Modelling Latent Attitudes From Ordinal Likert Data In SoSeC

Clemens Baldzuhn

August 14, 2025

Overview

Introduction

Introduction

Most of SoSeC data are of ordinal type organized in likert scales (e.g. trust in institutions).

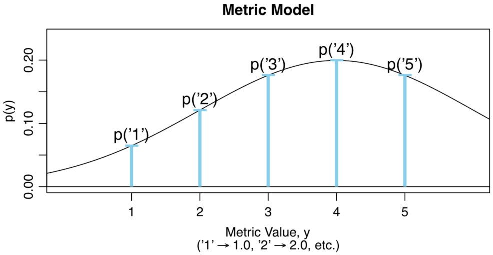
Likert scales work based on the idea that each respondent has a latent (metric) attitude that is expressed discretely within the amount of agreement/disagreement in the question block (scale).

Treating the response data as if it was metric is common practice but inaccurate and can lead to biased statistical analyses (Liddell & Kruschke, 2018):

- distorted effect size estimates (e.g. in OLS)
- greatly inflated false alarm rates and low correct detection rates (e.g. in group comparisons like t-test, ANOVA)

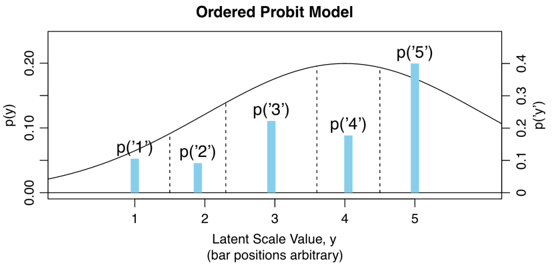
I'm providing a new* and robust GP-IRT model, which can be used to circumvent these biases and make your analyses publication-ready.

Biased approach



- **Aim**: Model parameters (e.g. μ) of the latent distribution.
- Probabilites for observed responses (y) in the metric model correspond to the probability density of the latent attitude at the corresponding value.
- Modeling the latent distribution like this is inaccurate since the distances between the response levels are not equally spaced in likert-type items.
- Estimated parameters are biased.

Unbiased approach



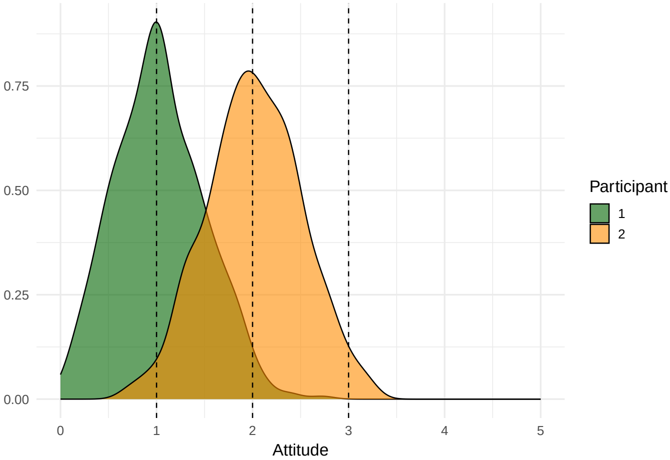
- Probabilities for the observed responses in the correct model correspond to the probability mass within a subinterval of the latent distribution.
- Estimation of this model now also encompasses thresholds for the subintervals (dashed lines).

Simulation Explainer: Why Modelling Thresholds is Indispensable

Scenario 1

Scenario 1

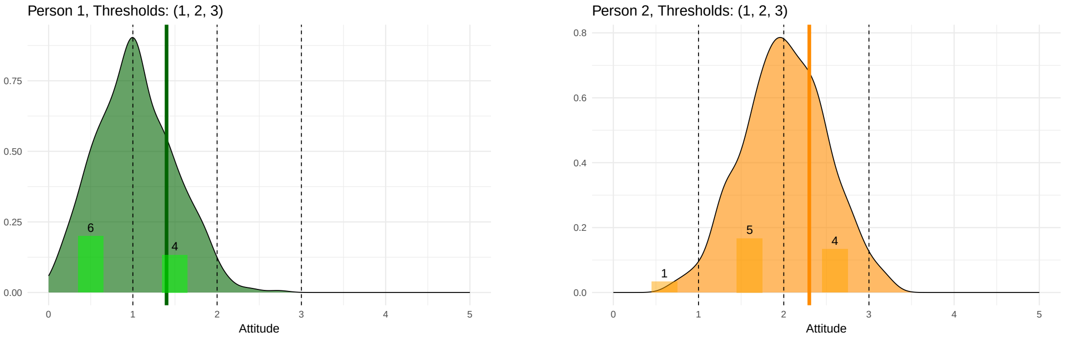
- Two participants with two latent attitudes on some subject matter: $\mu_1 = 0, \mu_2 = 1$.
- Item specific **threshold set 1**, $\tau = \{-3.5, -1.5, 0.5\}$.



