ha ca hierarchical Bayes do for you?

ha ca you do for he hierarchy?

Frederi Viens

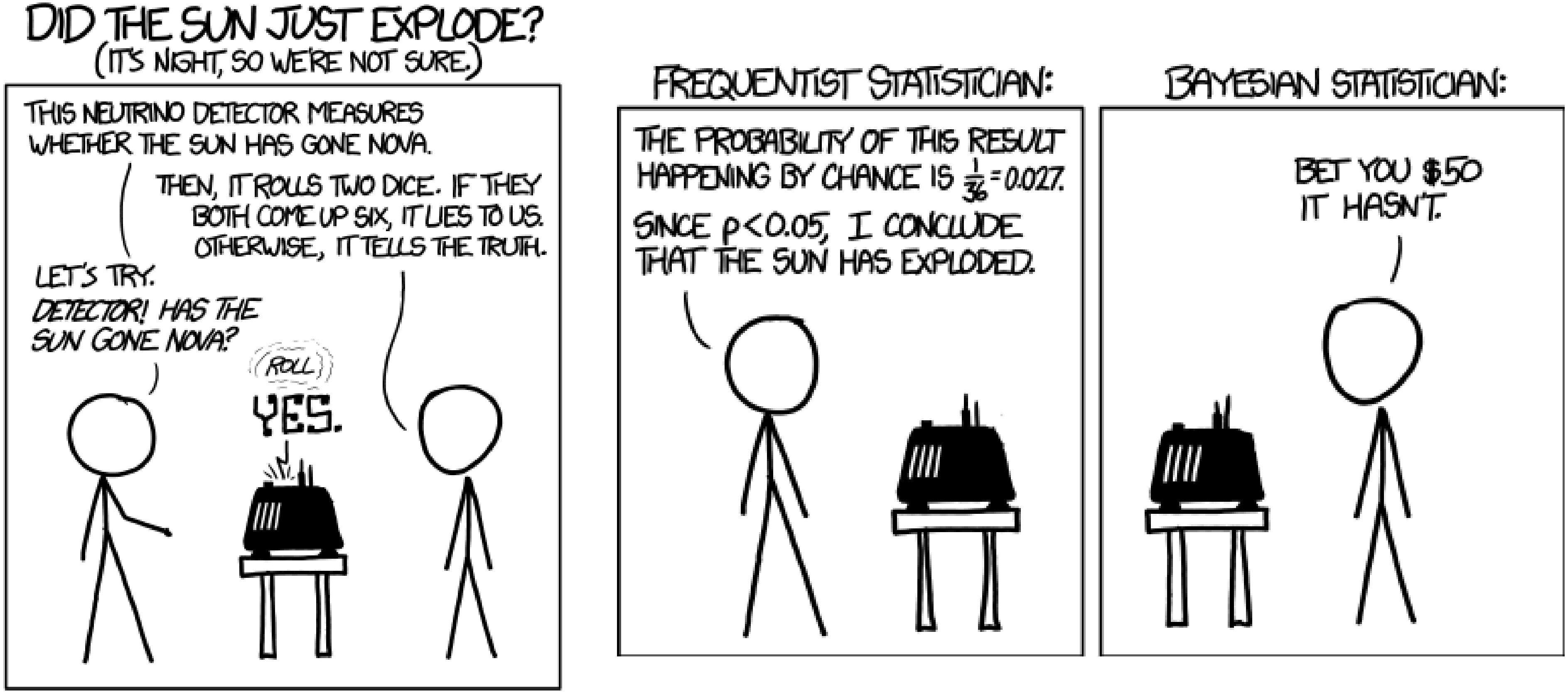
ichigan tate University

Information and statistics in nuclear experiment and theory I

i ersi y of or o

Bayesian vs. Classical Statistics

(in a nutshell…)



Source: http://xkcd.com/1132/

Bayesian Inference 101

- Bayesian approach is a complete inferential approach!
  - Inference includes point estimation and model testing
  - recise accounting for uncertainty: full param distribs. prediction
  - Decision thanks to strong probability statements about uncertainty which are ***conditional on the data***
- Probability statements: Bayesian vs. frequentist
  - Bayesian credible intervals: “conditional on ***all*** data, and based on probability modeling, ***we know that there is a 95% probability*** that the true value of a parameter is in the quoted credible interval.”
  - Classical confidence intervals: “By repeating a random experiment many times, and producing a confidence interval for a parameter each time, ***we know that 95% of these intervals*** will contain the true (population) parameter.”

Probability modeling replaces the need for repeated experiments!

Bayesian Inference 101 (continued)

- Bayes' theorem links conditional and marginal probabilities of stochastic events (or parameters):

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)}$$

But from whence do we get the models?

$$\propto p(B|A)\,p(A)$$

Posterior Distribution
updated information about unknown events/parameters given data and prior information

Likelihood of Data given possible events or parameter values
contains all information relevant for inference

Prior Distribution
summarizes the initial information on unknown event or parameter.
This is where an EXPERT OPINION can come in.

Bayesian Inference 101 (end)

- Bayes' theorem for random variable densities has exact same form, except probabilities are replaced by densities
- Denominator p(B) in Bayes' theorem is not really important for computing posteriors of unobserved parameter/variable when data B is fixed: just sample from the numerator p(B|A)p(A) and renormalize.
- **Hierarchical Bayes**? What on earth is that? It's easy:
  - non-hierarchical Bayes =
    - one probabilistic model (likelihood) for data B given unobservable A
    - another probabilistic (prior) model for unobservable A
  - ierarchical Bayes
    - Same as above, except there is a separate set of data, NOT related to B, which helps construct the prior model for A.
    - rior odel for contains
    - ore levels external data can contain model w more data
    - tc., etc. with more and more hierarchy levels if needed

ha ca hierarchical Bayes do for you ?

