An Empirical Bayes Approach to Nonparametric Regression with Correlated Errors

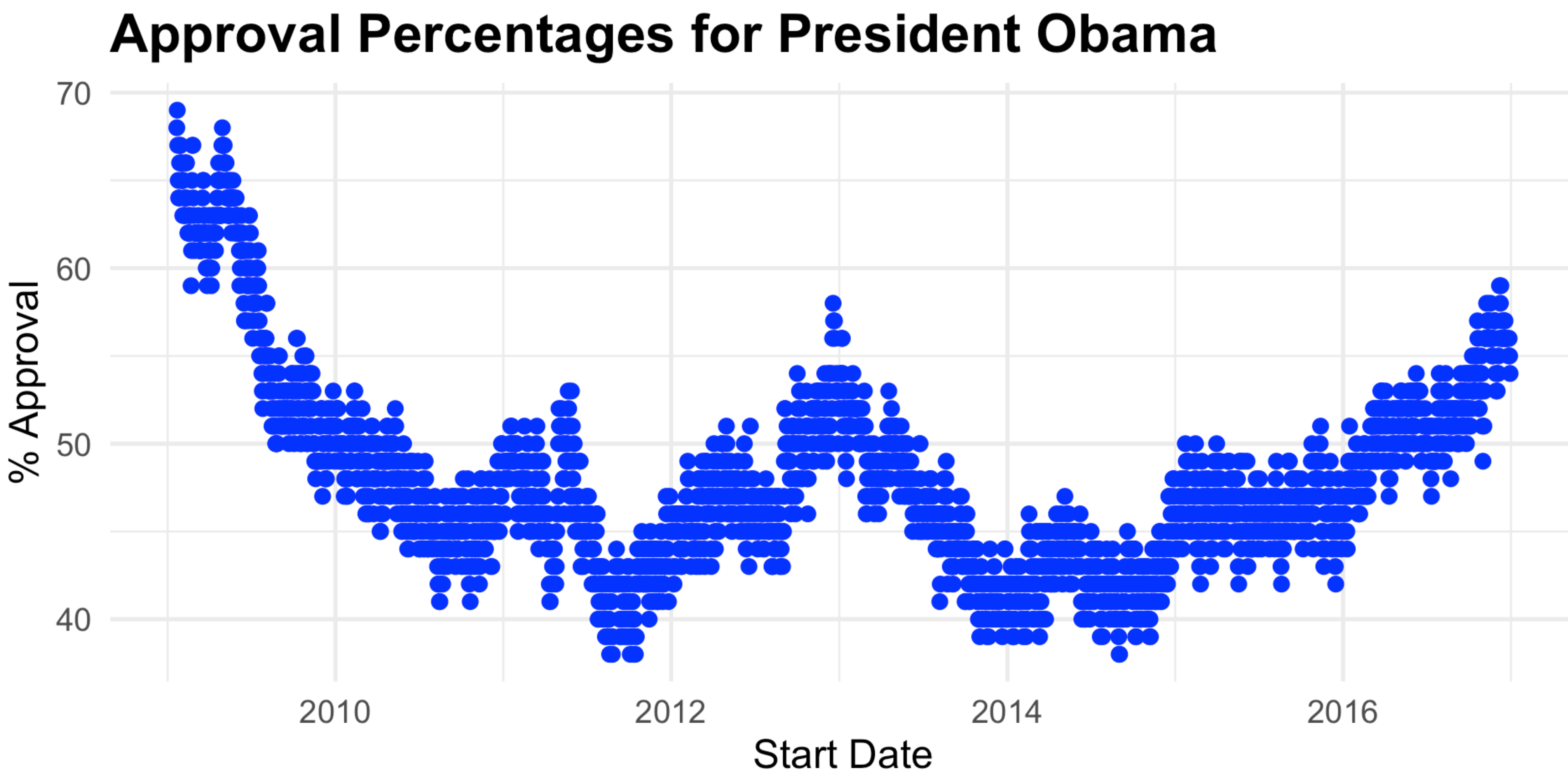
Research Presentation
Cal Poly, San Luis Obispo, CA

Ethan P. Marzban
Department of Statistics and Applied Probability
University of California, Santa Barbara

January 12, 2026

Presidential Approval Ratings

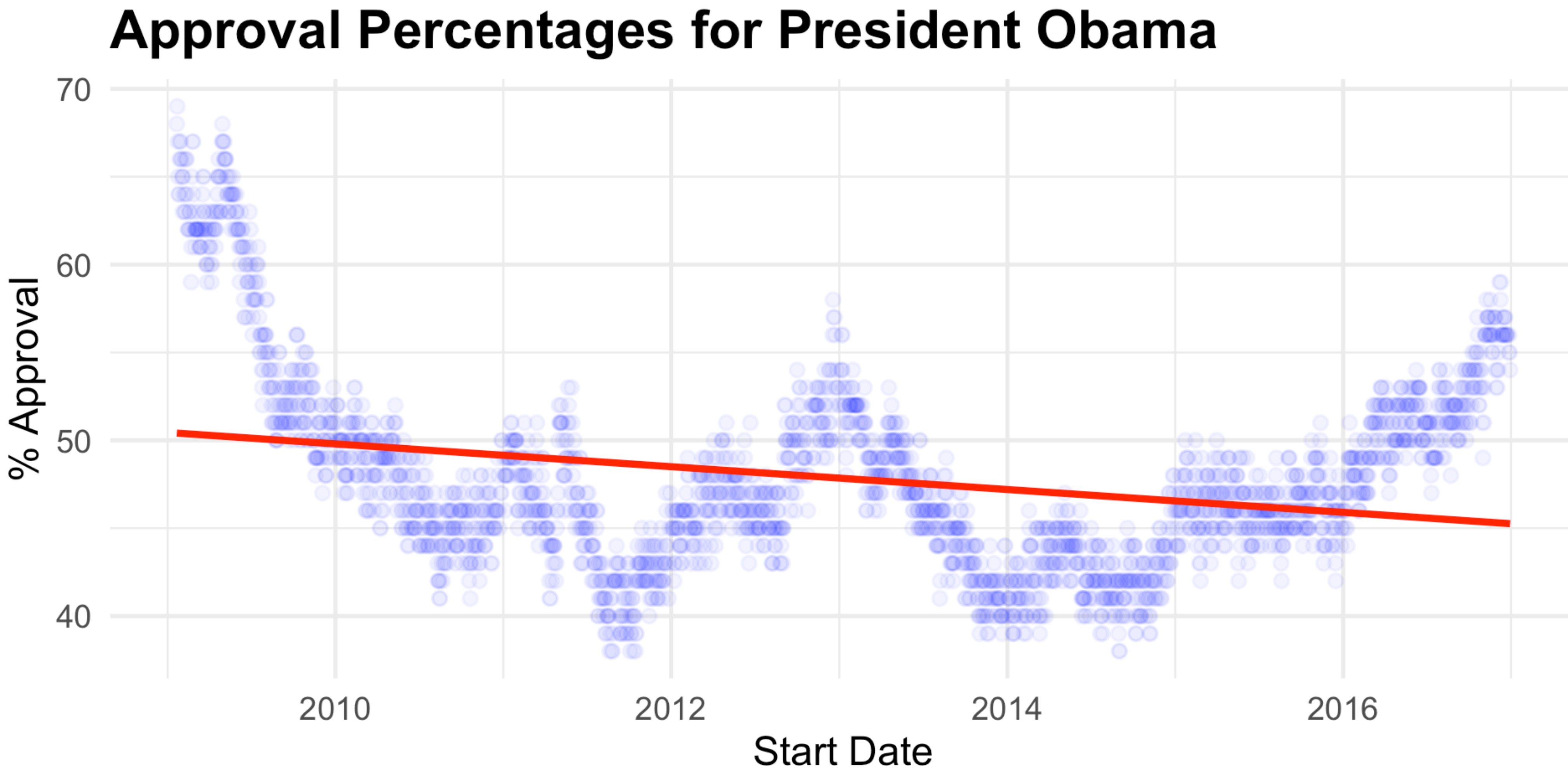
Visualization



Approval Ratings for President Obama across his two terms in office. Each point represents a three-day average.

Presidential Approval Ratings

Linear Fit



Approval Ratings for President Obama across his two terms in office. Each point represents a three-day average.

Presidential Approval Ratings

Nonparametric Regression

- A **nonparametric regression** problem
  - → I.e. can we estimate the true underlying **signal** function with as few assumptions as possible?
- Equally-spaced univariate data: $\mathcal{D} = \{(t, y_t)\}_{t=1}^{n}$
  - → $y_t =$ approval rating on day *t*
- **Model:** $y_t = f(t) + \omega_t$
  - → **Signal Function** $f(\cdot)$ is to be estimated
- Many pre-existing nonparametric regression techniques assume i.i.d. Gaussian **Noise:** ω_t; for this dataset, however, this is a *poor* assumption!
  - → Approval ratings have been reported as *three-day averages*

Main Ideas

An Overview of Our Procedure

- As such, we seek to develop an estimation procedure under the assumption of *correlated* noise.
- We leverage two main ideas in our approach:
    1. Utilizing a **Bayesian** approach (technically, an **Empirical Bayes**)
    2. Utilizing the **spectral domain** (**Fourier Transforms**)
- With these two ideas, we are able to reduce the estimation problem to a **Generalized Linear Model** (GLM).
- To ensure we're all on roughly equal footing, let's briefly discuss these topics in generality first

Bayesian Estimation

A High-Level Overview

- In a typical (**frequentist**) paradigm, parameters are treated as *deterministic* (i.e. nonrandom)
- Example: $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with σ^2 known; then μ, the unknown mean, is treated as fixed
  - → To account for variability across samples, we develop **estimators** (e.g. $\bar{X}$) to estimate the "true value" of μ.
- The Bayesian framework acknowledges that we are coming into our problem with certain *prior beliefs*, and, consequently, treats μ as a *random variable in itself*, following some prespecified **prior distribution**.
  - → Rather than looking at point estimators (like $\bar{X}$), we instead estimate μ with a **posterior distribution** $\pi(\mu \mid X_1, X_2, \cdots, X_n)$.

