

## Document Structure

- **Summary Tables of Outputs:** Consolidated tables listing all output files generated during the analysis.
  - **Summary Table of Key Model Posteriors:** A table summarizing the posterior densities for the key scaling parameters from each analysis.
  - **High-level Convergence Statistic Overview:** Plots summarizing standard MCMC convergence diagnostics.
  - **Trace Plots for All Analyses:** A list of all traceplots produced in the order they appear in the summary table of outputs.
  - **Analysis Workflow:** Detailed descriptions of data wrangling, modeling, and results validation.
    - **1. Data Wrangling:** Preparation of datasets used in the analyses.
    - **2. Model Fitting:** Bayesian modeling approaches and diagnostics.
    - **3. Results:** Extended tables and plots supporting the main manuscript.
    - **4. Supplemental Analyses:** Alternative model testing and additional robustness checks.
- 

## Summary Table of Outputs

### File Naming Conventions

Please refer to this table to understand the file naming conventions used in the R script and the subsequent tables containing the output file names.

File Name Convention	Description
post_summary_<model>.csv	Contains parameter summaries (e.g., mean, confidence intervals) for a model.
tplots_<model>.png	Trace plots of MCMC chains for the specified model.
geweke_<model>.csv	Geweke diagnostic results for the specified model.
grhat_<model>.csv	R-hat convergence diagnostic for the specified model.
resid_<model>.png	Residual diagnostic plots for the specified model.
waic_<model>.csv	WAIC model comparison metrics for the specified model.
loo_<model>.csv	Leave-One-Out cross-validation results for the specified model.
city_outliers_<model>.csv	Identifies cities considered outliers for the specified model.
scaling_posteriors.png	Posterior density plots for scaling parameters.
analysis_results_summary.csv	Summary of main results across all models.
lppd.csv <sup>*</sup>	Log Point-wise Predictive Density for all models and all observations.

**NOTE** In the tables that follow, we split repetitive filename prefixes in order to avoid excessive overflow from long filenames. The prefixes are in bold in the rows immediately preceding the relevant files that all share the given prefix.

### Main Analyses, Plot Files

Context	Script Section	Filename
Summary	Plot Scaling Param Posterior Densities	<b>scaling_...</b> posteriors.pdf
Summary	Plot Scaling Param Posterior Densities	posteriors.png
Model Diagnostic	First Analysis	<b>resid_...</b> allmonuments.pdf
Model Diagnostic	First Analysis	allmonuments.png
Model Diagnostic	Second Analysis	allwalls.pdf
Model Diagnostic	Second Analysis	allwalls.png
Model Diagnostic	Fourth Analysis	epigraphy.pdf
Model Diagnostic	Fourth Analysis	epigraphy.png
Model Diagnostic	Third Analysis	filtmonuments.pdf
Model Diagnostic	Third Analysis	filtmonuments.png
Model Diagnostic	Fifth Analysis	hnwi.pdf
Model Diagnostic	Fifth Analysis	hnwi.png
Model Diagnostic	Sixth Analysis	tallbuildings.pdf
Model Diagnostic	Sixth Analysis	tallbuildings.png
MCMC Diagnostic	First Analysis	<b>tplots_...</b> allmonuments.pdf
MCMC Diagnostic	First Analysis	allmonuments.png
MCMC Diagnostic	First Analysis	allmonuments_nozero.pdf
MCMC Diagnostic	First Analysis	allmonuments_nozero.png
MCMC Diagnostic	First Analysis	allmonuments_nozero_provinces.pdf
MCMC Diagnostic	First Analysis	allmonuments_nozero_provinces.png
MCMC Diagnostic	First Analysis	allmonuments_provinces.pdf
MCMC Diagnostic	First Analysis	allmonuments_provinces.png
MCMC Diagnostic	Second Analysis	allwalls.pdf
MCMC Diagnostic	Second Analysis	allwalls.png
MCMC Diagnostic	Second Analysis	allwalls_nozero.pdf
MCMC Diagnostic	Second Analysis	allwalls_nozero.png
MCMC Diagnostic	Second Analysis	allwalls_nozero_provinces.pdf
MCMC Diagnostic	Second Analysis	allwalls_nozero_provinces.png
MCMC Diagnostic	Second Analysis	allwalls_provinces.pdf
MCMC Diagnostic	Second Analysis	allwalls_provinces.png
MCMC Diagnostic	Fourth Analysis	epigraphy.pdf
MCMC Diagnostic	Fourth Analysis	epigraphy.png
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero.pdf
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero.png
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero_provinces.pdf
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero_provinces.png
MCMC Diagnostic	Fourth Analysis	epigraphy_provinces.pdf
MCMC Diagnostic	Fourth Analysis	epigraphy_provinces.png

Context	Script Section	Filename
MCMC Diagnostic	Third Analysis	filtmonuments.pdf
MCMC Diagnostic	Third Analysis	filtmonuments.png
MCMC Diagnostic	Third Analysis	filtmonuments_nozero.pdf
MCMC Diagnostic	Third Analysis	filtmonuments_nozero.png
MCMC Diagnostic	Third Analysis	filtmonuments_nozero_provinces.pdf
MCMC Diagnostic	Third Analysis	filtmonuments_nozero_provinces.png
MCMC Diagnostic	Third Analysis	filtmonuments_provinces.pdf
MCMC Diagnostic	Third Analysis	filtmonuments_provinces.png
MCMC Diagnostic	Fifth Analysis	hnwi.pdf
MCMC Diagnostic	Fifth Analysis	hnwi.png
MCMC Diagnostic	Sixth Analysis	tallbuildings.pdf
MCMC Diagnostic	Sixth Analysis	tallbuildings.png
<b>pa_...</b>		
nan	Visualization	gini_results.png
<b>point_...</b>		
nan	Visualization	scatters.pdf
nan	Visualization	scatters.png

### Main Analyses, Numeric Files (csv)

Context	Script Section	Filename
<b>analysis_...</b>		
Summary	Results Summary	results_summary.csv
<b>post_...</b>		
Summary	First Analysis	summary_allmonuments.csv
Summary	First Analysis	summary_allmonuments_nozero.csv
Summary	Second Analysis	summary_allwalls.csv
Summary	Second Analysis	summary_allwalls_nozero.csv
Summary	Fourth Analysis	summary_epigraphy.csv
Summary	Fourth Analysis	summary_epigraphy_nozero.csv
Summary	Third Analysis	summary_filtmonuments.csv
Summary	Third Analysis	summary_filtmonuments_nozero.csv
Summary	Fifth Analysis	summary_hnwi.csv
Summary	Sixth Analysis	summary_tallbuildings.csv
<b>rsquared.csv_...</b>		
Summary	Collate Main Results Table	rsquared.csv
<b>scaling_...</b>		
Summary	Plot Scaling Param Posterior Densities	samples.csv
<b>city_...</b>		
Model Diagnostic	First Analysis	outliers_allmonuments.csv

Context	Script Section	Filename
Model Diagnostic	Second Analysis	outliers_allwalls.csv
Model Diagnostic	Fourth Analysis	outliers_epigraphy.csv
Model Diagnostic	Third Analysis	outliers_filtmonuments.csv
Model Diagnostic	Fifth Analysis	outliers_hnwi.csv
Model Diagnostic	Sixth Analysis	outliers_tallbuildings.csv
		<b>geweke_...</b>
MCMC Diagnostic	First Analysis	allmonuments.csv
MCMC Diagnostic	First Analysis	allmonuments_nozero.csv
MCMC Diagnostic	Second Analysis	allwalls.csv
MCMC Diagnostic	Second Analysis	allwalls_nozero.csv
MCMC Diagnostic	Fourth Analysis	epigraphy.csv
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero.csv
MCMC Diagnostic	Third Analysis	filtmonuments.csv
MCMC Diagnostic	Third Analysis	filtmonuments_nozero.csv
MCMC Diagnostic	Fifth Analysis	hnwi.csv
MCMC Diagnostic	Sixth Analysis	tallbuildings.csv
		<b>grrhat_...</b>
MCMC Diagnostic	First Analysis	allmonuments.csv
MCMC Diagnostic	First Analysis	allmonuments_nozero.csv
MCMC Diagnostic	Second Analysis	allwalls.csv
MCMC Diagnostic	Second Analysis	allwalls_nozero.csv
MCMC Diagnostic	Fourth Analysis	epigraphy.csv
MCMC Diagnostic	Fourth Analysis	epigraphy_nozero.csv
MCMC Diagnostic	Third Analysis	filtmonuments.csv
MCMC Diagnostic	Third Analysis	filtmonuments_nozero.csv
MCMC Diagnostic	Fifth Analysis	hnwi.csv
MCMC Diagnostic	Sixth Analysis	tallbuildings.csv

## Supplementary Analyses, Plot Files

Context	Script Section	Filename
		<b>resid_...</b>
Model Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.pdf
Model Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.png
Model Diagnostic	Supplemental: Model Comparison	allwalls_linlog.pdf
Model Diagnostic	Supplemental: Model Comparison	allwalls_linlog.png
Model Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.pdf
Model Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.png
Model Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.pdf
Model Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.png
Model Diagnostic	Supplemental: Model Comparison	hnwi_linlog.pdf

Context	Script Section	Filename
Model Diagnostic	Supplemental: Model Comparison	hnwi_linlog.png
Model Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.pdf
Model Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.png
		<b>tplots_...</b>
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.pdf
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.png
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog_provinces.pdf
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog_provinces.png
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup.pdf
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup.png
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup_provinces.pdf
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup_provinces.png
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog.pdf
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog.png
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog_provinces.pdf
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog_provinces.png
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.pdf
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.png
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog_provinces.pdf
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog_provinces.png
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.pdf
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.png
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin_provinces.pdf
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin_provinces.png
MCMC Diagnostic	Supplemental: Model Comparison	hnwi_linlog.pdf
MCMC Diagnostic	Supplemental: Model Comparison	hnwi_linlog.png
MCMC Diagnostic	Supplemental: Outlier Effects	hnwi_sup.pdf
MCMC Diagnostic	Supplemental: Outlier Effects	hnwi_sup.png
MCMC Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.pdf
MCMC Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.png
		<b>point_...</b>
nan	Visualization	scatters_linear.pdf
nan	Visualization	scatters_linear.png
nan	Visualization	scatters_linear_log.pdf
nan	Visualization	scatters_linear_log.png

### Supplementary Analyses, Numeric Files (csv)

Context	Script Section	Filename
		<b>model_...</b>

Context	Script Section	Filename
Summary	Supplemental: Model Comparison	comparison.csv <b>post ...</b>
Summary	Supplemental: Model Comparison	summary_allmonuments_linlog.csv
Summary	Supplemental: Outlier Effects	summary_allmonuments_sup.csv
Summary	Supplemental: Model Comparison	summary_allwalls_linlog.csv
Summary	Supplemental: Model Comparison	summary_epigraphy_linlog.csv
Summary	Supplemental: Model Comparison	summary_filtmonuments_loglin.csv
Summary	Supplemental: Model Comparison	summary_hnwi_linlog.csv
Summary	Supplemental: Outlier Effects	summary_hnwi_sup.csv
Summary	Supplemental: Model Comparison	summary_tallbuildings_linlog.csv <b>waic.csv...</b>
Summary	Supplemental: Model Comparison	waic.csv <b>city...</b>
Model Diagnostic	Supplemental: Model Comparison	outliers_allmonuments_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	outliers_allwalls_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	outliers_epigraphy_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	outliers_filtmonuments_loglin.csv
Model Diagnostic	Supplemental: Model Comparison	outliers_hnwi_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	outliers_tallbuildings_linlog.csv <b>loo...</b>
Model Diagnostic	Supplemental: Model Comparison	est_allmonuments.csv
Model Diagnostic	Supplemental: Model Comparison	est_allmonuments_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	est_allwalls.csv
Model Diagnostic	Supplemental: Model Comparison	est_allwalls_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	est_epigraphy.csv
Model Diagnostic	Supplemental: Model Comparison	est_epigraphy_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	est_filtmonuments.csv
Model Diagnostic	Supplemental: Model Comparison	est_filtmonuments_loglin.csv
Model Diagnostic	Supplemental: Model Comparison	est_hnwi.csv
Model Diagnostic	Supplemental: Model Comparison	est_hnwi_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	est_tallbuildings.csv
Model Diagnostic	Supplemental: Model Comparison	est_tallbuildings_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	pw_allmonuments.csv
Model Diagnostic	Supplemental: Model Comparison	pw_allmonuments_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	pw_allwalls.csv
Model Diagnostic	Supplemental: Model Comparison	pw_allwalls_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	pw_epigraphy.csv
Model Diagnostic	Supplemental: Model Comparison	pw_epigraphy_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	pw_filtmonuments.csv
Model Diagnostic	Supplemental: Model Comparison	pw_filtmonuments_loglin.csv
Model Diagnostic	Supplemental: Model Comparison	pw_hnwi.csv

Context	Script Section	Filename
Model Diagnostic	Supplemental: Model Comparison	pw_hnwi_linlog.csv
Model Diagnostic	Supplemental: Model Comparison	pw_tallbuildings.csv
Model Diagnostic	Supplemental: Model Comparison	pw_tallbuildings_linlog.csv <b>lppd.csv_...</b>
Model Diagnostic	Supplemental: Model Comparison	lppd.csv <b>waic_...</b>
Model Diagnostic	Supplemental: Model Comparison	looic.csv <b>geweke_...</b>
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.csv
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup.csv
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog.csv
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.csv
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.csv
MCMC Diagnostic	Supplemental: Model Comparison	hnwi_linlog.csv
MCMC Diagnostic	Supplemental: Outlier Effects	hnwi_sup.csv
MCMC Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.csv <b>grrhat_...</b>
MCMC Diagnostic	Supplemental: Model Comparison	allmonuments_linlog.csv
MCMC Diagnostic	Supplemental: Outlier Effects	allmonuments_sup.csv
MCMC Diagnostic	Supplemental: Model Comparison	allwalls_linlog.csv
MCMC Diagnostic	Supplemental: Model Comparison	epigraphy_linlog.csv
MCMC Diagnostic	Supplemental: Model Comparison	filtmonuments_loglin.csv
MCMC Diagnostic	Supplemental: Model Comparison	hnwi_linlog.csv
MCMC Diagnostic	Supplemental: Outlier Effects	hnwi_sup.csv
MCMC Diagnostic	Supplemental: Model Comparison	tallbuildings_linlog.csv

## Summary Table of Key Model Posteriors

### Note

In the table below, the focal posterior summaries (lower and upper 95% CI, mean, and standard error) of parameters from the models are presented. For the archaeological data analyses, only the top-level scaling parameters in the model hierarchies are presented. This was intentional because the lower-level (group-level) parameters correspond to Roman provinces and there are 50+ for each model, making their presentation in a table unwieldy. Please see the csv files in the Output folder of the project repository (and the Summary Table of Outputs above) for the statistical summaries of the relevant group-level parameters. In line with the Methods section of the paper (and the R code) the parameters  $b_0$  and  $b_1$  refer to the intercept and focal scaling coefficients for the archaeological analyses.

## Consolidated Parameter Summary

model	param	lower	upper	mean	stdd
allmonuments	b0	2.88	5.19	4.12	0.45
allmonuments	b1	0.91	1.48	1.17	0.11
allmonuments_linlog	b0	2.81	5.09	4.06	0.44
allmonuments_linlog	b1	0.97	1.54	1.23	0.11
allmonuments_nozero	b0	3.06	5.33	4.27	0.44
allmonuments_nozero	b1	0.84	1.4	1.1	0.11
allmonuments_sup	b0	2.85	5.21	4.13	0.45
allmonuments_sup	b1	0.9	1.49	1.18	0.11
allwalls	b0	2.46	5.93	4.44	0.66
allwalls	b1	0.69	1.55	1.06	0.16
allwalls_linlog	b0	2.21	5.47	4.03	0.64
allwalls_linlog	b1	0.81	1.65	1.19	0.16
allwalls_nozero	b0	2.64	5.86	4.43	0.63
allwalls_nozero	b1	0.69	1.5	1.06	0.16
epigraphy	b0	3.26	5.85	4.64	0.5
epigraphy	b1	0.79	1.44	1.09	0.13
epigraphy_linlog	b0	2.76	5.78	4.36	0.57
epigraphy_linlog	b1	0.85	1.6	1.2	0.14
epigraphy_nozero	b0	2.13	6.1	4.19	0.78
epigraphy_nozero	b1	0.68	1.69	1.18	0.2
filtmonuments	b0	2.83	5.18	4.1	0.47
filtmonuments	b1	0.92	1.49	1.18	0.11
filtmonuments_loglin	b0	2.9	5.19	4.16	0.44
filtmonuments_loglin	b1	0.94	1.52	1.21	0.11
filtmonuments_nozero	b0	3.18	5.53	4.45	0.46
filtmonuments_nozero	b1	0.79	1.38	1.06	0.11
hnwi	intercept	-4.32	0.93	-1.73	1.01
hnwi	scaling	0.09	0.44	0.27	0.07
hnwi_linlog	intercept	-14.33	6.91	-4.19	4.16
hnwi_linlog	scaling	0.22	1.7	0.95	0.29
hnwi_sup	intercept	-4.26	0.97	-1.73	1.01
hnwi_sup	scaling	0.09	0.44	0.27	0.07
tallbuildings	intercept	-6.59	-2.53	-4.58	0.78
tallbuildings	scaling	0.41	0.69	0.55	0.05
tallbuildings_linlog	intercept	-18.51	7.59	-5.07	5.06
tallbuildings_linlog	scaling	1.7	3.72	2.71	0.39

## High-level Convergence Statistic Overview

In this section we present two plots below that summarize, at a high level, the convergence diagnostics from MCMC chains for all parameters and all models in our analysis. The first is a

histogram of  $\hat{R}$  statistics, while the second is a histogram of Geweke statistics. Together with the trace plots in the following section, these plots indicate good convergence throughout the analysis and suggest robust estimates of model posteriors. The Geweke histogram does show some extreme values, but these are only slightly more frequent than what would be expected by random chance at the 95% and 99% thresholds. Importantly, the  $\hat{R}$  statistics provide strong evidence of convergence, and the trace plots reveal no signs of drift. Taken together, these results suggest that the few extreme Geweke values do not raise significant concerns about the reliability of the analysis.

### rhat\_histogram.png

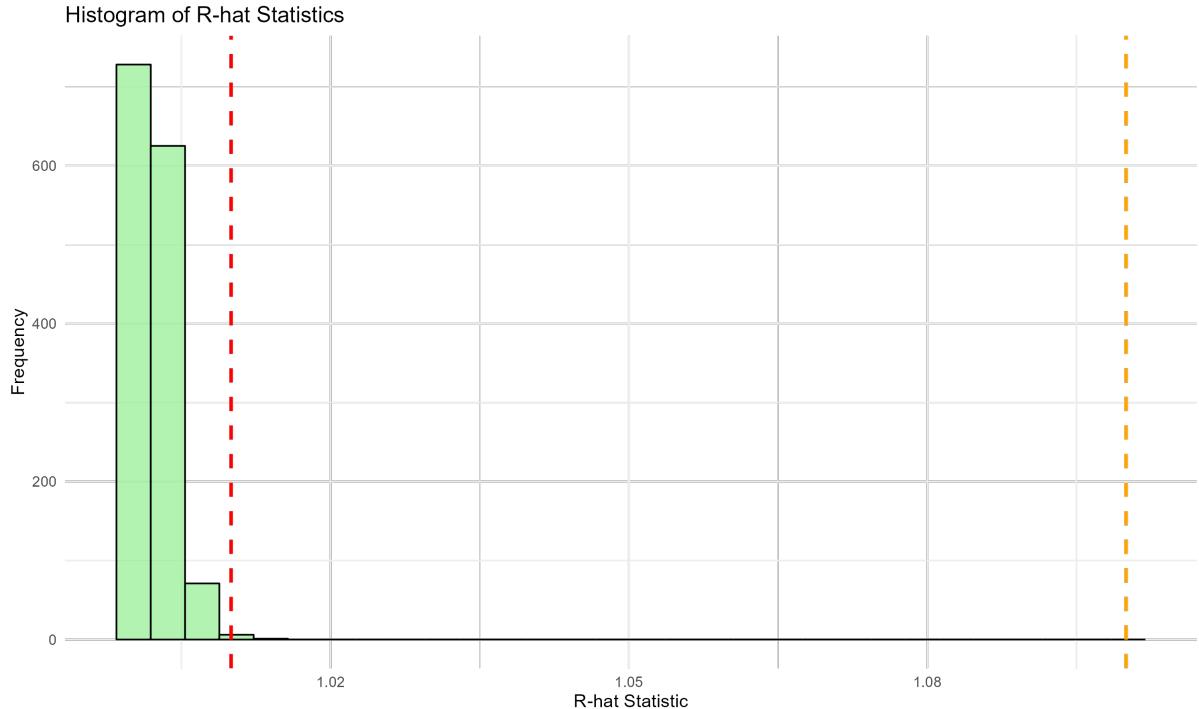


Figure 1: grrhat\_histogram.png

geweke\_histogram.png

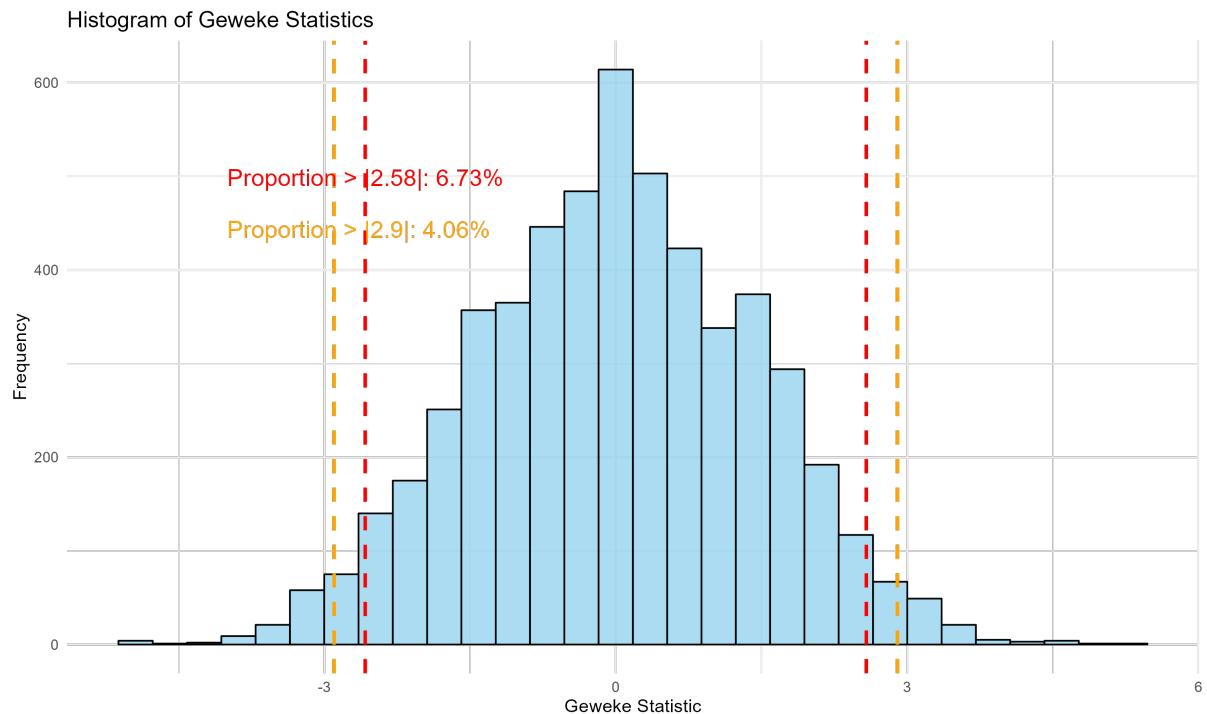


Figure 2: geweke\_histogram.png

---

## Trace Plots for All Analyses

### Trace Plots

This section contains the trace plots (`tplots_...`.png) for the MCMC diagnostics.

