

# Package ‘agregR’

May 7, 2026

**Encoding** UTF-8

**Type** Package

**URL** <https://rnmag.github.io/agregR/>, <https://github.com/rnmag/agregR>

**Title** Bayesian State-Space Aggregation of Brazilian Presidential Polls

**Description** A set of dynamic measurement models to estimate latent vote shares from noisy polling sources. The models build on Jackman (2009, ISBN: 9780470011546) and feature specialized methods for bias adjustment based on past performance and correction for asymmetric errors based on candidate political alignment.

**Version** 1.0.3

**Depends** R (>= 4.1.0), ggdist (>= 3.1.0)

**Imports** instantiate, tibble, dplyr, tidyr, purrr, readr, stringr, tidyselect, ggplot2, scales, lubridate, cli, ragg, sysfonts, showtext, stringi, grid, systemfonts

**Suggests** bayesplot, cmdstanr (>= 0.5.0), testthat (>= 3.0.0)

**Additional\_repositories** <https://mc-stan.org/r-packages/>

**SystemRequirements** CmdStan  
(<https://mc-stan.org/users/interfaces/cmdstan>)

**LazyData** true

**RoxygenNote** 7.3.3

**Config/testthat/edition** 3

**License** MIT + file LICENSE

**NeedsCompilation** yes

**Author** Rafael N. Magalhães [aut, cre]

**Maintainer** Rafael N. Magalhães <[rnunesmagalhaes@gmail.com](mailto:rnunesmagalhaes@gmail.com)>

**Repository** CRAN

**Date/Publication** 2026-03-06 13:30:09 UTC

## Contents

configurar_agregador . . . . .	2
configurar_grafico . . . . .	4
configurar_prioris . . . . .	5
grafico_agregador . . . . .	7
grafico_priori_posteriori . . . . .	8
grafico_vies . . . . .	9
historico_pesquisas_poder360 . . . . .	10
rodar_agregador . . . . .	12

<b>Index</b>	<b>16</b>
--------------	-----------

---

configurar\_agregador    *Configuration function for Poll Aggregator*

---

## Description

Defines configuration parameters for the poll aggregator, including Stan settings, and election details.

## Usage

```

configurar_agregador(
  pesquisas = NULL,
  resultado_eleicao_passada = NULL,
  resultado_eleicao_atual = NULL,
  historico_pesquisas = NULL,
  candidaturas_1t = NULL,
  candidaturas_2t = NULL,
  direita_eleicao_atual = NULL,
  direita_eleicao_passada = "Bolsonaro",
  esquerda_eleicao_atual = NULL,
  esquerda_eleicao_passada = "Lula",
  eleicao_passada_primeiro_turno = "2/10/2022",
  eleicao_passada_segundo_turno = "30/10/2022",
  stan_cores = pmin(parallel::detectCores(), 4),
  stan_chains = 4,
  stan_warmup = 500,
  stan_sampling = 500,
  stan_init = 0.1,
  stan_adapt_delta = 0.99,
  saida_bases_tratadas = "resultados_agregador/bases_tratadas",
  saida_modelos_brutos = "resultados_agregador/modelos_brutos"
)

```

**Arguments**

pesquisas	Path to a CSV file or URL containing current poll data. Defaults to a GitHub Raw URL.
resultado_eleicao_passada	Path to a CSV file containing results from the previous election. Defaults to a GitHub Raw URL.
resultado_eleicao_atual	Path to a CSV file containing results for the current election (useful for retrospective model). Defaults to a GitHub Raw URL.
historico_pesquisas	Path to a CSV/RDS file containing historical poll data. If NULL (default), uses the package's internal dataset.
candidaturas_1t	Character vector of candidates in the 1st round. If NULL, uses default candidates.
candidaturas_2t	Character vector of candidates in the 2nd round. If NULL, uses default candidates.
direita_eleicao_atual	Character vector of right-wing candidates in the current race. If NULL, uses default candidates. The model can compensate institute errors against right-wing candidates in the last election.
direita_eleicao_passada	Name of the right-wing candidate in the previous election.
esquerda_eleicao_atual	Character vector of left-wing candidates in the current race. If NULL, uses default candidates. The model can compensate institute errors against left-wing candidates in the last election.
esquerda_eleicao_passada	Name of the left-wing candidate in the previous election.
eleicao_passada_primeiro_turno	Date of the previous 1st round (e.g., "2/10/2022").
eleicao_passada_segundo_turno	Date of the previous 2nd round (e.g., "30/10/2022").
stan_cores	Number of CPU cores for Stan to use.
stan_chains	Number of MCMC chains.
stan_warmup	Number of warmup iterations per chain.
stan_sampling	Number of sampling iterations per chain.
stan_init	Initial value for Stan parameters.
stan_adapt_delta	The target acceptance rate for Stan's NUTS algorithm.
saida_bases_tratadas	Directory where treated data will be saved.
saida_modelos_brutos	Directory where raw model objects will be saved.