Bayesian Estimation

An Example

- **Example:** the so-called **normal-normal problem**:

$$\begin{aligned}(X_1, X_2, \cdots, X_n \mid \mu) &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2) \\ \mu &\sim \mathcal{N}(\nu, \tau^2)\end{aligned}$$

- With some work, the posterior distribution can be shown to be:

$$(\mu \mid X_1, X_2, \cdots, X_n) \sim \mathcal{N}(\mu_{\text{post}},\ \sigma^2_{\text{post}})$$

$$\begin{aligned}\mu_{\text{post}} &:= \left(\frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right) \nu + \left(\frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right) \bar{X} \\ \sigma^2_{\text{post}} &:= \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\end{aligned}$$

Bayesian Estimation

An Example

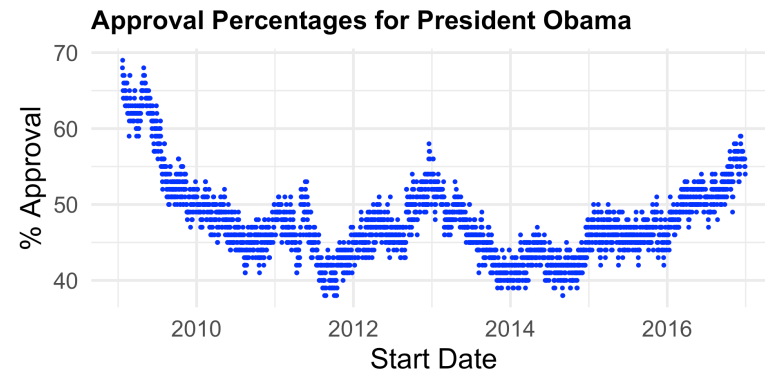
$$\mu_{\text{post}} := \left(\frac{\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \right) \nu + \left(\frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} \right) \bar{X}$$

- What we see is that the posterior mean is a weighted average of the *prior estimate* (ν) and the *frequentist estimator* ($\bar{X}$), with the weights proportional to the sums of the prior and model variances.
- A smaller prior variance (indicating more *certainty* in our choice of prior) leads to a higher weighting of ν over $\bar{X}$, and vice-versa.
- This assumes σ^2 to be known; if it were not, we could either assign a **hyperprior** (pure Bayesian) or estimate it using data (**Empirical Bayes**)

Presidential Approval Ratings

The Model

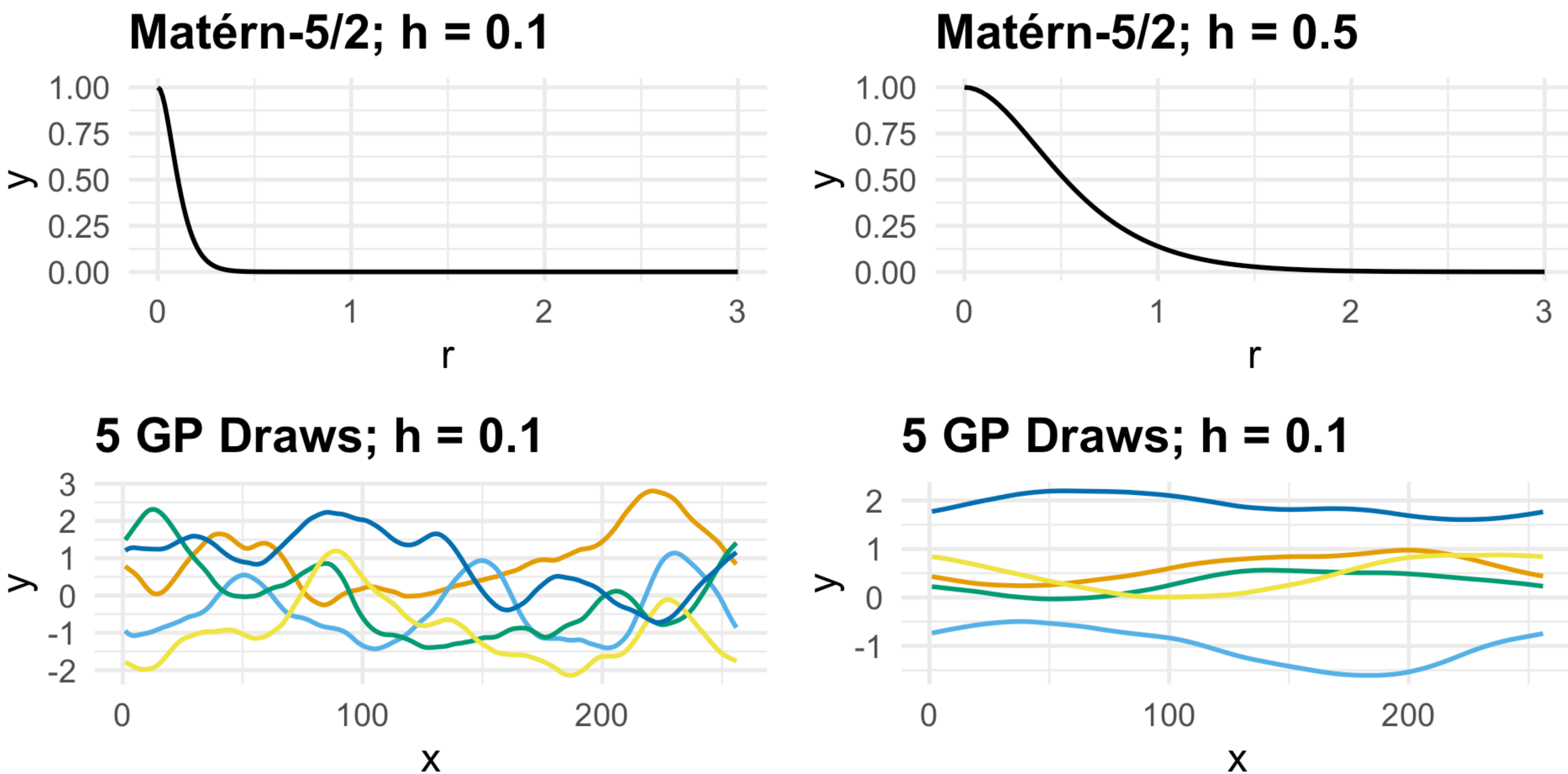
- **Model:** $y_t = f(t) + \omega_t$
  - → $y_t =$ approval rating on day *t*
  - → Signal function $f(\cdot)$ is to be estimated



- Adopting a Bayesian perspective, we assign a prior to the *signal function*
  - → The prior we use is called a **Gaussian Process**; think of it as a distribution such that draws from this distributions are *functions* (as opposed to *random vectors*).
    - ⤞ Prescribed by a mean *function* and a covariance *kernel* (akin to μ and σ^2, respectively)

Gaussian Process

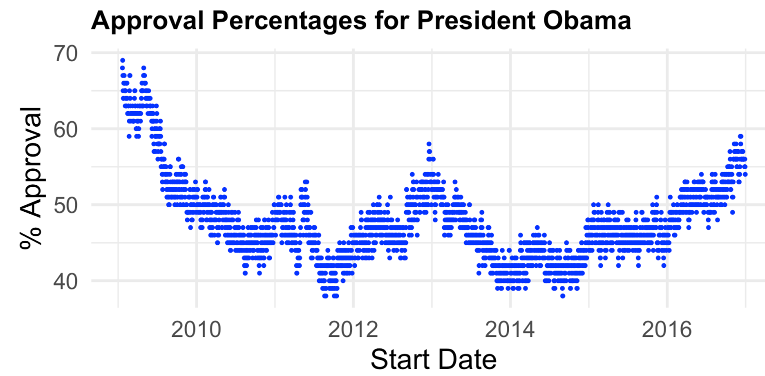
Example Draws



Presidential Approval Ratings

The Model

- **Model:** $y_t = f(t) + \omega_t$
  - → y_t = approval rating on day *t*
  - → Signal function $f(\cdot)$ is to be estimated



- **Prior:** $f \sim \mathcal{GP}(0,\ \tau^2 C(r))$ where $C(r)$ denotes the **Matérn-5/2 Kernel**