tplots\_allmonuments.png

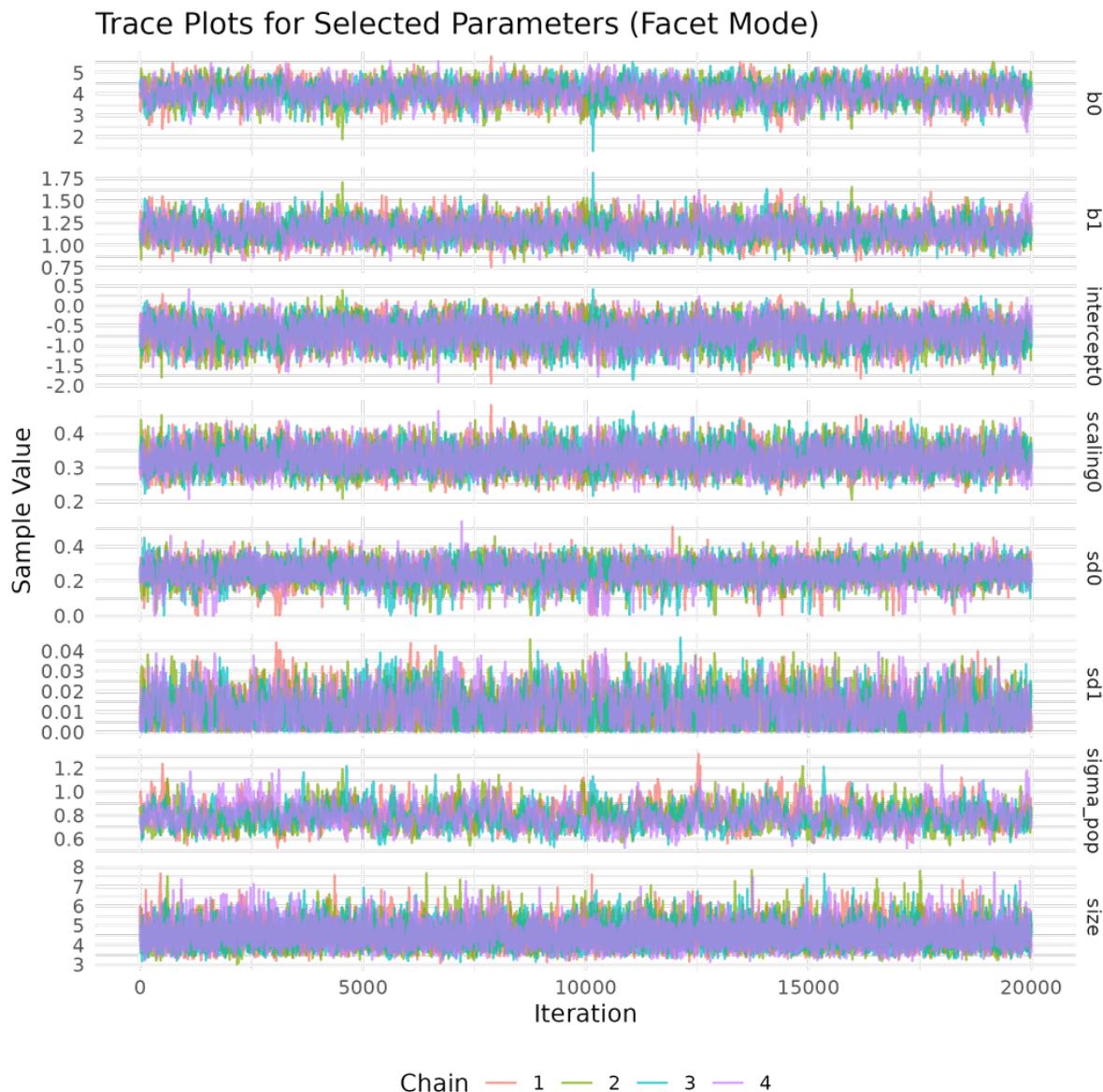


Figure 3: tplots\_allmonuments.png

tplots\_allmonuments\_nozero.png

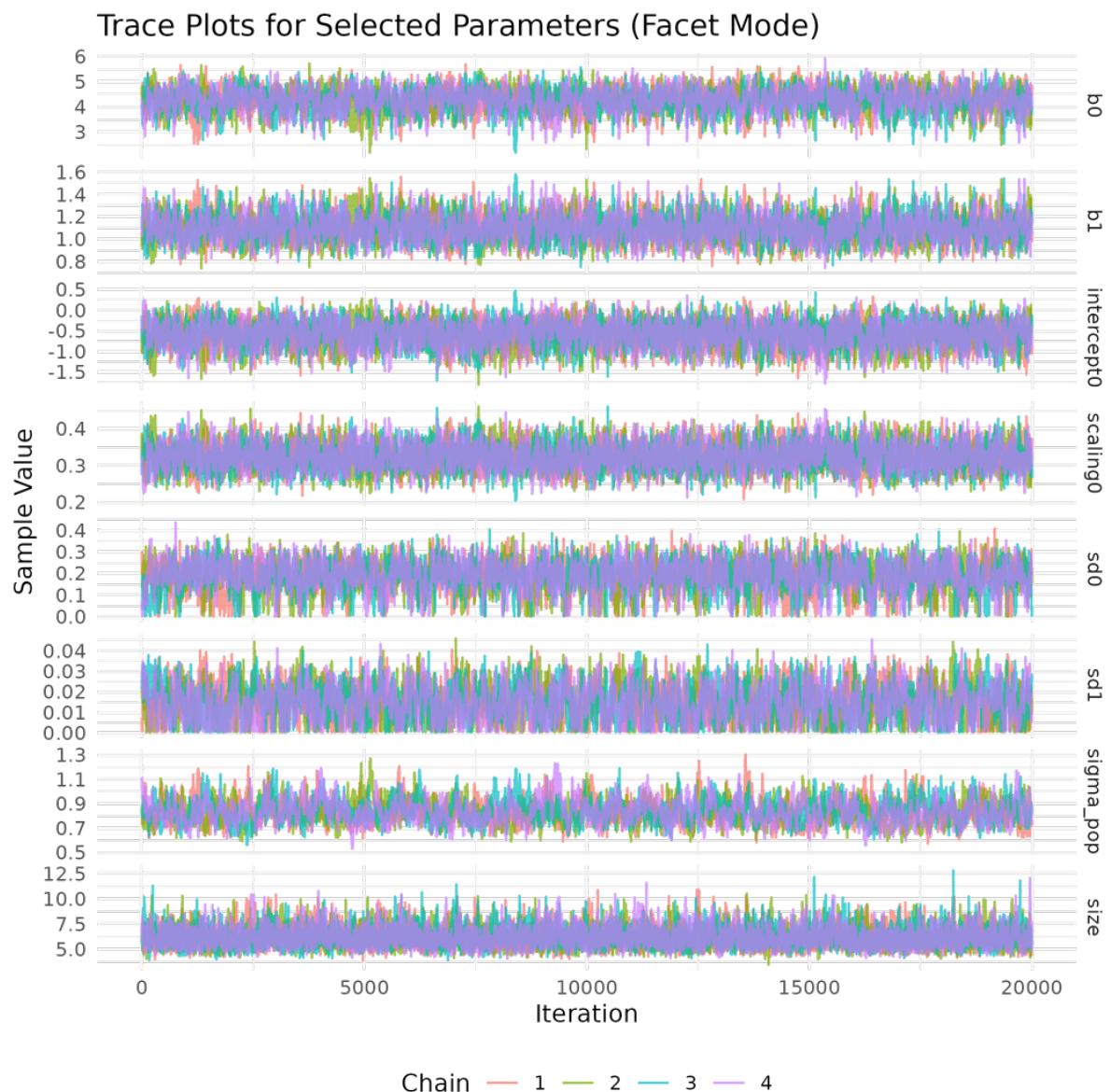


Figure 4: tplots\_allmonuments\_nozero.png

tplots\_allmonuments\_nozero\_provinces.png

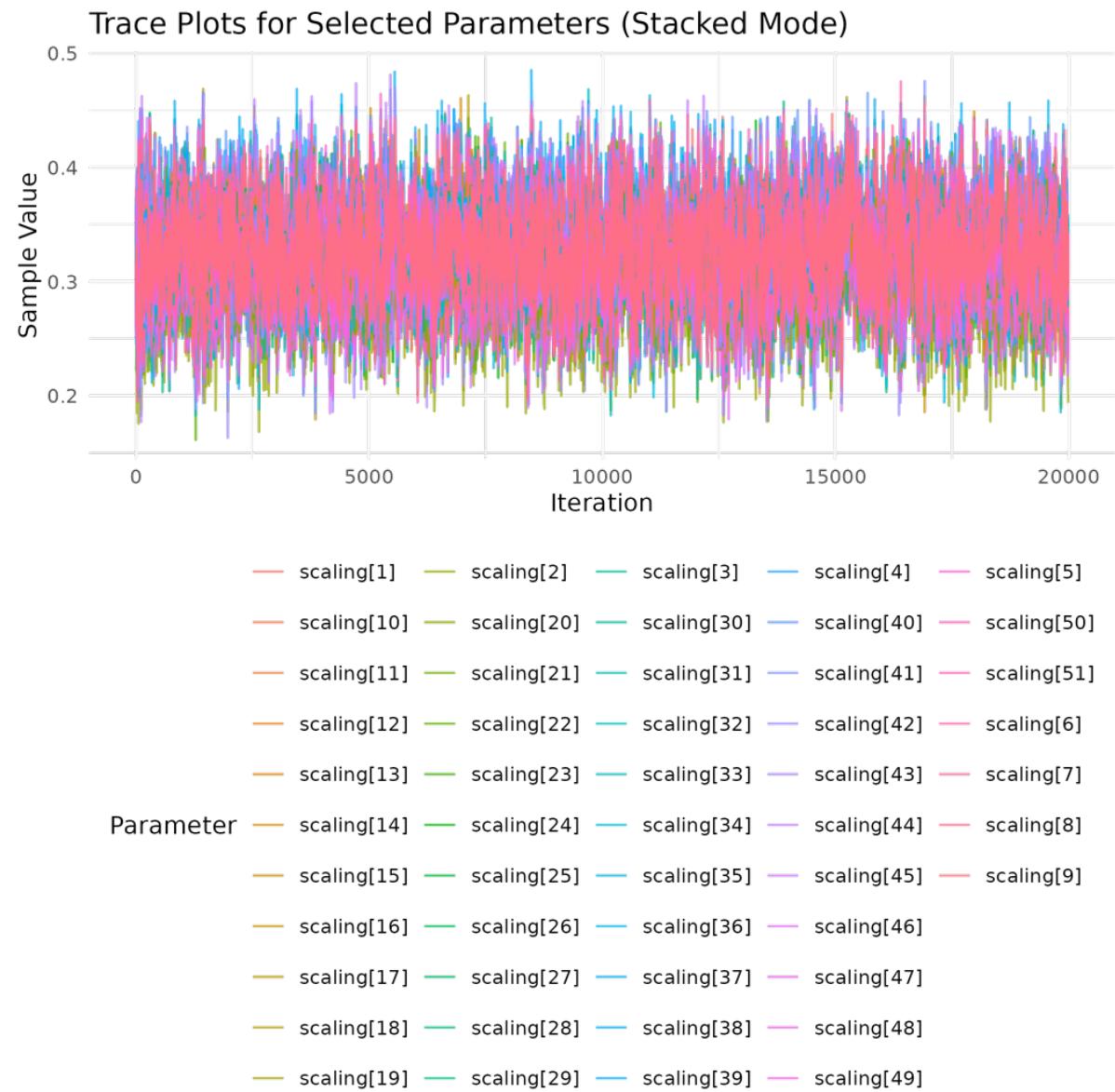


Figure 5: tplots\_allmonuments\_nozero\_provinces.png

tplots\_allmonuments\_provinces.png

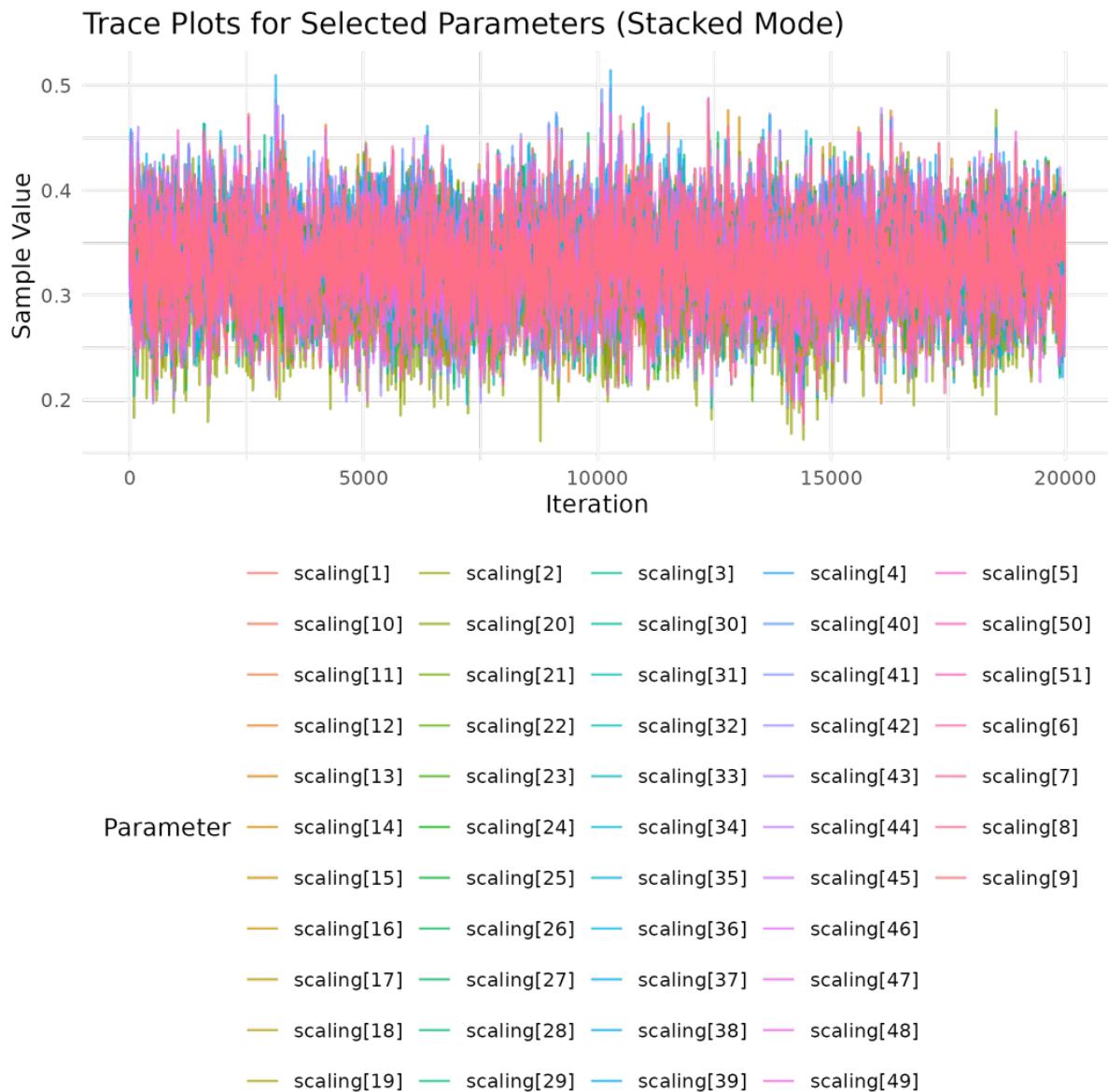


Figure 6: tplots\_allmonuments\_provinces.png

tplots\_allwalls.png

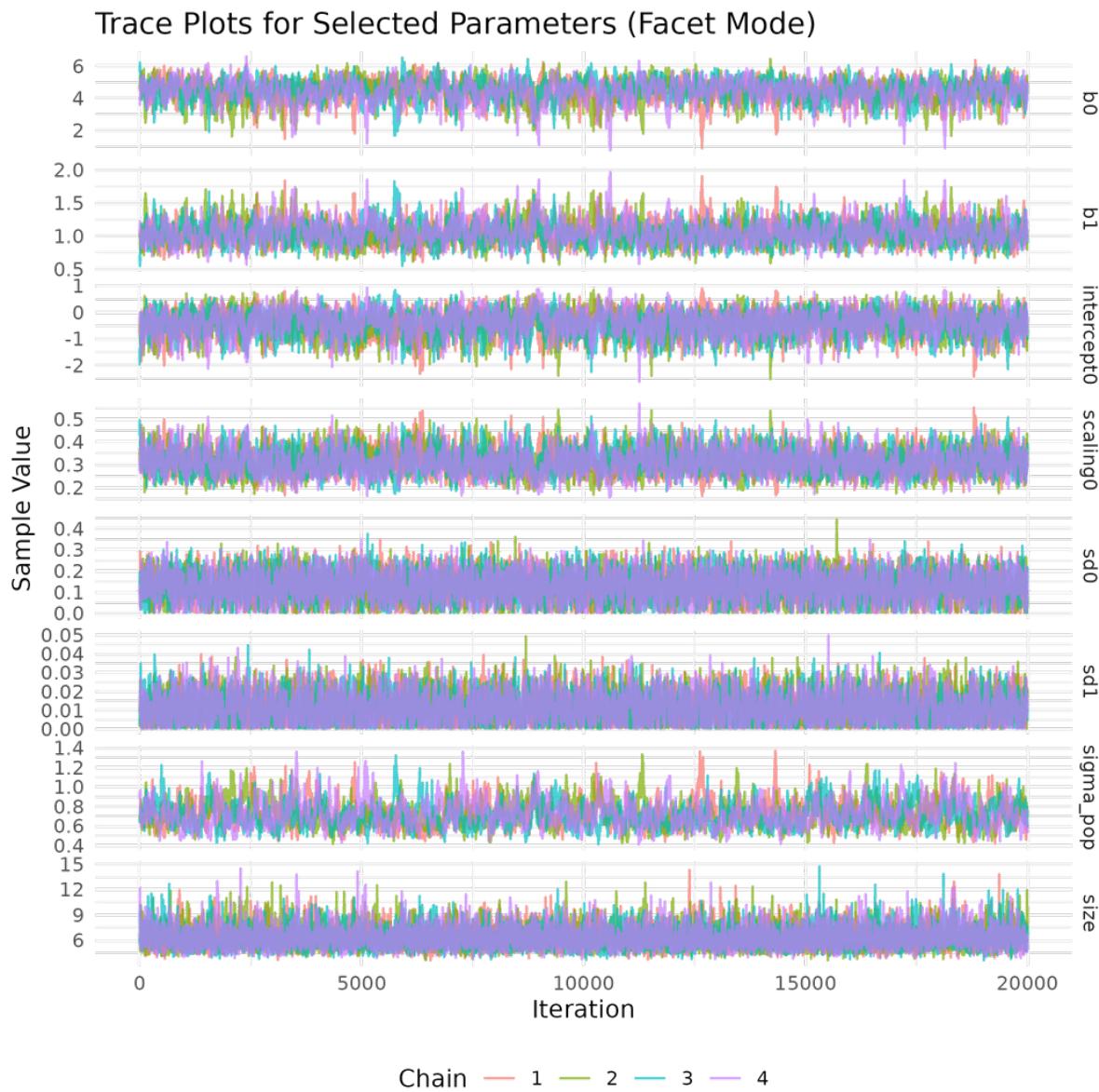


Figure 7: tplots\_allwalls.png

tplots\_allwalls\_nozero.png

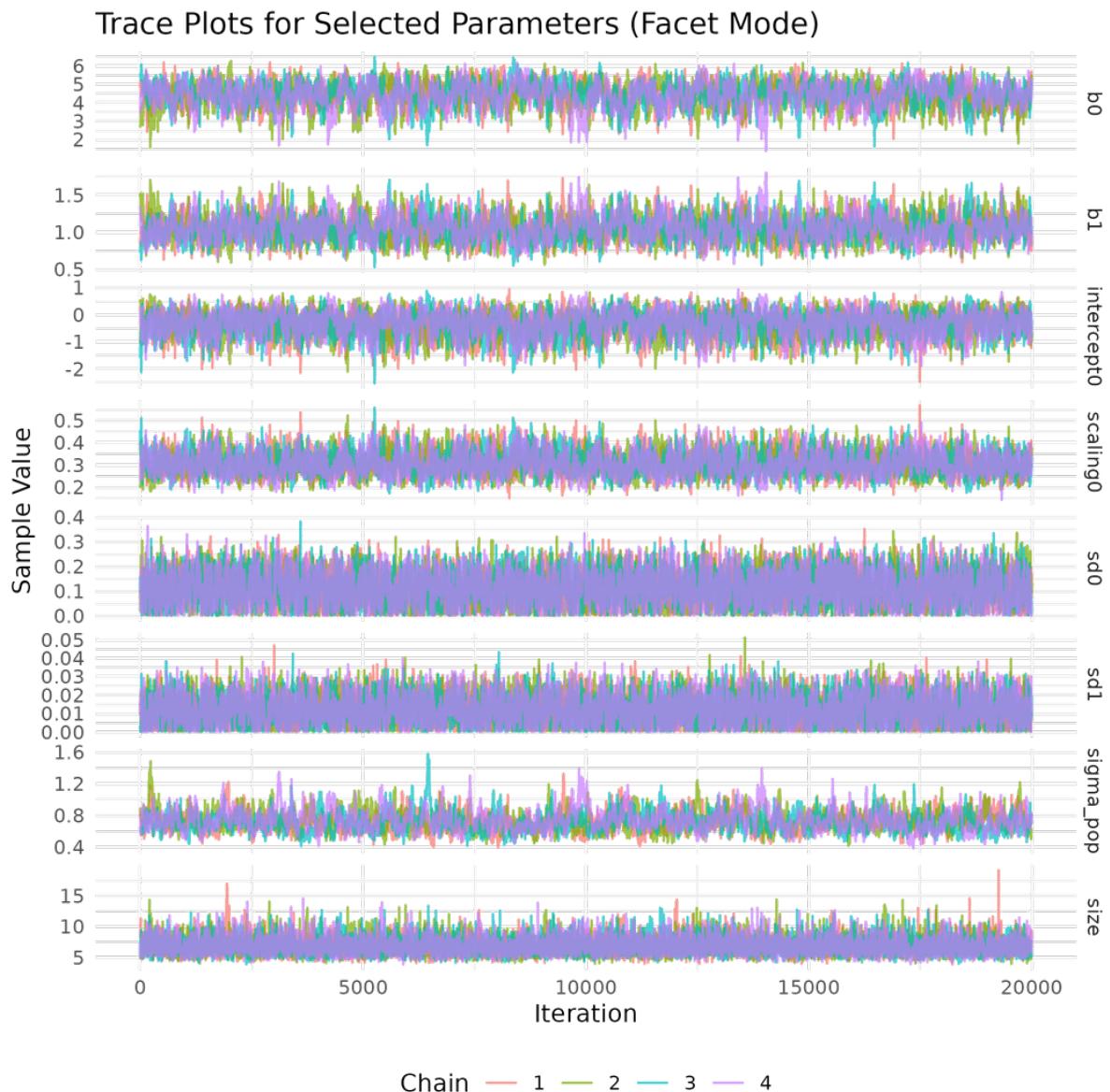


Figure 8: tplots\_allwalls\_nozero.png

tplots\_allwalls\_nozero\_provinces.png

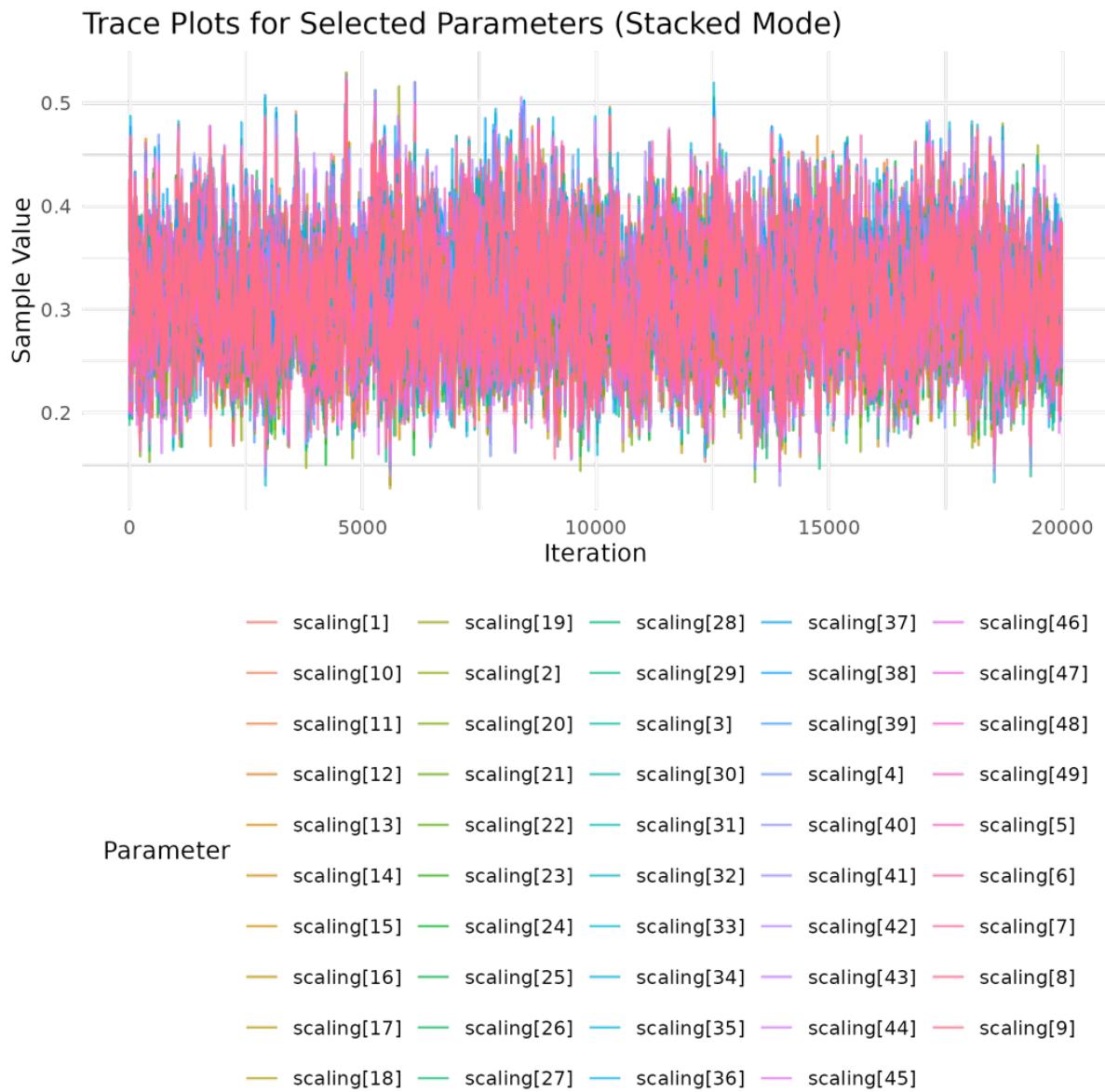


Figure 9: tplots\_allwalls\_nozero\_provinces.png

tplots\_allwalls\_provinces.png

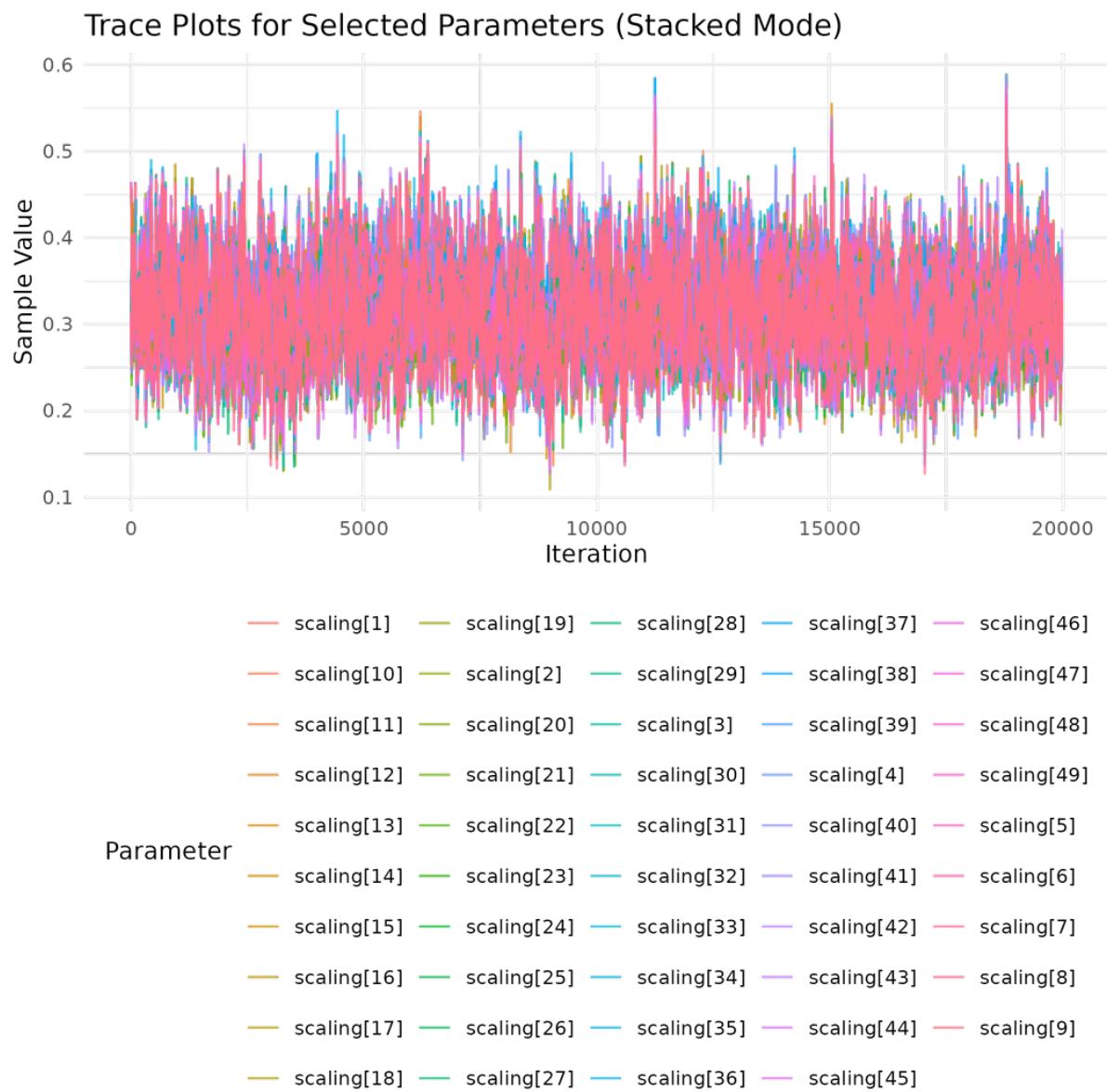


Figure 10: tplots\_allwalls\_provinces.png

tplots\_epigraphy.png

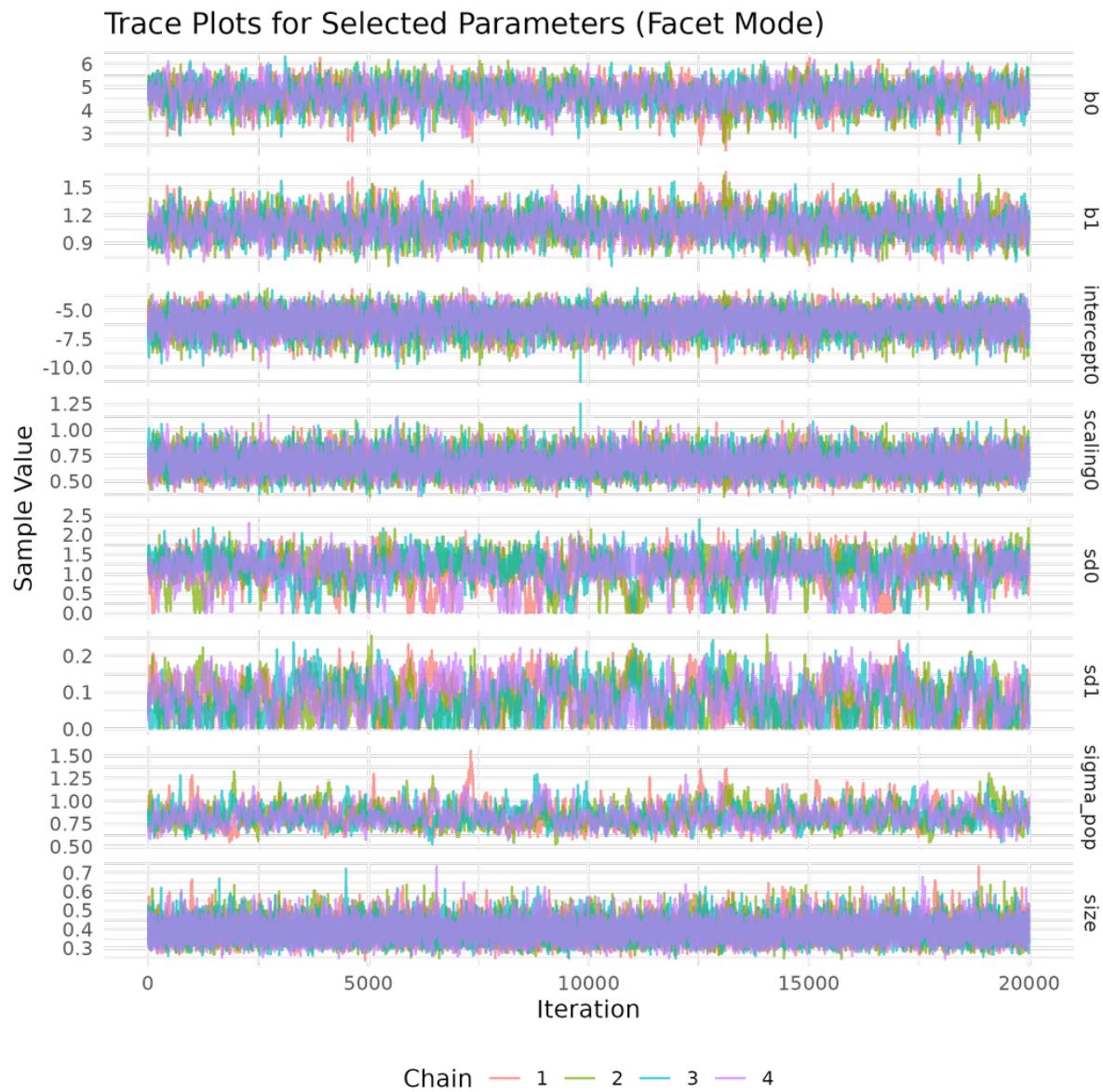


Figure 11: tplots\_epigraphy.png

tplots\_epigraphy\_nozero.png

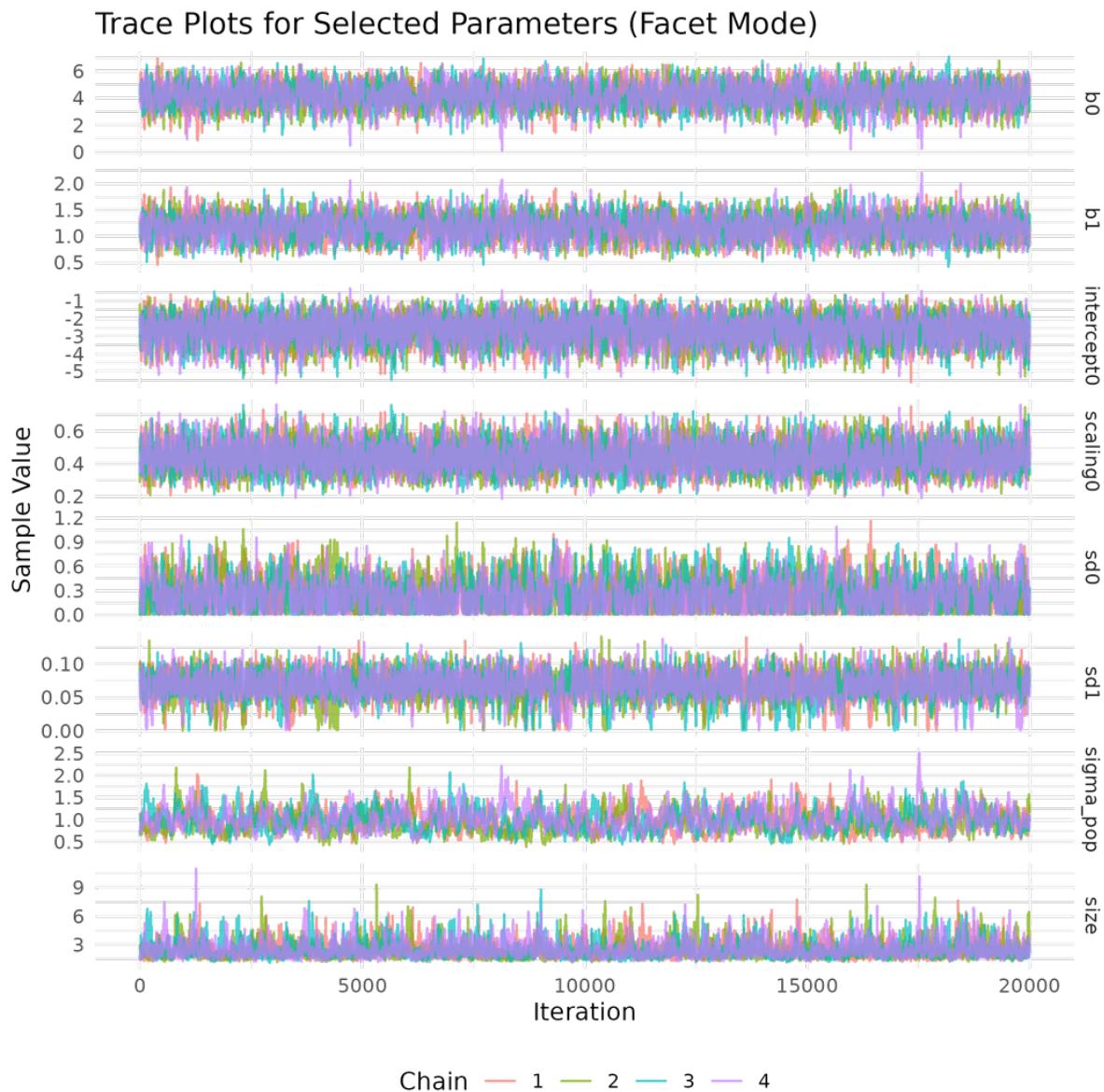


Figure 12: tplots\_epigraphy\_nozero.png

tplots\_epigraphy\_nozero\_provinces.png

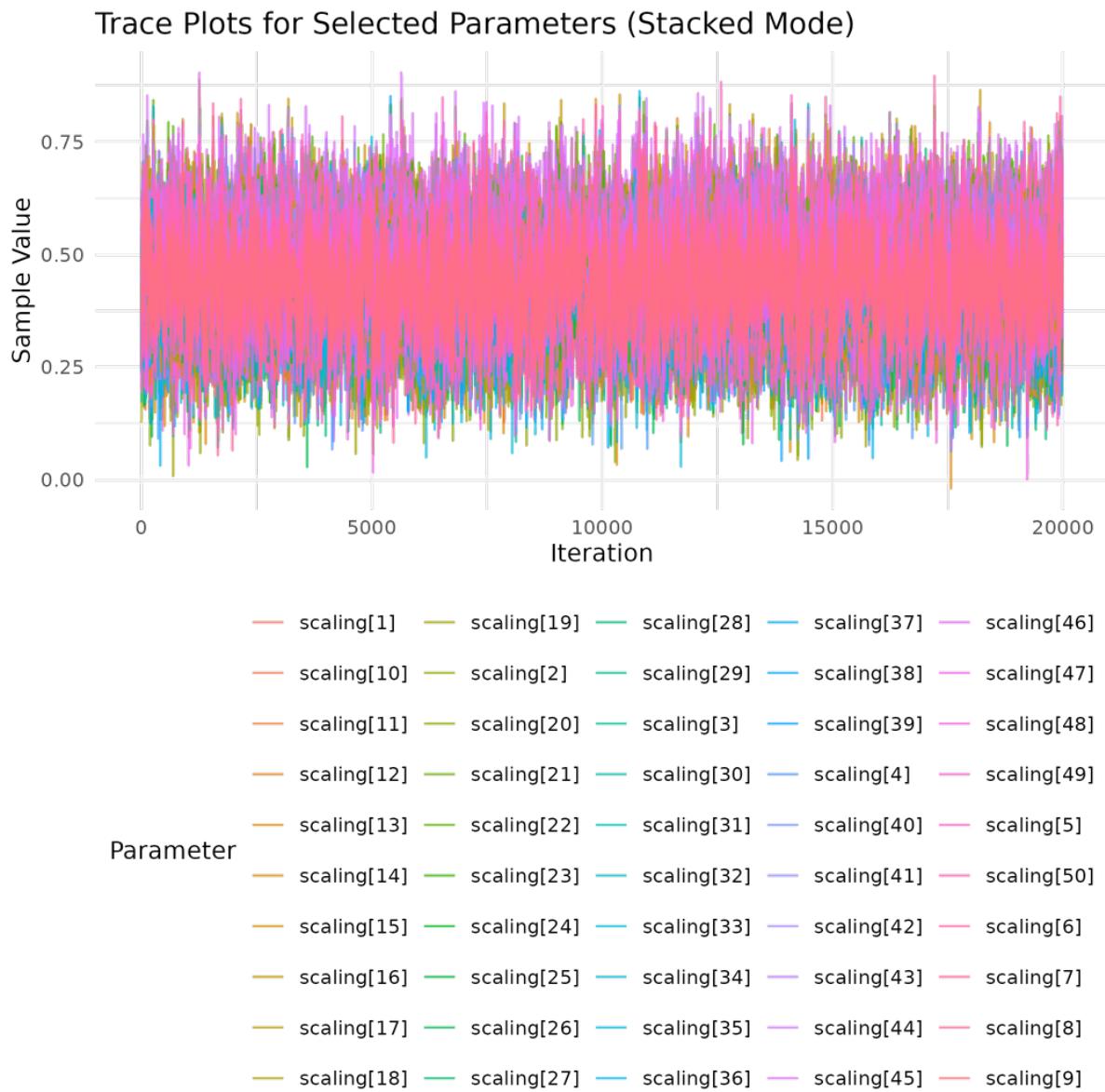


Figure 13: tplots\_epigraphy\_nozero\_provinces.png

tplots\_epigraphy\_provinces.png

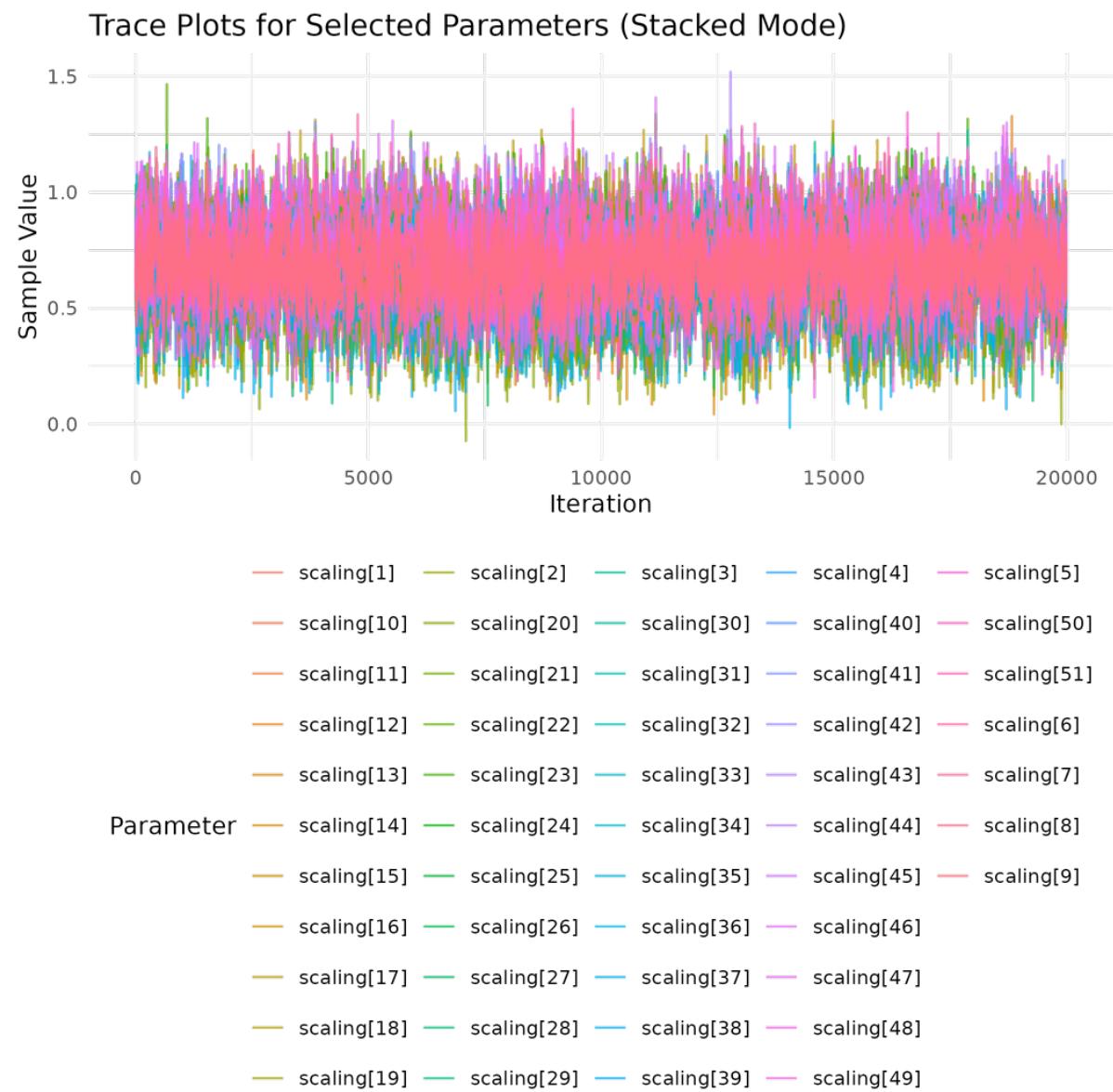


Figure 14: tplots\_epigraphy\_provinces.png

tplots\_filtmonuments.png

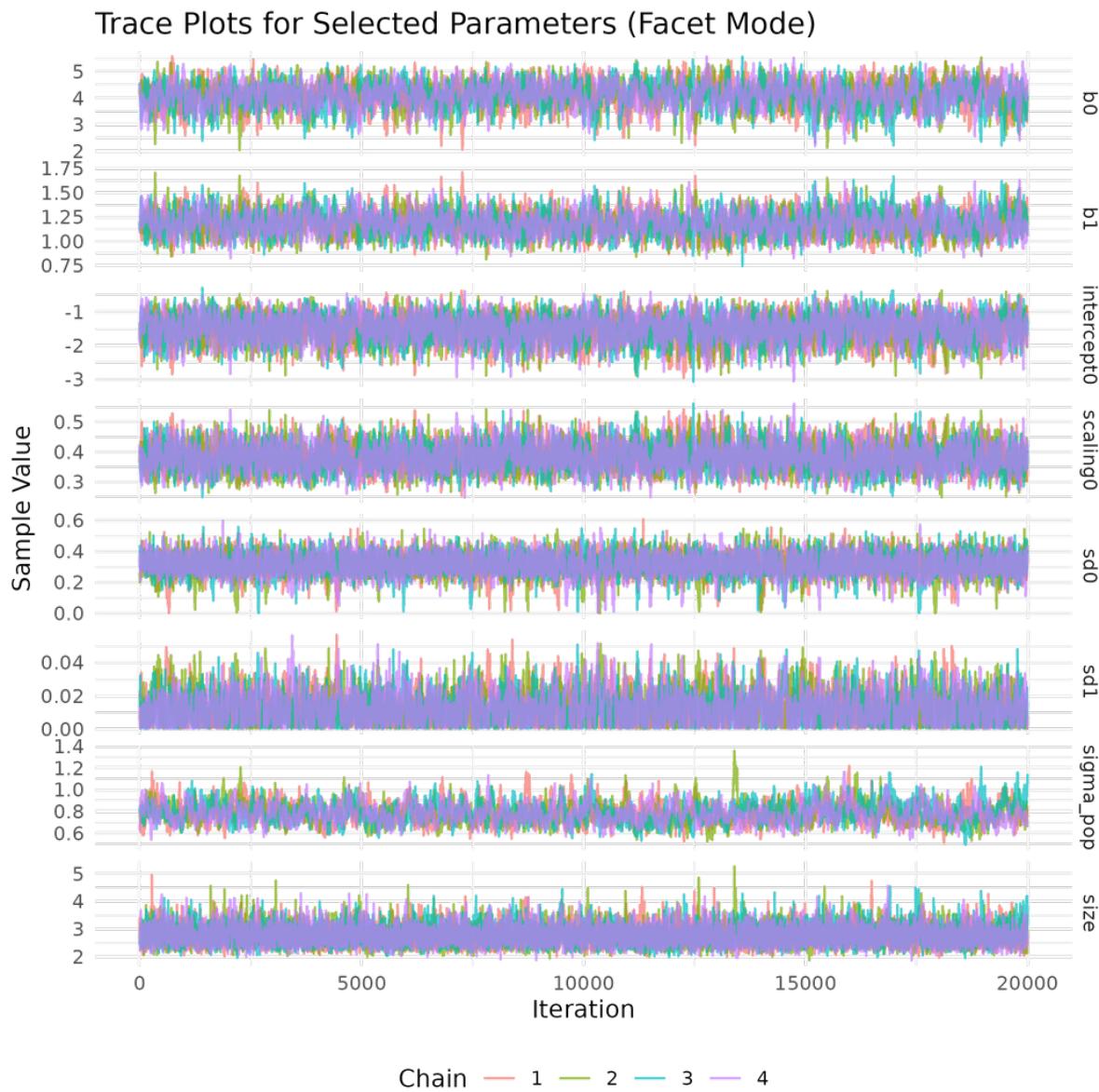


Figure 15: tplots\_filtmonuments.png

tplots\_filtmonuments\_nozero.png

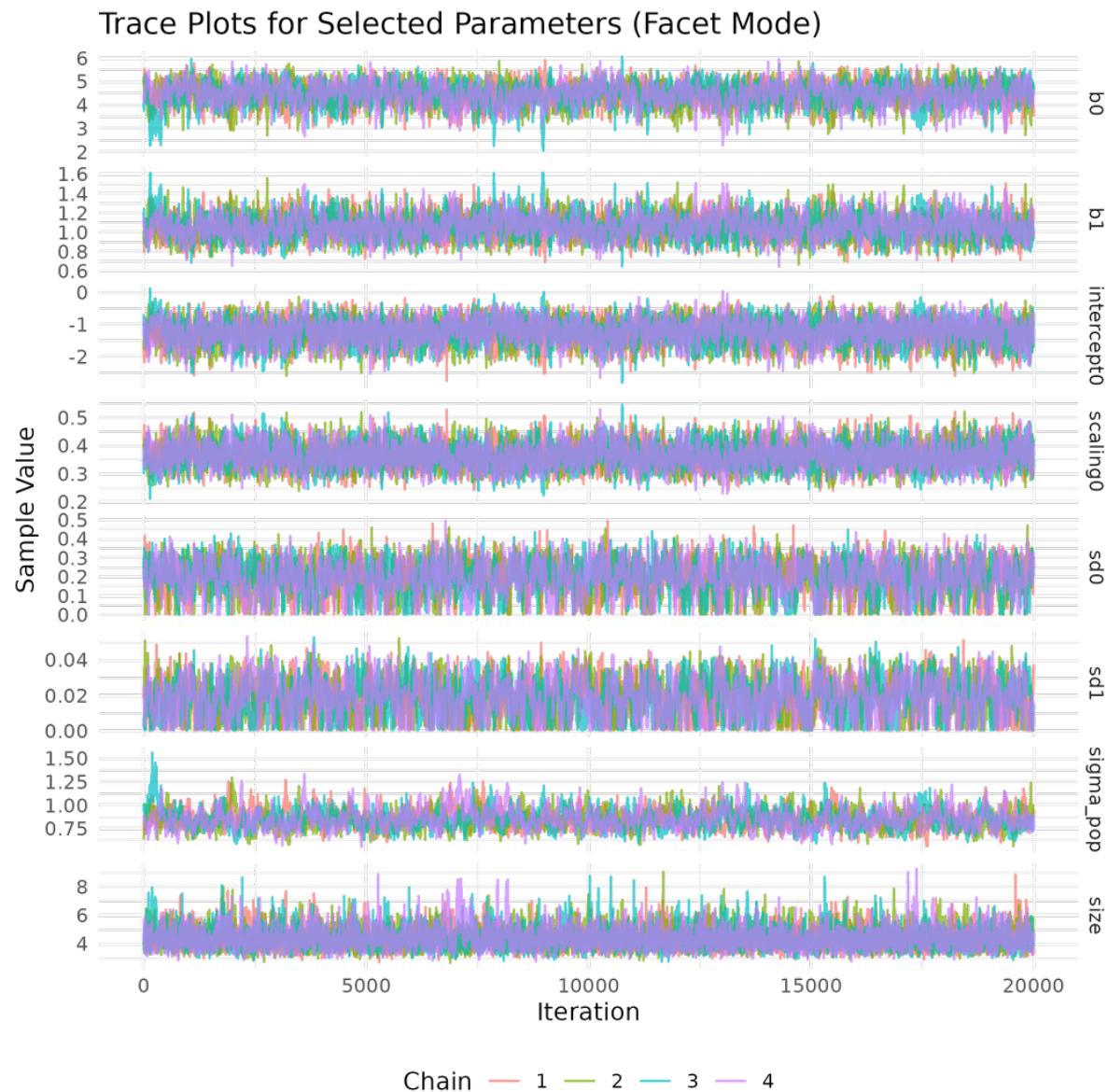


Figure 16: tplots\_filtmonuments\_nozero.png

tplots\_filtmonuments\_nozero\_provinces.png

### Trace Plots for Selected Parameters (Stacked Mode)

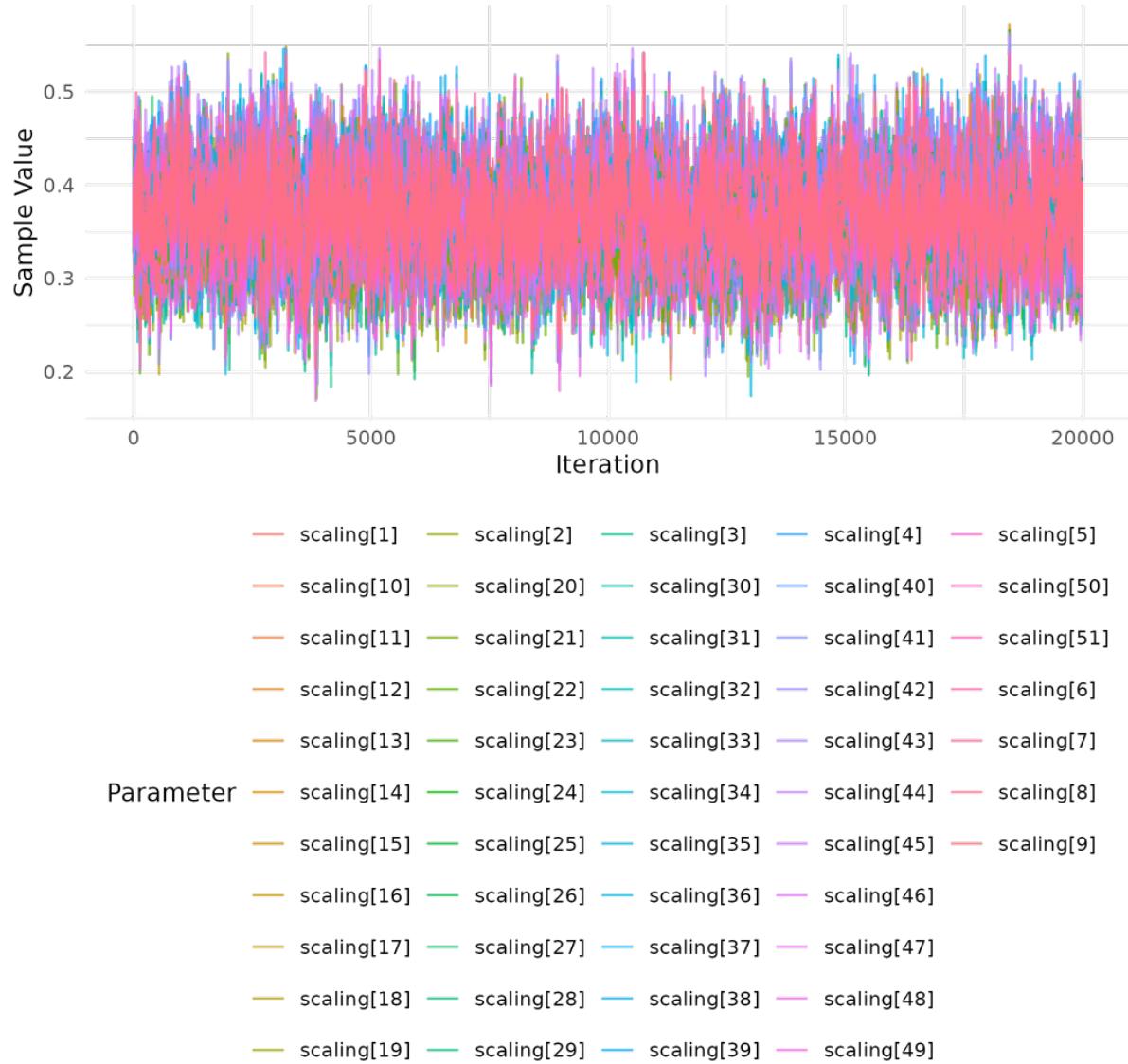


Figure 17: tplots\_filtmonuments\_nozero\_provinces.png

tplots\_filtmonuments\_provinces.png

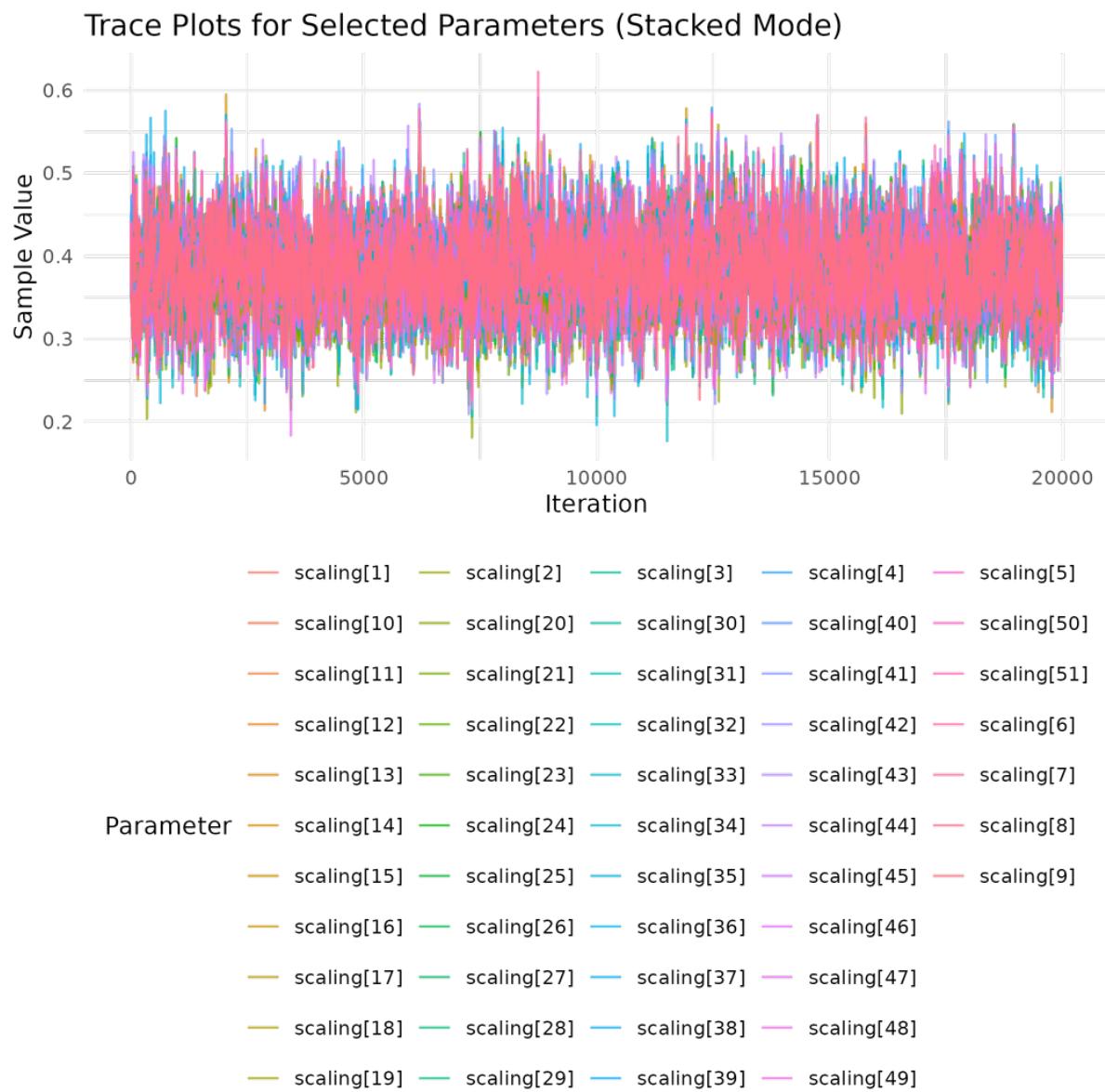


Figure 18: tplots\_filtmonuments\_provinces.png

tplots\_hnwi.png

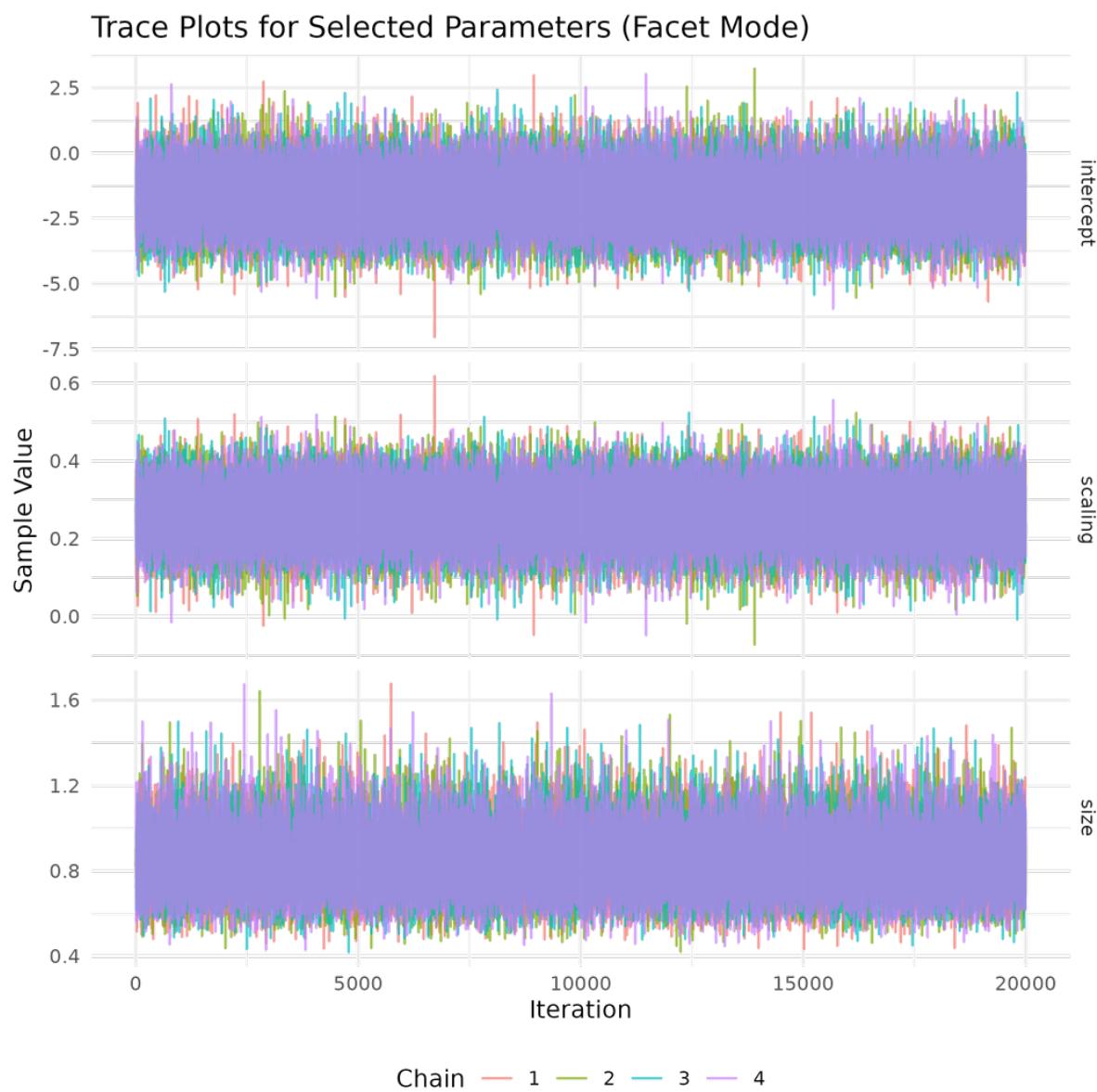


Figure 19: tplots\_hnwi.png

tplots\_tallbuildings.png

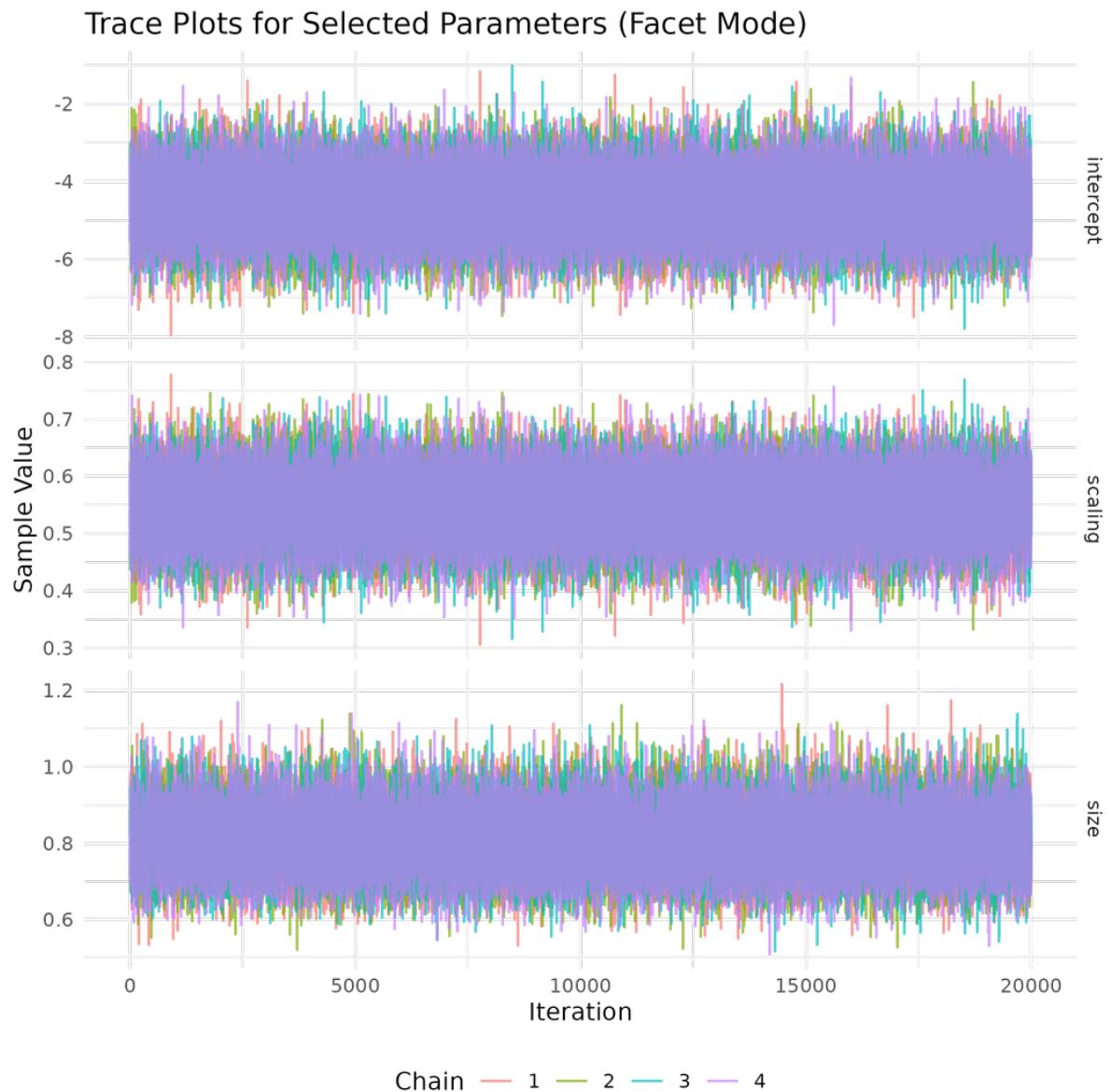


Figure 20: tplots\_tallbuildings.png

tplots\_allmonuments\_linlog.png

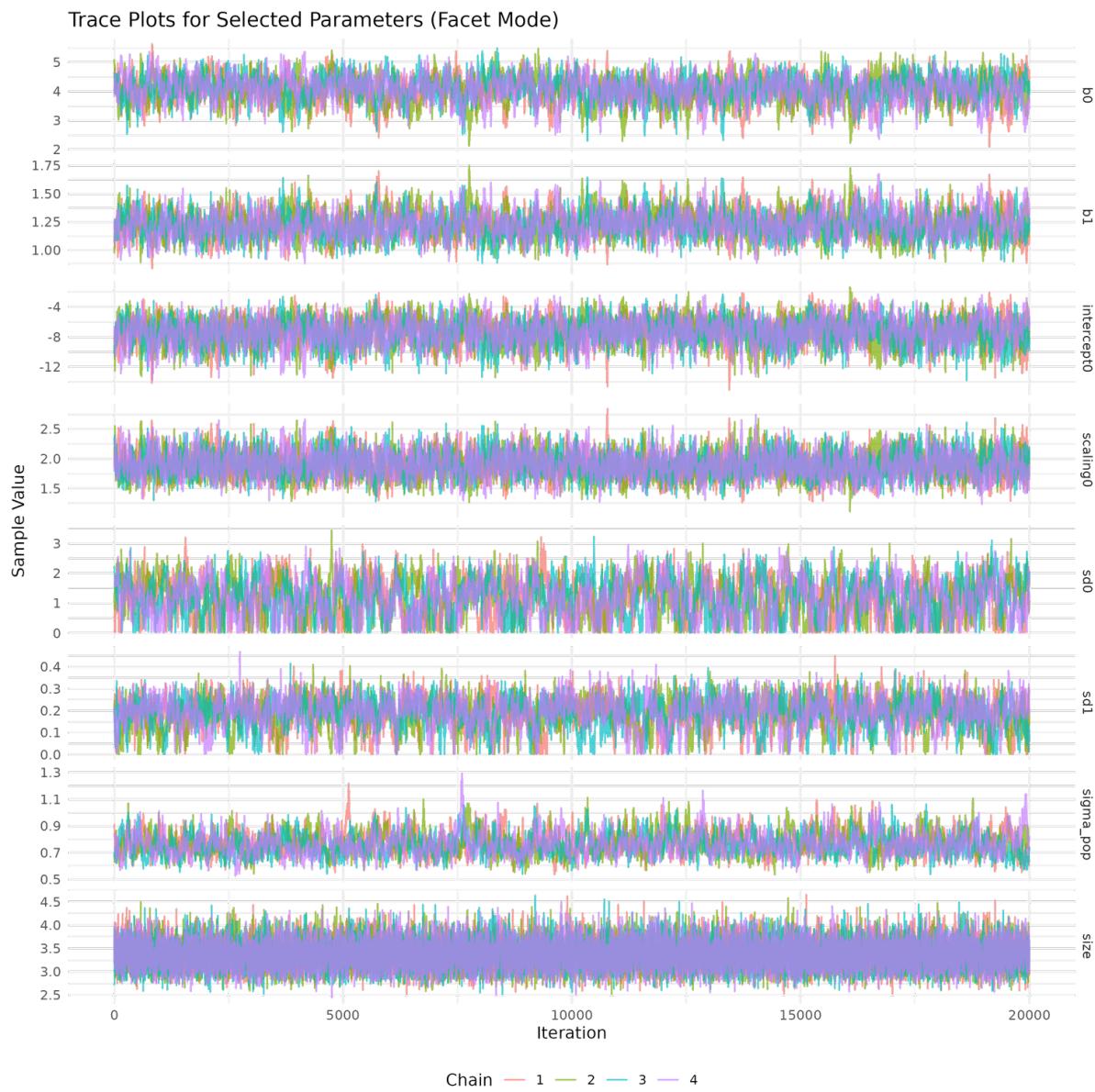


Figure 21: tplots\_allmonuments\_linlog.png

tplots\_allmonuments\_linlog\_provinces.png

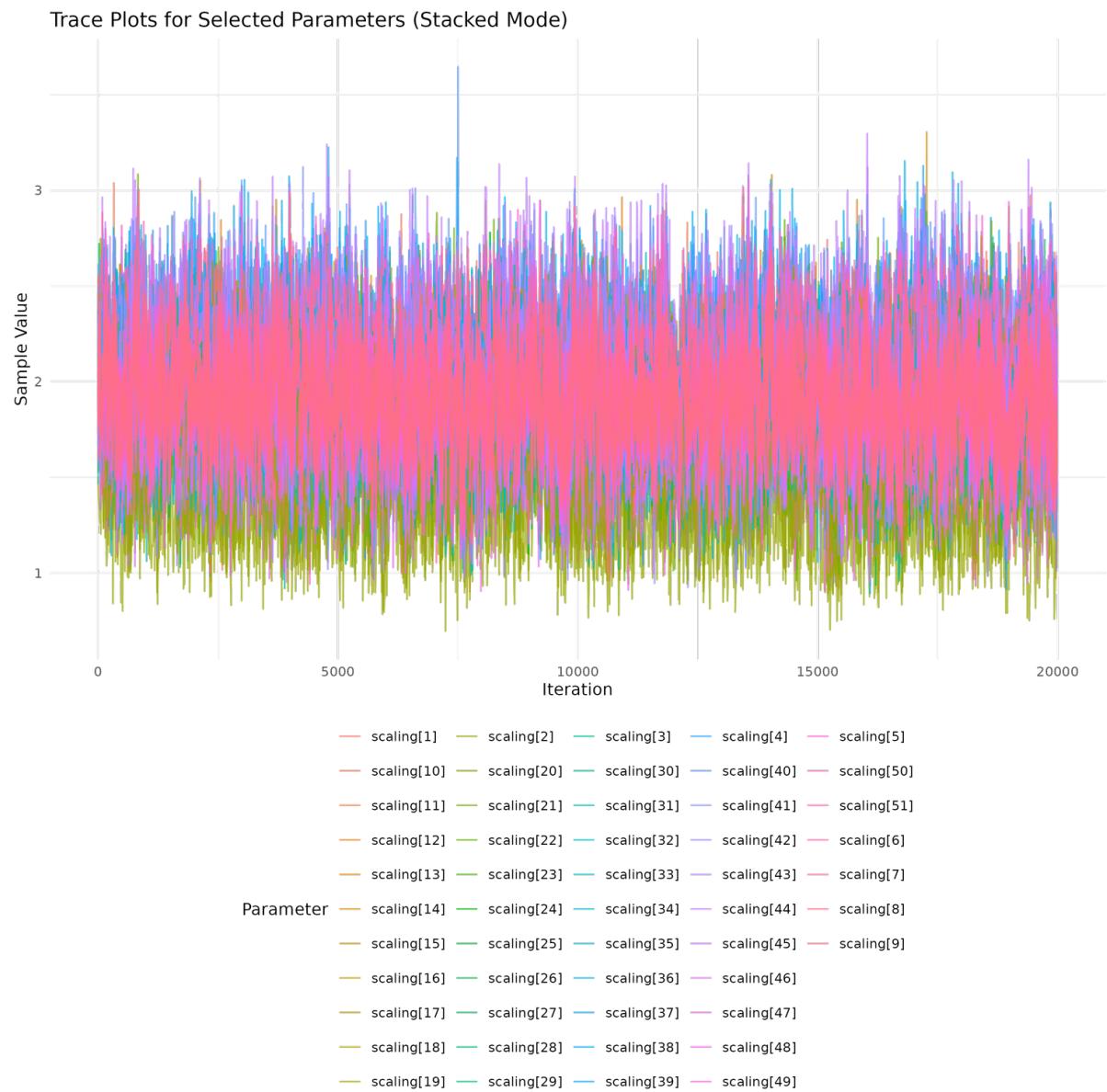


Figure 22: tplots\_allmonuments\_linlog\_provinces.png

tplots\_allmonuments\_sup.png

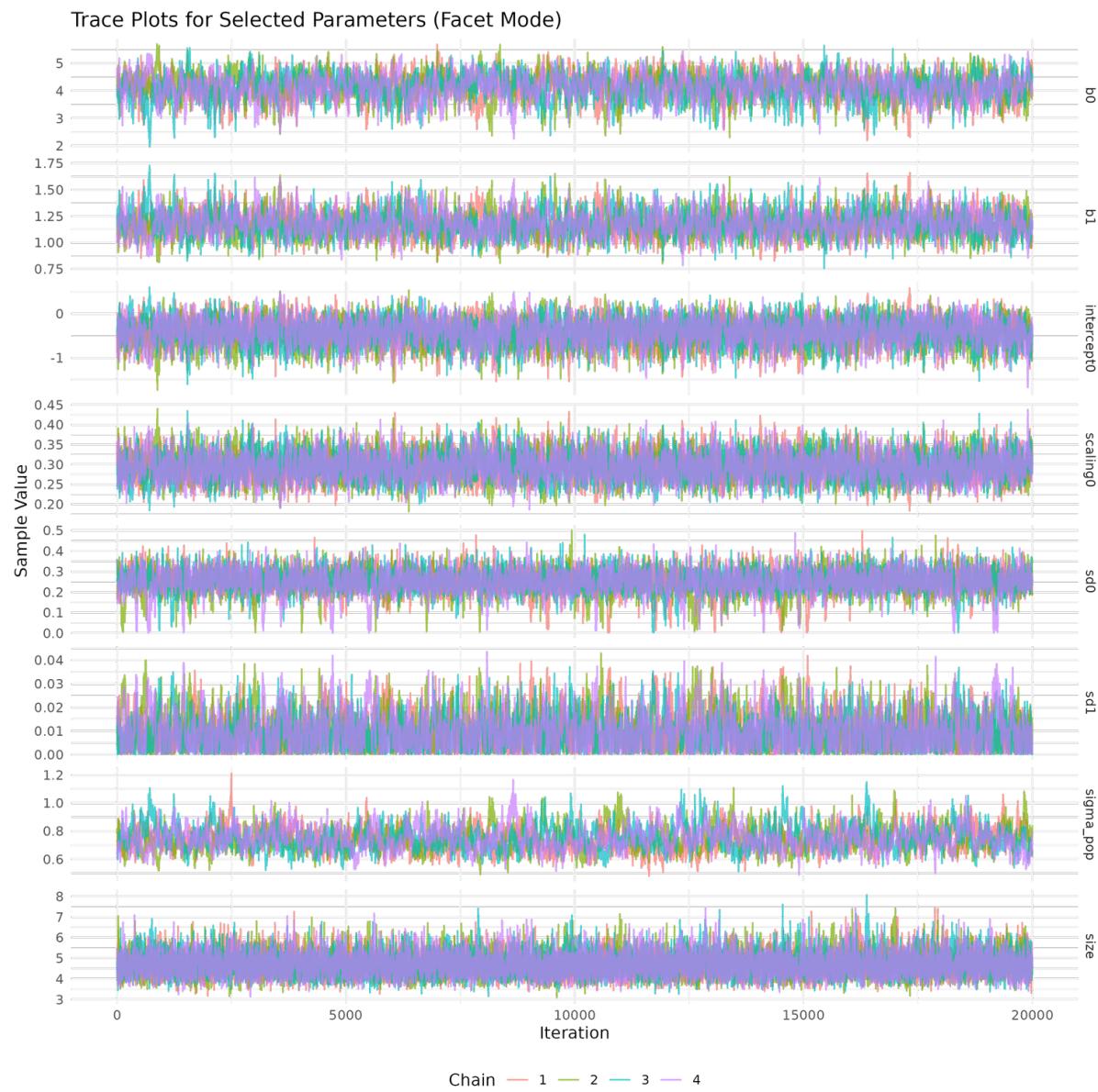


Figure 23: tplots\_allmonuments\_sup.png

tplots\_allmonuments\_sup\_provinces.png

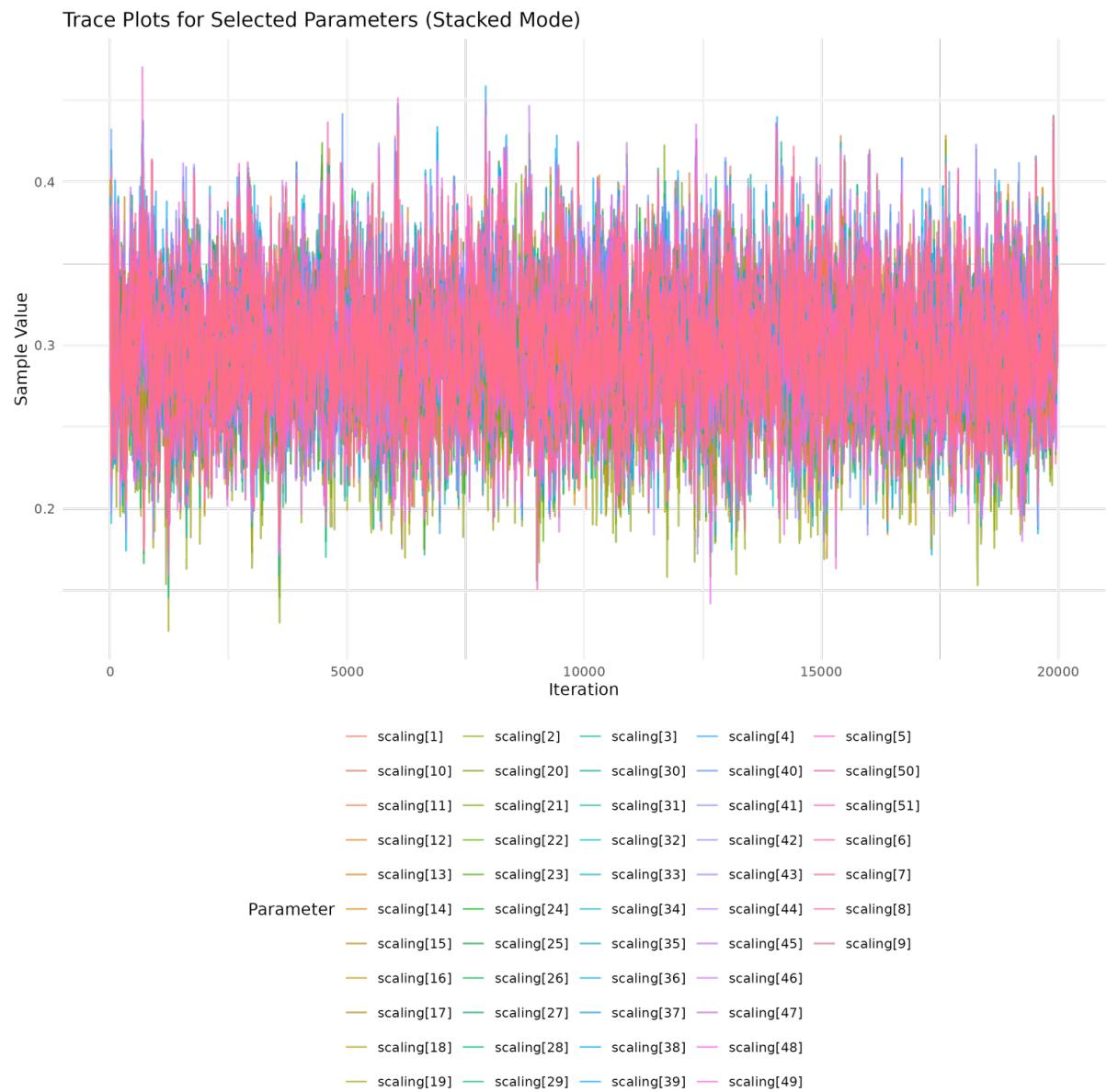


Figure 24: tplots\_allmonuments\_sup\_provinces.png