**Value**

A list of configuration parameters.

**Examples**

```
# Create custom Stan settings
cfg_custom <- configurar_agregador(
  stan_warmup = 100,
  stan_sampling = 100
)
```

---

configurar\_grafico      *Configuration for Graphics*

---

**Description**

Defines configuration parameters for graphics, including colors, fonts, and dimensions.

**Usage**

```
configurar_grafico(
  fonte = "Fira Sans",
  cores_candidaturas = NULL,
  simbolos = NULL,
  graf_largura = 2918,
  graf_altura = 1913,
  graf_unidade = "px",
  graf_dpi = 320,
  dir_grafico = "resultados_agregador/graficos"
)
```

**Arguments**

fonte	Font family (default: "Fira Sans").
cores_candidaturas	Named vector or list of colors for candidates. Can be a partial override.
simbolos	Named vector or list of symbols for methodologies. Can be a partial override.
graf_largura	Width of saved plots.
graf_altura	Height of saved plots.
graf_unidade	Unit for dimensions ("px", "in", "cm", "mm").
graf_dpi	DPI for saved plots.
dir_grafico	Directory to save plots.

**Value**

A list of graphic configuration parameters.

**Examples**

```
# Alternative colors for use in the config_grafico argument in a plot
config_custom <- configurar_grafico(
  cores_candidaturas = c(Lula = "darkred")
)
```

---

configurar\_prioris      *Configuration for Statistical Models*

---

**Description**

Defines hyperparameters for the specific Bayesian models.

**Usage**

```
configurar_prioris(nome = "Viés Relativo com Pesos", ...)
```

**Arguments**

nome	Name of the model. Options: "Viés Relativo com Pesos", "Viés Relativo sem Pesos", "Viés Empírico", "Retrospectivo" and "Naive".
...	Named arguments to override default hyperparameters (e.g., sd_tau_priori = 0.05).

**Value**

A list of model parameters.

**Priors Details**

These hyperparameters control the strength of assumptions regarding latent state evolution, institute bias, and non-sampling errors.

Variable names refer to the model notation described in <https://rnmag.github.io/agregR/index.html#conceptual-framework>

Recommended reading: <https://github.com/stan-dev/stan/wiki/prior-choice-recommendations>

**State Model - Level ( $\mu$ )**

- mu\_priori: Prior mean for the latent vote share at  $t = 1$ .
- sd\_mu\_priori: Prior uncertainty for the initial latent vote.
  - *Default values:*  $\mu$  starts with a flat prior of  $N(0.5, 0.5)$ , allowing data to quickly dominate inference.
- omega\_eta\_priori: Prior mean for the level volatility ( $\omega_\eta$ ).
- sd\_omega\_eta\_priori: Prior uncertainty for the level volatility.

- *Default values:* With  $\text{omega\_eta\_priori} = 0.002$  and  $\text{sd\_omega\_eta\_priori} = 0.0001$ , the model assumes a **baseline drift** of approx.  $\pm 2$  percentage points over a month ( $1.96 \times \sqrt{30} \times 0.002 \approx 0.02$ ).
- *Higher values:* The latent vote ( $\mu$ ) can jump more from one day to the next. The model adapts more quickly to new polls but becomes more "jittery".
- *Lower values:* The model assumes the public opinion level is more stable over time, resulting in smoother curves.