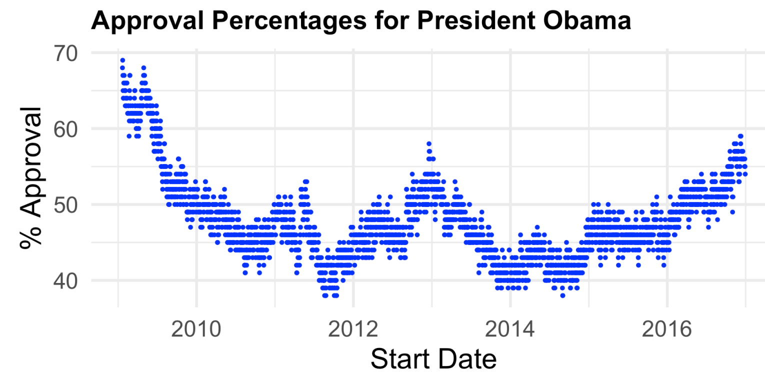
$$C(r) = \left(1 + \frac{r\sqrt{5}}{h} + \frac{5r^2}{3h^2} \right) \exp\left\{ -\frac{r\sqrt{5}}{h} \right\}$$

- **Noise:** **Moving Average** process of order *q*: $\omega_t = \varepsilon_t + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}$ where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ \sigma^2)$

Presidential Approval Ratings

The Model

- **Model:** $y_t = f(t) + \omega_t$
  - → y_t = approval rating on day t
  - → Signal function $f(\cdot)$ is to be estimated



- **Prior:** $f \sim \mathcal{GP}(0, \tau^2 C(r))$ where $C(r)$ denotes the **Matérn-5/2 Kernel**

$$C(r) = \left(1 + \frac{r\sqrt{5}}{h} + \frac{5r^2}{3h^2} \right) \exp\left\{ -\frac{r\sqrt{5}}{h} \right\}$$

- **Noise:** **Moving Average** process of order q: $\omega_t = \varepsilon_t + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}$ where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

Presidential Approval Ratings

The Problem

- Note that this model contains three parameters (often called **hyperparameters** in the Bayesian framework, since they are parameters not of particular interest to us in themselves beyond their link to the signal function).
- We adopt an Empirical Bayes framework, opting to estimate these parameters using *data* (as opposed to imposing further *hyperpriors*).
- A key proposition we make is to convert the estimation problem into the **spectral domain**.
  - → This is related to the notion of a **Fourier Transform**

Spectral Domain

- Recall that a vector space can be described using a set of **basis vectors**.
  - → These basis vectors allow any element of the vector space to be expressed as a series of **coordinates**, with respect to the selected basis vectors.
- A **function space** is essentially an extension of a vector space, where elements are now *functions*.
  - → Function spaces can be expressed in terms of **basis functions**, which then allow us to describe functions in term of coordinates wrt. the selected basis *functions*.
- It turns out that most functions can be expressed as some (infinite) combinations of sines and cosines of different periods.
- So, we can consider using a collection of sinusoids to form a basis, with coordinates expressed with respect to these sines and cosines.

Fourier Transforms

Intuition

- This is, in essence, what the **Fourier Transform** represents: a decomposition of a function into a superposition of sines and cosines of differing periods.
- For example, the **Sawtooth Wave**

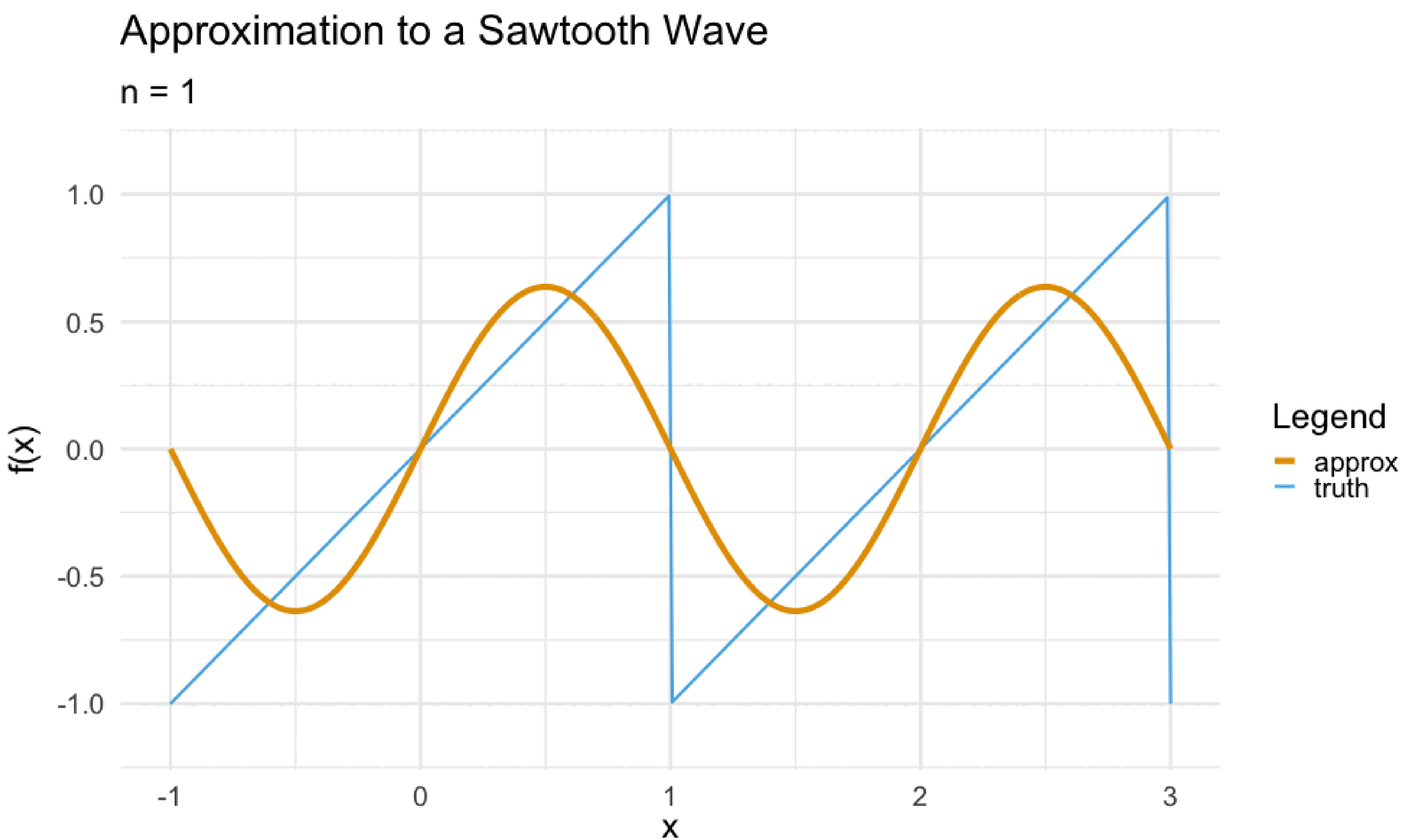
$$f(x) = x - 2 \cdot \mathbf{round}(x/2)$$

admits the following Fourier decomposition:

$$f(x) = \sum_{n=1}^{\infty} \frac{2}{n\pi} (-1)^{n+1} \sin(n\pi x)$$

Fourier Transforms

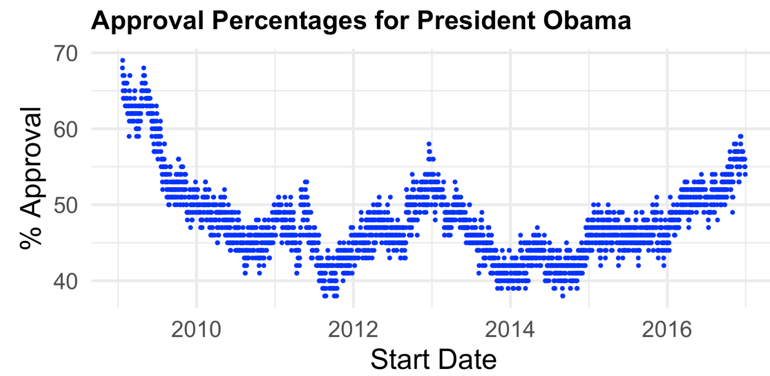
Example: Sawtooth Wave



Presidential Approval Ratings

The Model

- **Model:** $y_t = f(t) + \omega_t$
  - → $y_t =$ approval rating on day *t*
  - → Signal function $f(\cdot)$ is to be estimated



- Let $\{\eta_k\}_{k \geq 0}$ denote the Fourier coefficients of the signal function and let $\{\xi_k\}_{k \geq 0}$ denote the Fourier coefficients of the noise process. By standard results, these will be a collection of independent *complex*-normal random variables:

$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ v_k \mathbf{I}_2\right)$$
$$\xi_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

Presidential Approval Ratings

The Model

- **Model:** $y_t = f(t) + \omega_t$
  - → $y_t =$ approval rating on day t
  - → Signal function $f(\cdot)$ is to be estimated

$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ v_k \mathbf{I}_2\right)$$
$$\xi_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

- Given appropriate estimators $\hat{\eta}_k$ for η_k, we can construct an estimator for the signal function as

$$\widehat{f}(t) = \sum_{k=-\infty}^{\infty} \hat{\eta}_k e^{-2\pi i k t / n}$$

- Therefore, we're pretty much back to a normal-normal problem!

