Model Calibration and The Bootstrap

BEE 6940 Lecture 5

February 20, 2023

Table of Contents

Uncertainty Quantification and Model Calibration

What is "Uncertainty Quantification"?

Uncertainty quantification (UQ) is...

> *the full specification of likelihoods as well as distributional forms necessary to infer the joint probabilistic response across all modeled factors of interest...*
>
> *— Cooke, Experts in Uncertainty: Opinion and Subjective Probability in Science (1991)*

UQ vs. Uncertainty Characterization

By contrast, **uncertainty characterization** involves mapping how alternative representations of the stressors and form and function of modeled systems influence outcomes of interest.

UQ vs. Uncertainty Characterization

By contrast, **uncertainty characterization** involves mapping how alternative representations of the stressors and form and function of modeled systems influence outcomes of interest.

In short:

- UQ emphasizes probabilistic inference and projection,
- UC explores impact of alternative hypotheses and representations.

UQ vs. Sensitivity Analysis

Sensitivity analysis involves analyzing how uncertainty in the output is influenced by different sources of uncertainty in the input(s).

UQ vs. Sensitivity Analysis

Sensitivity analysis involves analyzing how uncertainty in the output is influenced by different sources of uncertainty in the input(s).

In short:

- UQ looks at the probabilistic representation of input(s) and output uncertainty;
- SA looks at the impact of variability in input(s) on outputs (for a number of purposes).

When is UQ Appropriate?

- Can specify a probability model to generate and/or draw inferences about data;
- Have sufficient data to constrain model adequately.
- Often need a model which is computationally inexpensive.

Model Calibration

A common UQ task is **model calibration**:

> *the selection of model parameters and structures to maximize the fidelity of the system model to observational data given model and computational constraints as calibration.*
>
> — *Oreskes et al (1994)*

We can calibrate *statistical* models or numerical *simulation* models.

Calibrating Simulation Models

Establishing some notation:

Let $F(\mathbf{x};\theta)$ be the simulation model:

- $\mathbf{x}$ are the "control variables";
- θ are the "calibration variables".

Data-Model Discrepancy

The other important consideration is **data-model discrepancy**.

If $\mathbf{y}$ are the "observations,"" we can model these as:

$$\mathbf{y} = z(\mathbf{x}) + \varepsilon,$$

where

- $z(\mathbf{x})$ is the "true" system state at control variable $\mathbf{x}$;
- ε are observation errors.

Data-Model Discrepancy

Then the *discrepancy* ζ between the simulation and the modeled system is:

$$\zeta(\mathbf{x};\theta) = z(\mathbf{x}) - F(\mathbf{x};\theta).$$

Data-Model Discrepancy

Then the *discrepancy* ζ between the simulation and the modeled system is:

$$\zeta(\mathbf{x};\theta) = z(\mathbf{x}) - F(\mathbf{x};\theta).$$

This implies that the "observations" can be written as

$$\mathbf{y} = F(\mathbf{x};\theta) + \zeta(\mathbf{x};\theta) + \varepsilon.$$

Model Calibration and Discrepancy

Model calibration then involves specifying a probability model for the data, which can include:

- Distributions/likelihoods for the calibration parameters;
- A probability model for the discrepancy and observation errors.

What Happens If We Neglect Discrepancy?

Neglecting discrepancy can result in biased inferences and, as a result, projections, as we will see over the next few weeks.

Review of Frequentist Statistics and Sampling Distributions

Frequentist vs. Bayesian UQ

Recall:

- Frequentist statistics assumes that parameters of a given probability model have a "fixed" value corresponding to those statistics.

Frequentist vs. Bayesian UQ

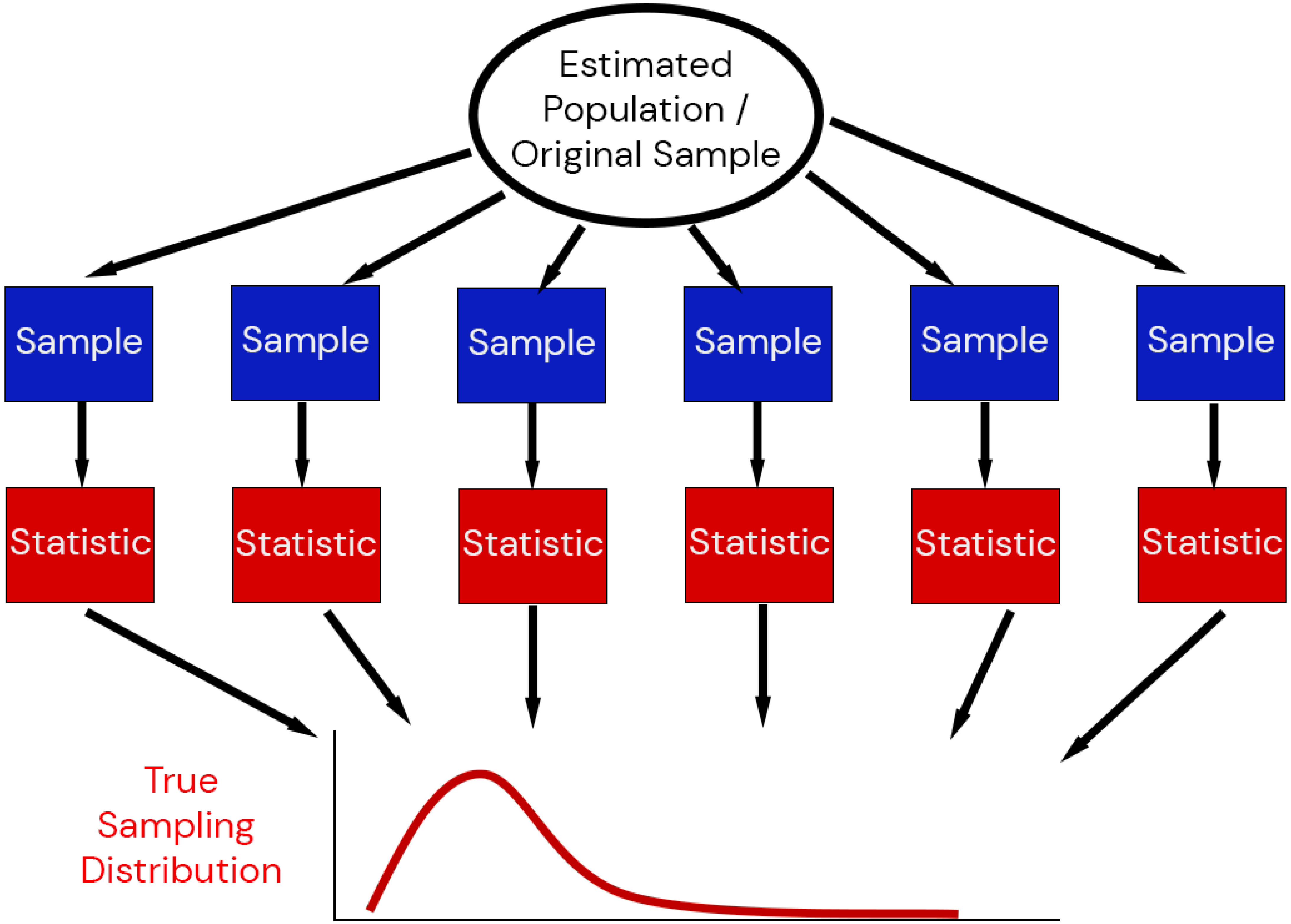
Recall:

- Frequentist statistics assumes that parameters of a given probability model have a "fixed" value corresponding to those statistics.
- Bayesian statistics assigns probabilities to parameters as part of the probability model based on the degree of belief about their consistency with the data (which is also random).

Sampling Distributions

A central concept in frequentist statistics (and to a lesser degree in Bayesian statistics) is the **sampling distribution** of a statistic, which captures the uncertainty associated with random samples.

Sampling Distributions



Frequentist UQ

Most frequentist UQ questions are centered around the sampling distribution, as it captures uncertainty from variations in the sample.

Frequentist UQ

Most frequentist UQ questions are centered around the sampling distribution, as it captures uncertainty from variations in the sample.