Scenario 1

Participant 1&2 respond to same likert item/same set of items measuring same concept 10 times and we observe the responses (bars).

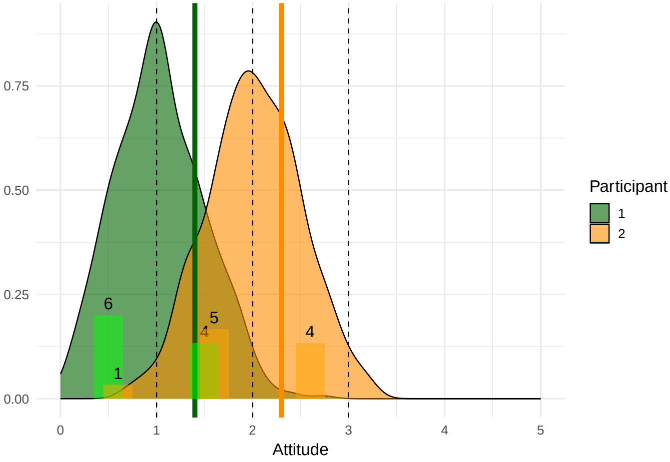
The are 4 response categories and – e.g. – participant 1 answered 8 times in category 3 and twice in category 4.



Scenario 1

ML estimators for μ work fine:

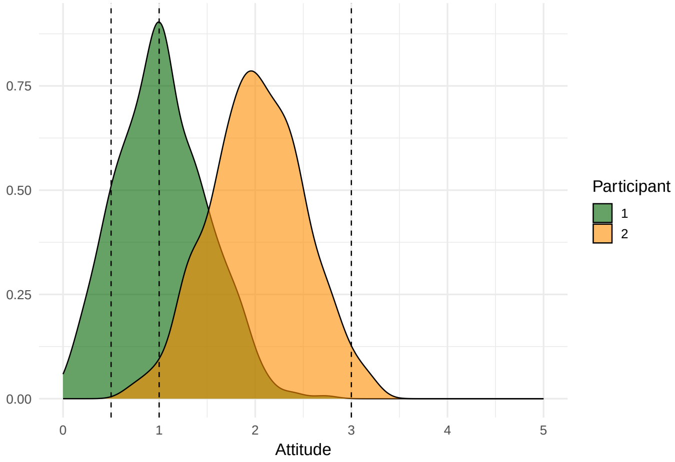
1. μ is estimated correctly.
2. Estimators correctly reflect the ordering between participants.



Simulation explainer: Scenario 2

Scenario 2

- Same latent attitudes as in scenario 1.
- Item specific **threshold set 2**, $\tau = \{0, 2, 4\}$

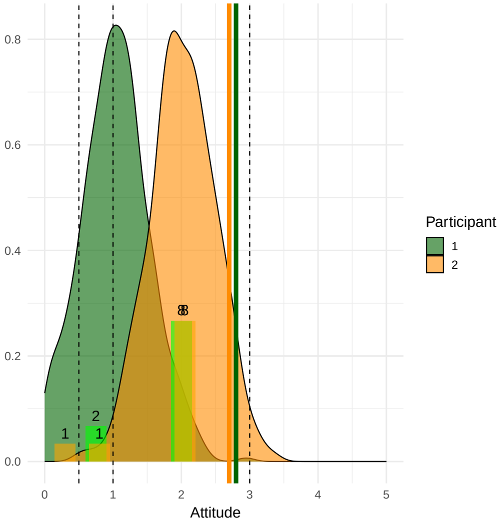


Scenario 2

ML estimators for μ biased:

1. μ for participant 2 *biased*.
2. Ordering between participants *exchanged*.

