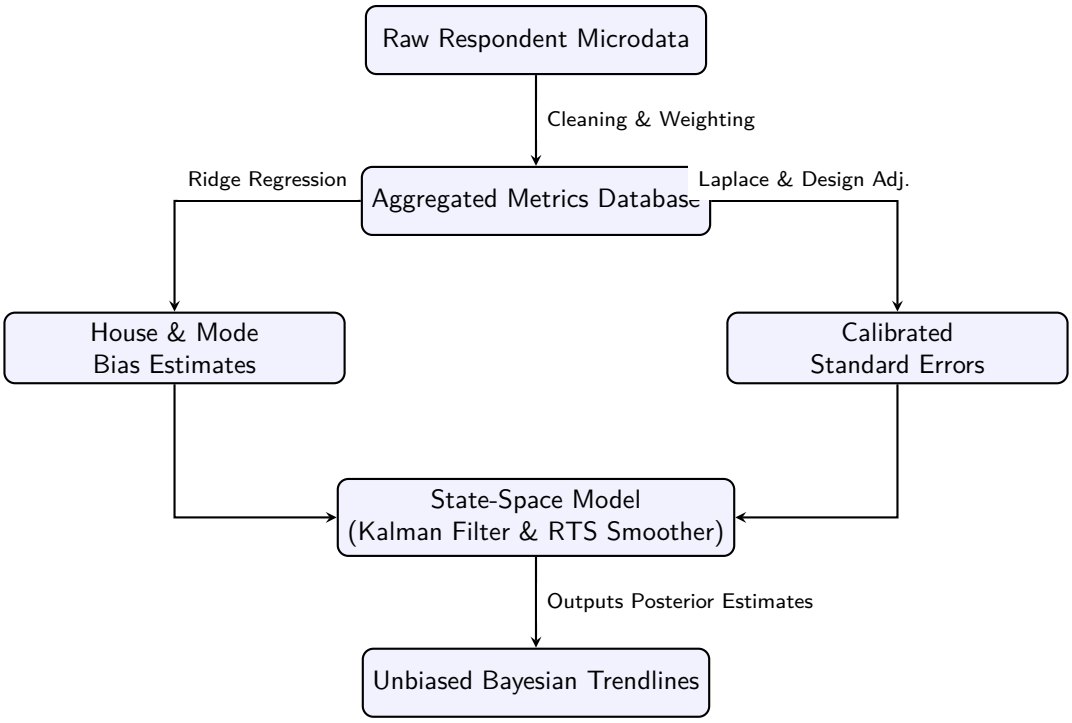
2026-08-01

Table of contents

1 Introduction & Overview

The **American Religion Polling Aggregator** is an interactive, data-driven platform designed to track opinion trends, religious practices, and demographic shifts among Latter-day Saints (LDS) and comparison religious groups in the United States from 2007 to the present.

Because polling data on small religious minorities is sparse, unevenly spaced, and subject to differing survey methodologies, simple averaging fails to represent trends accurately. This report details the multi-stage statistical pipeline designed to unify raw survey observations into a continuous, bias-corrected, and smoothed aggregate trendline.



---

2 Data Pipeline & Preprocessing

The pipeline begins with respondent-level microdata from major polling series, including the Cooperative Election Study (CES), Pew Religious Landscape Studies (RLS) and American Trends Panel (ATP), the General Social Survey (GSS), the Survey Center on American Life (SCOAL), and others.

2.1 Collapsing Granular Scales into Binary Categories

To maintain statistical precision across different survey formats, the aggregator utilizes a strategy of collapsing granular response scales into simple, binary categories. For example, in the education demographic break, the aggregator collapses multi-tier education variables (e.g., high school, some college, associate degree, bachelor's degree, post-graduate work) into a binary threshold: whether the respondent received a Bachelor's degree or higher. This cut-off is methodologically optimal because it is highly salient—respondents reliably report whether they have completed a four-year degree—minimizing reporting variance and measurement error across different survey administrations.

Mathematically, this decision is justified under a Mean Squared Error (MSE) minimization framework:

$$\text{MSE}(\hat{p}) = \text{Bias}(\hat{p})^2 + \text{Var}(\hat{p})$$

For small subgroups (where sample sizes typically hover around $n \approx 30$–100), the sampling variance $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$ dominates the total error. Collapsing response categories dramatically boosts the cell counts within categories, yielding a drastic reduction in sampling variance ($\text{Var}(\hat{p})$). In small-sample scenarios, this variance reduction vastly outweighs the minor discretization bias ($\text{Bias}(\hat{p})^2$) introduced by grouping adjacent response categories together, thereby optimizing the bias-variance trade-off.