The Model

Time- and Frequency-Domain Specifications

- **Time-Domain Model:**

  $y_t = f(t) + \omega_t$

  → $f \sim \mathcal{GP}(0,\ \sigma^2 C(\cdot))$

  → $\omega_t = \varepsilon_t + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j}$ where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ \sigma_0^2)$

- **Frequency-Domain Model:**

$$(\eta_k + \xi_k) \mid \eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(\eta_k,\ v_k \mathbf{I}_2\right)$$
$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

- The "best" Bayesian point estimator is the posterior mean:

$$\mathbb{E}[\eta_k \mid \eta_k + \xi_k] = \frac{v_k}{v_k + u_k}(\eta_k + \xi_k)$$

- Standard results tell us that the variances v_k and u_k are related to the **spectral densities** of the signal and noise processes, respectively.

The Model

Frequency-Domain

$$(\eta_k + \xi_k) \mid \eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(\eta_k,\ v_k \mathbf{I}_2\right) \qquad \mathbb{E}[\eta_k \mid \eta_k + \xi_k] = \frac{v_k}{v_k + u_k}(\eta_k + \xi_k)$$
$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

- For a Matérn-5/2 kernel, $\boldsymbol{v_k \propto \tau^2 h\left(1 + k^2 h^2\right)^{-3}}$
- For an MA(q) noise process, $\boldsymbol{u_k = \frac{\sigma^2}{n}\left[b_0 + \sum_{j=1}^{q} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]}$

Problems:

- η_k and ξ_k are **unknown** (they are like "population Fourier transforms")
- $\boldsymbol{v_k}$ and $\boldsymbol{u_k}$ depend on the unknown hyperparameters $\boldsymbol{\tau^2}$ (signal variance), $\boldsymbol{h}$ (signal bandwidth), and $\boldsymbol{\sigma^2}$ (noise variance).

The Model

Frequency-Domain

$$(\eta_k + \xi_k) \mid \eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(\eta_k,\ v_k \mathbf{I}_2\right) \qquad \mathbb{E}[\eta_k \mid \eta_k + \xi_k] = \frac{v_k}{v_k + u_k}(\eta_k + \xi_k)$$

$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

- For a Matérn-5/2 kernel, $v_k \propto \tau^2 h\left(1 + k^2 h^2\right)^{-3}$
- For an MA(q) noise process, $u_k = \frac{\sigma^2}{n}\left[b_0 + \sum_{j=1}^{q} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$

Resolution: Unknown η_k and ξ_k

- Since η_k and ξ_k appear in $\hat{\eta}_k$ only through their *sum*, we can replace their sum with $\check{y}_k$, the **Discrete Fourier Transform** (DFT) of the observed y_t values.

The Model

Frequency-Domain

$$(\eta_k + \xi_k) \mid \eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(\eta_k,\ v_k \mathbf{I}_2\right)$$
$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right)$$

$$\hat{\eta}_k := \left(\frac{v_k}{v_k + u_k} \right) \check{y}_k =: w_k \cdot \check{y}_k$$

- For a Matérn-5/2 kernel, $v_k \propto \tau^2 h \left(1 + k^2 h^2\right)^{-3}$
- For an MA(q) noise process, $u_k = \frac{\sigma^2}{n} \left[b_0 + \sum_{j=1}^{q} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right) \right]$

- We still have the “ultimate” problem of estimating the three hyperparameters τ^2, h, and σ^2.
  → This is a highly *nonlinear* estimation problem, making it somewhat challenging.

The Model

Estimating the Hyperparameters

- Recall that η_k and ξ_k are *complex* random variables, meaning they each have two components (real and imaginary).

$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right) \qquad v_k \propto \tau^2 h\left(1 + k^2 h^2\right)^{-3}$$

$$\xi_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ v_k \mathbf{I}_2\right) \qquad u_k = \tfrac{\sigma^2}{n}\left[b_0 + \textstyle\sum_{j=1}^{q} b_j \cos\left(j \cdot \tfrac{2\pi k}{n}\right)\right]$$

$$\implies (\eta_k + \xi_k) \sim \mathcal{CN}\Big(0,\ (u_k + v_k)\mathbf{I}_2\Big)$$

$$\implies \frac{\|\eta_k + \xi_k\|^2}{2} \sim \text{Exponential}\left(u_k + v_k\right)$$

The Model

Estimating the Hyperparameters

- Recall that η_k and ξ_k are *complex* random variables, meaning they each have two components (real and imaginary).

$$\eta_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ u_k \mathbf{I}_2\right) \qquad v_k \propto \tau^2 h\left(1 + k^2 h^2\right)^{-3}$$

$$\xi_k \overset{\text{ind}}{\sim} \mathcal{CN}\left(0,\ v_k \mathbf{I}_2\right) \qquad u_k = \frac{\sigma^2}{n}\left[b_0 + \textstyle\sum_{j=1}^{q} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

$$\underbrace{\frac{\|\eta_k + \xi_k\|^2}{2}}_{:=\theta_k} \sim \text{Exponential}\left[\tau^2 h\left(1 + k^2 h^2\right)^{-3} + \frac{\sigma^2}{n} b_0 + \sum_{j=1}^{q} \frac{\sigma^2}{n} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

The Model

Estimating the Hyperparameters

$$\underbrace{\frac{\|\eta_k + \xi_k\|^2}{2}}_{:=\theta_k} \sim \mathrm{Exponential}\left[\tau^2 h\left(1 + k^2 h^2\right)^{-3} + \frac{\sigma^2}{n} b_0 + \sum_{j=1}^{q} \frac{\sigma^2}{n} b_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

- Again, η_k and ξ_k appear only through their sum, which can be replaced with (estimated by) the DFT of the y_k values, $\check{y}_k$!
- **Proposition:** Fit a **Gamma Generalized Linear Model** (GLM) of the form

$$\frac{\|\check{y}_k\|^2}{2} \sim \mathrm{Exponential}\left[\delta_1\left(1 + k^2 h^2\right)^{-3} + \beta_0 + \sum_{j=1}^{q} \beta_k \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

The Model

Estimating the Hyperparameters

- **Proposition:** Fit a **Gamma Generalized Linear Model** (GLM) of the form

$$\frac{\|\check{y}_k\|^2}{2} \sim \mathrm{Exponential}\left[\delta_1\left(1 + k^2 h^2\right)^{-3} + \beta_0 + \sum_{j=1}^{q} \beta_k \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

- This is still nonlinear in the bandwidth; as such, we propose fitting the GLM for a fixed bandwidth and running a grid search over candidate bandwidth values to find the one that minimizes the model deviance.
  - → We are exploring more sophisticated techniques, like Newton-Raphson; research is ongoing (and perhaps one of you can help with this in the future!)

The Algorithm

1. Obtain the coefficients $\check{y}_k$ of the Discrete Fourier Transform of the y_t values, and use this to compute $\tilde{\theta}_k := (1/2)\|\check{y}_k\|^2$ for $k = 0, \cdots, B$ where $B = \lfloor (n-1)/2 \rfloor$ if n is odd and $B = n/2$ if n is even.
2. Identify the bandwidth $\widehat{h}$ that minimizes the deviance of the following Gamma Generalized Linear Model:

$$\tilde{\theta}_k \sim \mathbf{Exponential}\left[\delta_1 \left(1 + k^2 h^2\right)^{-3} + \beta_0 + \sum_{j=1}^{q} \beta_j \cos\left(j \cdot \frac{2\pi k}{n}\right)\right]$$

3. Utilizing the optimal bandwidth from step (2) above, refit the Gamma GLM and obtain fitted values $\hat{\theta}_k$, $k = 1, \cdots, B$ for θ_k. Estimates for v_k are computed as

$$\hat{v}_k := \hat{\theta}_k - \sum_{j=0}^{q} \widehat{\beta_j} \cos(2\pi jk/n)$$

The Algorithm (cont'd)

4. Multiply the $\check{\boldsymbol{y}}_k$ terms from step 1 by weights

$$\widehat{w}_k := \left(\frac{\hat{v}_k}{\hat{\theta}_k} \right)$$

taking care to not downweight the zeroth frequency $\check{\boldsymbol{y}}_{\boldsymbol{0}}$.