- hen a parameter summari es an entire section of a discipline, there might be data and models to support that parameter
- hat model, and those data, might seem physically unrelated to the problem at hand.
- By building that model and those data into your prior, you gain possible access to a critical look at that unrelated section of your discipline. ou might make friends, or enemies. *But at least you are doing hierarchical Bayes !!!*
- hen a prior model for the variable of interest seems contrived or otherwise grossly inadequate, given what people know, use data for that knowledge to build a prior model for the variable.
- It may be that people in your discipline consider a variable U which cannot be directly observed but which explains a lot. By building a model for U as it effects observables, another model for U as it is affected by external data, you have yourself a hierarchy where you can reconstruct this unobservable U !!

ha ca you do for he hierarchy ?

I I I I B I I

- **ea** with all your colleagues, friends and enemies, to discover what physics might inform portions of your modeling which don t fit nicely in your model.

- **ea** with the statisticians working with you to make sure that the priors they are using, including possible external models, make sense to you. sk about prior uncertainty levels. peak about your own tolerance for uncertainty. est your models or emulators for sensitivity or robustness with respect toparameters.

- *ortant i you ant to reconstruct a latent or ictitous aria le*
  If it s fictitious, support for its imaginary existence better be good! ust like in frequentist statistics, it s easy to overfit, but Bayesian stat can work with a lot less data than frequentist, so don t be afraid. se the model to check that uncertainty levels are realistic, honest

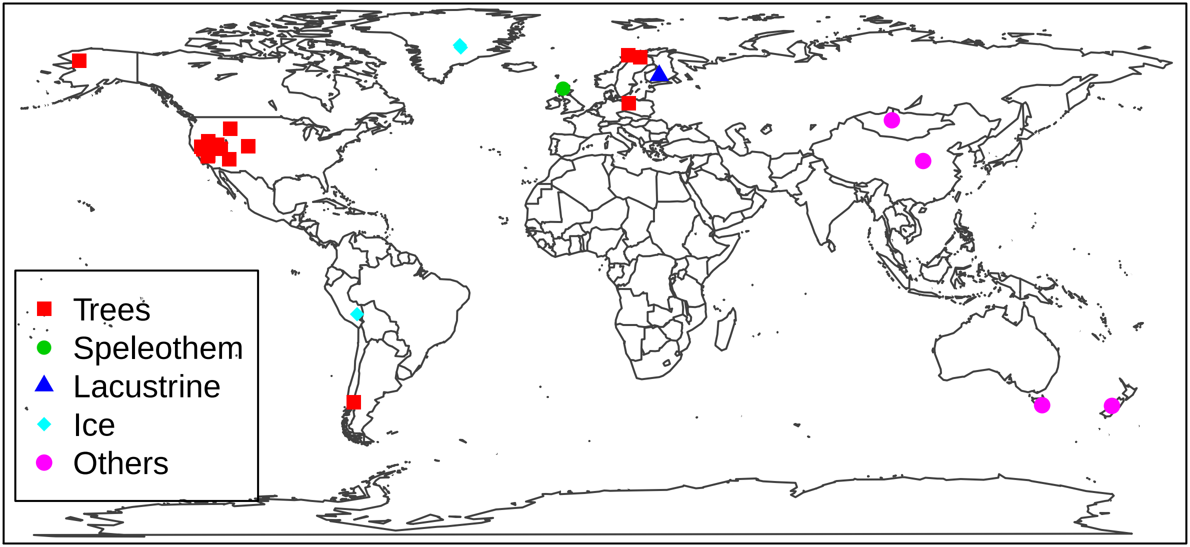
Bayesian Analysis: aleo e era ure reco s uc io usi ro ies a d forci s

- Design and implement a Bayesian hierarchical model linking past global temperatures with external data like and volcanism and with observed proxies like tree rings use a linear framework with normal errors

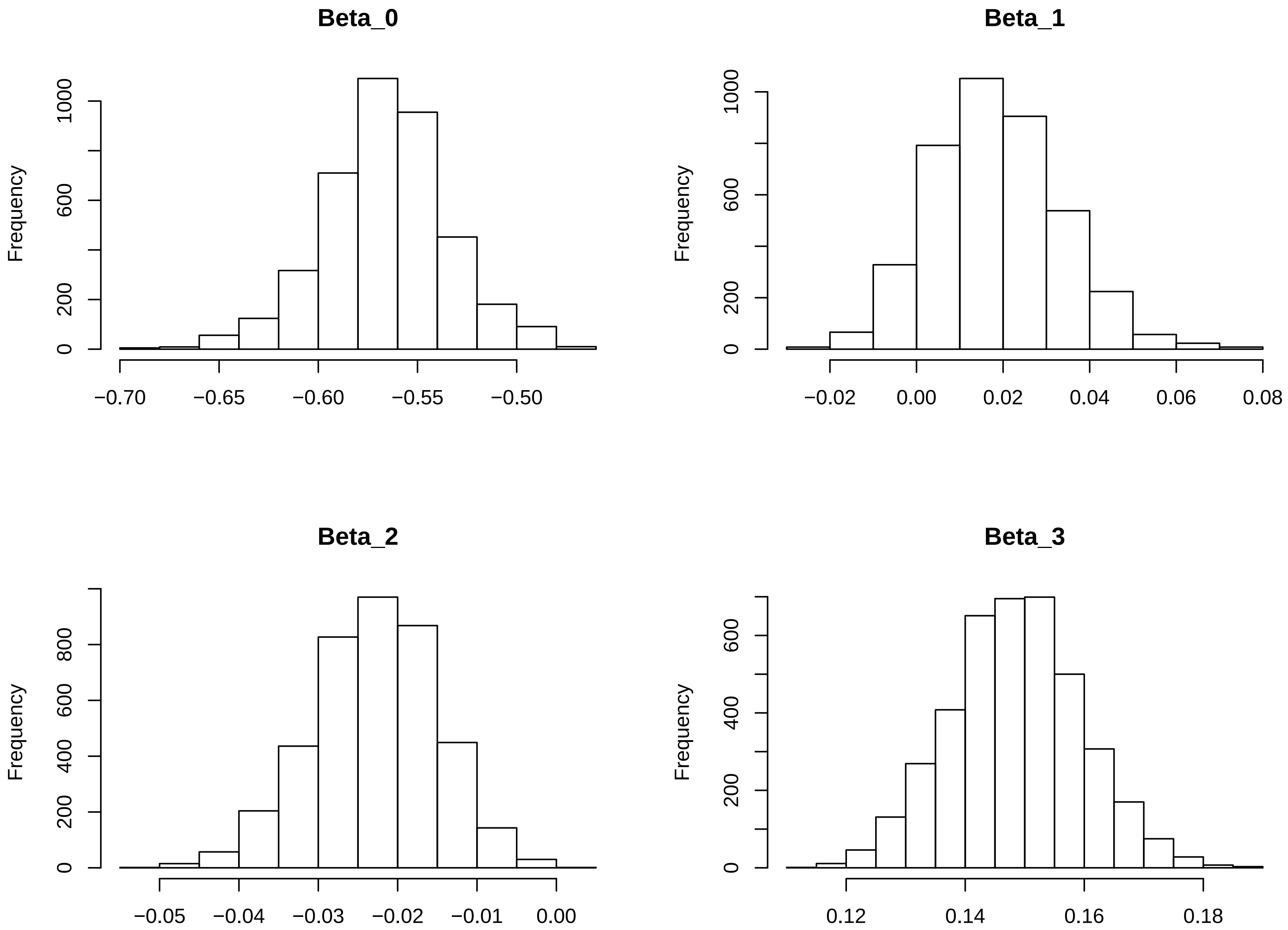
$$P_t = \alpha_0 + \alpha_1 T_t + \sigma_t \varepsilon_t$$