But what if we only have access to one sample?

In laboratory settings, can run multiple experiments, but this is not the case for the real world.

"Classical" Approaches

Classical statistical approaches involve relying on asymptotics or a special case distribution:

Assume an approximation to the sampling distribution and analyitcally derive test statistics, confidence intervals, etc.

The Bootstrap Principle

What if we do not want to make these assumptions? Efron (1979) suggested combining estimation with simulation: the **bootstrap**.

The key idea: use the data to define a data-generating mechanism.

Source: Wikipedia

Monte Carlo vs Bootstrapping

Monte Carlo: If we have a *generative* probability model, simulate new samples from the model and estimate the sampling distribution.

Bootstrap: assumes the existing data is representative of the "true" population, and can simulate based on properties of the data itself.

Why Does The Bootstrap Work?

Efron's key insight: due to the Central Limit Theorem, the **differences** between estimates drawn from the sampling distribution and the true value converge to a normal distribution.

- Use the bootstrap to approximate the sampling distribution through re-sampling and re-estimation.
- Can draw asymptotic quantities (bias estimates, confidence intervals, etc) from the differences between the sample estimate and the bootstrap estimates.

What Can We Do With The Bootstrap Sampling Distribution?

Let t_0 the "true" value of a statistic, $\hat{t}$ the estimate of the statistic from the sample, and $(\tilde{t}_i)$ the bootstrap estimates.

- Estimate Variance: $\mathrm{Var}[\hat{t}] \approx \mathrm{Var}[\tilde{t}]$
- Bias Correction: $\mathbb{E}[\hat{t}] - t_0 \approx \mathbb{E}[\tilde{t}] - \hat{t}$
- Compute *basic* α-confidence intervals:

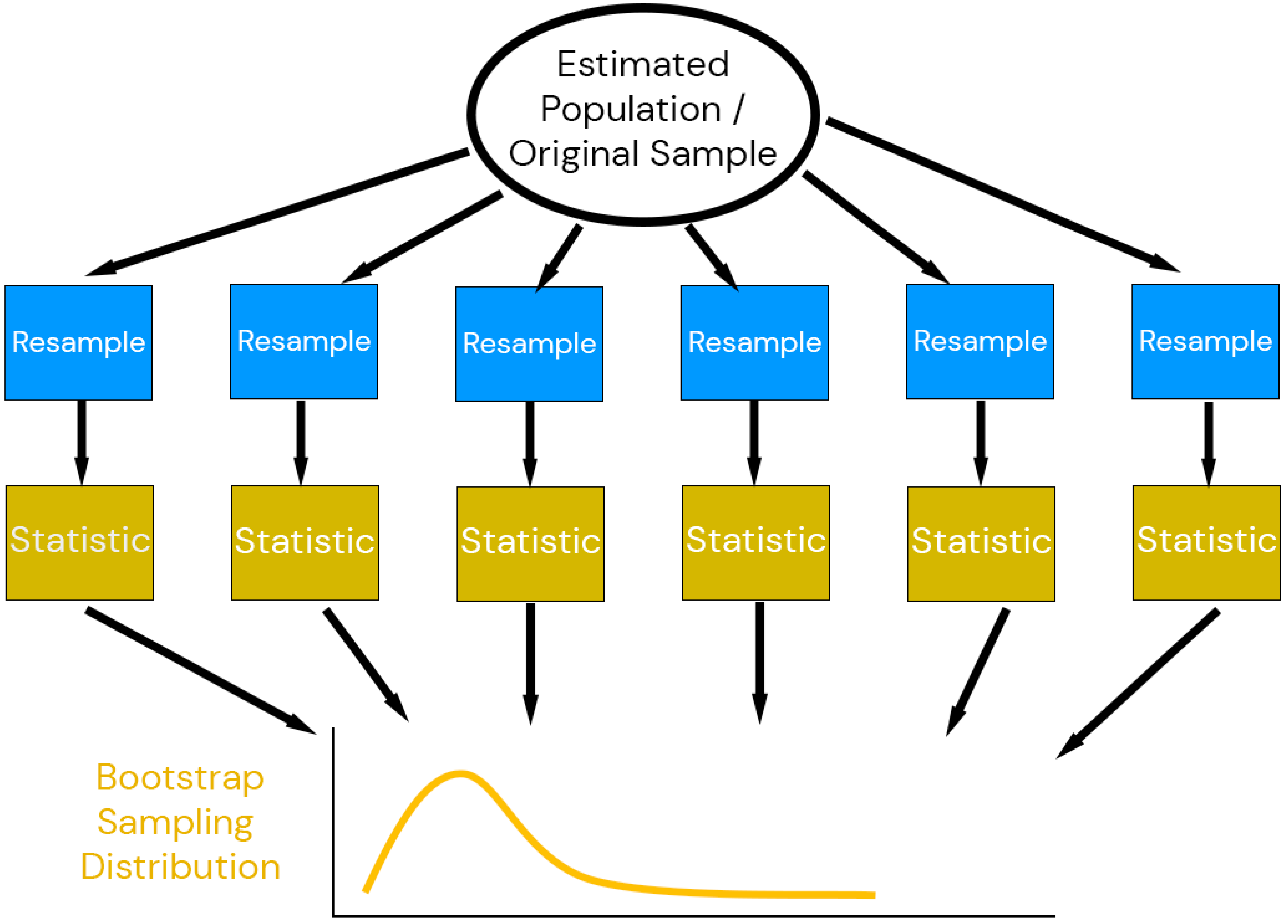
$$\left(\hat{t} - (Q_{\tilde{t}}(1-\alpha/2) - \hat{t}), \hat{t} - (Q_{\tilde{t}}(\alpha/2) - \hat{t})\right)$$

The Non-Parametric Bootstrap

The Non-Parametric Bootstrap

The non-parametric bootstrap is the most "naive" approach to the bootstrap: **resample-then-estimate**.

Non-Parametric Bootstrap Scheme



Simple Example: Is A Coin Fair?

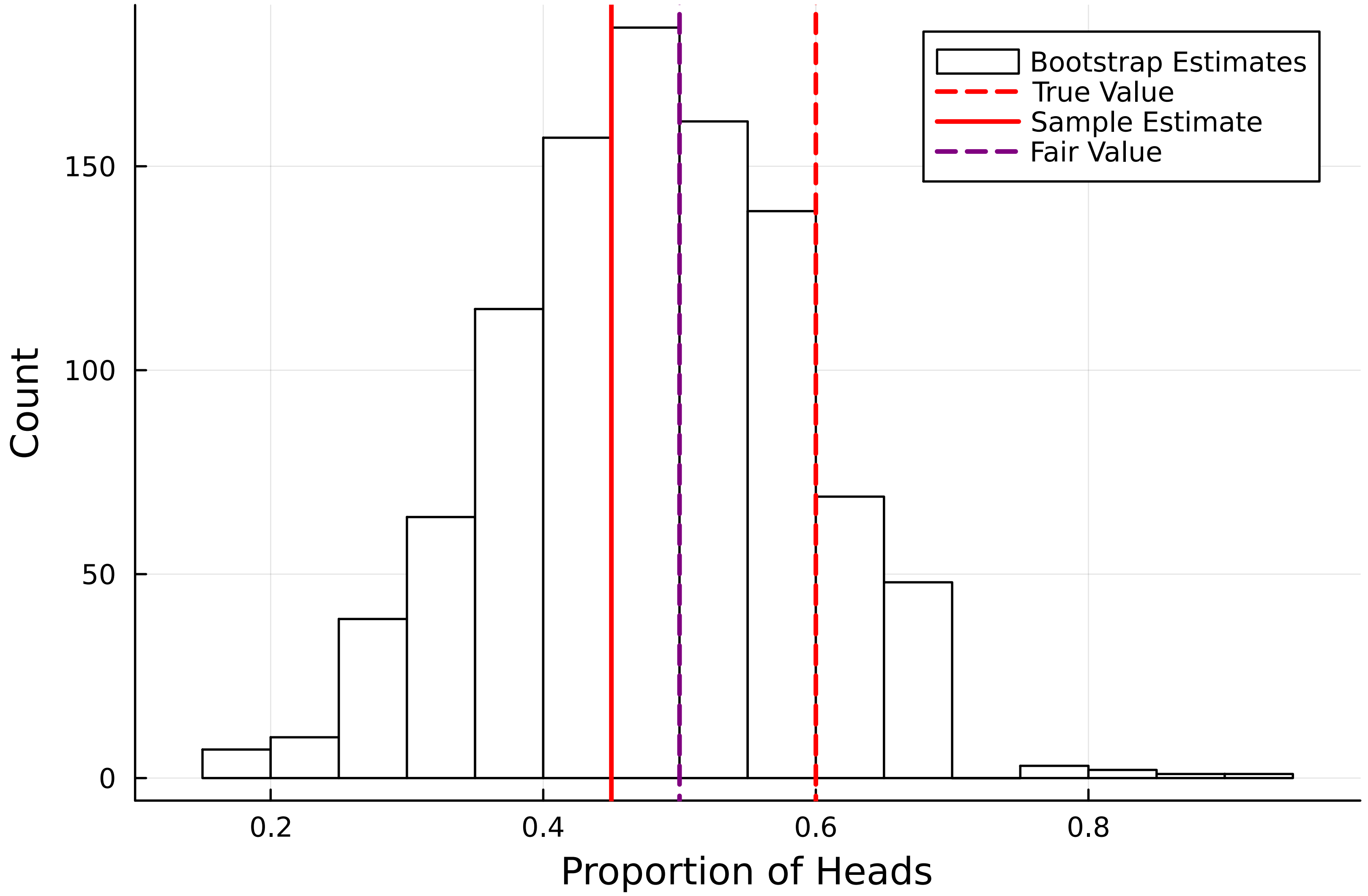
Suppose we have observed twenty flips with a coin, and want to know if it is weighted.