tplots\_allwalls\_linlog.png

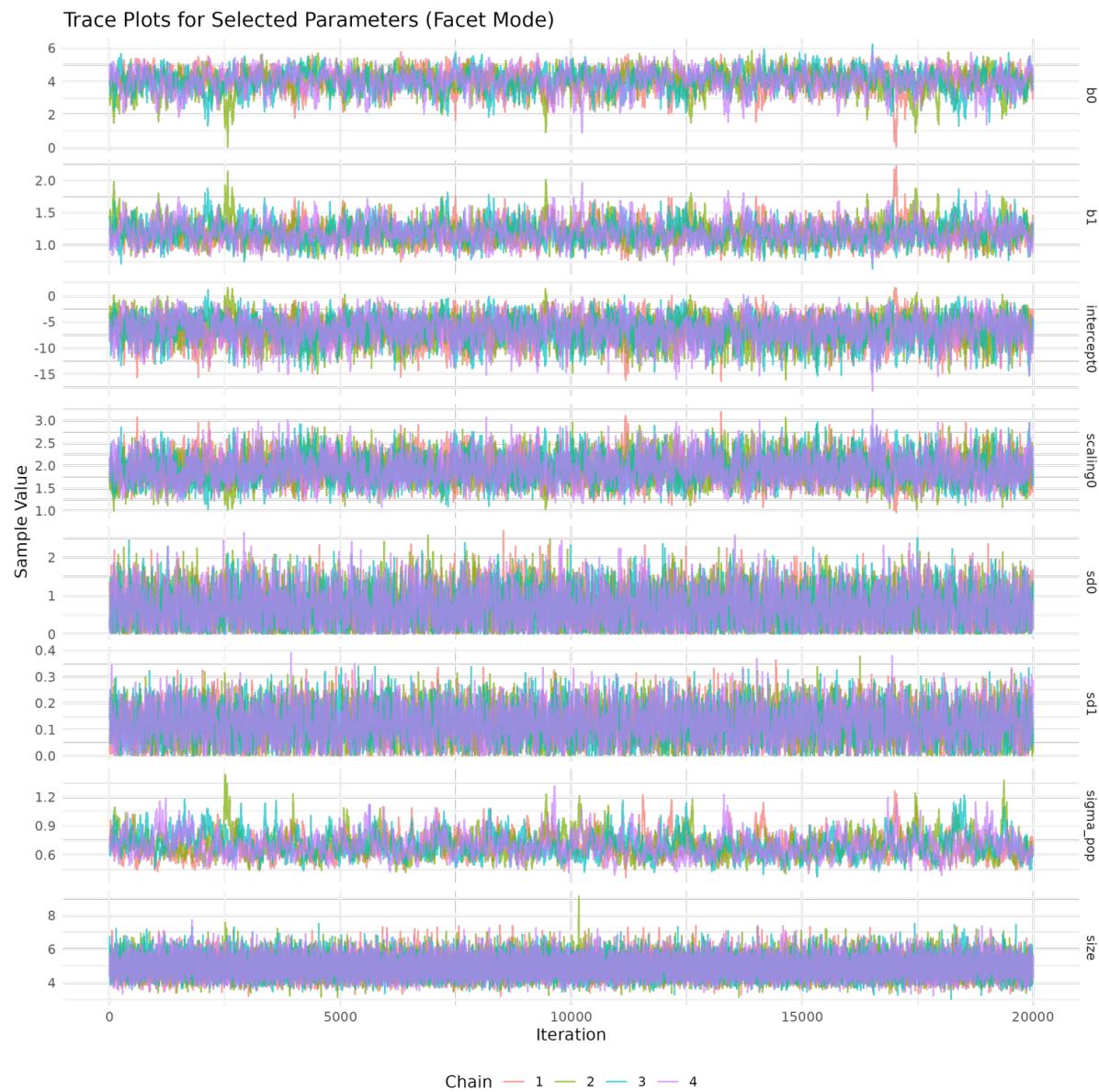


Figure 25: tplots\_allwalls\_linlog.png

tplots\_allwalls\_linlog\_provinces.png

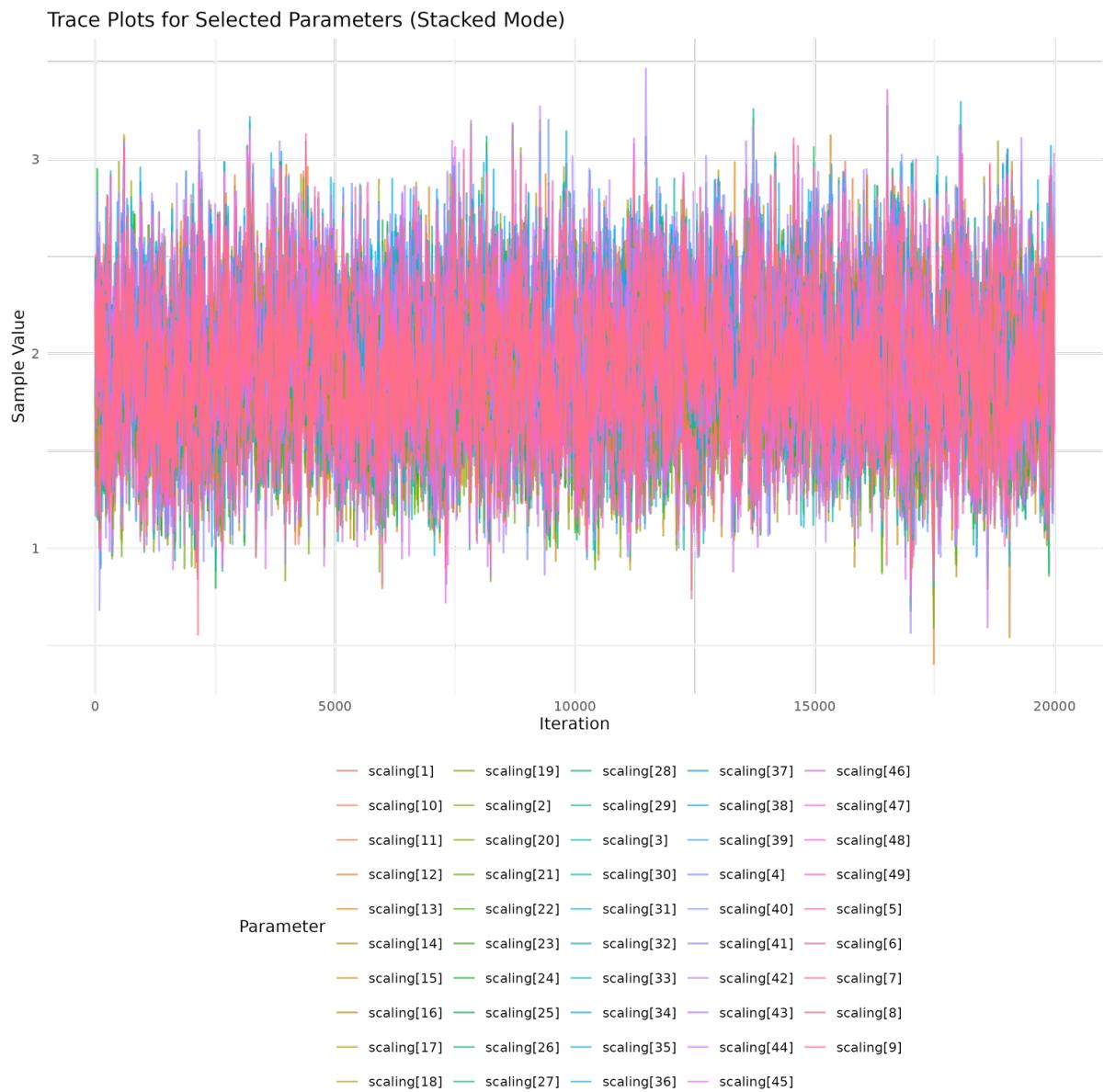


Figure 26: tplots\_allwalls\_linlog\_provinces.png

tplots\_epigraphy\_linlog.png

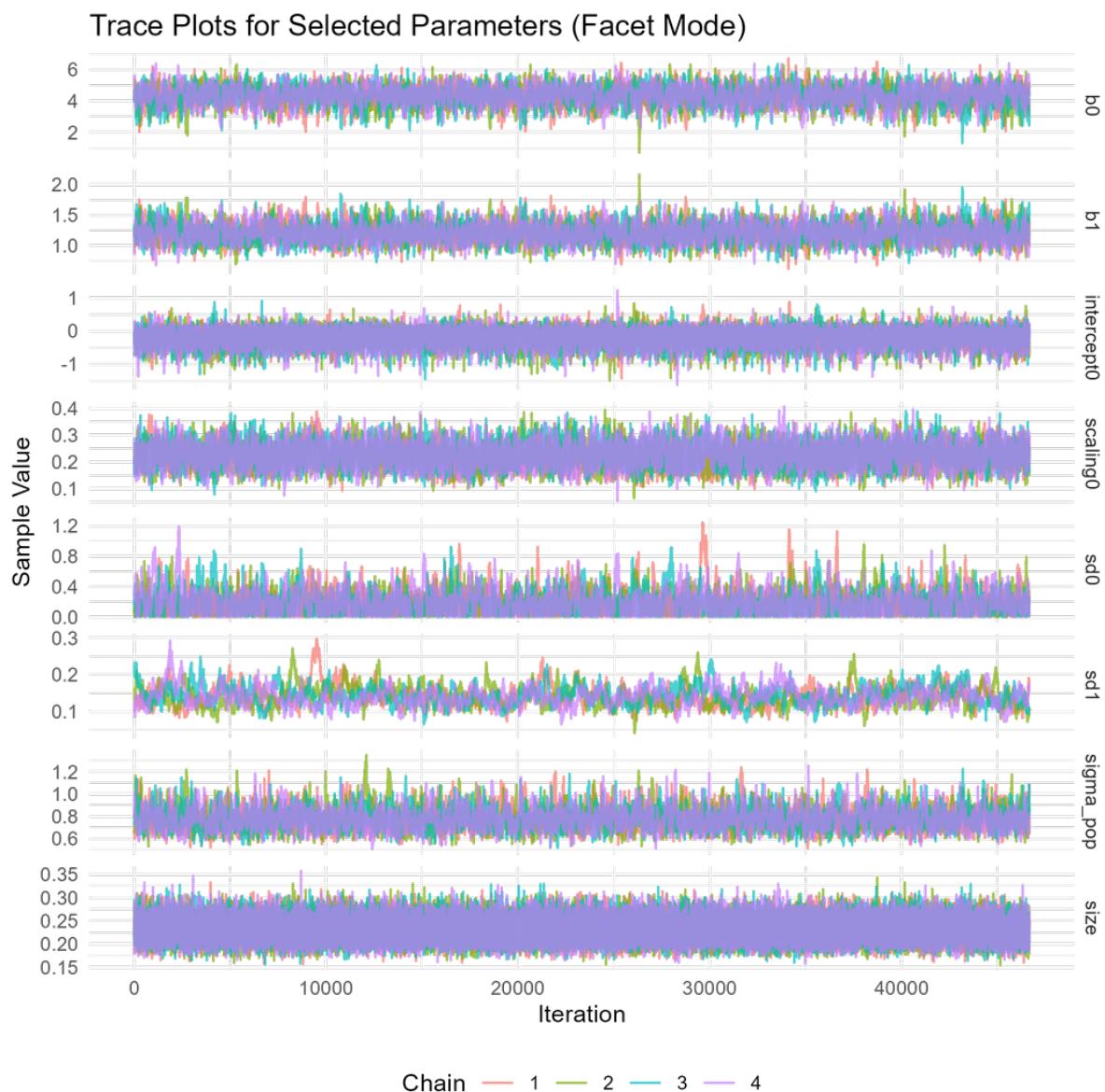


Figure 27: tplots\_epigraphy\_linlog.png

tplots\_epigraphy\_linlog\_provinces.png

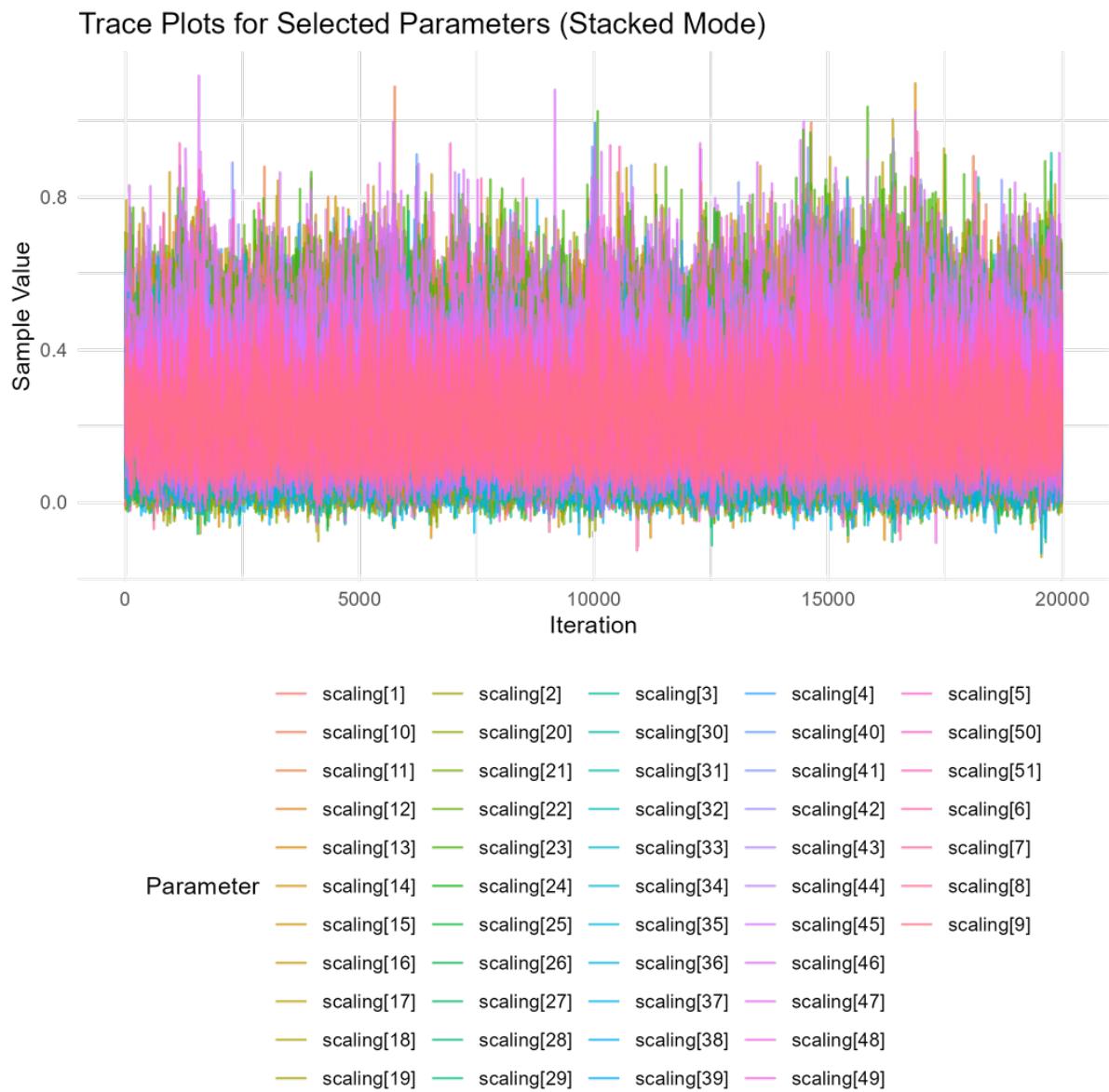


Figure 28: tplots\_epigraphy\_linlog\_provinces.png

tplots\_filtmonuments\_loglin.png

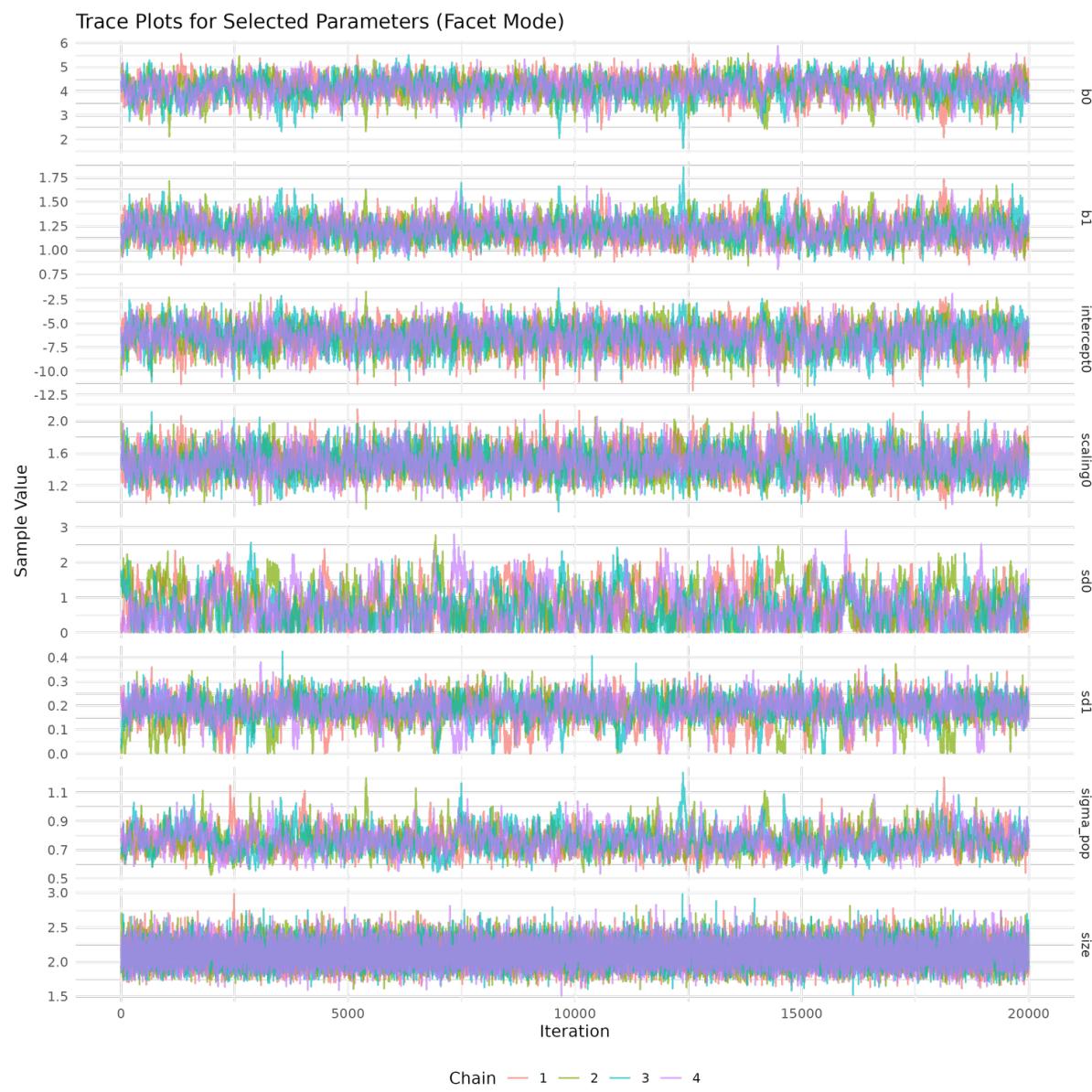


Figure 29: tplots\_filtmonuments\_loglin.png

tplots\_filtmonuments\_loglin\_provinces.png

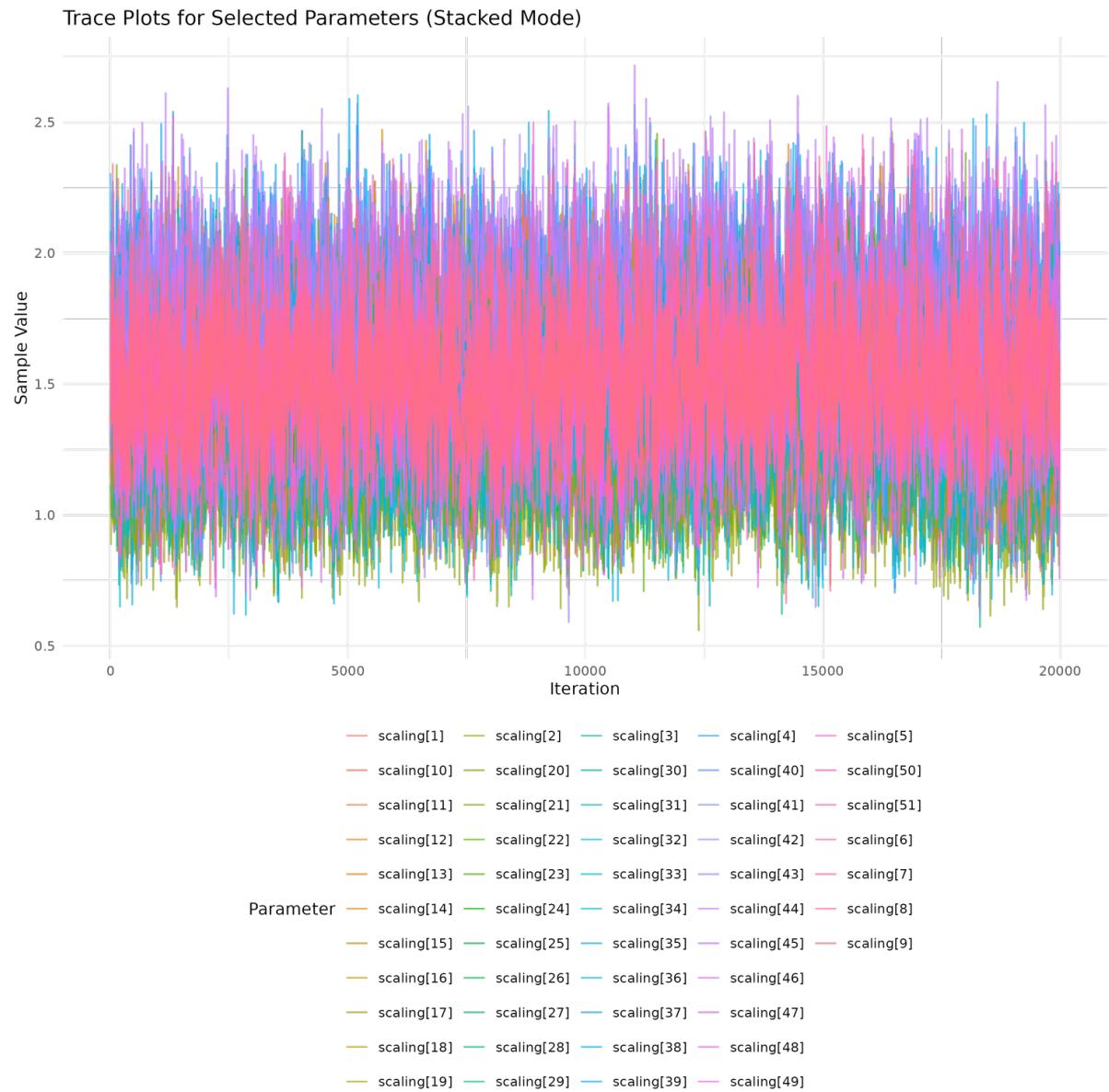


Figure 30: tplots\_filtmonuments\_loglin\_provinces.png

tplots\_hnwi\_linlog.png

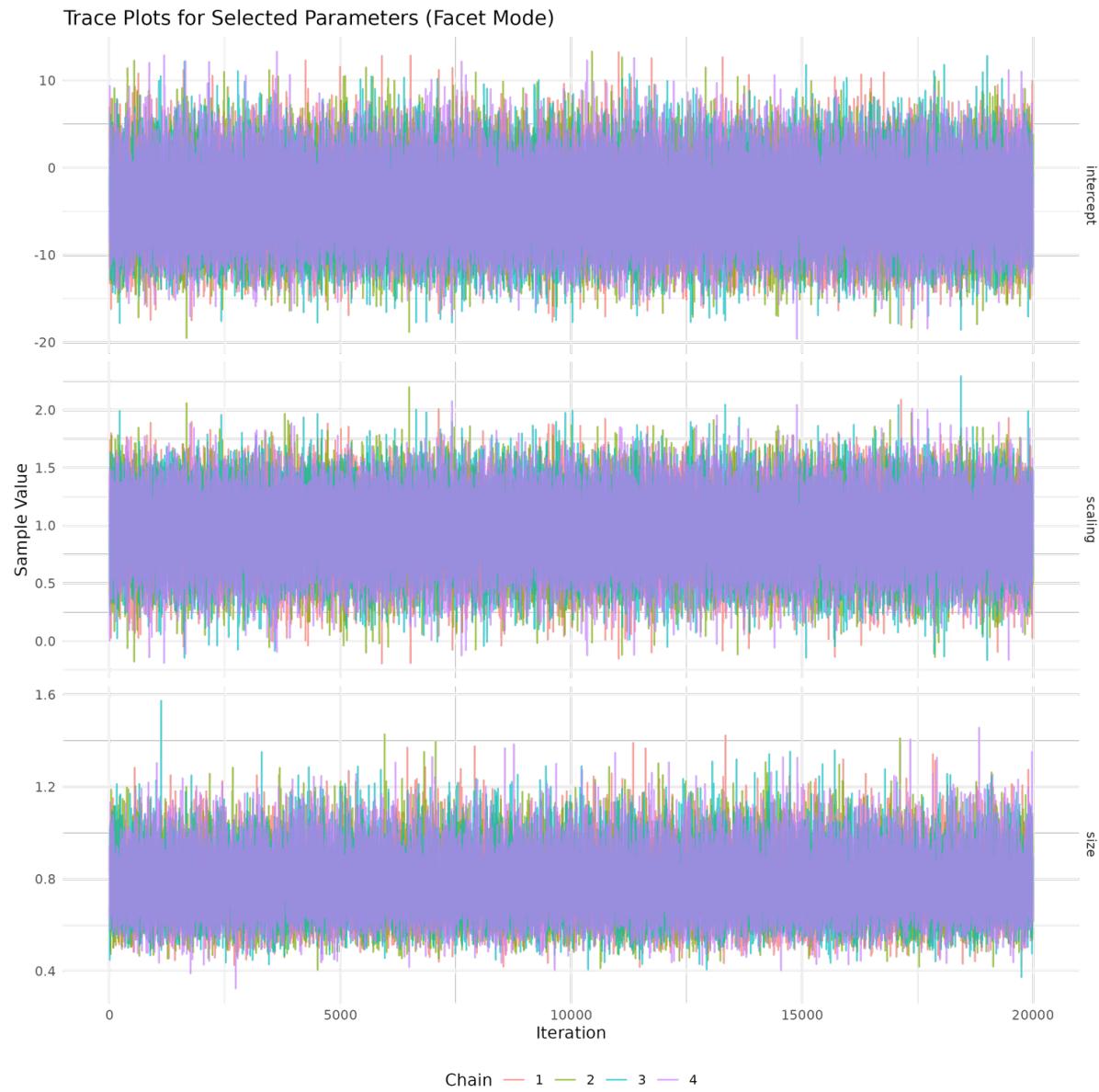


Figure 31: tplots\_hnwi\_linlog.png

## tplots\_hnwi\_sup.png

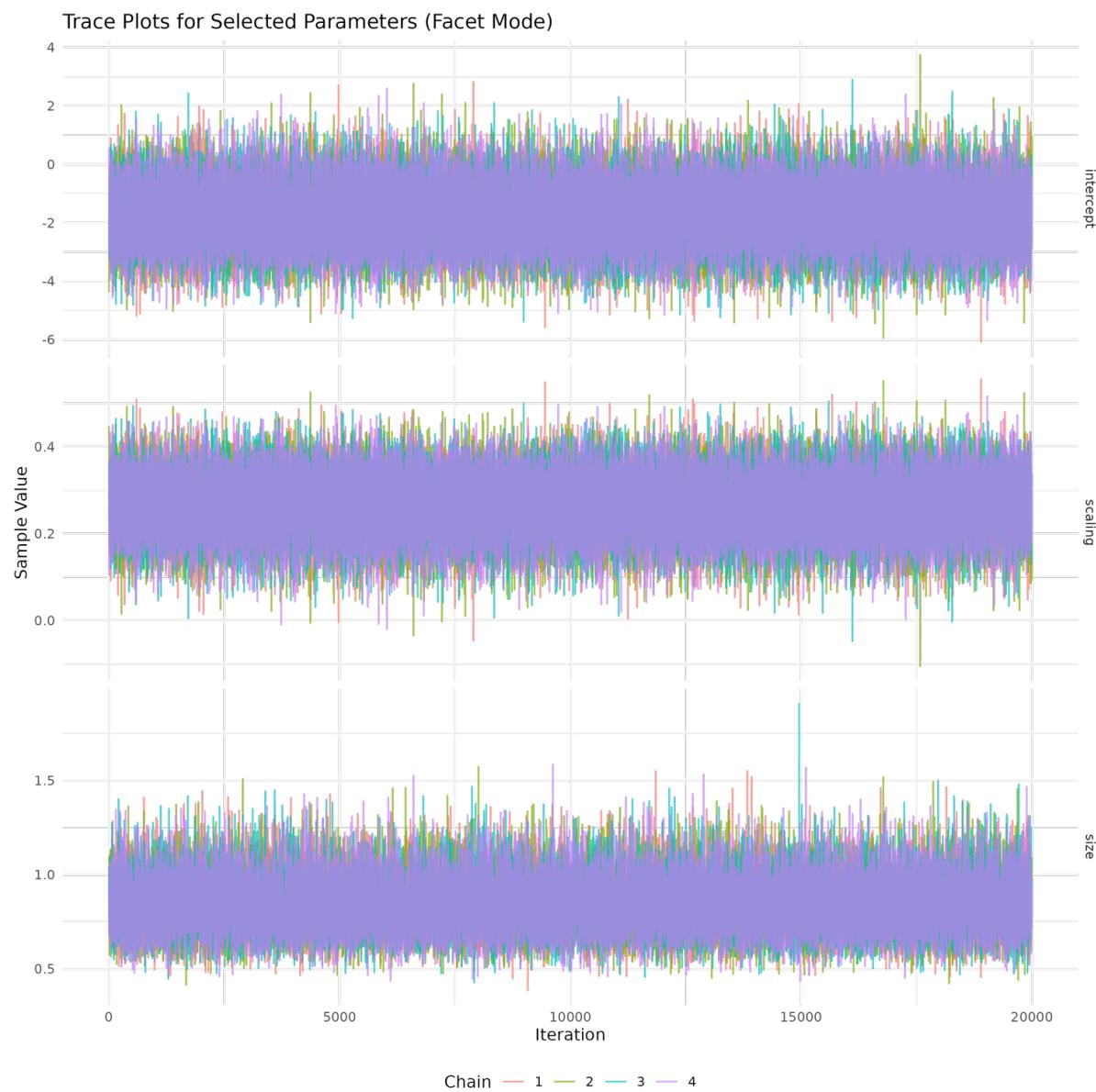


Figure 32: tplots\_hnwi\_sup.png

tplots\_tallbuildings\_linlog.png

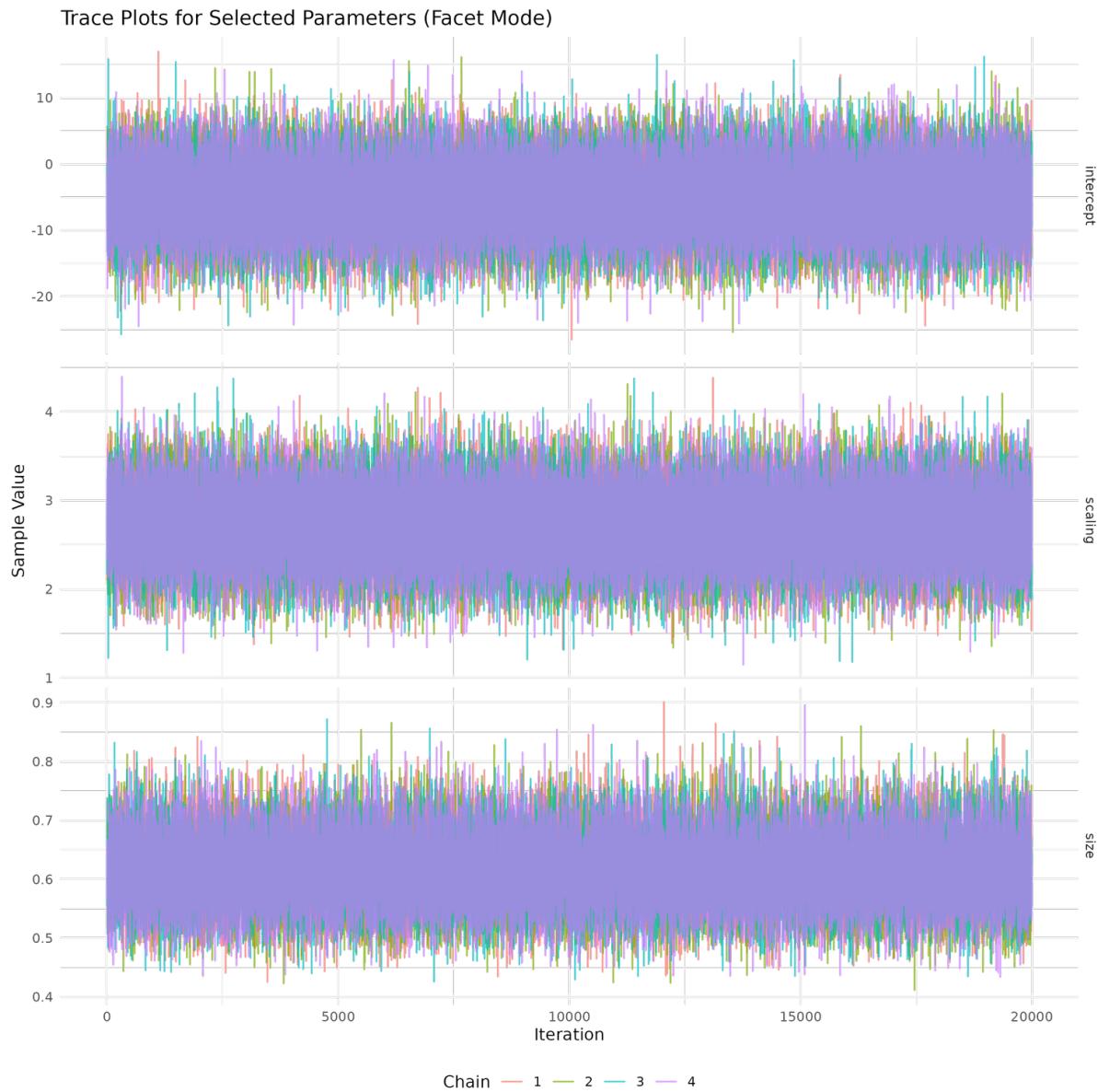


Figure 33: tplots\_tallbuildings\_linlog.png

---

## Analysis Workflow

The following sections describe the analytical steps taken in relation to the section headings in the primary R script (`Src/urban_wealth_scale.R`). Each section below briefly describes the aim of the relevant section of the script along with the data sources, key operations, and outputs.

### 1. Data Wrangling

This section processes data from various sources to prepare it for analysis, including Roman cities, epigraphic records, high-net-worth individuals (HNWI), and tall buildings.

## 1.1 Roman Cities Data

- **Sources:**
  - Population and area data: `Data/Hanson2016_CitiesDatabase_0xREP.xlsx`
  - Population estimates: `Data/population_size.csv`
- **Key Operations:**
  - Merge population estimates with area data.
  - Filter monumental constructions (excluding walls and non-monumental structures).
  - Extract relevant columns: Area, monuments, province, coordinates.
  - Identify wall-determined areas and compute sample size (`n_roman`).

### Outputs:

- Cleaned dataset: `RomanUrban`
  - Columns:
    - `Area`, `Province`, `Longitude`, `Latitude`, `Monuments`, `Monuments_filt`
- 

## 1.2 Epigraphic Data

- **Source:**
  - Epigraphic inscriptions database: <https://zenodo.org/records/4888168>
- **Key Operations:**
  - Extract building/dedicatory inscriptions.
  - Validate and extract geographic coordinates.
  - Calculate distances from cities to inscriptions.
  - Count inscriptions within each city's radius.

### Outputs:

- Updated dataset: `RomanUrban`
    - New column: `InscriptionCount`
- 

## 1.3 HNWI Data

- **Source:**
  - High-net-worth individuals: `Data/hnwi_by_city.xlsx`