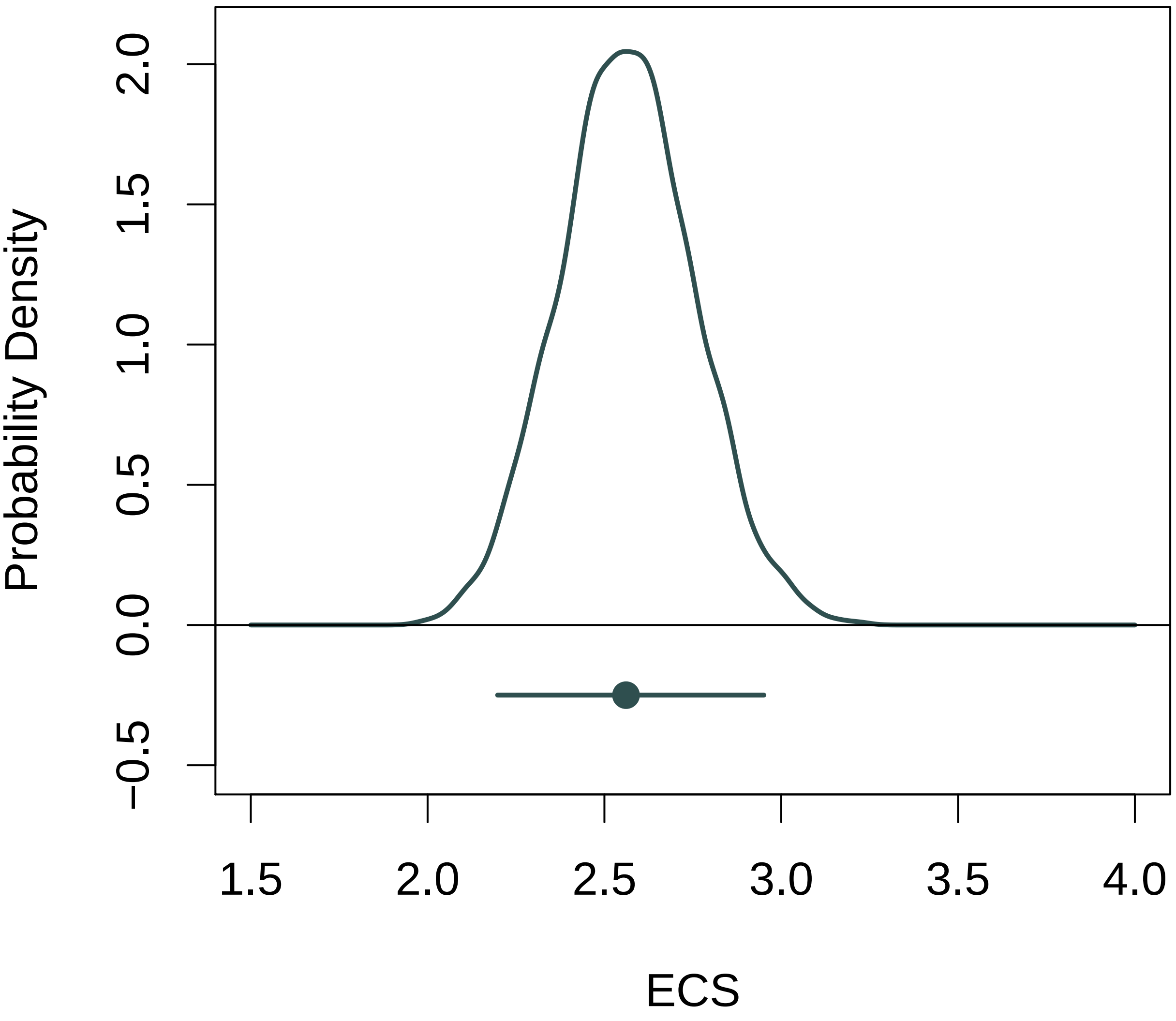
$$T_t = b_0 + b_1 V_t + b_2 S_t + b_3 C_t + \mu_t \eta_t$$

- Here P is an aggregate of proxies, T is the global mean temperature, t is from year 1000 to 2000, and V, S, C are external data: volcanism, solar irradiance, greenhouse gases; σ_t ,μ_t noise intensities; ε,η = noises.