T, H, T, T, T, T, T, H, H, H, H, T, T, H, T, H, H, T, T, H

Based on this sample, the frequency of heads is 45%.

Simple Example: Is A Coin Fair?

If we generate 1000 samples:

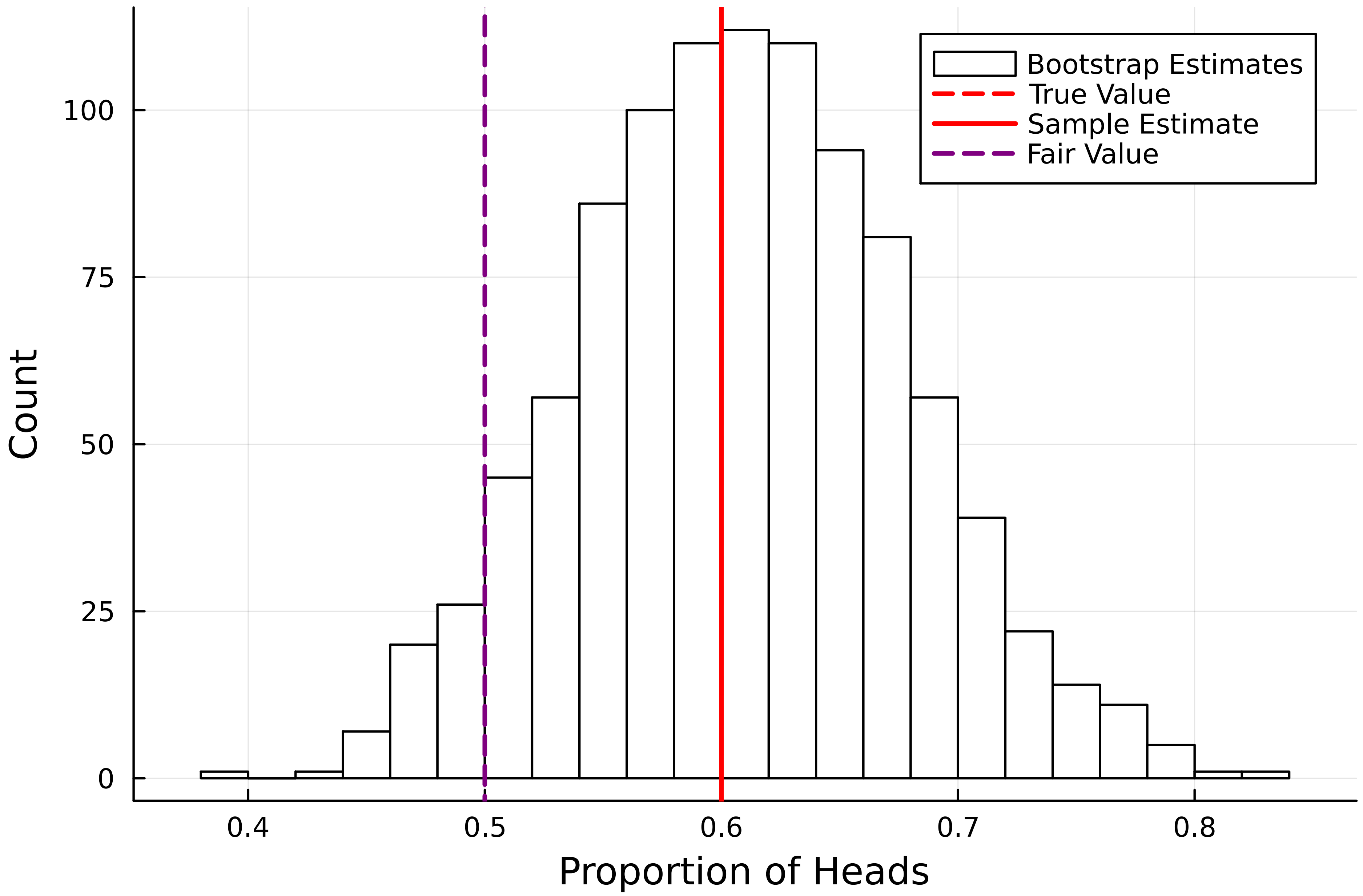


The 95% basic confidence interval is $(0.25, 0.65)$.

However, if we wanted to use this sample to bias-correct, we'd get an estimated bias of about 0, when the "true" bias is $0.6 - 0.45 = 0.15$.

Simple Example: Is A Coin Fair?

What if we redo this with fifty realizations?