  - Population data: `Data/worldcities.csv`
- **Key Operations:**
  - Match population data with HNWI cities.
  - Manually resolve inconsistencies (e.g., “Hong Kong (SAR China)” -> “Hong Kong”).
  - Remove rows with missing population values.
  - Assign zero values to missing billionaire counts.

### Outputs:

- Cleaned dataset: `global_hnwi`
    - Key columns: `City`, `Country_cleaned`, `population`, `Billionaires`
-

## 1.4 Tall Buildings Data

- **Source:**
  - Dataset of tall buildings: `Data/MPI_150m_Dataset.xlsx`
- **Key Operations:**
  - Filter rows with missing population data.

### Outputs:

- Cleaned dataset: `tall_buildings`
- 

### Cleanup

Unnecessary objects are removed to optimize the workspace: - Retained objects: `RomanUrban`, `global_hnwi`, `tall_buildings`, `walls_idx`.

## 2. Defining the Core Bayesian (Nimble) Models

This section implements Bayesian models using the Nimble framework to estimate scaling parameters while accounting for uncertainty in population sizes and count data. The models are defined for both archaeological (Roman) and modern datasets and involve comparisons between power-law and linear-log scaling relationships.

### 2.1 General Approach

- **Objective:** Simultaneously estimate missing population sizes and scaling parameters.
  - **Innovations:**
    - Use of a negative binomial distribution to model count data, addressing overdispersion.
    - Flexible parameters to account for differences across Roman provinces.
  - **MCMC Parameters:**
    - Iterations: `niter = 80000`
    - Burn-in: `nburnin = 20000 (25%)`
    - Thinning: `thin = 3`, resulting in 20000 samples per chain.
    - Chains: `nchains = 4`
- 

### 2.2 Archaeological Models

#### 2.2.1 Power-Law Model

- **Purpose:** Estimate monument scaling parameters based on Roman cities data.
- **Key Features:**
  - Province-level variations in intercept and scaling parameters ( $v$  as province index).
  - Negative binomial distribution to handle count data.
  - Log-log population-area linking.

### Outputs:

- Posterior distributions for parameters:
  - Intercepts: `intercept0`, `intercept_raw[k]`
  - Scaling coefficients: `scaling0`, `scaling_raw[k]`

- Dispersion parameter: `size`
- Population linking coefficients: `b0`, `b1`, `sigma_pop`
- Predictions and fitted values: `y_hat[n]`

### 2.2.2 Linear-Log Model

- **Purpose:** Provide an alternative to the power-law model.
- **Key Features:**
  - Similar to the power-law model but with linear-log population scaling.

#### Outputs:

- Posterior distributions analogous to the power-law model.
- 

## 2.3 Modern Models

### 2.3.1 Power-Law Model

- **Purpose:** Estimate scaling relationships in modern datasets (HNWI, tall buildings).