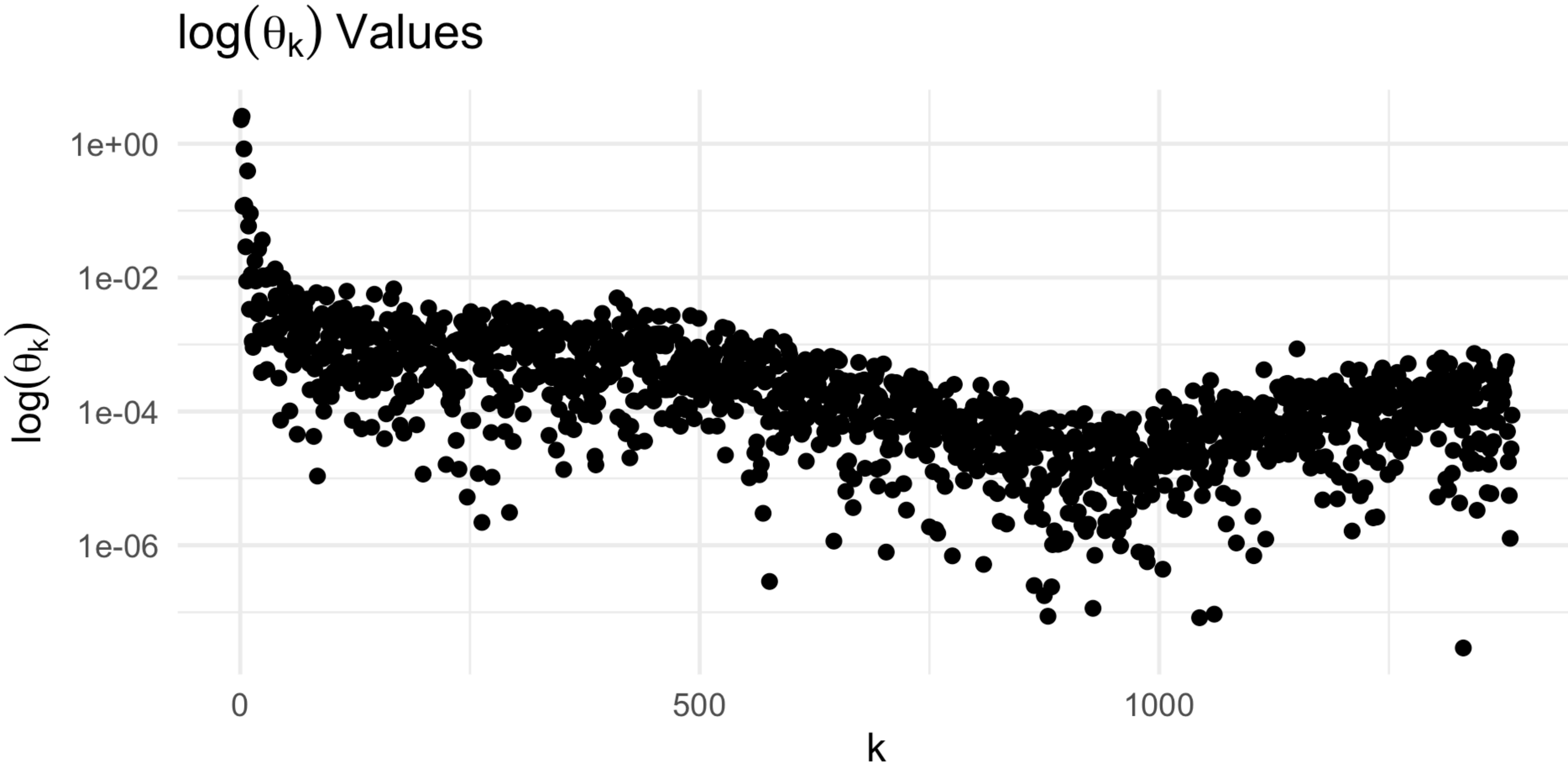
5. Apply the inverse Fourier transform to the weighted coefficients from step (4) above to obtain fitted values $\widehat{\boldsymbol{y}}$.

- Let us explore the application of this algorithm to the Presidential Approval Ratings dataset from the beginning of this presentation.