To build a continuous time-series across disparate survey houses and waves, variables measuring the same underlying concepts must be reconciled. The aggregator classifies question harmonization into two hierarchical tiers based on alignment precision and the mathematical transformations required:

2.1.1 Tier 1: Direct / High-Fidelity Match

- **Definition**: Questions that share near-identical core phrasing, semantics, and scale structure across survey waves.
- **Conversion Needed**: Minimal to none (e.g., standardizing missing-value codes such as mapping `-99` to `NA`, or renaming variables to a unified schema).
- **Example**:
  - *Survey A*: "Do you approve or disapprove of…" (Approve / Disapprove)
  - *Survey B*: "Do you approve or disapprove of…" (Approve / Disapprove)
- **Statistical Signal**: Gold Standard. Zero statistical transformation bias is introduced.

2.1.2 Tier 2: Scale-Recoded Match

- **Definition**: Questions measuring the exact same underlying concept but using different response categories or scale lengths.

- **Conversion Needed**: Mathematical rescaling or binning (e.g., collapsing multi-point frequency scales to a unified binary threshold). Midpoint or neutral options are either assigned proportionally or treated as a neutral category where appropriate.
- **Example**:
  - *Survey A (5-point scale)*: "Several times a day", "Once a day", "A few times a week", "Once a week", "Seldom/Never"
  - *Survey B (6-point scale)*: "Several times a day", "Once a day", "A few times a week", "Once a week", "A few times a month", "Seldom/Never" → Both rescaled and harmonized to a binary indicator representing "**Daily or more**" by collapsing the top two categories ("Several times a day" and "Once a day").
- **Statistical Signal**: High Rigor. Methodologically sound mathematical transformation with standard, transparent cut-point rules.

In the premium tier of the interactive dashboard, users can hover over any raw survey point in the trend charts to see a detailed tooltip displaying the exact question wording fielded in that specific wave, a description of how the response options were rescaled, and the harmonization tier (Tier 1 or Tier 2) used for that survey item.

2.2 Religion Classification & Harmonization Heuristics

To establish consistent religious categories across diverse polling organizations over a two-decade span, the aggregator implements a standardized categorization heuristic to group respondents into eight mutually exclusive categories: *Latter-day Saints (LDS)*, *Evangelical Protestant*, *Mainline Protestant*, *Black Protestant*, *Catholic*, *Jewish*, *No Religion*, and *Other*.

Because different survey houses use varying classification standards and detail levels in their questionnaires, the coding pipeline applies the following hierarchical rule scheme:

2.2.1 Pre-Coded RELTRAD Standard

Wherever available (e.g., in Pew Research Center datasets that include the pre-calculated `reltrad` or `RELTRAD` classification variable), the aggregator defaults to this gold-standard classification system. In these cases:

- `RELTRAD` value 1 maps to *Evangelical Protestant*.
- `RELTRAD` value 2 maps to *Mainline Protestant*.
- `RELTRAD` value 3 maps to *Black Protestant*.

2.2.2 Dynamic Protestant Demarcation Heuristics (Born-Again & Race)

When a pre-calculated `reltrad` variable is absent, but the survey includes Protestant denominational detail, born-again status, and racial identity, the aggregator dynamically constructs the Protestant subgroup splits. This classification relies on standard sociological definitions (e.g., Steensland et al., 2000):

- **Black Protestant**: Any respondent identifying as Protestant/Christian who also identifies their race as Black or African American.
- **Evangelical Protestant**: Any non-Black respondent identifying as Protestant or non-denominational Christian who responds "yes" to whether they identify as a "born-again or evangelical Christian."
- **Mainline Protestant**: Any non-Black respondent identifying as Protestant or non-denominational Christian who responds "no" to the "born-again or evangelical Christian" self-identification question.

2.2.3 Unification of Non-Religious Cohorts ("No Religion")

Questionnaires differ significantly in how they capture unaffiliated populations—some surveys offer separate, detailed options for "Atheist," "Agnostic," and "Nothing in particular," while others offer only a single "None" or "No religion" response option. To prevent artificial structural breaks in the aggregate time-series, the aggregator pools all these cohorts (including Atheists, Agnostics, Nones, and Unaffiliated respondents) into a single, unified **No Religion** category.