Thresholds and latent distribution have to be estimated jointly!



Formal Derivation

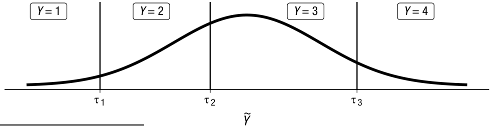
Model derivation I

Consider a latent continuous variable $\tilde{Y}$ and the K ordered categories within which it can realize.

The probability that the observed response Y lies within a specific subinterval k can be described in terms of the latent thresholds τ_k and the distribution function $F_{\tilde{Y}}$ of $\tilde{Y}$.

$$P(Y = k) = P(\tau_{k-1} < \tilde{Y} \leq \tau_k) = F_{\tilde{Y}}(\tau_k) - F_{\tilde{Y}}(\tau_{k-1}) \quad (1)$$

$$P(Y \leq k) = P(\tilde{Y} \leq \tau_k) = F_{\tilde{Y}}(\tau_k) \quad (2)$$



[1]

[1](Bürkner & Vuorre, 2019)

Model derivation II

The goal is to predict (parameters of) the latent distribution $\tilde{Y}_i$ with a linear predictor η_i. ϵ describes the random measurement error and follows some distribution F.

$$\tilde{Y}_i = \eta_i + \epsilon_i$$

In this regression formulation, the probability to observe some response $\leq k$ hence becomes:

$$\begin{aligned} P(Y_i \leq k|\eta_i) &\overset{(2)}{=} P(\tilde{Y}_i \leq \tau_k|\eta_i) \\ &= P(\eta_i + \epsilon_i \leq \tau_k) \\ &= P(\epsilon_i \leq \tau_k - \eta_i) \\ &= F(\tau_k - \eta_i) \end{aligned}$$

or equivalently

$$P(Y_i = k|\eta_i) = F(\tau_k - \eta_i) - F(\tau_{k-1} - \eta_i)$$

Model derivation III

Choosing F to be the logistic distribution function leads to:

$$
\begin{aligned}
P(Y_i \leq k | \eta_i) &= \text{logit}^{-1}(\tau_k - \eta_i) \\
&= \frac{\exp(\tau_k - \eta_i)}{1 + \exp(\tau_k - \eta_i)} \\
\iff \ln\left(\frac{P(Y_i \leq k|\eta_i)}{1 - P(Y_i \leq k|\eta_i)}\right) &= \tau_k - \eta_i \\
\ln\left(\frac{P(Y_i \leq k|\eta_i)}{P(Y_i > k|\eta_i)}\right) &= \tau_k - \eta_i
\end{aligned}
$$

This cumulative logit/graded response model estimates the odds of observing a response $\leq k$ versus observing a response $> k$.

Model derivation IV

In the spirit of IRT models:

1. Respondents $i = 1, ...n$ have unique attitudes θ_i.
2. Items $j = 1, ..., p$ to have varying discrimination parameters α_j.
3. Thresholds are item-specific, thereby indicating item difficulty.

- θ_i can be interpreted as the central tendency of the continuous latent attitude. Higher values of θ_i imply higher success probabilities regardless of the administered item.
- α_j determines how strongly the latent attitude θ_i influences the probability of selecting a higher category.

The linear predictor hence becomes:

$$\eta_i = \alpha_j \theta_i$$

Estimation

Parameters are estimated in Stan using a Bayesian approach. Note that in this first formulation the model is estimated for a single wave only ($t = 1$).

Full model for a single wave:

$$\ln\left(\frac{P(Y_{ij} \leq k \mid \theta_i, \alpha_j)}{P(Y_{ij} > k \mid \theta_i, \alpha_j)}\right) = \tau_{kj} - \alpha_j \theta_i \quad \text{for} \begin{cases} k = 1, \dots, K-1 \\ i = 1, \dots, n \\ j = 1, \dots, p \end{cases}$$

$$\begin{aligned} \tau_{kj} &\sim \mathcal{N}(\mu_k, \sigma_k) \quad \text{s.t. } \tau_{1j} < \tau_{2j} < \cdots < \tau_{(K-1)j} \\ \mu_k &\sim \mathcal{N}(0, 5) \\ \sigma_k &\sim \text{Cauchy}^+(0, 5) \\ \alpha_j &\sim \mathcal{N}(1, 0.5) \\ \theta_i &\sim \mathcal{N}(0, 1) \end{aligned}$$

Incorporating the panel structure

Making individual attitudes time-dependent ($\theta_i \to \theta_i(t)$) permits incorporating the assumption that attitudes change with time (Chen et al., 2025).