Presidential Approval Ratings

Step 1

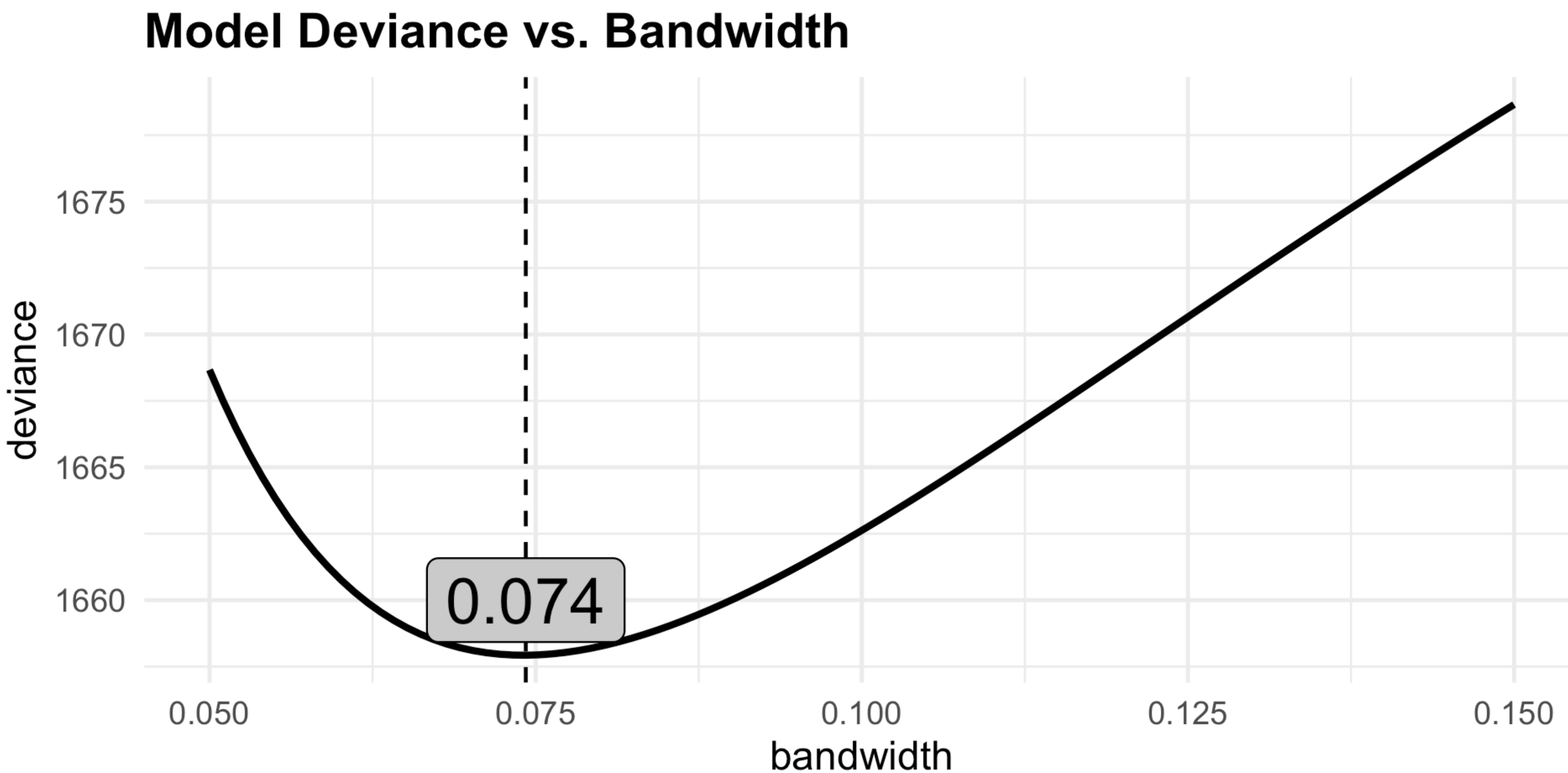
▶ Code



Presidential Approval Ratings

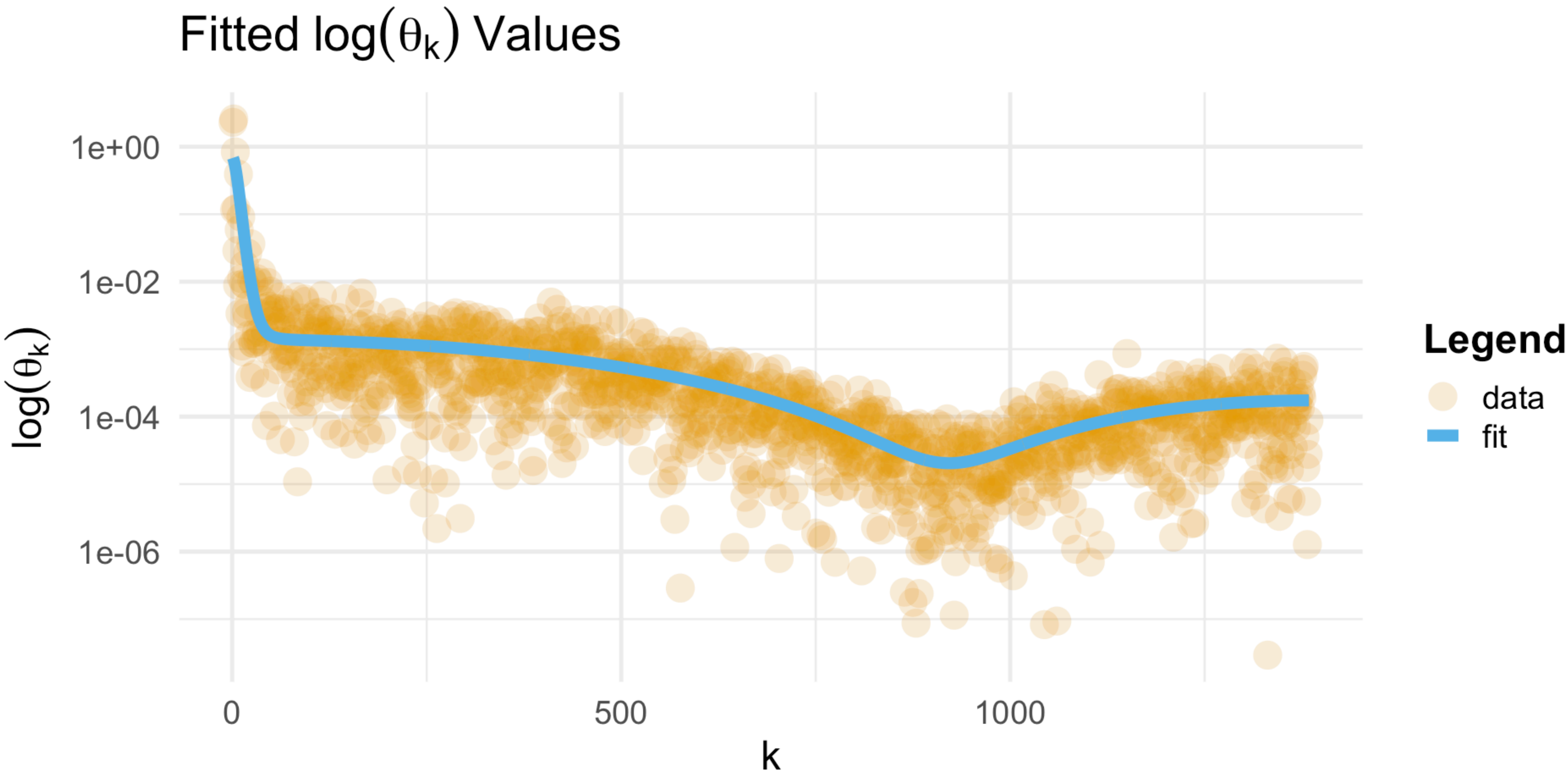
Step 2

▶ Code



Presidential Approval Ratings

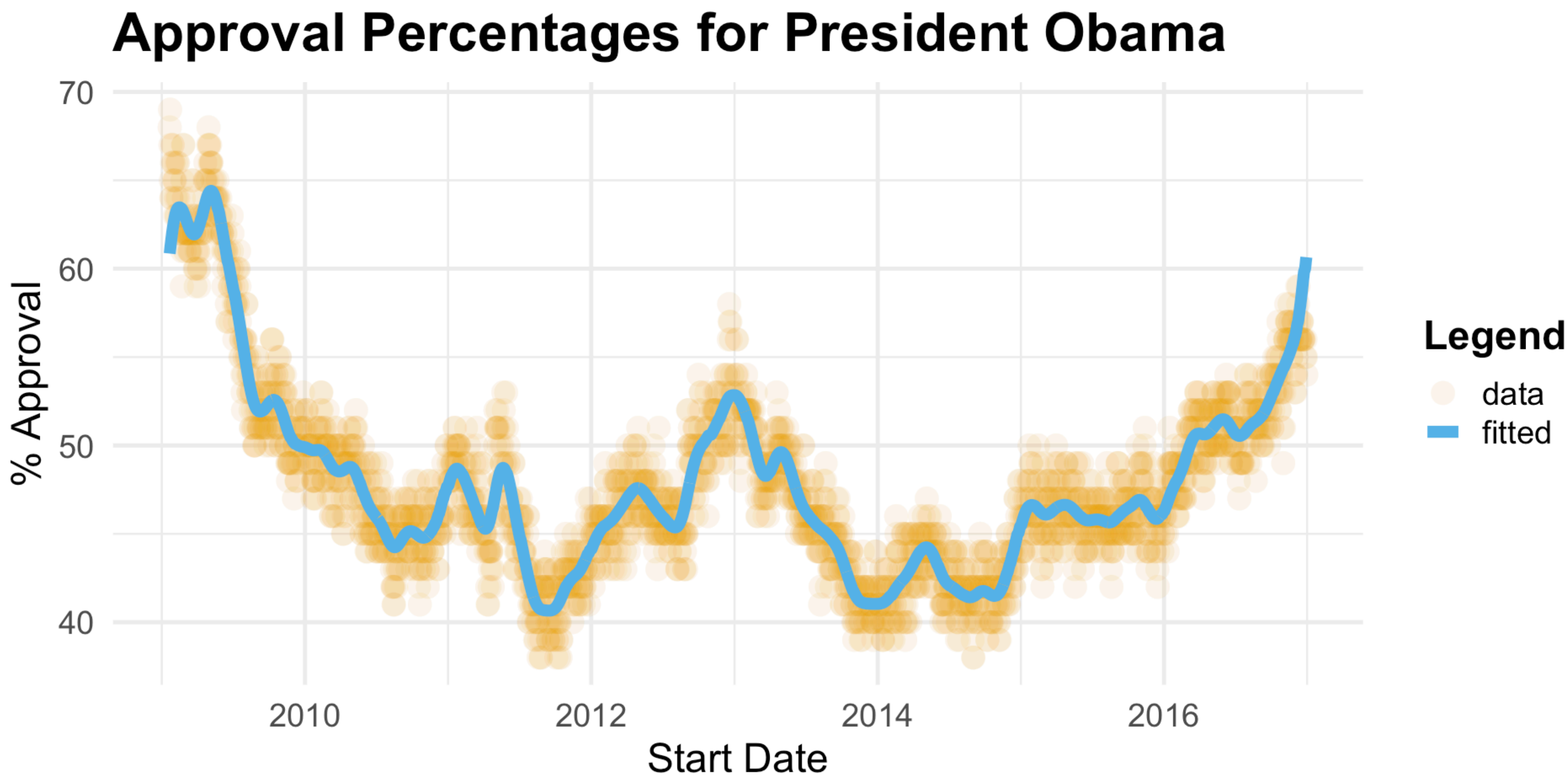
Step 3



Presidential Approval Ratings

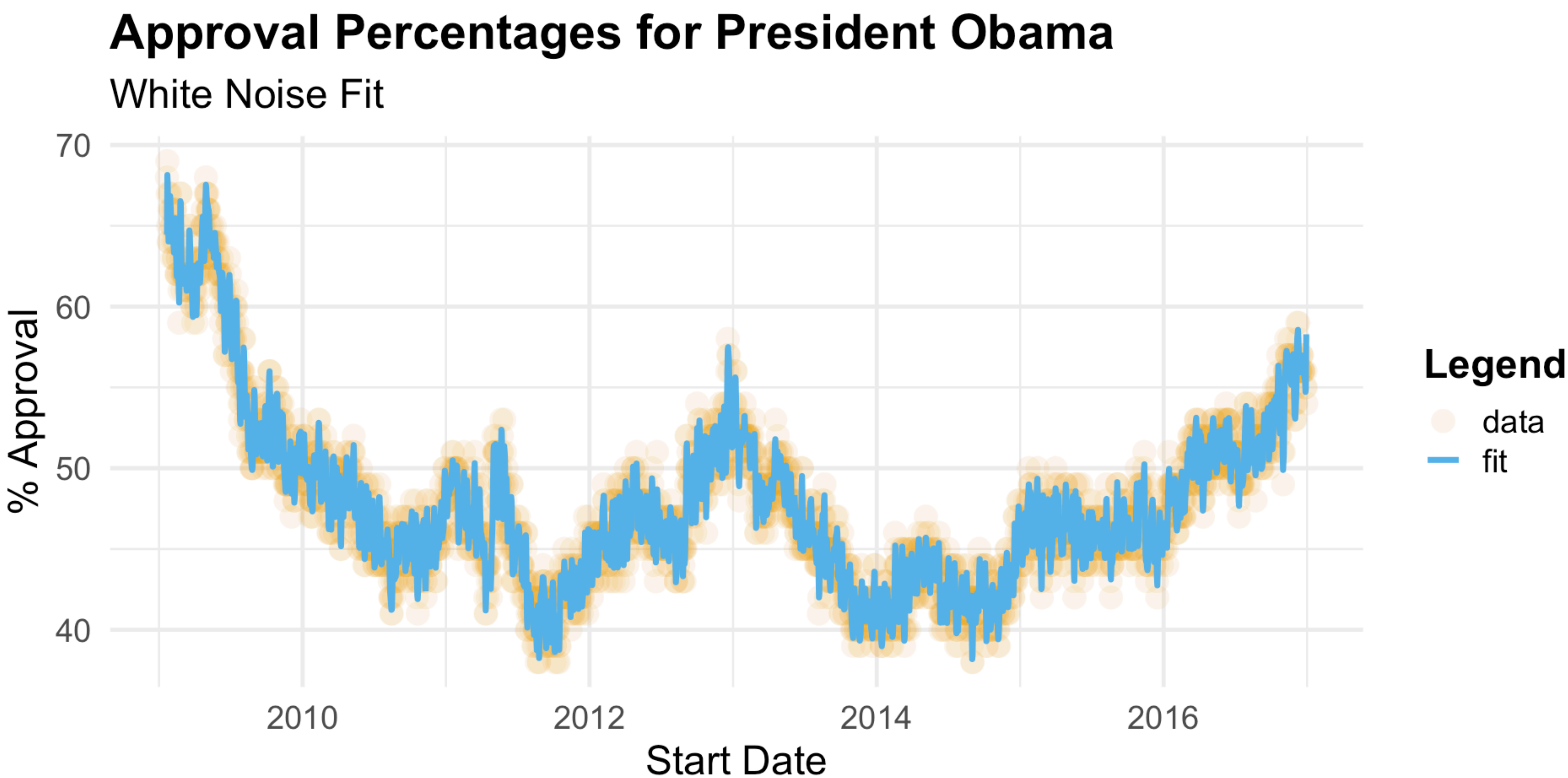
Steps 4 and 5

▶ Code



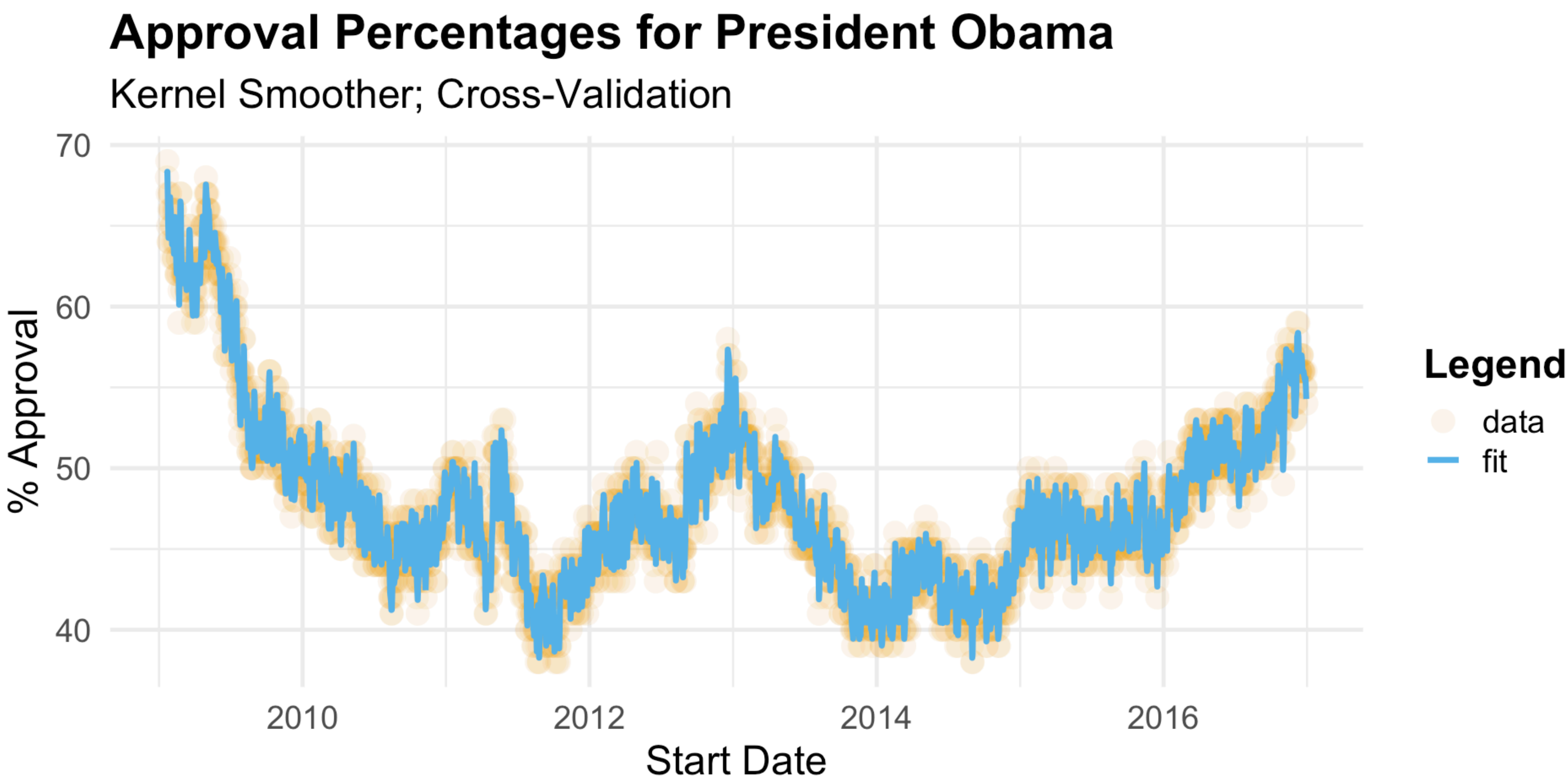
Presidential Approval Ratings

White Noise Fit



Presidential Approval Ratings

Kernel Smoothing Fit



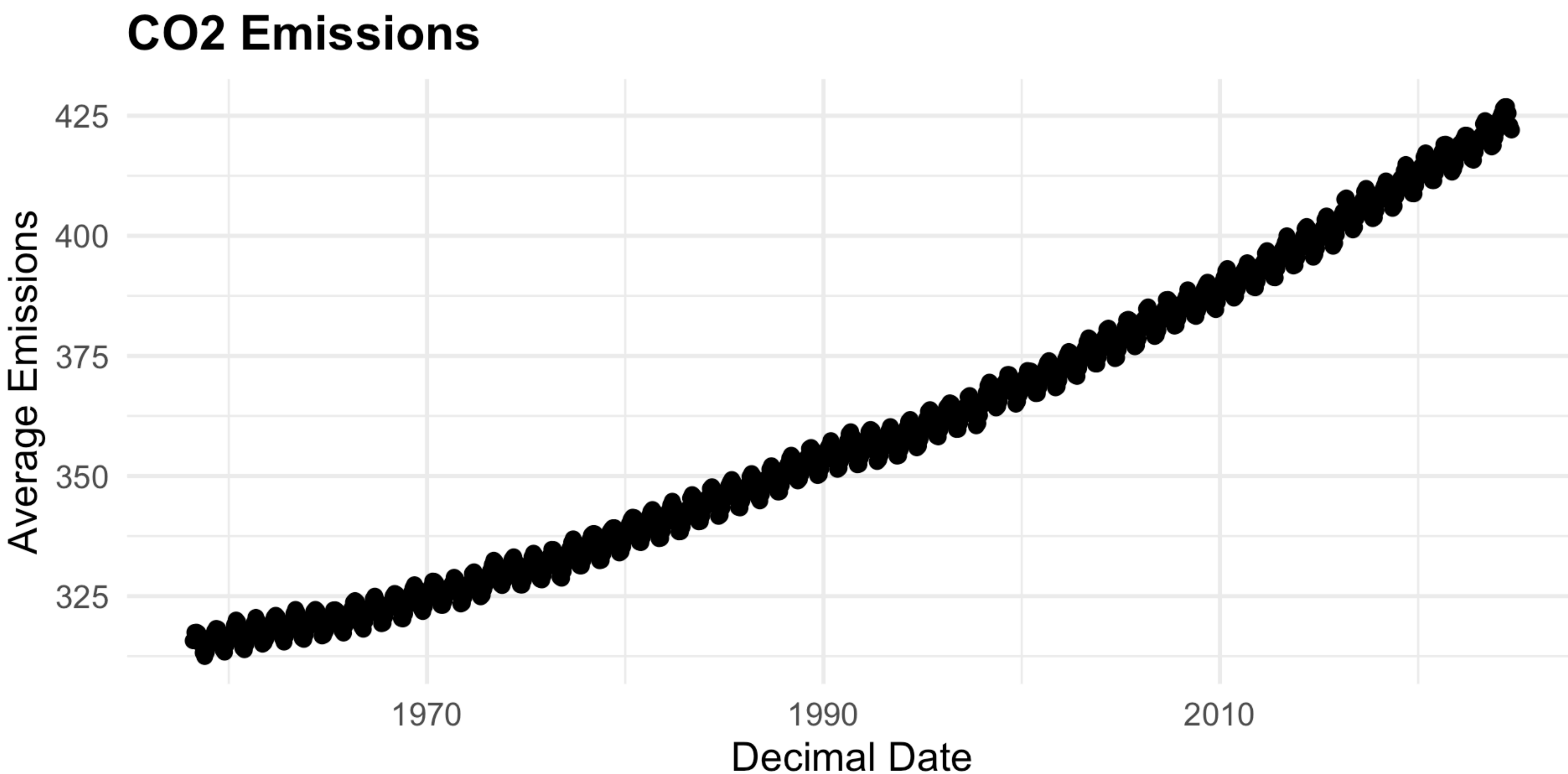
Trends

An Overview

- Fourier-based methods assume an inherent amount of *periodicity* in the signal function.
  - → As such, most Fourier-based estimators suffer from "edge effects" where the starting and ending fitted values are constrained to be equal, even if this is not desired.
- Our algorithm, admittedly, is no exception, which poses an additional challenge for datasets with a noticeable trend.
  - → As a somewhat "famous" example, we consider the CO_2 Emissions recorded at the Mauna Loa observatory in Hawai'i.