- **Key Features:**
  - Intercept and scaling parameters on a log-log scale.
  - Negative binomial distribution for count data.

#### Outputs:

- Posterior distributions for parameters:
  - `intercept`
  - `scaling`
  - Dispersion parameter: `size`
- Predictions and fitted values: `y_hat[n]`

### 2.3.2 Linear-Log Model

- **Purpose:** Provide a linear-log alternative for modern datasets.

#### Outputs:

- Posterior distributions analogous to the power-law model.
- 

## 2.4 Model Outputs

All models produce:

- **Posterior distributions** for scaling parameters (`Output/posterior_*`).
- **MCMC Diagnostics:**
  - Trace plots: `Output/mcmc_trace_*`
  - Effective sample size: `Output/effective_sample_size_*`
- **Model Fit Metrics:**
  - WAIC and LOOIC values for comparisons (`Output/waic.csv`, `Output/looic.csv`).

---

## Notes

- The models are structured to allow future extensions or modifications (e.g., adding hierarchical effects).
- Additional thinning is applied during plotting to accommodate memory constraints.

### 3. Utility Functions

This section outlines the custom utility functions developed to assist with the analysis, including visualization, diagnostics, and output generation.

#### 3.1 Visualization Functions

##### `stacked_traceplot`

- **Purpose:** Creates trace plots of MCMC chains with options for faceted or stacked visualizations.
- **Inputs:**
  - `mcmc_obj`: MCMC output object.
  - `params`: Parameters to plot.
  - `mode`: Visualization mode (`facet` or `stacked`).
  - `thin`: Thinning interval for MCMC samples.
- **Output:** A `ggplot2` object displaying MCMC trace plots.

##### `plot_residual_intervals`

- **Purpose:** Visualizes credible intervals for model residuals.
  - **Inputs:**
    - `mcmc_obj`: MCMC output object.
    - `y`: Observed data.
    - `y_hat_param`: Predicted values from the model.
    - `cr_level`: Credible interval level (default 0.95).
    - `outlier_indices`: Highlight outliers (optional).
  - **Output:** A plot with residual intervals and optional outlier annotations.
- 

#### 3.2 Diagnostics Functions

##### `output_geweke`

- **Purpose:** Outputs Geweke diagnostic scores for convergence.
- **Inputs:**
  - `g`: Geweke diagnostic object.
  - `modelname`: Name of the model.
  - `outfolder`: Path to save the output file.
- **Output:** A CSV file with Geweke statistics for model parameters.

##### `output_gr_rhat`

- **Purpose:** Outputs Gelman-Rubin `R_hat` diagnostics.
- **Inputs:**
  - `mcmc_out`: MCMC output object.
  - `modelname`: Name of the model.

- `outfolder`: Path to save the output file.
- **Output:** A CSV file summarizing `R_hat` diagnostics.