### State Model - Trend ( $\nu$ )

- $\text{nu\_priori}$ : Prior mean for the initial trend (daily growth rate).
- $\text{sd\_nu\_priori}$ : Prior uncertainty for the initial trend.
  - *Default values:* With  $\text{nu\_priori} = 0$  and  $\text{sd\_nu\_priori} = 0.001$ , the model expects an initial trend within  $\pm 0.2$  percentage points per day ( $1.96 \times 0.001 \approx 0.002$ ).
- $\text{omega\_zeta\_priori}$ : Prior mean for the trend volatility ( $\omega_\zeta$ ).
- $\text{sd\_omega\_zeta\_priori}$ : Prior uncertainty for the trend volatility.
  - *Default values:* With  $\text{omega\_zeta\_priori} = 0$  and  $\text{sd\_omega\_zeta\_priori} = 0.00001$ , the model assumes a linear evolution, allowing the trend to shift rapidly (accelerations) only under strong evidence.
  - *Higher values:* The trend ( $\nu$ ) can change direction or magnitude rapidly.
  - *Lower values:* The trend is assumed to be more constant over time (more linear evolution).

### Institute Bias ( $\delta$ )

- $\text{delta\_priori}$ : Mean expected bias for institutes. Default is 0, except in "Viés Empírico" where it is anchored on past performance.
- $\text{sd\_delta\_priori}$ : Scale of the bias prior.
  - *Default values:* With  $\text{delta\_priori} = 0$  and  $\text{sd\_delta\_priori} = 0.02$ , the model assumes that 95% of institutes have a bias within  $\pm 4$  percentage points ( $1.96 \times 0.02 \approx 0.04$ ).
  - *Higher values:* Allow for larger, more variable biases across institutes.
  - *Lower values:* Constrain institutes to have similar biases (shrinkage toward the anchor).

### Non-Sampling Error ( $\tau$ )

- $\text{tau\_priori}$ : Mean expected magnitude of errors not explained by sampling or house effects. In weighted models, this is replaced by the empirical RMSE from past elections.
- $\text{sd\_tau\_priori}$ : Prior uncertainty for non-sampling error.
  - *Default values:* With  $\text{tau\_priori} = 0.02$  and  $\text{sd\_tau\_priori} = 0.02$ , the model assumes a **baseline** of  $\pm 4$  percentage points of "noise" in each poll, allowing it to spread closer to  $\pm 7$  percentage points.
  - *Higher values:* The model treats polls as less precise, widening the credible intervals of the latent state.
  - *Lower values:* The model trusts polling precision more, leading to tighter intervals and potentially more sensitivity to outliers.

## Examples

```
# Get default parameters for the "Naive" model
naive_params <- configurar_prioris(nome = "Naive")

# Get parameters for "Naive" and override a default value
custom_params <- configurar_prioris(nome = "Naive", sd_mu_priori = 0.2)
```

---

grafico\_agregador      *Plot Aggregator Results*

---

## Description

Generates a plot of the aggregated poll results over time.

## Usage

```
grafico_agregador(
  bd,
  salvar = FALSE,
  config_grafico = configurar_grafico(),
  dir_saida = NULL,
  ...
)
```

## Arguments

bd	The results object returned by <code>rodar_agregador()</code> .
salvar	Logical. If TRUE, saves the plot to disk.
config_grafico	A list of graphic parameters created by <code>configurar_grafico()</code> .
dir_saida	Output directory for the saved plot if <code>salvar = TRUE</code> .
...	Additional arguments.

## Value

A ggplot2 object.

## Examples

```
if (instantiate::stan_cmdstan_exists()) {
  result <- rodar_agregador(
    data_inicio = "01/01/2025",
    turno = 2,
    cenario = "Lula vs Bolsonaro"
  )

  # Standard plot
  std_plot <- grafico_agregador(result)
```

```

# Altering candidate colors
custom_plot <- grafico_agregador(
  result,
  config_grafico = configurar_grafico(
    cores_candidaturas = c(Lula = "yellow")
  )
)
}

```

---

grafico\_priori\_posteriori

*Plot Prior vs Posterior*


---

### Description

Generates a plot comparing prior and posterior distributions for candidates or bias.