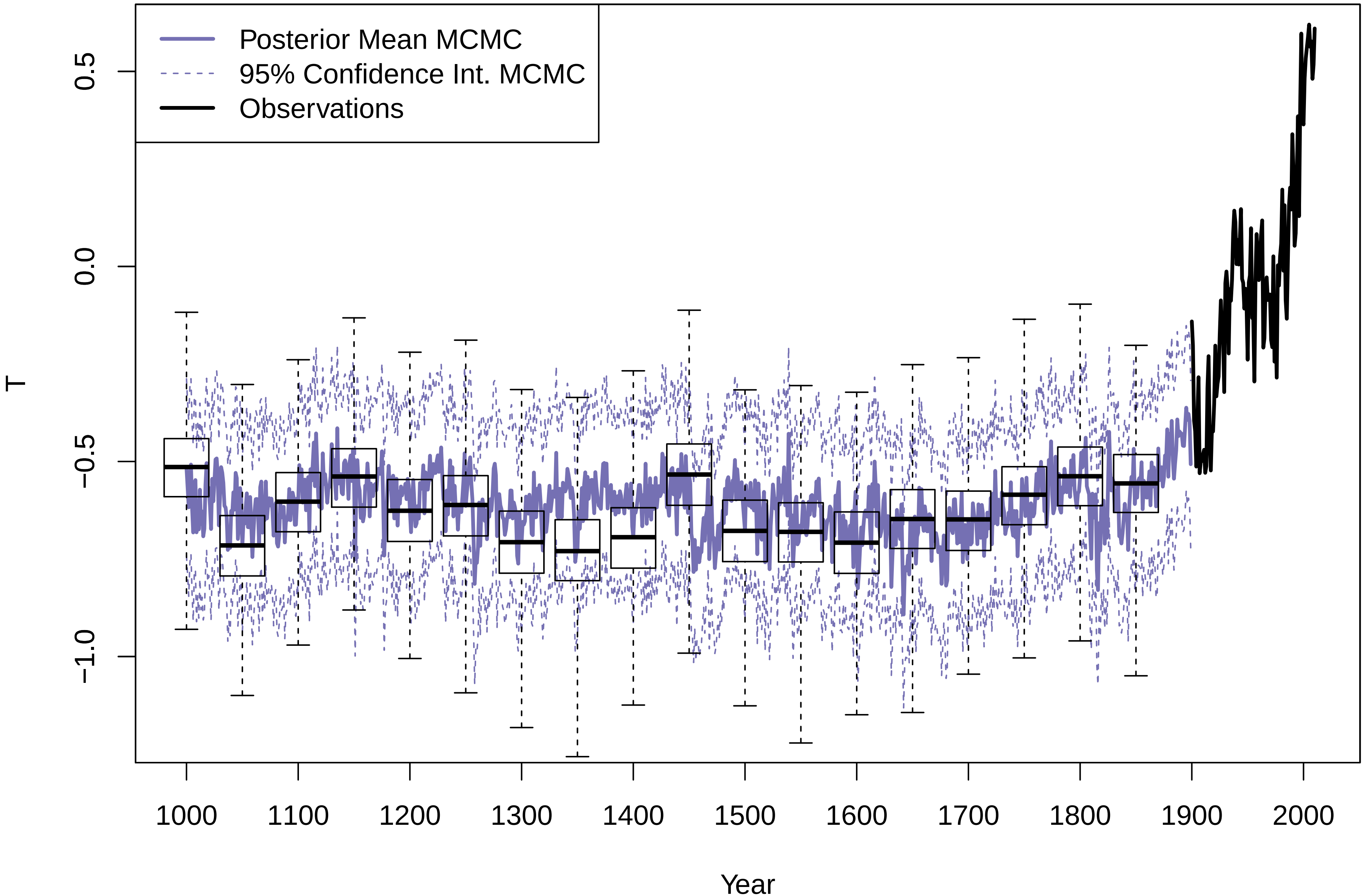


Trees
Speleothem
Lacustrine
Ice
Others

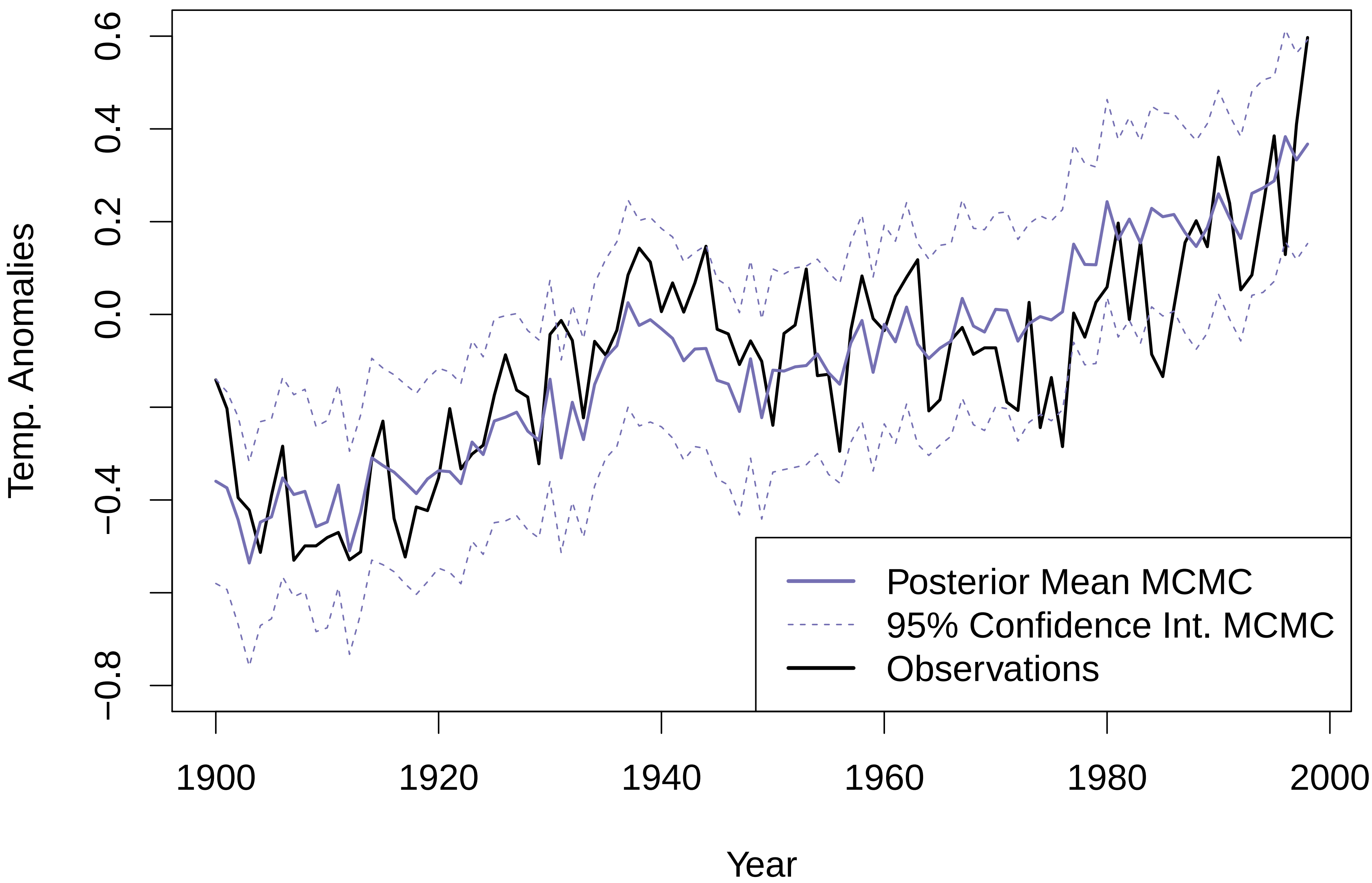

Beta_0
Frequency
0
200
600
1000
−0.70
−0.65
−0.60
−0.55
−0.50
Beta_1
Frequency
0
200
600
1000
−0.02
0.00
0.02
0.04
0.06
0.08
Beta_2
Frequency
0
200
400
600
800
−0.05
−0.04
−0.03
−0.02
−0.01
0.00
Beta_3
Frequency
0
200
400
600
0.12
0.14
0.16
0.18