Instead of estimating latent attitudes at different time periods independently we use a Gaussian process prior on θ.

1. Each latent trajectory is modelled jointly through time.
2. GP hyperparameters have useful interpretations.
3. We can use the model to impute latent attitudes robustly.

$$\theta_i(t) \sim \mathcal{GP}\left(0,\ K^{(i)}_{\mathrm{M}5/2}(t, t')\right)$$

GP Hyperpriors

Gaussian process kernel (Matern 5/2):

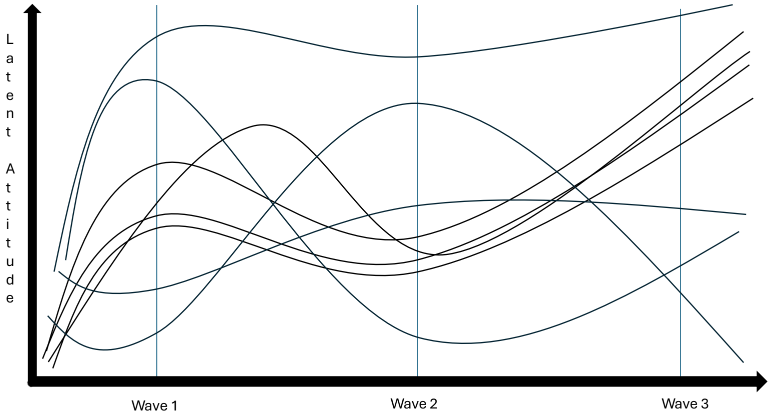
$$
\begin{aligned}
k^{(i)}_{\text{M52}}(t, t') &= \sigma^2_{f,i} \left(1 + \frac{\sqrt{5} r_{tt'}}{\ell_i} + \frac{5 r^2_{tt'}}{3\ell_i^2}\right) \exp\left(-\frac{\sqrt{5} r_{tt'}}{\ell_i}\right) \\
r_{tt'} &= |t - t'|
\end{aligned}
$$

GP hyperpriors (per participant):

$$
\begin{aligned}
\log \sigma_{f,i} &\sim \mathcal{N}(\mu_{\log \sigma_f}, \sigma_{\log \sigma_f}) \\
\log \ell_i &\sim \mathcal{N}(\mu_{\log \ell}, \sigma_{\log \ell}) \\
\mu_{\log \sigma_f} &\sim \mathcal{N}(0, 0.5) \\
\sigma_{\log \sigma_f} &\sim \text{Normal}^+(0.3, 0.2) \\
\mu_{\log \ell} &\sim \mathcal{N}(\log 5, 0.3) \\
\sigma_{\log \ell} &\sim \text{Normal}^+(0.2, 0.1)
\end{aligned}
$$

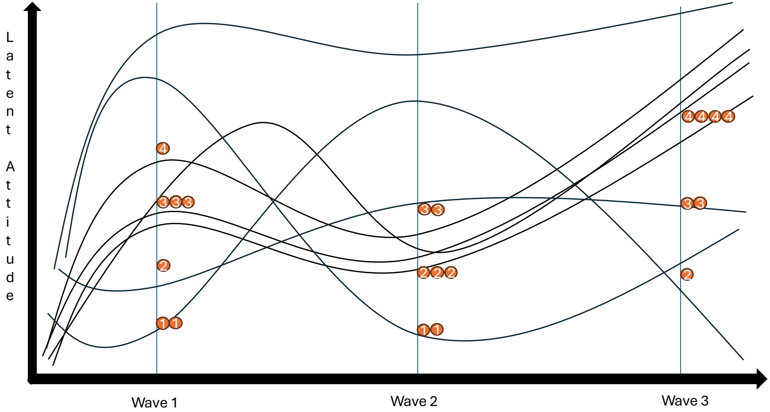
Gaussian process priors: Intuition

Intuitively, GP priors can be understood as a distribution over functions of the latent attitude.