### Usage

```

grafico_priori_posteriori(
  bd,
  candidaturas,
  tipo = "Viés",
  salvar = FALSE,
  config_agregador = configurar_agregador(),
  config_grafico = configurar_grafico(),
  config_prioris = configurar_prioris(bd$nome_modelo),
  dir_saida = NULL
)

```

### Arguments

bd	The results object returned by <a href="#">rodar_agregador()</a> .
candidaturas	A character vector of candidate names to include in the plot.
tipo	The type of da to plot: "Viés" (for institute bias) or "Percentual" (for candidate voting share).
salvar	Logical. If TRUE, saves the plot to disk.
config_agregador	A list of configuration parameters created by <a href="#">configurar_agregador()</a> .
config_grafico	A list of graphic parameters created by <a href="#">configurar_grafico()</a> .
config_prioris	A list of model hyperparameters created by <a href="#">configurar_prioris()</a> .
dir_saida	Output directory for the saved plot if salvar = TRUE.

### Value

A ggplot2 object.

## Examples

```
if (instantiate::stan_cmdstan_exists()) {
  result <- rodar_agregador(
    data_inicio = "01/01/2025",
    turno = 2,
    cenario = "Lula vs Bolsonaro"
  )

  # Prior vs Posterior plot for institute bias
  std_plot <- grafico_priori_posteriori(
    result,
    tipo = "Viés",
    candidaturas = c("Lula", "Bolsonaro")
  )

  # Altering candidate colors
  custom_plot <- grafico_priori_posteriori(
    result,
    candidaturas = c("Lula", "Bolsonaro"),
    config_grafico = configurar_grafico(
      cores_candidaturas = c(Lula = "yellow")
    )
  )
}
```

---

grafico\_vies

*Plot Institute Bias*

---

## Description

Generates a plot visualizing the bias of polling institutes.

## Usage

```
grafico_vies(
  bd,
  candidaturas,
  salvar = FALSE,
  config_grafico = configurar_grafico(),
  dir_saida = NULL,
  ...
)
```

## Arguments

bd	The results object returned by <code>rodar_agregador()</code> .
candidaturas	A character vector of candidate names to include in the plot.
salvar	Logical. If TRUE, saves the plot to disk.

`config_grafico` A list of graphic parameters created by `configurar_grafico()`.  
`dir_saida` Output directory for the saved plot if `salvar = TRUE`.  
... Additional arguments.

### Value

A `ggplot2` object.

### Examples

```
if (instantiate::stan_cmdstan_exists()) {  
  result <- rodar_agregador(  
    data_inicio = "01/01/2025",  
    turno = 2,  
    cenario = "Lula vs Bolsonaro"  
  )  
  
  # Standard bias plot  
  std_plot <- grafico_vies(  
    result,  
    candidaturas = c("Lula", "Bolsonaro")  
  )  
  
  # Altering candidate colors  
  custom_plot <- grafico_vies(  
    result,  
    candidaturas = c("Lula", "Bolsonaro"),  
    config_grafico = configurar_grafico(  
      cores_candidaturas = c(Lula = "yellow")  
    )  
  )  
}
```

---

historico\_pesquisas\_poder360

*Historical Polls by Poder360*

---

### Description

A dataset containing historical electoral polls compiled by Poder360. This dataset is used to calculate empirical priors for the models.

### Usage

historico\_pesquisas\_poder360

### Format

A data frame with columns:

**ano** Election year  
**cargo** Office being contested  
**condicao** Condition (e.g., stimulated)  
**contratante** Entity that paid for the poll  
**data** Date of the poll  
**data\_referencia** Reference date for the poll  
**descricao\_cenario** Description of the electoral scenario  
**id\_candidato\_poder360** Unique ID for the candidate  
**id\_cenario** Unique ID for the scenario  
**id\_pesquisa** Unique ID for the poll  
**instituto** Name of the polling institute  
**margem\_mais** Upper margin of error  
**margem\_menos** Lower margin of error  
**nome\_candidato** Candidate name  
**nome\_municipio** City name (if applicable)  
**numero\_registro** Official registration number  
**orgao\_registro** Entity where the poll was registered  
**percentual** Voting intention percentage  
**quantidade\_entrevistas** Sample size  
**sigla\_partido** Political party abbreviation  
**sigla\_uf** State abbreviation  
**tipo** Poll type  
**tipo\_voto** Vote type (Total, Valid, etc.)  
**turno** Election round (1 or 2)

### Source

Poder360 via [Base dos Dados](#)

---

rodar_agregador	<i>Run Poll Aggregator</i>
-----------------	----------------------------

---

### Description

Main function to run the state-space model for poll aggregation.