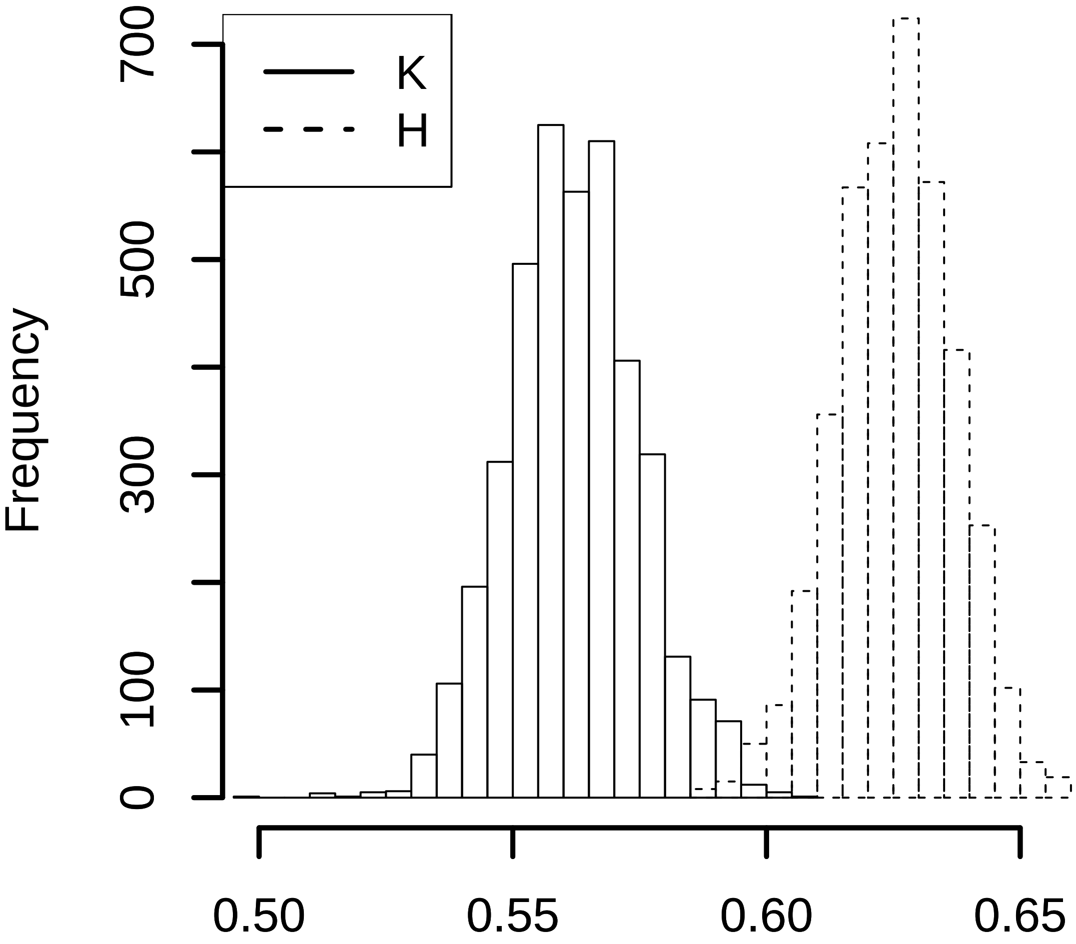


Probability Density
2.0
1.5
1.0
0.5
0.0
−0.5
1.5
2.0
2.5
3.0
3.5
4.0
ECS

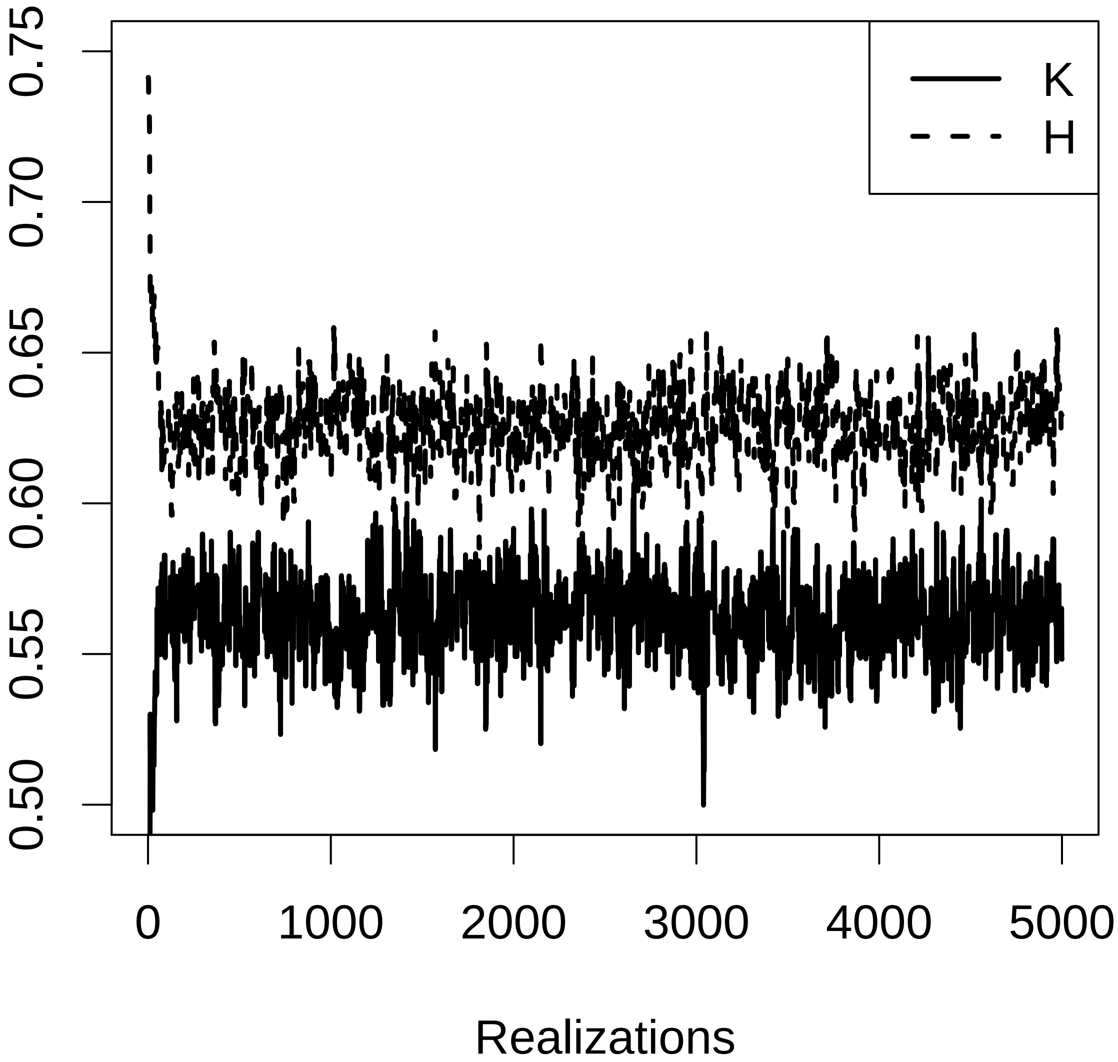


Posterior Mean MCMC
95% Confidence Int. MCMC
Observations
T
0.5
0.0
−0.5
−1.0
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
Year


Temp. Anomalies
−0.8
−0.4
0.0
0.2
0.4
0.6
1900
1920
1940
1960
1980
2000
Year
Posterior Mean MCMC
95% Confidence Int. MCMC
Observations