2.3 Two-Stage Estimation Framework: Survey Weighting and Aggregate Modeling

To model public opinion shifts efficiently across dozens of survey series, the aggregator utilizes a **two-stage estimation framework** that separates individual-level demographic weighting from temporal state-space modeling:

1. **Stage 1: Respondent-Level Survey Weighting** Within each individual survey wave, we process the raw respondent microdata. For every target question, we calculate weighted proportions for each demographic subgroup using the survey's official respondent-level weights (w_i). The weights account for non-random sampling frames, differential non-response rates, and demographic post-stratification (e.g., age, sex, education, race) designed by the original polling organizations. This stage collapses the microdata into a set of discrete, aggregated survey points:

$$p_j = \text{weighted proportion for survey wave } j$$

2. **Stage 2: Modeling on Aggregated Survey Points** Rather than running a massive hierarchical model directly on millions of individual-level microdata rows (which is computationally prohibitive given the amount of demographic cuts and metrics we process), the state-space and bias-correction models are fit to the compiled, aggregated survey points $\{p_j\}_{j=1}^{M}$ produced in Stage 1. This decoupling allows us to apply regularized Ridge regression to estimate global house/mode biases and feed the calibrated, bias-corrected point estimates into the quarterly Kalman filter and RTS smoother to produce clean, continuous trend lines.

2.4 Small Group Aggregation Strategy

A primary methodological challenge in tracking religious demographics is obtaining statistically robust sample sizes for small groups. Latter-day Saints (LDS), Jewish, and Black Protestant respondents constitute relatively small proportions of the United States population (approximately 2% each for LDS and Jewish cohorts, and 6–8% for Black Protestants). Consequently, a standard national survey of $1,000$ respondents yields very few respondents from these religious cohorts, resulting in high margins of error and unstable point estimates. A key advantage of the American Religion Polling Aggregator is its ability to pool disparate survey sources to improve statistical precision for these small groups.

To achieve reliable sample sizes, we utilize varying levels of temporal aggregation depending on the density and sample size of the survey source. The guiding principle is to apply the minimum amount of temporal aggregation necessary to yield a reasonable sample size of at least a few hundred respondents per aggregated point:

- **General Social Survey (GSS)**: Because the GSS has a relatively small biennial sample size, we pool GSS waves into a **10-year rolling average** for LDS, Jewish, and Black Protestant cohorts.
- **PRRI Religion News Survey (RNS)**: For PRRI RNS waves fielded between 2010 and 2012, surveys are **aggregated by year** to compile sufficient respondent counts.
- **Pew American Trends Panel (ATP)**: Because the ATP is a longitudinal panel survey where the same respondents are interviewed repeatedly across multiple waves, simply pooling waves within a year would result in double-counting and artificial deflation of standard errors. To resolve this panel structure bias, we apply a rigorous deduplication and reweighting protocol:

    1. *Annual Deduplication*: We aggregate ATP waves **yearly** and **deduplicate** respondents within each calendar year, retaining only one response per unique panelist per year (prioritizing the most recent wave response in the case of multiple survey submissions). This deduplication is applied uniformly to **all demographic groups** (both small groups and larger reference groups alike).

2. *Weight Normalization*: We normalize the weights (w_i) of the deduplicated respondents within each year to sum to the deduplicated sample size N_{dedup}.
3. *Trimming*: To mitigate the impact of outlier weights that would otherwise inflate the variance of our estimates, we apply a weight-trimming protocol, capping weights at the 99th percentile and floor-limiting them at the 1st percentile of the normalized weight distribution.
4. *Kish's Effective Sample Size (n_{eff})*: To adjust the sample size for the design effect ($Deff_w$) introduced by unequal respondent weights, we compute the Kish effective sample size (Kish, 1965, 1992) for each deduplicated year:

$$n_{\text{eff}} = \frac{(\sum w_i)^2}{\sum w_i^2}$$

This calculated n_{eff} is then used as the sample size denominator (n_i) in the standard error estimation to prevent overconfidence in weighted estimates.

- **Nationscape**: Given its massive weekly sample sizes ($n \approx 6,000$), Nationscape waves are pooled to a **monthly aggregation** level.
- **Standard/Large Surveys**: For major series with very large sample sizes (e.g., Pew Landscape Studies, AP VoteCast, and the Cooperative Election Study), points represent the exact end date of the survey administration with **no aggregation** applied.
- **Larger Reference Groups**: For major religious cohorts and the overall US population, sample sizes are inherently large enough that **no aggregation** is necessary, and estimates are computed at the individual wave level.