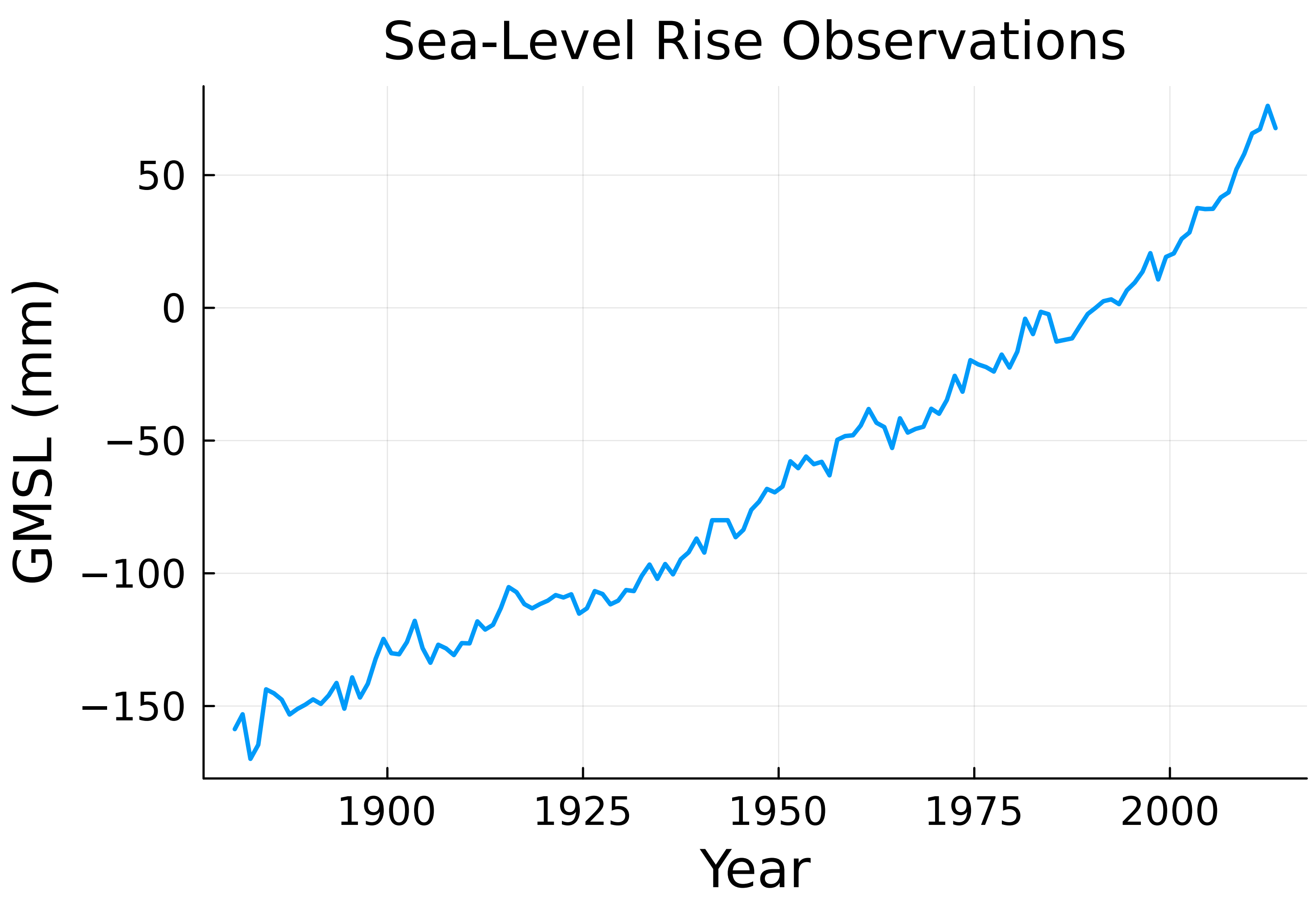


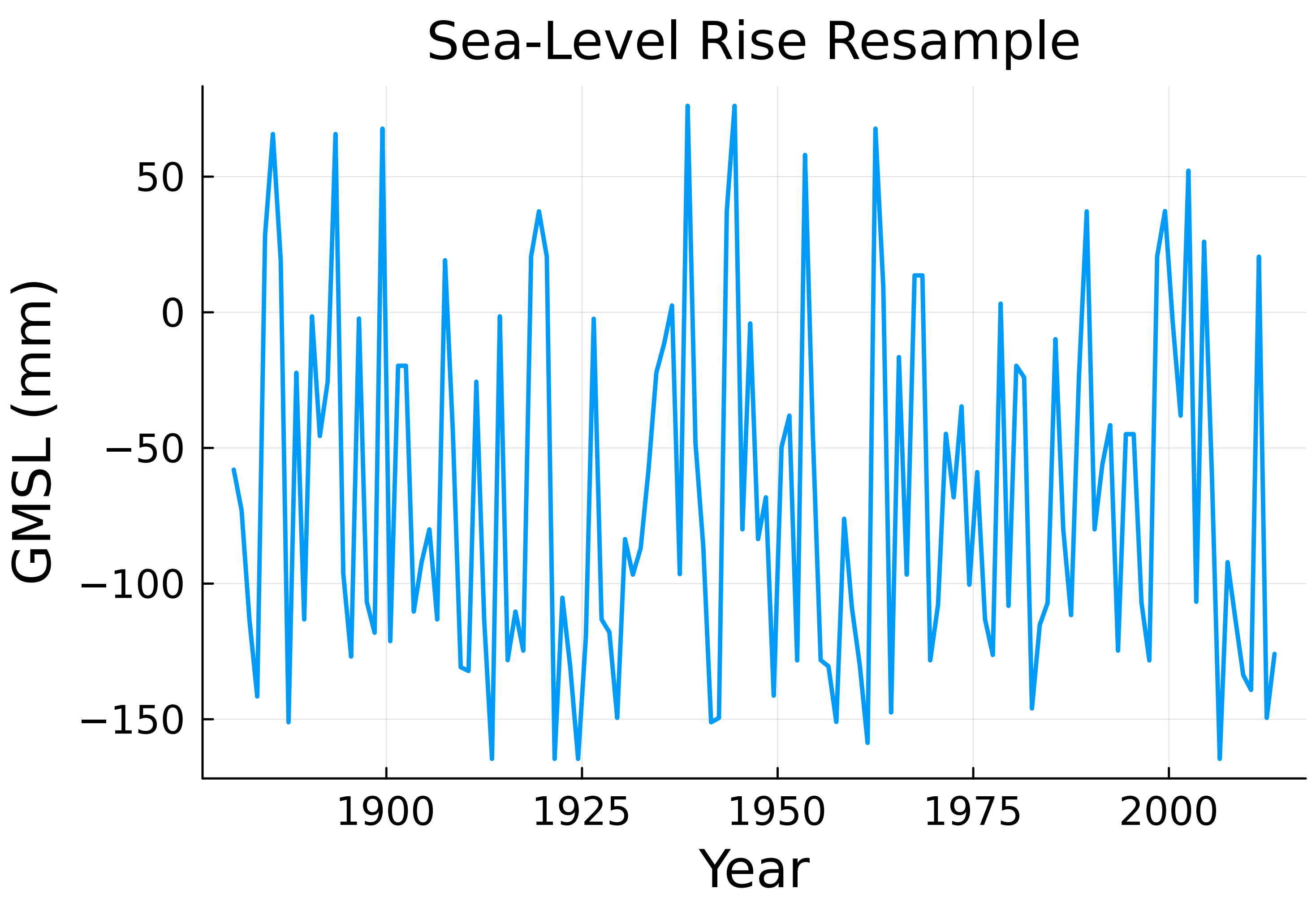
Simple Example: Is A Coin Fair?

This illustrates the centrality of the assumption that the sample is representative of the population, given the lack of any parametric structure.

Bootstrapping with Structured Data

The naive non-parametric bootstrap that we just saw doesn't work if data has structure, e.g. spatial or temporal dependence.





Bootstrapping with Structured Data

For some structures (such as correlations), can transform the data to an uncorrelated sample, then resample and re-transform back.

For time series data, need to be more clever.