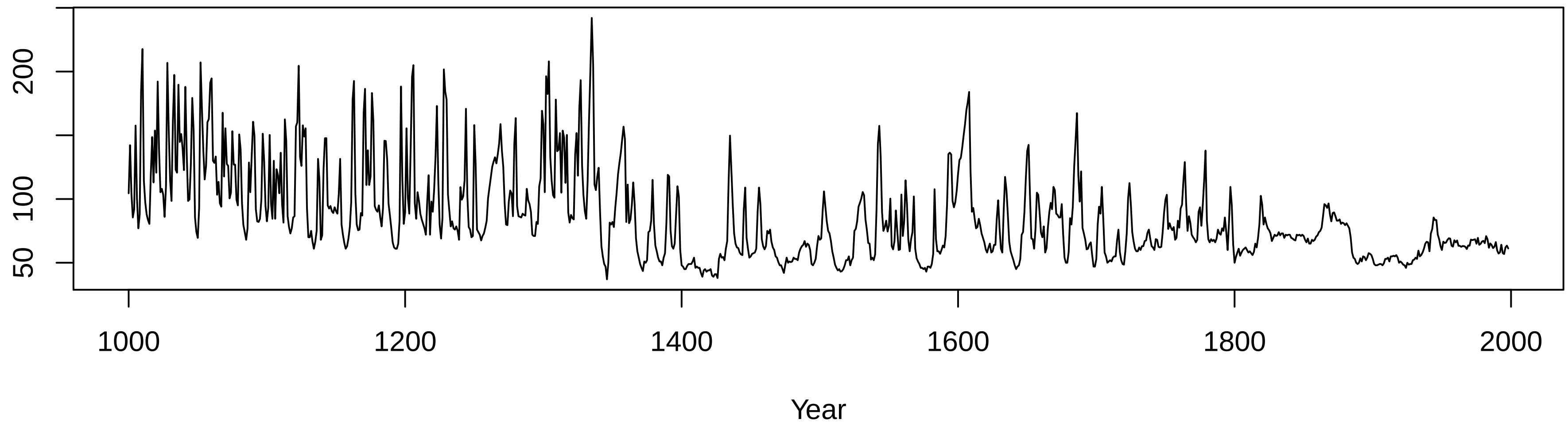


K
H
Frequency
0
100
300
500
700
0.50
0.55
0.60
0.65

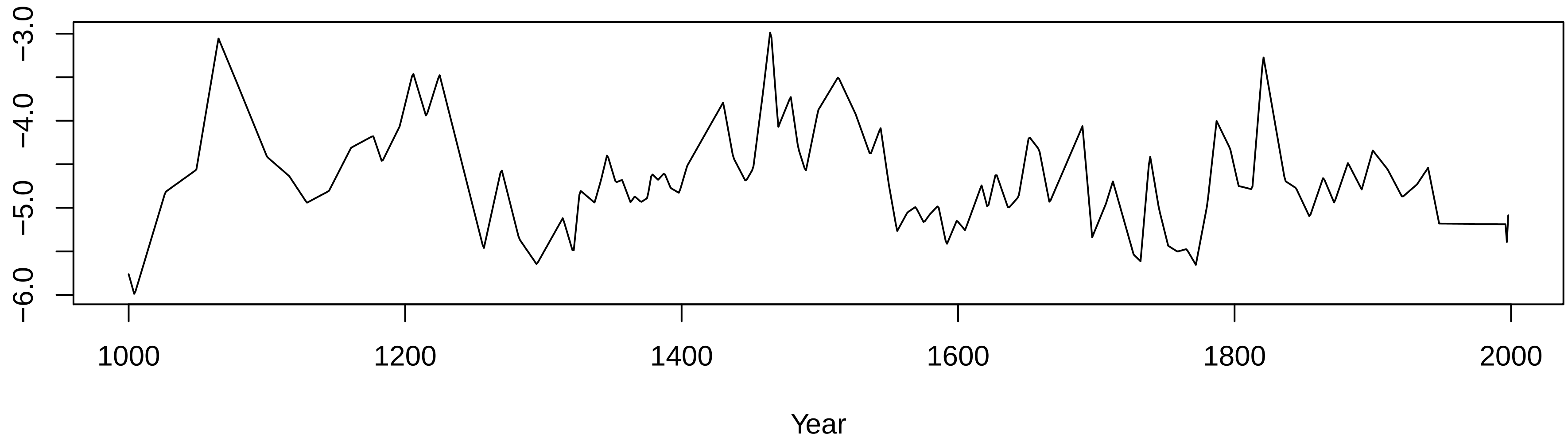


0.50
0.55
0.60
0.65
0.70
0.75
0
1000
2000
3000
4000
5000
Realizations
K
H

Lacustrine (moy_2002_ageredcolor)



Speleothem (lee_thorpe_2001_c13)



Bayesian Analysis:
U.S. Public R&D spending and TFP growth

- Design and implement a Bayesian hierarchical model linking TFP growth to R&D stocks, and stocks to R&D spending: use a linear framework with normal errors

$$T_t = \alpha_0 + \alpha_1 RD_t + \alpha_2 CI_t + \alpha_3 t + \varepsilon_{T_t}\sigma_T$$

$$RD_t = \sum_{i=0}^{49} \beta_{RD,i} XD_{t-i} + \varepsilon_{RD_t}\sigma_{RD}$$

T_t	Agricultural total factor productivity
RD_t	R&D knowledge capital stocks
CI_t	Corn moisture stress index
t	Time trend
XD_t	R&D spending