2.4.1 Kish Effective Sample Size (n_{eff}) for Aggregated Surveys

For all survey sources that use temporal aggregation or pooling (including the General Social Survey's 10-year rolling pools, the American Family Survey's 5-year rolling pools, the PRRI Tracker's annual pools, the Nationscape monthly pools, and the Pew American Trends Panel's annual pools), the sample size denominator (n_i) used in the uncertainty calibration for the three small subgroups (Latter-day Saints, Jewish, and Black Protestant cohorts) is replaced by Kish's Effective Sample Size (n_{eff}). This adjustment accounts for the design effect ($Deff_w$) introduced by unequal respondent weights across the pooled waves, preventing the state-space model from overestimating the precision of these aggregated small-sample estimates.

2.4.2 Asymmetric Temporal Aggregation Rationale

The use of asymmetric temporal aggregation windows (ranging from monthly pools for Nationscape to 10-year pools for the GSS) is methodologically deliberate. For small religious subgroups, the 10-year rolling GSS averages are not intended to capture high-frequency shifts

or act as short-term trend drivers. Instead, they serve as long-term, structural baseline level anchors that stabilize the model's intercept over decades.

During the Kalman filtering process, these rolling averages are assigned inflated measurement variances because their sample sizes represent pooled rather than single-point variance, and the model accounts for the temporal overlap between rolling waves. This prevents them from "dragging" or lagging high-frequency updates from modern, dense tracking polls (like the PRRI Tracker or weekly Nationscape waves), ensuring the RTS smoother remains highly responsive to contemporary inflections while anchored by long-term historical GSS levels.

2.5 Non-Response Filtering (Correcting Denominator Bias)

A key source of bias in survey aggregation is **non-response deflation**. If a survey contains respondents who skipped a target question, dividing the count of positive responses by the *total subgroup size* systematically deflates the resulting proportion.

To resolve this, the data pipeline filters out non-respondents to isolate individuals who provided valid answers. For any demographic subgroup S, the proportion (p) is calculated as:

$$p = \frac{\sum_{i \in S} y_i \cdot w_i \cdot \mathbf{1}(\text{valid}_i)}{\sum_{i \in S} w_i \cdot \mathbf{1}(\text{valid}_i)}$$

where:

- $y_i \in \{0, 1\}$ is the binary response of individual i.
- w_i is the survey respondent weight.
- $\mathbf{1}(\text{valid}_i)$ is an indicator function evaluating to 1 if respondent i answered the question and 0 if they skipped, refused, or answered "don't know".

---

3 Uncertainty Calibration & Standard Errors

Each poll point is weighted in the aggregator based on its statistical precision. Because raw surveys underestimate uncertainty due to design effects and non-sampling errors, the aggregator refines standard errors through three adjustments, following the total survey error framework outlined by Shirani-Mehr, Rothschild, Goel, and Gelman (2018).

3.1 Laplace Smoothing

To prevent standard errors from collapsing to zero when a small demographic subgroup registers 0% or 100% on a metric, the proportion is smoothed using a Laplace correction:

$$p_{\text{smooth}} = \frac{p \cdot n + 2}{n + 4}$$

where n is the active sample size of respondents who answered the question.

3.2 Design Effect Multiplier

Most surveys use complex sampling frames and post-stratification weighting that increase variance (Kish, 1965, 1992). The aggregator applies a method-tiered design effect multiplier (DEFT_i) to account for variance inflation:

- **Tier 1 (Probability-based sampling or live interviewer phone surveys):** $\text{DEFT}_i = 1.15$ (representing a variance inflation factor of 1.32).
- **Tier 2 (Opt-in online panels or non-probability mixed-mode surveys):** $\text{DEFT}_i = 1.35$ (representing a variance inflation factor of 1.82).

3.3 Non-Sampling Noise Floor

Even with infinite sample size, polling is subject to systematic flaws (timing, question wording, context, house effects). Following empirical non-sampling error estimates from Shirani-Mehr et al. (2018), we incorporate a method-tiered poll-level non-sampling noise floor ($\sigma_{\text{ns},i}$):

- **Tier 1:** $\sigma_{\text{ns},i} = 1.5\%$.
- **Tier 2:** $\sigma_{\text{ns},i} = 2.5\%$.