Parametric Bootstrap

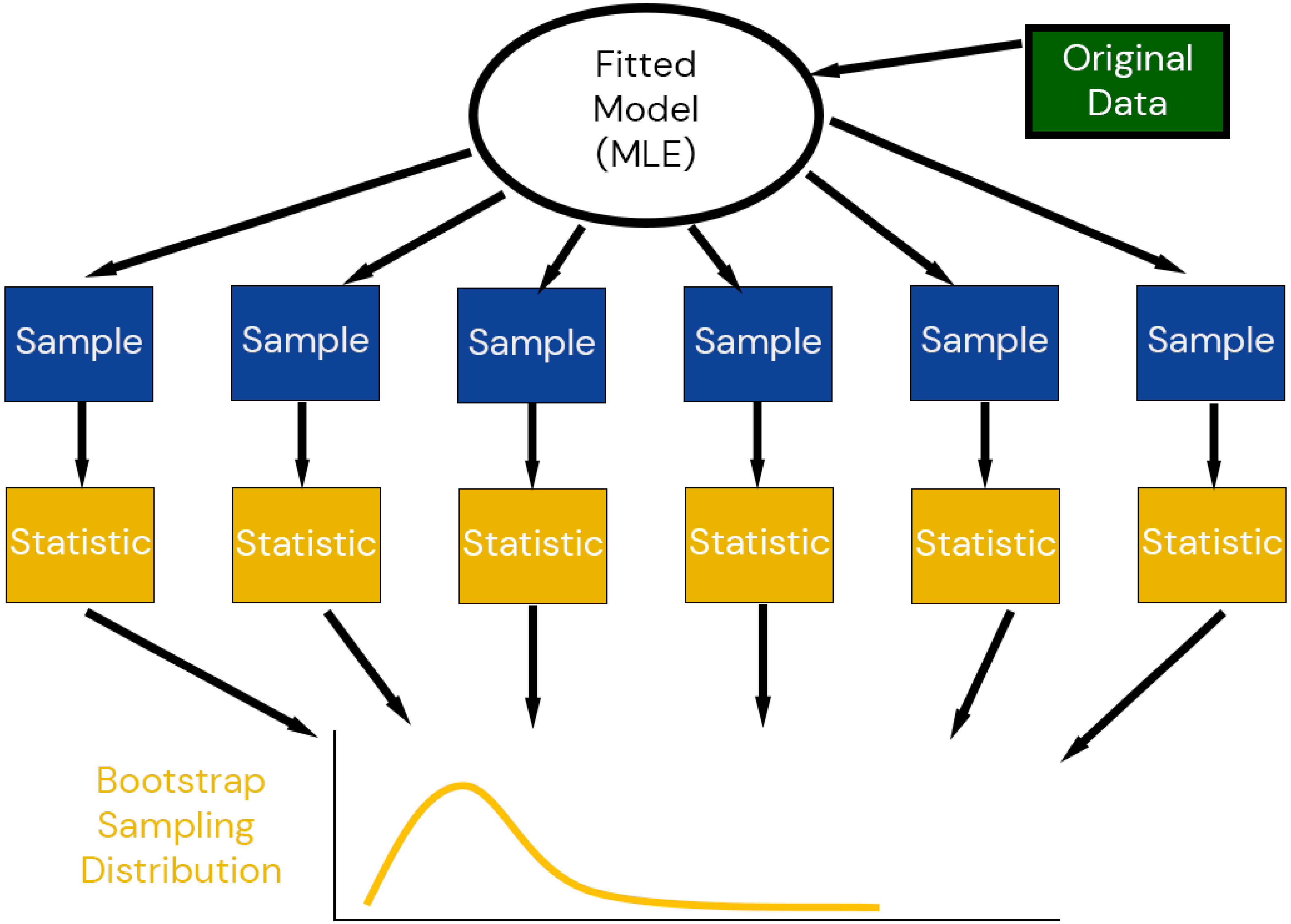
The Parametric Bootstrap

The parametric bootstrap is more interesting (and powerful) than the non-parametric bootstrap.

- **Non-Parametric Bootstrap**: Resample directly from the data.
- **Parametric Bootstrap**: Fit a model to the original data and simulate new samples, then calculate bootstrap estimates.

This lets us use additional information, such as a simulation or statistical model.

Parametric Bootstrap Scheme



Parametric Bootstrap For Time Series

A simple approach to using the parametric bootstrap for time-series data:

1. Specify and fit a trend model.
2. Calculate the residuals from the trend.
3. Resample from the residuals to generate new time-series.
4. Refit trend model to capture uncertainty.

Parametric Bootstrap For Time Series

A simple approach to using the parametric bootstrap for time-series data:

1. Specify and fit a trend model.
2. Calculate the residuals from the trend.
3. Resample from the residuals to generate new time-series.
4. Refit trend model to capture uncertainty.

This approach can be used for model calibration to capture uncertainty in parameter estimates.

Parametric Bootstrap and SLR Data

Using the discrepancy notation, this is:

$$\mathbf{y} = F(\mathbf{x}; \theta) + \zeta(\mathbf{x}) + \cancel{\varepsilon}.$$

where:

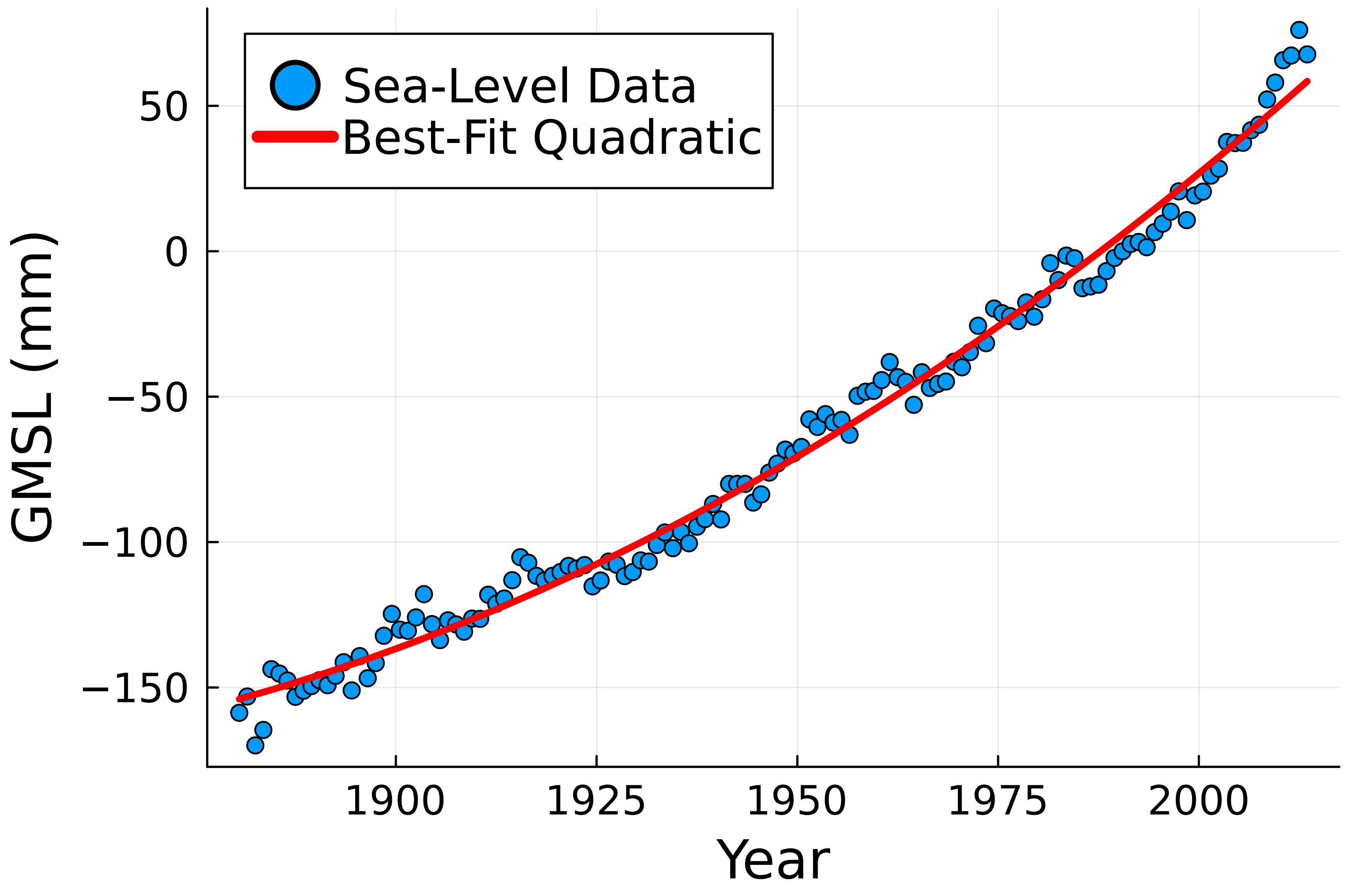
- F is the trend model (we'll use a quadratic);
- ζ is the data-generating model for residuals.
- We do not explicitly consider observation errors ε, as these are part of the residual process.