#### `output_ess`

- **Purpose:** Calculates and outputs Effective Sample Size (ESS).
  - **Inputs:**
    - `mcmc_out`: MCMC output object.
    - `modelname`: Name of the model.
    - `outfolder`: Path to save the output file.
  - **Output:** A CSV file with ESS values for each parameter.
- 

### 3.3 Summary and Model Evaluation Functions

#### `output_posterior_summary`

- **Purpose:** Outputs posterior summaries (mean, standard deviation, and credible intervals).
- **Inputs:**
  - `mcmc_out`: MCMC output object.
  - `modelname`: Name of the model.
  - `outfolder`: Path to save the output file.
- **Output:** A CSV file summarizing posterior distributions.

#### `rsquared`

- **Purpose:** Calculates pseudo R-squared values for model fit.
- **Inputs:**
  - `y`: Observed values.
  - `y_hat`: Predicted values.
- **Output:** R-squared values.

#### `output_rsquared`

- **Purpose:** Outputs pseudo R-squared values.
  - **Inputs:**
    - `y`: Observed values.
    - `mcmc_out`: MCMC output object.
    - `y_hat_param`: Predicted parameter.
  - **Output:** A CSV file summarizing R-squared values for models.
- 

### 3.4 Model Comparison Functions

#### `output_waic`

- **Purpose:** Outputs WAIC scores for model comparison.
- **Inputs:**
  - `waic`: WAIC object from model fit.
  - `modelname`: Name of the model.
- **Output:** A CSV file with WAIC scores.

### `output_lppd`

- **Purpose:** Outputs log-posterior predictive density (LPPD) values.
  - **Inputs:**
    - `waic_dets`: Detailed WAIC results.
    - `modelname`: Name of the model.
  - **Output:** A CSV file summarizing LPPD values.
- 

## 3.5 Scaling Parameter Functions

### `output_scaling`

- **Purpose:** Extracts scaling parameter samples and saves them in a long format for plotting.
- **Inputs:**
  - `mcmc_out`: MCMC output object.
  - `modelname`: Name of the model.
  - `parameter`: Target scaling parameter.
- **Output:** A CSV file with scaling parameter samples.

## 4. Core Analysis Wrapper

The `run_scaling_analysis` function orchestrates the analysis, from model building to output generation. This section outlines the steps involved, along with the file-naming conventions for each output type, to guide users in locating relevant results within the repository.

---

### 4.1 Overview

The wrapper function performs the following tasks: 1. Constructs and compiles the Bayesian model. 2. Configures and runs the MCMC process. 3. Generates diagnostics and summaries. 4. Saves results and visualizations to the `Output` directory.

---

### 4.2 Key Analytical Steps and Output Patterns

#### 4.2.1 Model Construction and Compilation

- **Steps:**
    - Defines the Nimble model using `scalingCode`.
    - Configures MCMC sampling, including special samplers for correlated parameters.