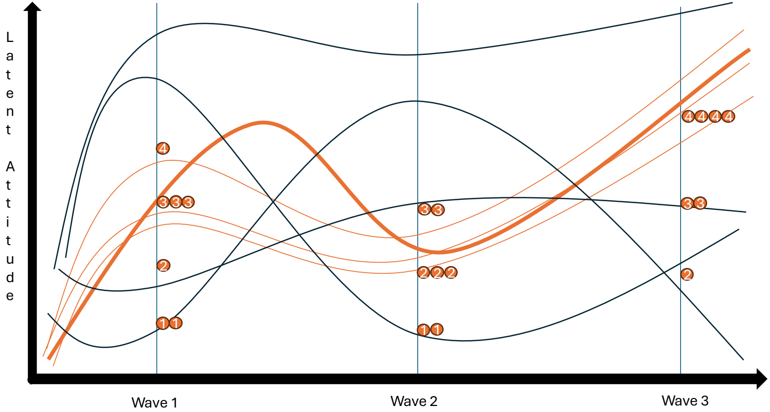
Gaussian process priors: Intuition

Taking them as priors in a Bayesian estimation setting means that observing data ...



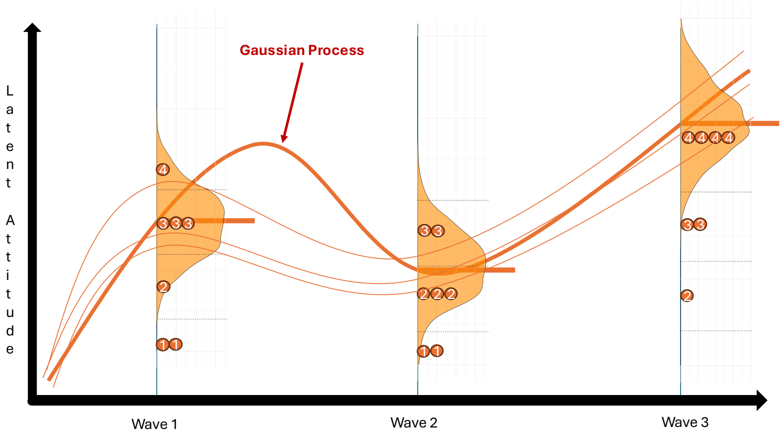
Gaussian process priors: Intuition

... makes some of these functions more likely.



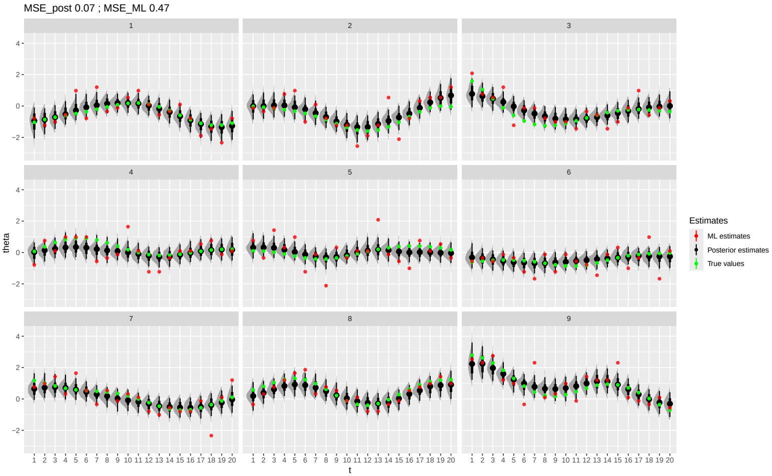
Gaussian process priors: Intuition

Marginally (i.e. fixing a specific timepoint), the GP is normally distributed.