Parametric Bootstrap and SLR Data

Let's start by fitting a quadratic function

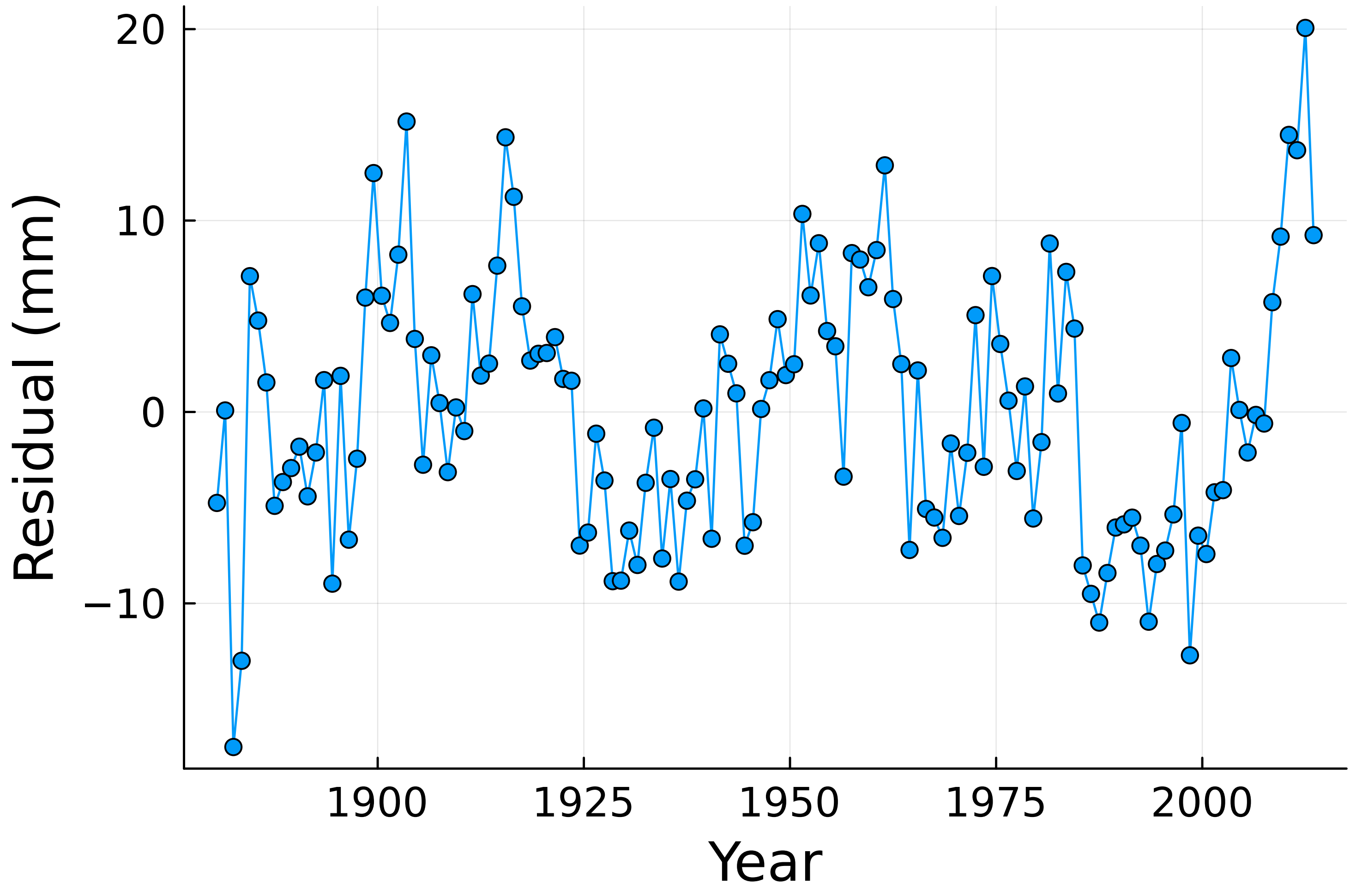
$$a(t - t_0)^2 + b(t - t_0) + c$$

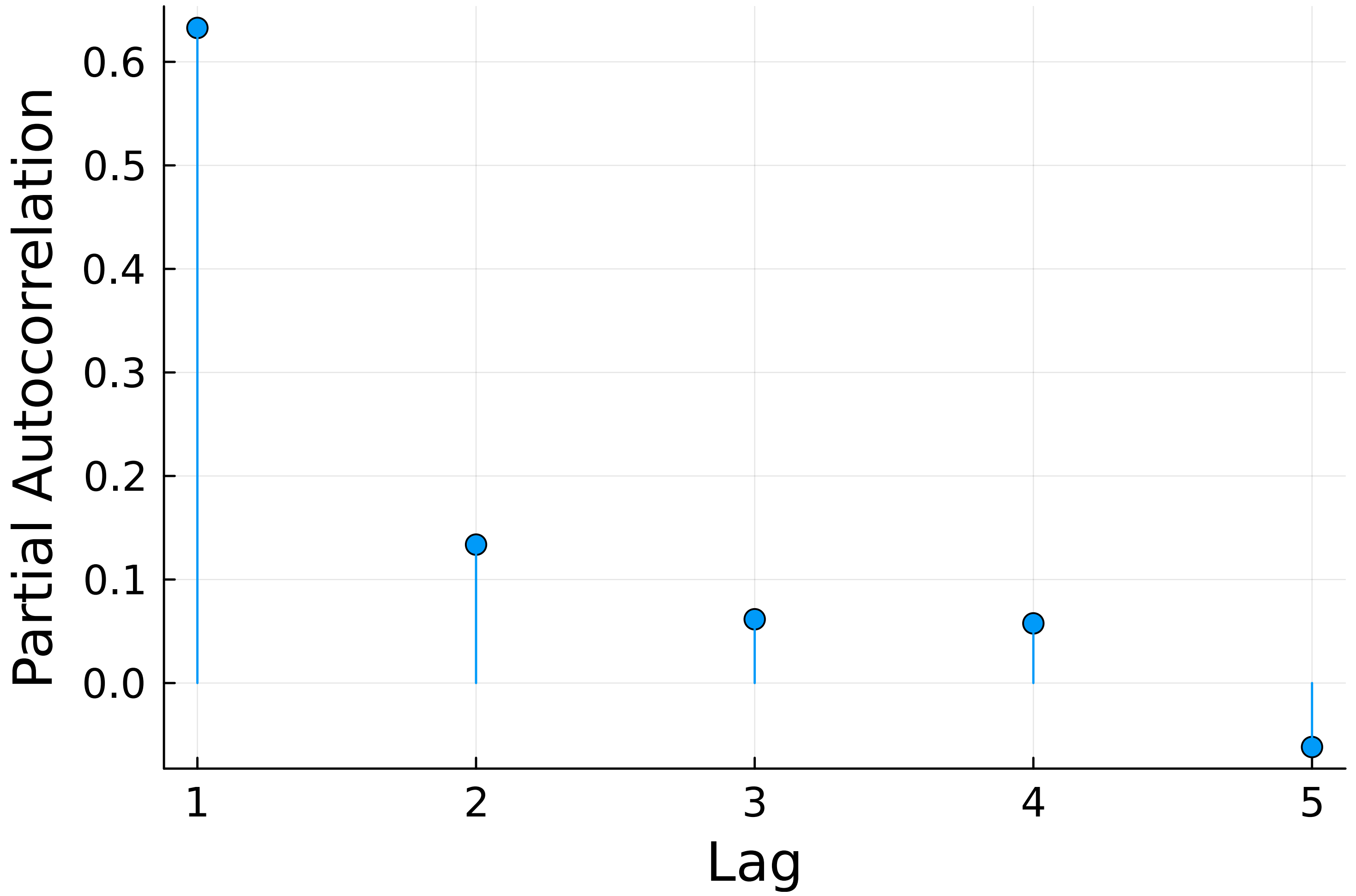
to the data.