$\varepsilon_{T_t}\sigma_T, \varepsilon_{RD_t}\sigma_{RD}$	Error terms

Bayesian Analysis:
U.S. Public R&D spending and TFP growth

- Following Alston et al (2010), we adapt a gamma lag distribution structure with a 50-year lag span, under which R&D lags are parameterized by two factors: λ, δ
- λ helps determine how fast the distribution peters out,
- δ helps determine the distribution's shape and maximum

$$\beta_{RD,i} = \frac{(i+1)^{\delta/1-\delta}\left(\lambda\right)^{i}}{\sum_{i=0}^{L}(i+1)^{\delta/1-\delta}\lambda^{i}} \qquad \sum_{i=0}^{49}\beta_{RD,i} = 1$$

Bayesian Analysis:
U.S. Public R&D spending and TFP growth

- Steps for implementation and estimation
  - Define the prior distribution for model parameters

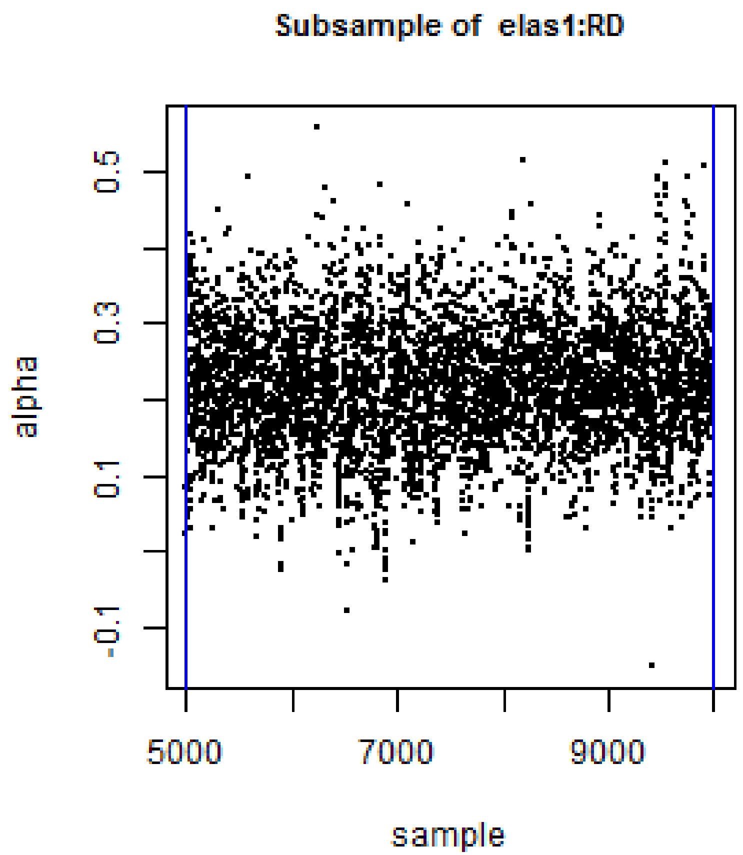
$$\boldsymbol{\alpha} \sim N(0,1); \boldsymbol{\sigma} \sim IG(2, 0.1) \quad \lambda, \delta$$