  - **Outputs:**
    - Intermediate model objects (`compiled_model`, `compiled_mcmc`), not saved to files.
- 

#### 4.2.2 Running MCMC

- **Steps:**
  - Runs the MCMC chains with specified iterations, burn-in, and thinning.
  - Monitors parameters of interest and computes WAIC along with other diagnostic stats.

- **Output Patterns:**
    - **WAIC Scores:** Output/waic\_<modelname>.csv
    - **Pointwise LPPD (Log Posterior Predictive Density):**
      - \* Details: Output/loo\_pw\_<modelname>.csv
      - \* Estimates: Output/loo\_est\_<modelname>.csv
    - **Outliers with Pareto k Diagnostics:** Output/city\_outliers\_<modelname>.csv
- 

#### 4.2.3 Visual Diagnostics

- **Steps:**
    - Generates trace plots for monitored parameters.
    - Optionally, produces province-specific trace plots for archaeological models.
  - **Output Patterns:**
    - **Trace Plots:**
      - \* General: Output/tplots\_<modelname>.pdf and Output/tplots\_<modelname>.png
      - \* Province-Specific: Output/tplots\_<modelname>\_provinces.pdf and Output/tplots\_<modelname>\_provinces.png
- 

#### 4.2.4 Convergence and Posterior Summaries

- **Steps:**
    - Computes convergence diagnostics (Geweke, Gelman-Rubin R\_hat).
    - Extracts posterior summaries, including credible intervals and means.
  - **Output Patterns:**
    - **Geweke Diagnostic:** Output/geweke\_<modelname>.csv
    - **Gelman-Rubin R\_hat:** Output/grrhat\_<modelname>.csv
    - **Posterior Summaries:** Output/post\_summary\_<modelname>.csv
- 

#### 4.2.5 Model Fit and Residual Diagnostics

- **Steps:**
    - Calculates pseudo R-squared for model fit.
    - Generates residuals vs. predicted plots with credible intervals.
  - **Output Patterns:**
    - **Pseudo R-Squared:** Output/rsquared\_<modelname>.csv
    - **Residual Diagnostics:**
      - \* Plot: Output/resid\_<modelname>.pdf and Output/resid\_<modelname>.png
- 

#### 4.2.6 Scaling Parameter Samples

- **Steps:**
    - Extracts MCMC samples for scaling parameters for later cross-model comparisons.
  - **Output Pattern:**
    - Output/scaling\_samples.csv
-

### 4.3 Additional Notes

- **Custom Cleanup:**
  - Objects not required for subsequent analyses are removed to optimize memory usage.
- **Parallel Execution:**
  - The function supports running multiple chains in parallel for computational efficiency. This allowed us to run multiple shorter chains, but the complexity of the model means that the Geweke statistic was less useful for determining convergence and we relied, as a result, more heavily on  $\hat{R}$ —this only made a difference in the epigraphic linear-log model where the individual chains contained more autocorrelation for the `sd_0` parameter.