Analyzing Residuals

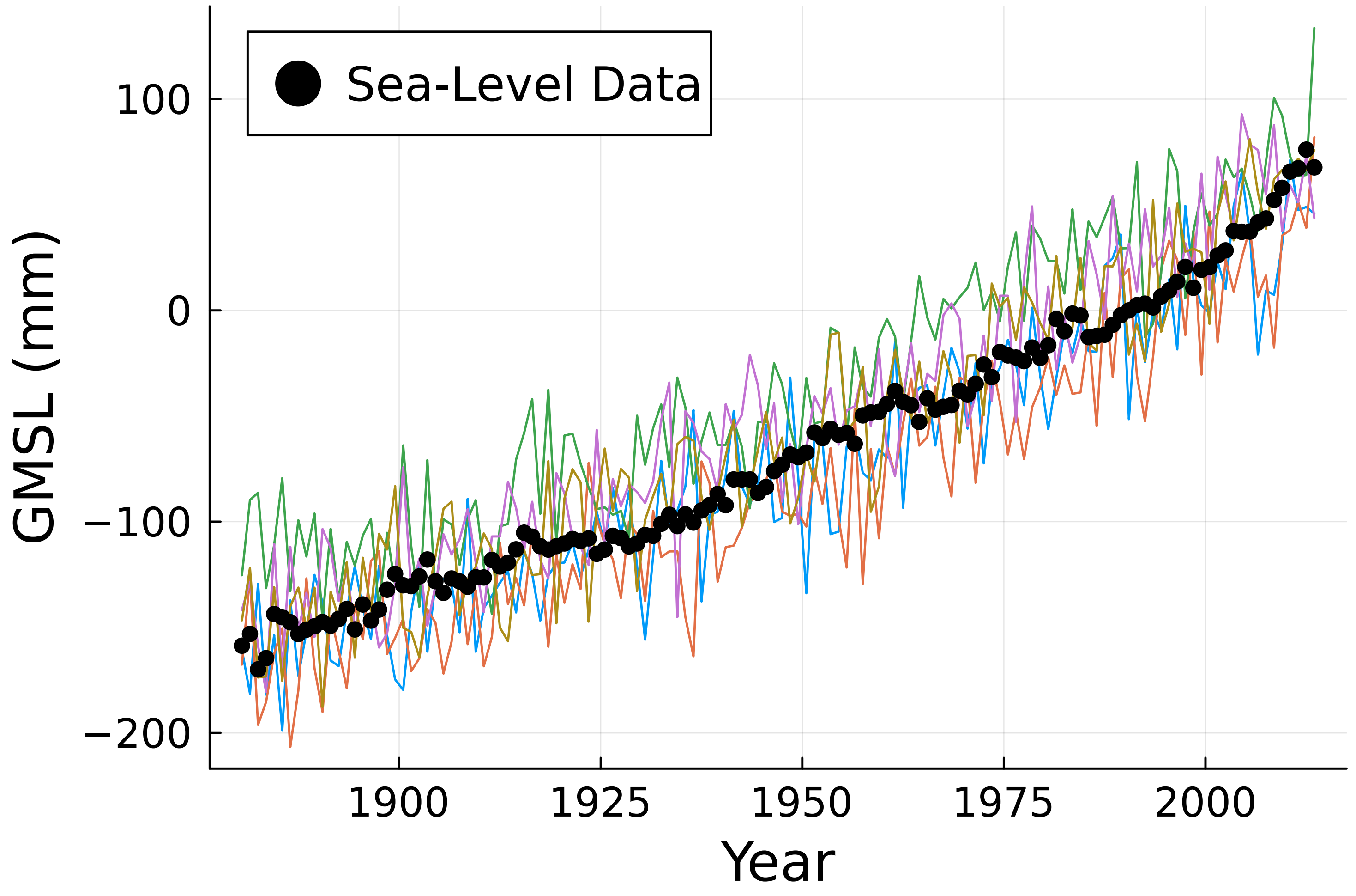
Now, let's analyze the residuals to see what probability model might be appropriate.