The final calibrated standard error (σ_i) for poll i is calculated as:

$$\sigma_i = \sqrt{\frac{p_{\text{smooth}}(1 - p_{\text{smooth}})}{n_i} \cdot (\text{DEFT}_i)^2 + (\sigma_{\text{ns},i})^2}$$

---

4 House & Mode Bias Correction (Ridge Regression)

Different survey houses (firms) and administration modes (online panels vs. telephone surveys) exhibit systematic deviations. To isolate these biases from actual public opinion shifts, we fit a regularized regression of poll deviations.

4.1 Deviations from Slice Means

For each poll i belonging to a metric-demographic slice s (e.g. *"% married"* among *LDS: Male*), we calculate its deviation (d_i) from the mean of that slice:

$$d_i = y_i - \mu_s$$

4.2 Constrained Ridge Regression

We model the deviations as a function of the polling firm and the methodology mode:

$$d_i = \mathbf{X}_{i,\text{firm}}\alpha + \mathbf{X}_{i,\text{mode}}\beta + \epsilon_i$$

To prevent rank deficiency and enforce that the adjustments are relative, we apply zero-sum constraints ($\sum \alpha_k = 0$ for all firms, and $\sum \beta_m = 0$ for all modes) and add a regularization factor ($\gamma = 0.5$) to prevent overfitting on sparse firms:

$$\text{Minimize } \|\mathbf{d} - \mathbf{X}\theta\|_2^2 + \gamma\|\theta\|_2^2 \quad \text{subject to } \sum \alpha_k = 0, \sum \beta_m = 0$$

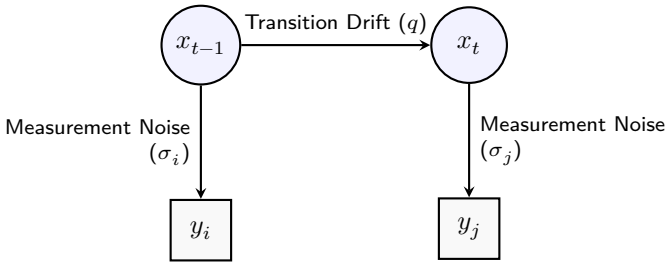
Solving this augmented system yields the firm-specific house effects (α) and mode-specific effects (β).[1] These estimates are subtracted from the raw poll values to align all points on a level playing field.

---

[1] To ensure longitudinal stability and prevent retroactive shifts in historical trend lines (e.g., when raw data weights are updated or new demographic variables are introduced), the regression solver anchors estimated coefficients using static baseline priors. These baseline priors are updated and recalibrated on an annual basis. Incorporating a dynamic, rolling-window Ridge regression to capture time-varying biases is flagged as a planned extension in Version 2.0 of the aggregator.

5 State-Space Model & Kalman Smoothing

For standard metrics with ample historical data, the aggregator tracks public opinion using a linear-Gaussian **State-Space Model (SSM)**, following Jackman (2005) for latent public opinion tracking and dynamic polling aggregation.



5.1 Model Specifications

- **Transition Equation**: The hidden state x_t (true public opinion at quarter t) is modeled as a random walk:
$$x_t = x_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, q^2)$$
where q is the transition drift parameter.
- **Observation Equation**: The adjusted poll points y'_i are observations of the hidden state at their respective grid times $t(i)$:
$$y'_i = x_{t(i)} + v_i, \quad v_i \sim \mathcal{N}(0, \sigma_i^2)$$
where σ_i is the calibrated standard error of the poll.

5.2 Adaptive Transition Noise

The drift q determines the smoothness of the line. The model scales q dynamically based on the standard deviation of the raw polls (std_{raw}), clamped to prevent erratic jumps:

$$q = \max\left(0.001, \min\left(0.006, 0.04 \cdot \text{std}_{\text{raw}}\right)\right)$$

These bounds ($q \in [0.001, 0.006]$) function as an informative Bayesian prior on quarterly state stability. This prior reflects macro-sociological domain limits, enforcing the constraint that true latent public opinion or demographic proportions rarely undergo a shift greater than 0.6% in a single quarter.