This function ensures all relevant outputs are systematically saved, enabling reproducibility and detailed exploration of results.

## 5. Scatter Plots of the Data

This section visualizes the data analyzed in the study through a series of scatter plots. Three types of plots are produced: log-log, linear, and linear-log, with offsets applied where necessary to handle zero values.

---

### 5.1 Log-Log Scatter Plots

#### Description:

- Plots log-transformed values of key variables to explore scaling relationships.
- Offset of +1 applied to avoid issues with zero values in logarithmic transformations while still illustrating the basic patterns in the data.

#### Included Variables:

- **RomanUrban:**
  - Area vs. Monuments
  - Area vs. Monuments (filtered)
  - Area vs. InscriptionCount
- **global\_hnwi:**
  - Population vs. Billionaires
- **tall\_buildings:**
  - Population vs. 150 m+ Buildings

#### Output File Naming Pattern:

- Combined plots saved as:
    - PDF: `Output/point_scatters.pdf`
    - PNG: `Output/point_scatters.png`
- 

### 5.2 Linear Scatter Plots

### Description:

- Plots raw (untransformed) values of the same variables.
- Useful for identifying linear trends or deviations.

### Included Variables:

- **RomanUrban:**
  - Area vs. Monuments
  - Area vs. Monuments (filtered)
  - Area vs. InscriptionCount
- **global\_hnwi:**
  - Population vs. Billionaires
- **tall\_buildings:**
  - Population vs. 150 m+ Buildings

### Output File Naming Pattern:

- Combined plots saved as:
    - PDF: Output/point\_scatters\_linear.pdf
    - PNG: Output/point\_scatters\_linear.png
- 

## 5.3 Linear-Log Scatter Plots

### Description:

- Plots raw values for the x-axis and log-transformed values for the y-axis.
- Offset of +1 applied to y-axis values to handle zeros.

### Included Variables:

- **RomanUrban:**
  - Area vs. log(Monuments + 1)
  - Area vs. log(Monuments (filtered) + 1)
  - Area vs. log(InscriptionCount + 1)
- **global\_hnwi:**
  - Population vs. log(Billionaires + 1)
- **tall\_buildings:**
  - Population vs. log(150 m+ Buildings + 1)

### Output File Naming Pattern:

- Combined plots saved as:
    - PDF: Output/point\_scatters\_linear\_log.pdf
    - PNG: Output/point\_scatters\_linear\_log.png
- 

These scatter plots provide a comprehensive visual overview of the data relationships and scaling patterns explored in the study.

## 6. First Analysis: All Monuments

This section presents the analysis of the relationship between urban area and the number of monuments across all Roman cities. The analysis is conducted twice: 1. Using the full dataset. 2. Excluding cities with zero monuments.

---

### 6.1 Analysis with Full Dataset

#### Description:

- Examines the scaling relationship between urban area and monument count using all available data.

#### Model Parameters:

- **Name:** `allmonuments`
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Number of monuments (`Monuments`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

#### Constants and Data:

- `N`: Number of cities in the dataset.
- `K`: Number of unique provinces.
- `v`: Province indices.
- `y`: Monument counts.
- `x`: Log-transformed area.
- `pop`: Log-transformed population estimates.

#### Initialization:

- Parameters initialized to plausible starting values (e.g., scaling coefficients, intercepts).

#### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_allmonuments.csv`
    - Posterior Summaries: `Output/post_summary_allmonuments.csv`
    - Residuals: `Output/resid_allmonuments.pdf` and `Output/resid_allmonuments.png`
    - Trace Plots: `Output/tplots_allmonuments.pdf` and `Output/tplots_allmonuments.png`
- 

### 6.2 Analysis Excluding Cities with Zero Monuments

#### Description:

- Repeats the analysis excluding cities with zero reported monuments to check the robustness of the scaling relationship.

### Model Parameters:

- **Name:** allmonuments\_nozero
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Number of monuments (`Monuments > 0`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

### Constants and Data:

- Similar to the full dataset analysis but using only cities where `Monuments > 0`.

### Initialization:

- Same initialization values as the full dataset analysis.

### Outputs:

- WAIC and posterior summaries for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_allmonuments_nozero.csv`
    - Posterior Summaries: `Output/post_summary_allmonuments_nozero.csv`
    - Trace Plots: `Output/tplots_allmonuments_nozero.pdf`
    - and `Output/tplots_allmonuments_nozero.png`
- 

## 7. Second Analysis: Walled Cities Only

This section analyzes the relationship between urban area and the number of monuments, focusing exclusively on cities where urban areas are defined by walls. As with the first analysis, this is conducted twice: 1. Using all walled cities. 2. Excluding cities with zero monuments.

---

### 7.1 Analysis with All Walled Cities

#### Description:

- Examines scaling relationships for cities where walls were used to define urban areas.
- Helps to explore whether walled cities exhibit different scaling dynamics compared to the full dataset and accounts for the possible bias in using walled areas (or less discrete archaeological indicators of site size) to estimate the population-area relationships.

### Model Parameters:

- **Name:** allwalls
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Number of monuments (`Monuments`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

### Constants and Data:

- $N$ : Number of walled cities in the dataset.
- $K$ : Number of unique provinces.
- $v$ : Province indices.
- $y$ : Monument counts.
- $x$ : Log-transformed area.
- $\text{pop}$ : Log-transformed population estimates.

### Initialization:

- Parameters initialized with values identical to the first analysis.

### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_allwalls.csv`
    - Posterior Summaries: `Output/post_summary_allwalls.csv`
    - Residuals: `Output/resid_allwalls.pdf` and `Output/resid_allwalls.png`
    - Trace Plots: `Output/tplots_allwalls.pdf` and `Output/tplots_allwalls.png`
- 

## 7.2 Analysis Excluding Cities with Zero Monuments

### Description:

- Focuses on walled cities with at least one monument, removing cities with zero reported monuments.

### Model Parameters:

- Name: `allwalls_nozero`
- Type: Power-law model
- Inputs:
  - **Dependent Variable**: Number of monuments (`Monuments > 0`).
  - **Predictor**: Log-transformed area (`log(Area)`).
  - **Population Proxy**: Log-transformed population estimates (`log(pop_est)`).
  - **Province Index**: Integer-encoded province identifiers.

### Constants and Data:

- Similar to the first analysis but filtered to include only cities where `Monuments > 0`.

### Initialization:

- Parameters initialized with the same values as the full walled city analysis.

### Outputs:

- WAIC and posterior summaries for the scaling model.
- Output file naming pattern:
  - WAIC: `Output/waic_allwalls_nozero.csv`
  - Posterior Summaries: `Output/post_summary_allwalls_nozero.csv`

- Trace Plots: `Output/tplots_allwalls_nozero.pdf` and `Output/tplots_allwalls_nozero.png`
- 

### 7.3 Notes

- Walled cities provide a subset of the data that may exhibit distinct scaling properties due to their architectural and social significance.
- By excluding zeros, the robustness of scaling patterns is further examined within this subset.
- As before, all outputs are saved in the `Output` directory, with consistent naming conventions.

This analysis delves deeper into the dynamics of monument scaling, focusing on a distinctive subset of urban settlements.

## 8. Third Analysis: Above Ground Monuments Only

This section focuses on monuments explicitly identified as “above ground” in the dataset, excluding features like walls, forums, and other non-symbolic constructions. The analysis is conducted twice: 1. Using all cities. 2. Excluding cities with zero filtered monuments.

---

### 8.1 Analysis with All Cities

#### Description:

- Explores scaling relationships using a filtered count of symbolic monuments (`Monuments_filt`), focusing solely on above-ground, non-defensive structures.

#### Model Parameters:

- Name: `filtmonuments`
- Type: Power-law model
- Inputs:
  - **Dependent Variable:** Number of symbolic monuments (`Monuments_filt`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

#### Constants and Data:

- N: Number of cities in the dataset.
- K: Number of unique provinces.
- v: Province indices.
- y: Filtered monument counts.
- x: Log-transformed area.
- pop: Log-transformed population estimates.

#### Initialization:

- Identical to the first two analyses.

### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_filtmonuments.csv`
    - Posterior Summaries: `Output/post_summary_filtmonuments.csv`
    - Residuals: `Output/resid_filtmonuments.pdf` and `Output/resid_filtmonuments.png`
    - Trace Plots: `Output/tplots_filtmonuments.pdf` and `Output/tplots_filtmonuments.png`
- 

## 8.2 Analysis of Visible Above-Ground Monuments

### Description:

- Re-examines the scaling relationship after filtering the monument data (resulting in adjusted counts per city) to exclude cases from the monunents database without significant above-ground expression (e.g., city grid, roads, fora, etc.).

### Model Parameters:

- Name: `filtmonuments_nozero`
- Type: Power-law model
- Inputs:
  - **Dependent Variable:** Number of symbolic monuments (`Monuments_filt > 0`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

### Constants and Data:

- Similar to the full dataset analysis but filtered to exclude zeros.

### Initialization:

- Same as the previous analyses.

### Outputs:

- WAIC and posterior summaries for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_filtmonuments_nozero.csv`
    - Posterior Summaries: `Output/post_summary_filtmonuments_nozero.csv`
    - Trace Plots: `Output/tplots_filtmonuments_nozero.pdf` and `Output/tplots_filtmonuments_nozero.png`
- 

## 8.3 Notes

This analysis provides additional insights into scaling relationships specific to symbolic, above-ground monuments.

## 9. Fourth Analysis: Epigraphy

This section investigates the scaling relationships for inscriptions (epigraphy) found in or near cities. The analysis is conducted twice: 1. Using all cities with epigraphic data. 2. Excluding cities with zero recorded inscriptions.

---

### 9.1 Analysis with All Cities

#### Description:

- Examines scaling relationships using the count of inscriptions (`InscriptionCount`).
- Inscriptions are treated as a proxy for cultural or symbolic activity within cities.

#### Model Parameters:

- **Name:** `epigraphy`
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Number of inscriptions (`InscriptionCount`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

#### Constants and Data:

- `N`: Number of cities in the dataset.
- `K`: Number of unique provinces.
- `v`: Province indices.
- `y`: Inscription counts.
- `x`: Log-transformed area.
- `pop`: Log-transformed population estimates.

#### Initialization:

- Identical to the earlier analyses.

#### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_epigraphy.csv`
    - Posterior Summaries: `Output/post_summary_epigraphy.csv`
    - Residuals: `Output/resid_epigraphy.pdf` and `Output/resid_epigraphy.png`
    - Trace Plots: `Output/tplots_epigraphy.pdf` and `Output/tplots_epigraphy.png`
- 

### 9.2 Analysis Excluding Cities with Zero Inscriptions

#### Description:

- Re-analyzes scaling relationships for cities with at least one recorded inscription.

### Model Parameters:

- **Name:** `epigraphy_nozero`
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Number of inscriptions (`InscriptionCount > 0`).
  - **Predictor:** Log-transformed area (`log(Area)`).
  - **Population Proxy:** Log-transformed population estimates (`log(pop_est)`).
  - **Province Index:** Integer-encoded province identifiers.

### Constants and Data:

- Similar to the full dataset analysis but filtered to exclude cities with zero inscriptions.

### Initialization:

- Same as the previous analyses.

### Outputs:

- WAIC and posterior summaries for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_epigraphy_nozero.csv`
    - Posterior Summaries: `Output/post_summary_epigraphy_nozero.csv`
    - Trace Plots: `Output/tplots_epigraphy_nozero.pdf`
    - and `Output/tplots_epigraphy_nozero.png`
- 

### 9.3 Notes

- This analysis provides insights into the relationship between city size and epigraphic activity specifically tied to monument dedications (the product of euregetism).
- Excluding zeros ensures that scaling relationships are robust and not influenced by cities lacking any recorded inscriptions (which are anyway highly unlikely to represent reality as explained in the main text).

This analysis provides a robustness check on using monuments of unknown sponsorship or funding.

## 10. Fifth Analysis: High Net Worth Individuals (HNWI)

This section examines the scaling relationship between city population and the count of billionaires residing in cities. The analysis uses data for modern cities with available high net worth individual (HNWI) information.

---

### 10.1 Analysis Description

#### Description:

- Explores the scaling of billionaires (HNWI) with city population size.
- This analysis focuses exclusively on modern cities with population and billionaire count data.

### Model Parameters:

- **Name:** hnwi
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Count of billionaires in each city (`Billionaires`).
  - **Predictor:** Log-transformed population size (`log(population)`).

### Constants and Data:

- N: Number of cities in the dataset.
- y: Count of billionaires.
- pop: Log-transformed population size.

### Initialization:

- **Scaling Parameter (scaling):** Initialized to 0.
- **Intercept Parameter (intercept):** Initialized to 0.
- **Dispersion Parameter (size):** Initialized to 1.

### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: `Output/waic_hnwi.csv`
    - Posterior Summaries: `Output/post_summary_hnwi.csv`
    - Residuals: `Output/resid_hnwi.pdf` and `Output/resid_hnwi.png`
    - Trace Plots: `Output/tplots_hnwi.pdf` and `Output/tplots_hnwi.png`
- 

## 10.2 Notes

- This analysis provides insights into the distribution of extreme wealth relative to urban population size in modern cities.
- The results are critical for understanding wealth concentration patterns in contemporary urban systems.

## 11. Sixth Analysis: Tall Buildings

This section investigates the scaling relationship between city population and the number of skyscrapers (buildings taller than 150 meters). Skyscrapers are often proxies for urban wealth and engineering capacity, symbolizing status and economic power.

---

### 11.1 Analysis Description

#### Description:

- Explores the scaling of tall buildings (`150 m+ Buildings`) with city population size.
- This analysis uses a modern dataset exclusively focused on cities with recorded skyscrapers.

### Model Parameters:

- **Name:** tallbuildings
- **Type:** Power-law model
- **Inputs:**
  - **Dependent Variable:** Count of skyscrapers taller than 150 meters (150 m+ Buildings).
  - **Predictor:** Log-transformed population size ( $\log(\text{Population})$ ).

### Constants and Data:

- N: Number of cities in the dataset.
- y: Count of skyscrapers.
- pop: Log-transformed population size.

### Initialization:

- **Scaling Parameter (scaling):** Initialized to 0.
- **Intercept Parameter (intercept):** Initialized to 0.
- **Dispersion Parameter (size):** Initialized to 1.

### Outputs:

- WAIC, LPPD, posterior summaries, and diagnostics for the scaling model.
  - Output file naming pattern:
    - WAIC: Output/waic\_tallbuildings.csv
    - Posterior Summaries: Output/post\_summary\_tallbuildings.csv
    - Residuals: Output/resid\_tallbuildings.pdf and Output/resid\_tallbuildings.png
    - Trace Plots: Output/tplots\_tallbuildings.pdf and Output/tplots\_tallbuildings.png
- 

## 11.2 Notes

- This analysis connects the physical manifestation of wealth concentration (skyscrapers) with urban population scaling.

This analysis complements the broader project by addressing a modern architectural dimension of elite wealth scaling.

## 12. Plot Scaling Parameter Posterior Densities

This section visualizes the posterior densities of the scaling parameter (**scaling**) across all analyses. The plot aggregates results from the different subsets of data analyzed earlier, providing a comparative overview of scaling behavior across diverse urban proxies and contexts.

---

### 12.1 Analysis Description

#### Purpose:

- To compare the posterior distributions of the scaling parameter for all the analyses, highlighting differences or consistencies across datasets and models.

### **Visualization:**

- **Density Plot:** Displays posterior densities for the scaling parameter across analyses.
- **Vertical Reference Lines:**
  - Red dashed lines at  $x = 0.15$  and  $x = 0.85$ , indicating the predicted scaling coefficient values from Settlement Scaling Theory for average per-capita wealth and infrastructure, respectively.

### **Aesthetic Features:**

- **Color Palette:** A high-contrast palette to distinguish between analyses.
  - **Patterns:** Alternating patterns (e.g., solid, striped) to enhance readability in both color and grayscale formats.
- 

## **12.2 Output Description**

### **Plot:**

- File Names:
  - PDF: `Output/scaling_posteriors.pdf`
  - PNG: `Output/scaling_posteriors.png`

### **Inputs:**

- Aggregated scaling parameter samples stored in `Output/scaling_samples.csv`.
  - The `analysis` column distinguishes the dataset/model associated with each sample.
- 

## **13. Collate Main Results Table**

This section aggregates the top-level results from the various analyses into a single summary table for easy reference. The table includes the posterior summaries for the scaling parameter and other relevant metrics like the R-squared value for each analysis.

---

## **13.1 Analysis Description**

### **Purpose:**

- To create a concise summary table (`analysis_results_summary.csv`) that allows readers to quickly understand and compare the key outputs of each analysis. The results are presented in the main paper.

### **Contents:**

- **Columns:**
  - `analysis`: The name of the analysis (e.g., “allmonuments”, “hnwi”).
  - `mean`: The posterior mean of the scaling parameter.
  - `lower`: The lower bound of the 99% highest posterior density (HPD) interval.
  - `upper`: The upper bound of the 99% HPD interval.
  - `std`: The posterior standard deviation.

- `r_squared`: The pseudo R-squared value for the analysis.
  - **Rows:**
    - Each row corresponds to one analysis.
- 

## 13.2 Example Table Structure

Below is an example of the table structure with a few blank rows as placeholders:

analysis	mean	lower	upper	std	r_squared
allmonuments	...	...	...	...	...
hnwi	...	...	...	...	...
tallbuildings	...	...	...	...	...
...	...	...	...	...	...

---

## 13.3 Output Description

**File Name:**

- `Output/analysis_results_summary.csv`
- 

## 13.4 Notes

- The `analysis` column provides a direct link to the corresponding subset of data and analysis.
- Readers can use this table to identify high-level patterns or investigate specific analyses in greater detail by referring to the outputs from earlier steps.
- This summary table is a crucial component for interpreting the results presented in the main text and supplementary materials.

## Supplemental Analyses Workflow

This section describes the supplemental analyses conducted to compare model fits and examine outlier effects. It includes both linear-log model comparison and outlier robustness checks.

---

### 1. Model Comparison

**Purpose:**

- Compare the fit of the power-law model against a linear-log alternative for various datasets.

### Steps:

#### 1. Prepare Data:

- For each dataset (`allmonuments`, `allwalls`, `filtmonuments`, `epigraphy`, `hnwi`, and `tallbuildings`), prepare the inputs: response variable (`y`), predictor (`x` or `pop`), and constants (`Consts`).

#### 2. Run Linear-Log Models:

- Use `scalingCode_linlog` and `run_scaling_analysis()` to analyze each dataset.

### Output File Patterns:

- Posterior summaries: `Output/post_summary_<modelname>.linlog.csv`
- WAIC and LOOIC values: `Output/waic.csv` (merged with `Output/loo_est_<modelname>.csv`)
- Scaling parameter samples: `Output/scaling_samples.csv`

### Analysis Mapping:

Analysis	Model Name	Key Output Files
All Monuments	<code>allmonuments_linlog</code>	<code>post_summary_allmonuments_linlog.csv</code>
Walled Only	<code>allwalls_linlog</code>	<code>post_summary_allwalls_linlog.csv</code>
Above Ground	<code>filtmonuments_loglin</code>	<code>post_summary_filtmonuments_loglin.csv</code>
Epigraphy	<code>epigraphy_linlog</code>	<code>post_summary_epigraphy_linlog.csv</code>
HNWI	<code>hnwi_linlog</code>	<code>post_summary_hnwi_linlog.csv</code>
Tall Buildings	<code>tallbuildings_linlog</code>	<code>post_summary_tallbuildings_linlog.csv</code>

## 2. WAIC/LOOIC Summary

### Purpose:

- Aggregate and compare WAIC and LOOIC values for model evaluation.

### Output File:

- `Output/waic_looic_csv`: Combines WAIC and LOOIC values into a single summary table.

## 3. Outlier Effects

### Purpose:

- Assess the impact of extreme outliers on scaling parameter estimates for the primary datasets.

### Steps:

1. Identify outlier cities using `Output/city_outliers_<modelname>.csv`.
2. Exclude outliers from the datasets and rerun the analyses with the power-law model.

### **Output File Patterns:**

- **Posterior summaries:** Output/post\_summary\_<modelname>.sup.csv

### **Analysis Mapping:**

Analysis	Model Name	Key Output Files
All Monuments	allmonuments_sup	post_summary_allmonuments_sup.csv
HNWI	hnwi_sup	post_summary_hnwi_sup.csv
Tall Buildings	tallbuildings_sup	post_summary_tallbuildings_sup.csv

---

### **Supplemental Notes**

- **File Organization:** All outputs are stored in the Output/ subdirectory.
- **Scaling Parameter Summary:** The file Output/scaling\_samples.csv aggregates scaling parameter samples across analyses for posterior density plots.
- **Model Comparison Summary:** The file Output/waic\_looic\_csv includes WAIC and LOOIC values for both power-law and linear-log models.

### **MCMC Convergence Diagnostic Summary Plots**

This section contains code for plotting summary distributions of two standard convergence diagnostics for the MCMC posterior chains. One is the Gelman-Rubin  $\hat{R}$  statistic and the other is the Geweke z-score diagnostic.

#### **Purpose:**

- Aggregate and visualize standard MCMC parameter chain convergence diagnostics across all models and parameters.
- Provides a high-level, analysis-wide overview of MCMC convergence.

#### **Output File:**

- Output/geweke\_histogram.png: A histogram of Geweke z-scores across the entire analysis.
- Output/grrhat\_histogram.png: A histogram of Gelman-Rubin  $\hat{R}$  values across the entire analysis.