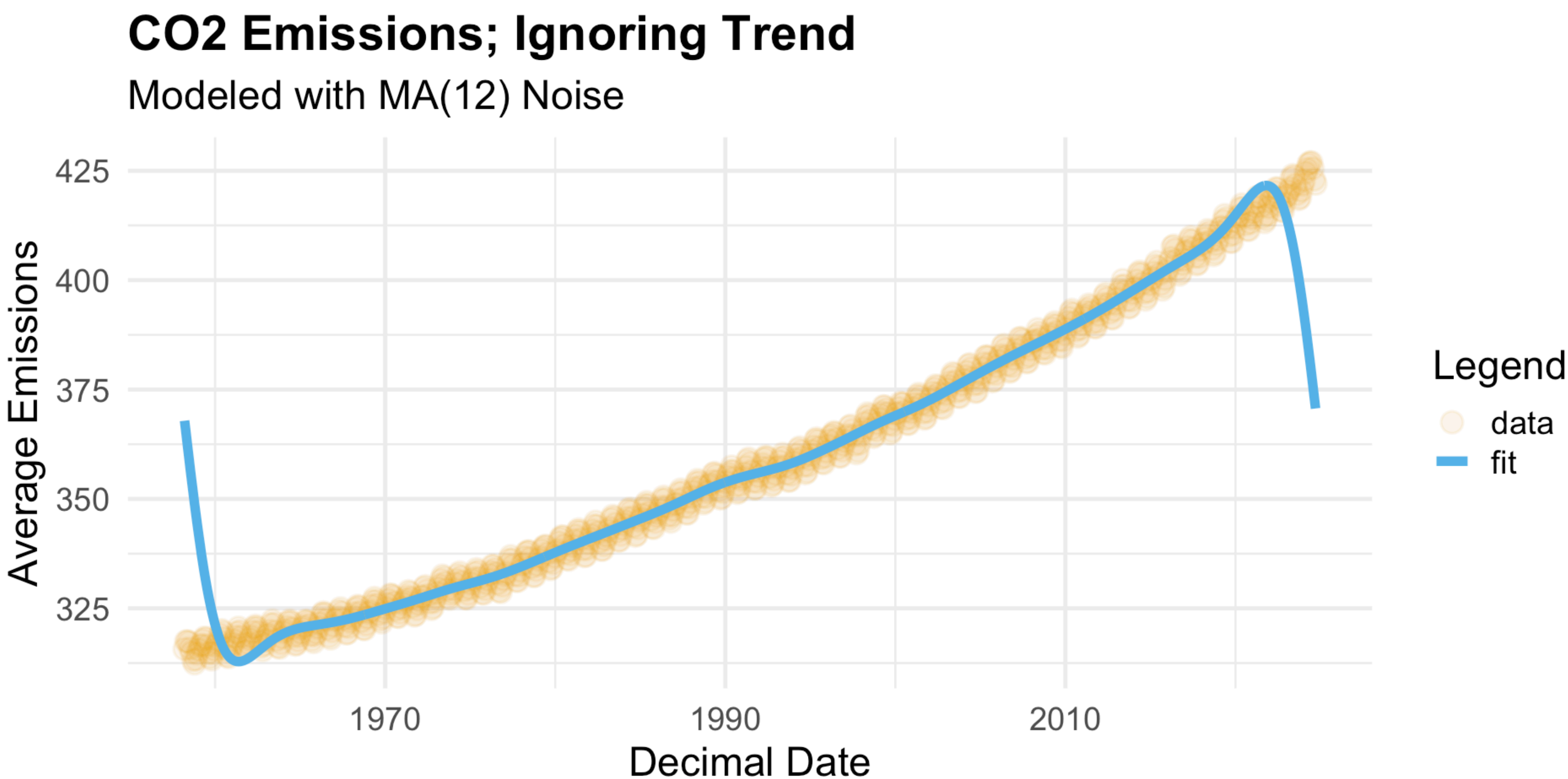
CO_2 Dataset

Visualization



CO_2 Dataset

Initial Fit



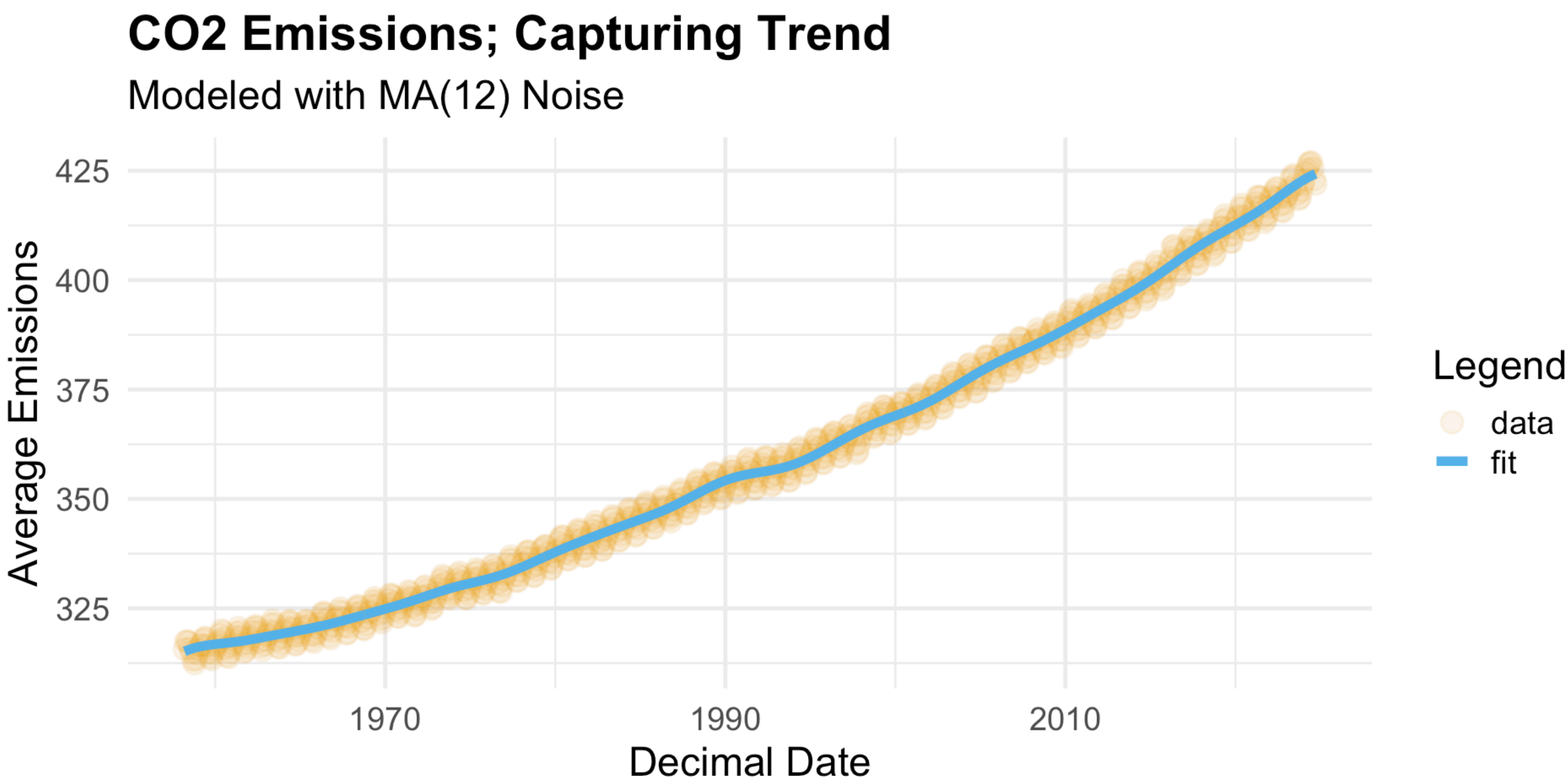
Trends

Our Approach

- Our current approach is to first remove the linear portion of the trend by way of a **simple linear regression**, and then apply our algorithm to the resulting residuals.
- Admittedly, this induces some heteroskedasticity which must be accounted for.
    - → Preliminary results indicate that the magnitude of this heteroskedasticity is relatively small, though, and can be ignored.
    - → A future direction of research that is of interest to us is finding a more systematic way to handle this heteroskedasticity.

CO_2 Dataset

Optimal Fit



Credible Intervals

A Work In Progress

- Finally, as a statistician, I would be remiss to not at least *mention* credible intervals.
- We are currently in the process of finalizing our treatment of credible intervals for our predicted signal values $f(t)$.
  - → We anticipate our treatment to be relatively sparse, with the intention of pursuing a more rigorous framework in future work.
- A preprint (containing further details of our algorithm and also our work on credible intervals) is expected to be completed in the coming months.

Conclusion

Summary

- In conclusion, we propose an Empirical Bayes estimation procedure for use in models with correlated data, with a Gaussian Process prior.
- A key aspect of our estimation procedure is the utilization of the spectral domain to carry out the estimation of necessary hyperparameters.
  - → This allows the hyperparameter estimation to be carried out by way of an Exponential Generalized Linear Regression, which is fast and efficient.
  - → Final fitted values are found by weighting the original spectrum, and inverting the Fourier transform.
- In cases where a noticeable trend is present, our algorithm can be applied to the residuals resulting from a Simple Linear Regression.

Conclusion

Future Work

- Three main avenues of future work immediately present themselves.
- From a computational standpoint, we would like to adopt a slightly more sophisticated procedure for estimating the bandwidth (beyond a simple grid-search).
- We would also like to explore the notion of credible intervals as they pertain to our final fitted values; work on this is ongoing.
  - → A quick note: we are also planning on exploring the asymptotics of our algorithm in the coming months.
- I would also like to explore whether it is feasible to extend our algorithm to *bivariate* data (though it is unclear whether this is feasible or not).

Acknowledgments

- I am deeply grateful for the tutelage of my advisor, Dr. Andrew Carter, and the other members of my doctoral committee (Dr. Yuedong Wang and Dr. Mengyang Gu)
- The Presidential Approval Ratings dataset has been accessed through the American Presidency Project at the University of California, Santa Barbara
- The CO_2 dataset has been accessed through the Global Monitoring Laboratory, a part of the National Oceanic and Atmospheric Administration
- This slide deck has been created using Quarto, in the RStudio IDE. Graphics have been produced with the `ggplot2` package, with colorblind-friendly Okabe-Ito Palette functionality added through the `ggokabeito` package.
  - → Slide icons have been sourced from `fontawesome`, and have been integrated into this slide deck using the `fontawesome` Quarto Extension.
  - → The base theme for this slide deck is the Quarto Clean Theme