### Usage

```
rodar_agregador(
  bd = NULL,
  data_inicio = NULL,
  data_fim = Sys.Date(),
  cargo = "Presidente",
  ambito = "Brasil",
  cenario = NULL,
  turno,
  modelo = "Viés Relativo com Pesos",
  config_agregador = NULL,
  config_prioris = NULL,
  salvar = FALSE,
  dir_saida = NULL
)
```

### Arguments

bd	Dataframe or path to a CSV file containing poll data.
data_inicio	Start date for the analysis (mandatory).
data_fim	End date for the analysis.
cargo	The office/position being contested (e.g., "Presidente"). Current data only contains presidential polls, but the package supports expansion for other offices.
ambito	The geographical scope (e.g., "Brasil"). Current data only contains national polls, but the package supports expansion for state races.
cenario	The specific electoral scenario. Mandatory for second round.
turno	The election round (1 or 2).
modelo	The name of the model to run. Options: "Viés Relativo com Pesos" (default), "Viés Relativo sem Pesos", "Viés Empírico", "Retrospectivo" and "Naive".
config_agregador	A list of configuration parameters created by <a href="#">configurar_agregador()</a> . If NULL, uses defaults.
config_prioris	A list of model hyperparameters created by <a href="#">configurar_prioris()</a> . If NULL, uses defaults based on modelo.
salvar	Logical. If TRUE, saves the results to disk.
dir_saida	Output directory for saved files if salvar = TRUE.

**Value**

A list containing the model name, estimated votes, institute bias, and the raw model object.

**Model Details**

The aggregator supports five types of Bayesian state-space models, each with specific assumptions about institute bias and non-sampling errors:

**1. Viés Relativo com Pesos (Default)**

- **Assumption:** Institute biases are relative to the average of all institutes (latent "truth" is anchored to the consensus).
- **Bias ( $\delta$ ):** Calculated relative to the mean bias of all institutes.
- **Weights ( $\tau$ ):** Uses past election performance to weight the non-sampling error. Institutes with larger historical errors have less influence on the current estimate.
- **Use case:** Best for general forecasting when historical data is available.

**2. Viés Relativo sem Pesos**

- **Assumption:** Same as above, but treats all institutes as having equal potential quality a priori.
- **Bias ( $\delta$ ):** Calculated relative to the mean bias.
- **Weights ( $\tau$ ):** None. All institutes share the same prior for non-sampling error.
- **Use case:** When historical data is unreliable or when a "fresh start" assumption is desired.

**3. Viés Empírico**

- **Assumption:** Institute biases are anchored to their specific historical performance.
- **Bias ( $\delta$ ):** Prior means are set to the bias observed in the previous election (directional error).
- **Weights ( $\tau$ ):** Uses past performance for non-sampling error, similar to the "Com Pesos" model.
- **Use case:** When institutes are expected to repeat their specific past directional errors (e.g., consistently underestimating a specific wing).

**4. Retrospectivo**

- **Assumption:** The true election result is known and used as the final anchor for the state-space model.
- **Method:** Runs the model "backwards" or constrained by the final result to estimate the true path of public opinion.
- **Use case:** Post-election analysis to diagnose institute performance and calculate accurate biases for future calibration.

**5. Naive**

- **Assumption:** Polls have no bias and no non-sampling error.
- **Method:** A random walk model where the only source of uncertainty is the sampling error ( $\sigma$ ).
- **Use case:** Baseline comparison. Assumes "polls are perfect" within their margin of error.

## Priors Details

The `config_priors` argument allows customization of the model's hyperparameters with the `configurar_priors()` function.

These hyperparameters control the strength of assumptions regarding latent state evolution, institute bias, and non-sampling errors.