Simulation Study

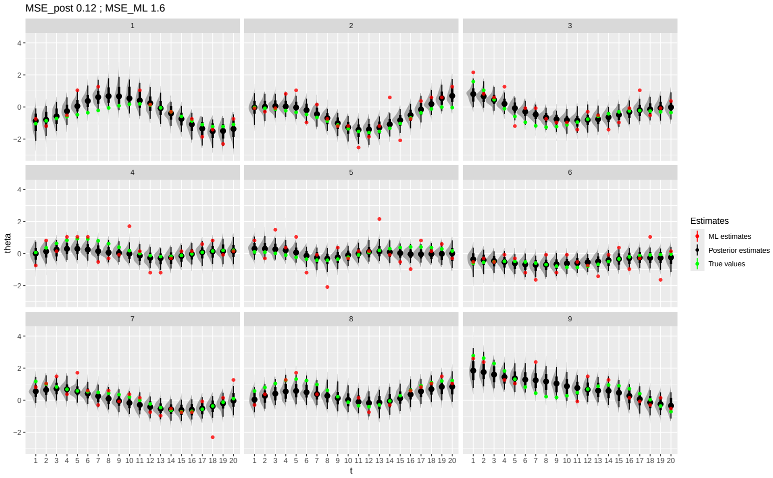
No missings



MSE_post 0.07 ; MSE_ML 0.47
theta
t
Estimates
ML estimates
Posterior estimates
True values

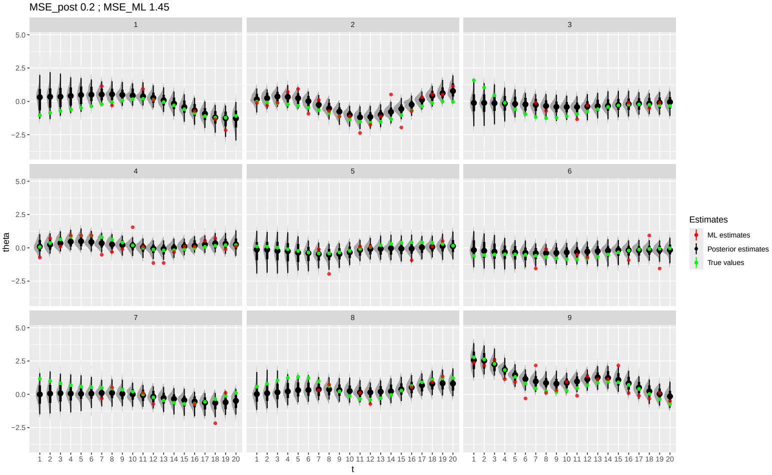
Random missings I

33% of participants (1,8,9) lack responses at 33% of waves (3,6,8,9,15).



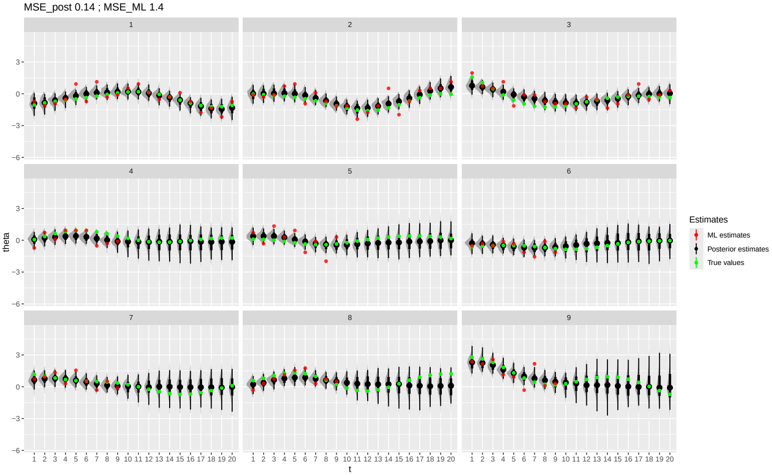
Random missings II

66% of participants lack responses at 66% of waves.



Sequential missings

4 sequential missings for 33% of participants.



Simulation study: Takeaways

- Model performance in terms of estimating latent attitudes is excellent in general.
- The model can be used for imputation: Missing values of several types get estimated robustly.
- In SoSeC, permitting one sequential missing more increases the number of respondents linearly by about 20%.