5.3 Kalman Filter (Forward Pass)

For each quarter $t \in [1, T]$, we predict the state prior:

$$\theta_{t|t-1} = \theta_{t-1|t-1}$$

$$P_{t|t-1} = P_{t-1|t-1} + q^2$$

If polls are observed at time step t, we update the predictions sequentially for each poll i using the Kalman gain K:

$$K = \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_i^2}$$

$$\theta_{t|t} = \theta_{t|t-1} + K(y_i' - \theta_{t|t-1})$$

$$P_{t|t} = (1 - K)P_{t|t-1}$$

5.4 Rauch-Tung-Striebel Smoother (Backward Pass)

Once the forward pass is complete, we smooth the state estimates backward from $t = T - 2$ down to 0 using future data:

$$C_t = \frac{P_{t|t}}{P_{t+1|t}}$$

$$\theta_{t|T} = \theta_{t|t} + C_t(\theta_{t+1|T} - \theta_{t+1|t})$$

$$P_{t|T} = P_{t|t} + C_t^2(P_{t+1|T} - P_{t+1|t})$$

The 95% Confidence Intervals for the trend line are calculated as:

$$95\%\ \text{CI} = \left[\theta_{t|T} - 1.96\sqrt{P_{t|T}},\quad \theta_{t|T} + 1.96\sqrt{P_{t|T}}\right]$$

5.5 Role of Calibrated Standard Errors in the Bayesian Update

The standard error calibration (σ_i) plays a crucial role in the Bayesian update equations during the filtering pass:

1. **Defining the Measurement Likelihood**: In Bayesian terms, σ_i^2 represents the variance of the observation likelihood function $p(y_i'|x_t) \sim \mathcal{N}(x_t, \sigma_i^2)$. Adjusting the standard errors (via design effects and the non-sampling noise floor) inflates σ_i^2, which widens the likelihood function.

2. **Determining the Kalman Gain (K)**: The Kalman Gain controls the weight given to a new poll relative to the historical prior:

$$K = \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_i^2}$$

   - If a poll has a small sample size or is penalized by our adjustments, its inflated σ_i^2 results in a smaller K (close to 0). The model updates conservatively, relying more heavily on the historical prior $\theta_{t|t-1}$.
   - If a poll is highly precise, its smaller σ_i^2 results in a larger K (approaching 1), pulling the trendline strongly towards the new poll y_i'.

3. **Preventing Overconfidence**: By adding the method-tiered non-sampling noise floor, we guarantee that $\sigma_i^2 \geq 0.000225$ for Tier 1 polls and $\sigma_i^2 \geq 0.000625$ for Tier 2 polls. This acts as a regularization barrier, preventing the posterior variance $P_{t|t} = (1 - K)P_{t|t-1}$ from collapsing to zero even when encountering massive sample sizes. This preserves realistic uncertainty bands throughout history.

---

6 Sparse Metrics Handling (Weighted Least Squares)

The dynamic Kalman smoother requires a continuous and diverse set of survey observations to construct dynamic state transitions. For experimental metrics (which are flagged with a caution symbol in the dashboard's metric selection dropdown and typically feature ≤ 3 unique survey series), the model falls back on **Weighted Least Squares (WLS) linear regression**.

6.1 WLS Formulation

The linear trend is represented as $\hat{y}_t = \beta_0 + \beta_1 \cdot t$. We solve:

$$\beta = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$$

where:

- $\mathbf{W}$ is a diagonal weight matrix of inverse-variances ($W_{ii} = 1/\sigma_i^2$).
- $\mathbf{X}$ is the design matrix containing the column of ones and dates.

The analytical standard error of the trend line prediction at grid time t is:

$$\sigma_t = \sqrt{\mathbf{X}_t^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}_t}$$

7 Citation & Data Attribution

If you cite this methodology or utilize estimates from the American Religion Polling Aggregator (ARPA) in academic publications, policy briefs, or media reporting, please use the following citations:

7.1 APA Format

Bass, A. (2026). *American Religion Polling Aggregator: Technical Methodology Report* (Version 1.2) [Technical Report]. Bass Empirical. https://arpa.bassempirical.com

7.2 BibTeX Entry

```
@techreport{bass2026arpa,
  author      = {Bass, Alex},
  title       = {{American Religion Polling Aggregator: Technical Methodology Report}},
  institution = {Bass Empirical},
  year        = {2026},
  type        = {Methodology Report},
  number      = {v1.2},
  url         = {https://arpa.bassempirical.com},
  note        = {Accessed: 2026-08-10}
}
```

8 References

Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40(4), 499–517.

Kish, L. (1965). *Survey Sampling*. John Wiley & Sons.

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics*, 8(2), 183–200.

Shirani-Mehr, H., Rothschild, D., Goel, S., & Gelman, A. (2018). Disentangling bias and variance in election polls. *Journal of the American Statistical Association*, 113(522), 607–614.

Steensland, B., Park, J. Z., Regnerus, M. D., Robinson, L. D., Wilcox, W. B., & Woodberry, R. D. (2000). The measure of American religion: Toward improving the state of the art. *Social Forces*, 79(1), 291–318.