Variable names refer to the model notation described in <https://rnmag.github.io/agregR/index.html#conceptual-framework>

Recommended reading: <https://github.com/stan-dev/stan/wiki/prior-choice-recommendations>

### State Model - Level ( $\mu$ )

- `mu_priori`: Prior mean for the latent vote share at  $t = 1$ .
- `sd_mu_priori`: Prior uncertainty for the initial latent vote.
  - *Default values*:  $\mu$  starts with a flat prior of  $N(0.5, 0.5)$ , allowing data to quickly dominate inference.
- `omega_eta_priori`: Prior mean for the level volatility ( $\omega_\eta$ ).
- `sd_omega_eta_priori`: Prior uncertainty for the level volatility.
  - *Default values*: With `omega_eta_priori = 0.002` and `sd_omega_eta_priori = 0.0001`, the model assumes a **baseline drift** of approx.  $\pm 2$  percentage points over a month ( $1.96 \times \sqrt{30} \times 0.002 \approx 0.02$ ).
  - *Higher values*: The latent vote ( $\mu$ ) can jump more from one day to the next. The model adapts more quickly to new polls but becomes more "jittery".
  - *Lower values*: The model assumes the public opinion level is more stable over time, resulting in smoother curves.

### State Model - Trend ( $\nu$ )

- `nu_priori`: Prior mean for the initial trend (daily growth rate).
- `sd_nu_priori`: Prior uncertainty for the initial trend.
  - *Default values*: With `nu_priori = 0` and `sd_nu_priori = 0.001`, the model expects an initial trend within  $\pm 0.2$  percentage points per day ( $1.96 \times 0.001 \approx 0.002$ ).
- `omega_zeta_priori`: Prior mean for the trend volatility ( $\omega_\zeta$ ).
- `sd_omega_zeta_priori`: Prior uncertainty for the trend volatility.
  - *Default values*: With `omega_zeta_priori = 0` and `sd_omega_zeta_priori = 0.00001`, the model assumes a linear evolution, allowing the trend to shift rapidly (accelerations) only under strong evidence.
  - *Higher values*: The trend ( $\nu$ ) can change direction or magnitude rapidly.
  - *Lower values*: The trend is assumed to be more constant over time (more linear evolution).

### Institute Bias ( $\delta$ )

- `delta_priori`: Mean expected bias for institutes. Default is 0, except in "Viés Empírico" where it is anchored on past performance.
- `sd_delta_priori`: Scale of the bias prior.

- *Default values:* With  $\text{delta\_priori} = 0$  and  $\text{sd\_delta\_priori} = 0.02$ , the model assumes that 95% of institutes have a bias within  $\pm 4$  percentage points ( $1.96 \times 0.02 \approx 0.04$ ).
- *Higher values:* Allow for larger, more variable biases across institutes.
- *Lower values:* Constrain institutes to have similar biases (shrinkage toward the anchor).

### Non-Sampling Error ( $\tau$ )

- $\text{tau\_priori}$ : Mean expected magnitude of errors not explained by sampling or house effects. In weighted models, this is replaced by the empirical RMSE from past elections.
- $\text{sd\_tau\_priori}$ : Prior uncertainty for non-sampling error.
  - *Default values:* With  $\text{tau\_priori} = 0.02$  and  $\text{sd\_tau\_priori} = 0.02$ , the model assumes a **baseline** of  $\pm 4$  percentage points of "noise" in each poll, allowing it to spread closer to  $\pm 7$  percentage points.
  - *Higher values:* The model treats polls as less precise, widening the credible intervals of the latent state.
  - *Lower values:* The model trusts polling precision more, leading to tighter intervals and potentially more sensitivity to outliers.

### Examples

```
# Running the default model for a second round scenario
if (instantiate::stan_cmdstan_exists()) {
  result <- rodar_agregador(
    data_inicio = "01/01/2025",
    turno = 2,
    cenario = "Lula vs Bolsonaro"
  )

# Tuning Stan, changing the model and altering specific priors
custom_result <- rodar_agregador(
  data_inicio = "01/01/2025",
  turno = 2,
  cenario = "Lula vs Bolsonaro",
  modelo = "Viés Relativo sem Pesos",
  config_agregador = list(stan_chains = 1, stan_warmup = 200),
  config_priors = list(tau_priori = 0.01)
)
}
```

# Index

## \* datasets

historico\_pesquisas\_poder360, 10

configurar\_agregador, 2

configurar\_agregador(), 8, 12

configurar\_grafico, 4

configurar\_grafico(), 7, 8, 10

configurar\_prioris, 5

configurar\_prioris(), 8, 12, 14

grafico\_agregador, 7

grafico\_priori\_posteriori, 8

grafico\_vies, 9

historico\_pesquisas\_poder360, 10

rodar\_agregador, 12

rodar\_agregador(), 7-9