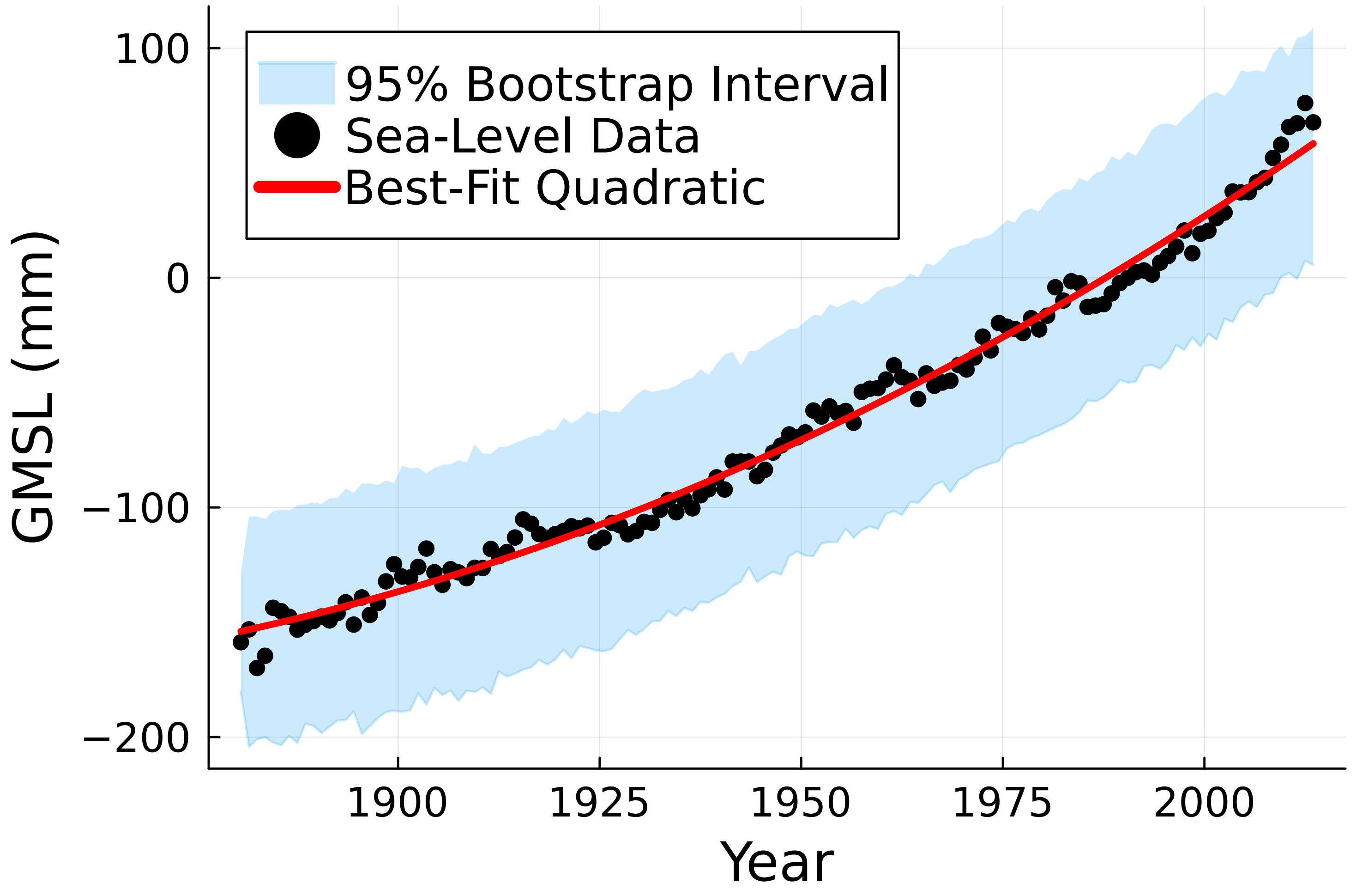




Parametric Bootstrap and SLR Data

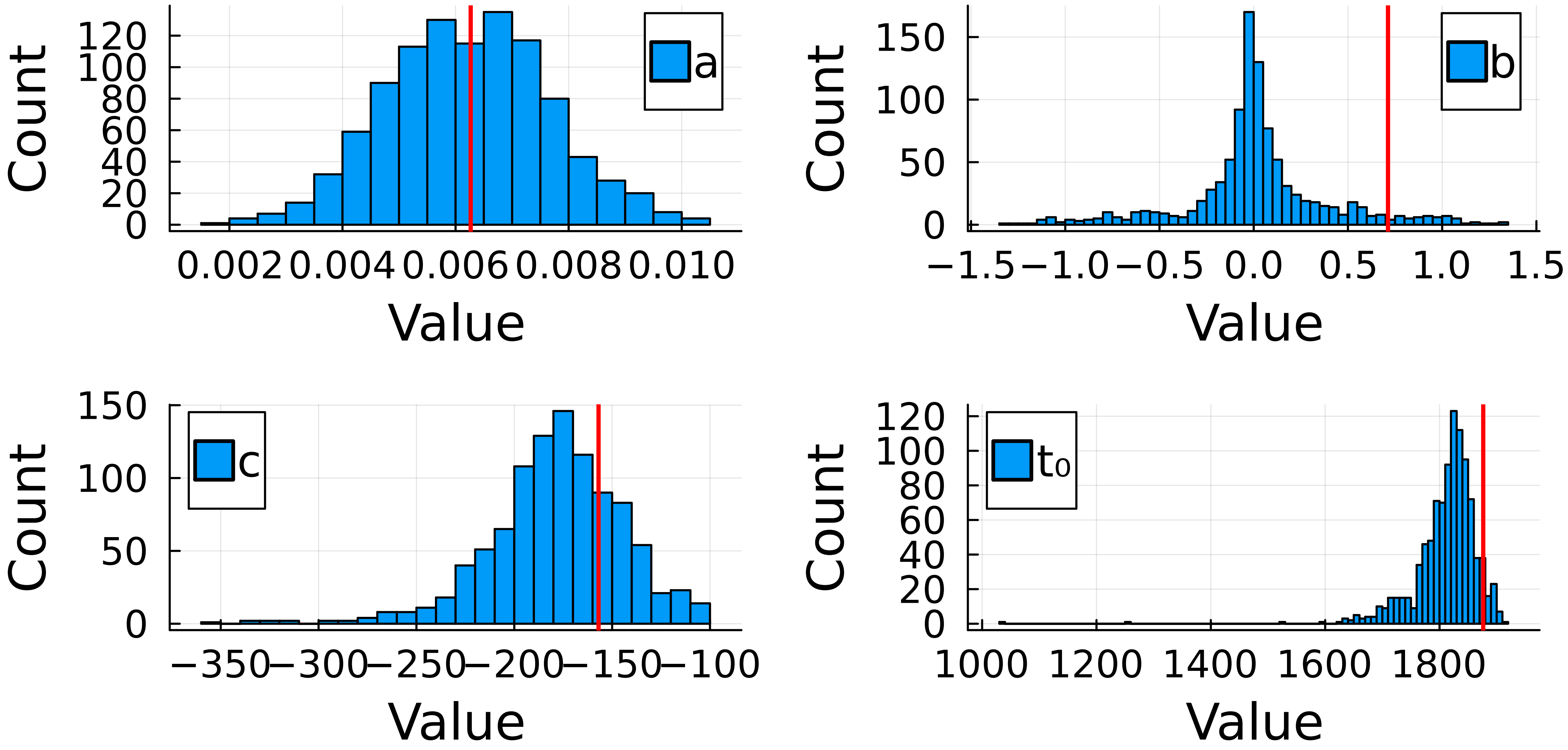
We can then fit an autoregressive model to the residuals and use that to generate new realizations.





Parametric Bootstrap and SLR Data

Finally, we refit the trend model to each realization, which yields sampling distributions for each parameter.



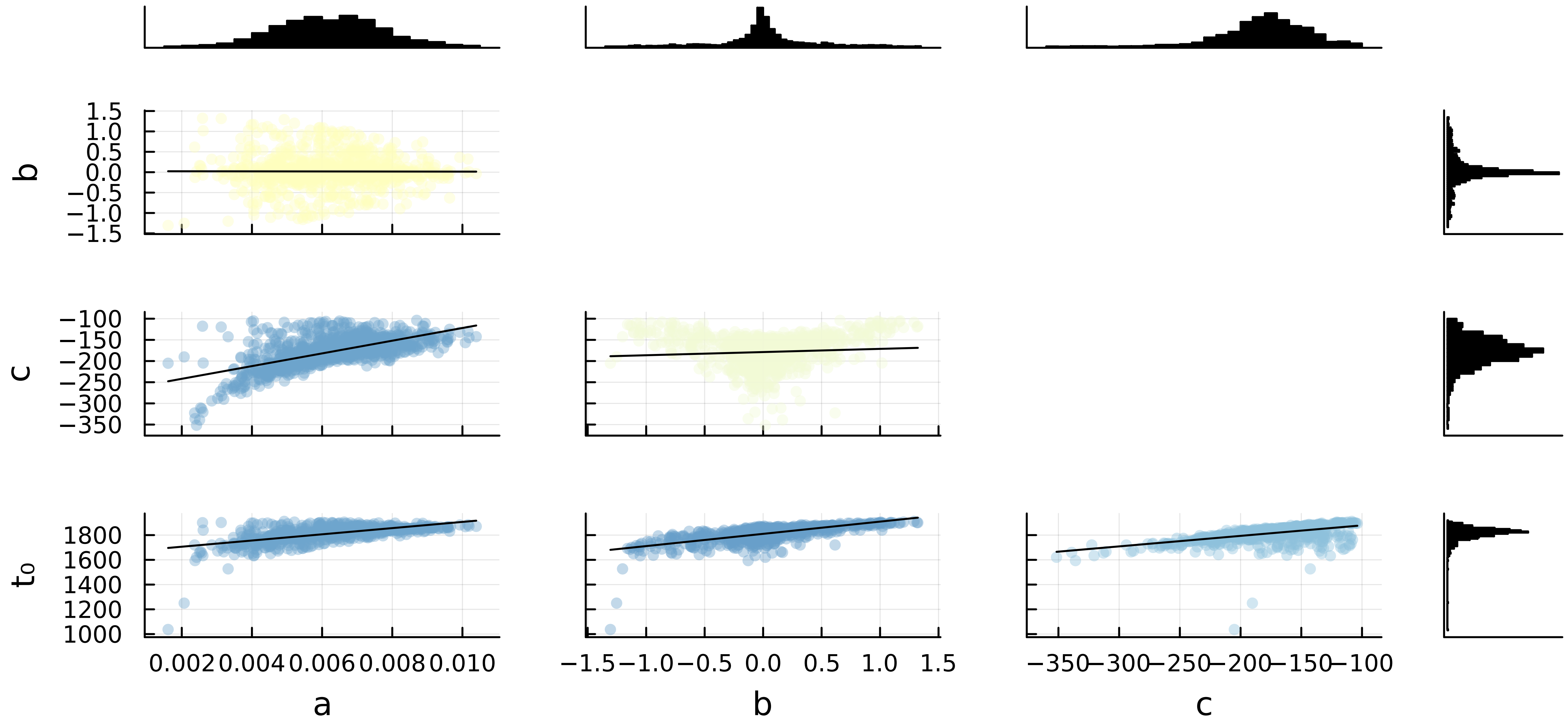
Parametric Bootstrap and SLR Data

The 95% confidence intervals for each value:

Variable	Sample Estimate	Bootstrap Confidence Interval
a	6×10^{-3}	$(3 \times 10^{-3}, 9 \times 10^{-3})$
b	0.7	$(0.5, 2.3)$
c	-157	$(-57, -198)$
$t0$	1876	$(1859, 2066)$

Parametric Bootstrap and SLR Data

We can also look for correlations across parameters with a pairs or corner plot:



Sources of Error for the Bootstrap

The bootstrap has three potential sources of error:

1. **Sampling error**: error from using finitely many replications
2. **Statistical error**: error in the bootstrap sampling distribution approximation
3. **Specification error** (parametric): Error in the data-generating model

Key Takeaways

Key Takeaways: Calibration

- Model calibration is an important UQ task.
- **Goal**: identify model structures and parameters which are consistent with the data and our prior beliefs (in a Bayesian setting);
- Important to capture the model discrepancy in an appropriate fashion!

Key Takeaways: The Bootstrap

- The bootstrap is a powerful method to combine simulation and estimation.
- Several sources of error: of particular concern are statistical error and specification error (but this is always a concern).
- There are more sophisticated versions of the bootstrap for different data structures, e.g. the *(moving) block bootstrap* for time series data.
- The parametric bootstrap provides a method for UQ through the sampling distribution approximation.

Upcoming Schedule

Upcoming Schedule

Wednesday: Discuss Ezer & Corlett (2012) and lab on using the bootstrap for sea-level data.

Next Monday: Introduction to Bayesian statistics and computation.