Application: Trust in Institutions

Trust in Institutions

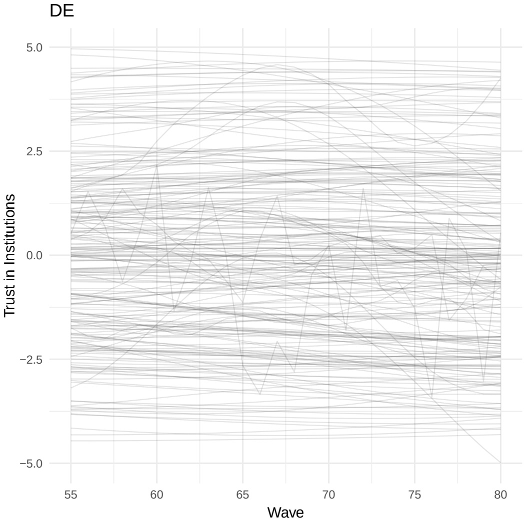
In accordance with Gestefeld and Lorenz (2023), 6 items were selected to conceptualize trust in institutions:

- Bundeswehr/Armed Forces (F5A2_1)
- Polizei/Police (F5A3_1)
- Parlament/Congress (F5A4_1)
- Regierung/Presidency (F5A5_1)
- Bundesverfassungsgericht/Supreme Court (F5A6_1)
- Verwaltung/Public Administration (F5A7_1)

The items have high internal consistency (Cronbach's $\alpha = 0.89$).

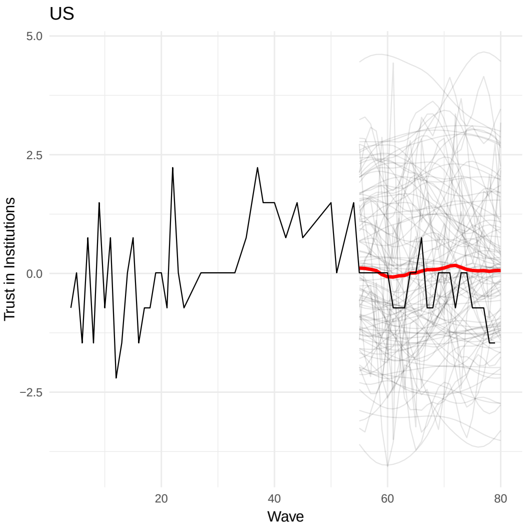
Results I

- Essentially, the model provides a smoothing procedure: Fitting the model gives smooth attitude-trajectories for all individuals.
- GP-curves come with two parameters: Lenght-scale and amplitude.



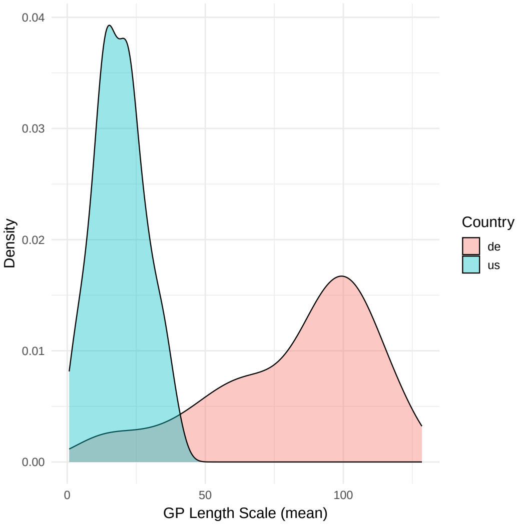
Results II: Words of caution

- Models with less NA's obviously perform better.
- Mean of the posterior estimates (red) is not a population estimate (black) but reflect prior behaviour!
  - Model most suitable for group comparisons.



Results III

- The length parameter is very different across the two panels.
- At least for waves 55-70 there is a tendency for US-Americans to have more strongly fluctuating attitudes than Germans.
- Length parameter is probably an interesting covariate for post hoc regression models .



Conclusion

- Estimating latent concepts from likert-scaled data can result in statistical biases.
- Most statistically sound way to do so is to estimate thresholds and latent distributions jointly.
- Beyond that, this GP-IRT model has the following advantages:
    1. Robust imputation procedure, which can be used to almost double the number of datapoints.
    2. Full posterior uncertainty quantification in the estimates through Bayesian estimation.
    3. Interesting parameter interpretations for length and amplitude.

References I

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101.

Chen, Y., Montgomery, J., & Garnett, R. (2025, April 3). *A Dynamic, Ordinal Gaussian Process Item Response Theoretic Model*. https://doi.org/10.48550/arXiv.2504.02643

Gestefeld, M., & Lorenz, J. (2023). Calibrating an Opinion Dynamics Model to Empirical Opinion Distributions and Transitions. *Journal of Artificial Societies and Social Simulation*, *26*(4), 9. https://doi.org/10.18564/jasss.5204

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong? *Journal of Experimental Social Psychology*, *79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009