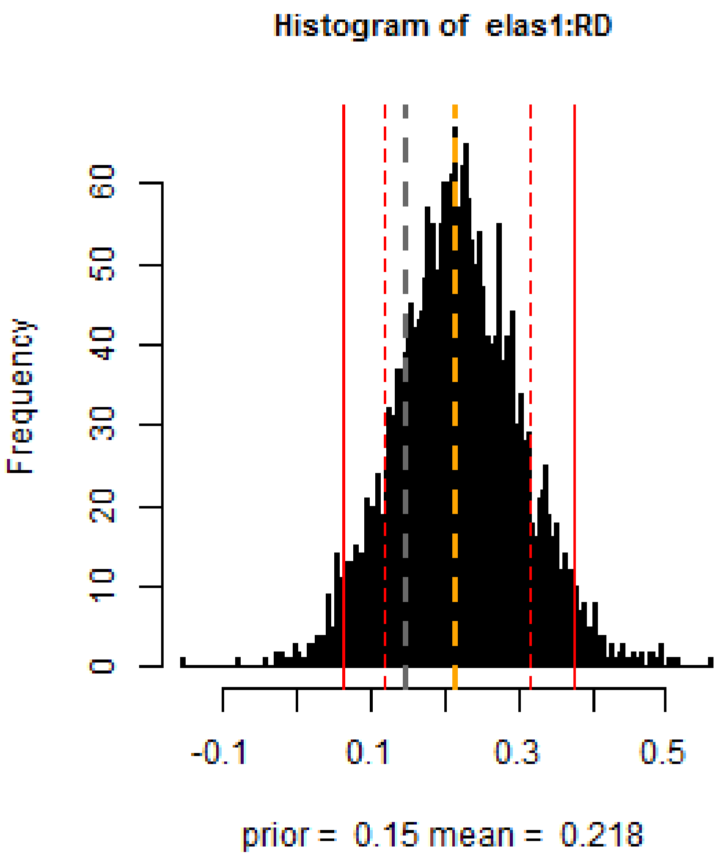
Truncated Normal with means and ranges from Alston et al (2010)

  - Compute (mathematically) posterior distribution of each variable given every other variable, thanks to prior and likelihood models in hierarchy, and data.
  - **Gibbs Sampler**: iteratively update each posterior distribution by repeatedly sampling a large number of times (computational technique easy in R)

Preliminary Results:
U.S. Public R&D experience : 1949-2011 using USDA data

- Estimated point elasticity of U.S. Ag. TFP with respect to R&D stocks (Alston et al 2010 in grey vs. mean in yellow)

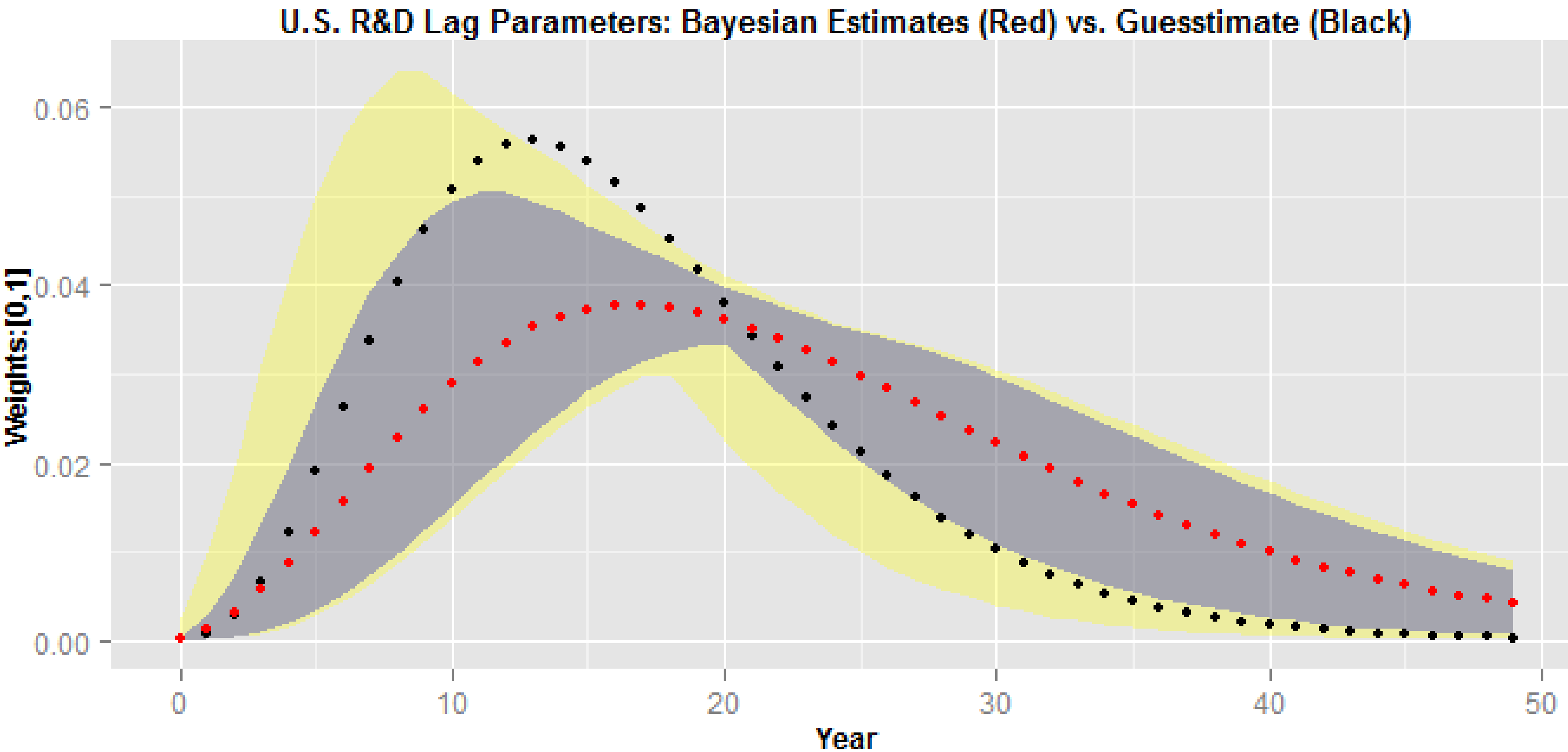




Preliminary Results:

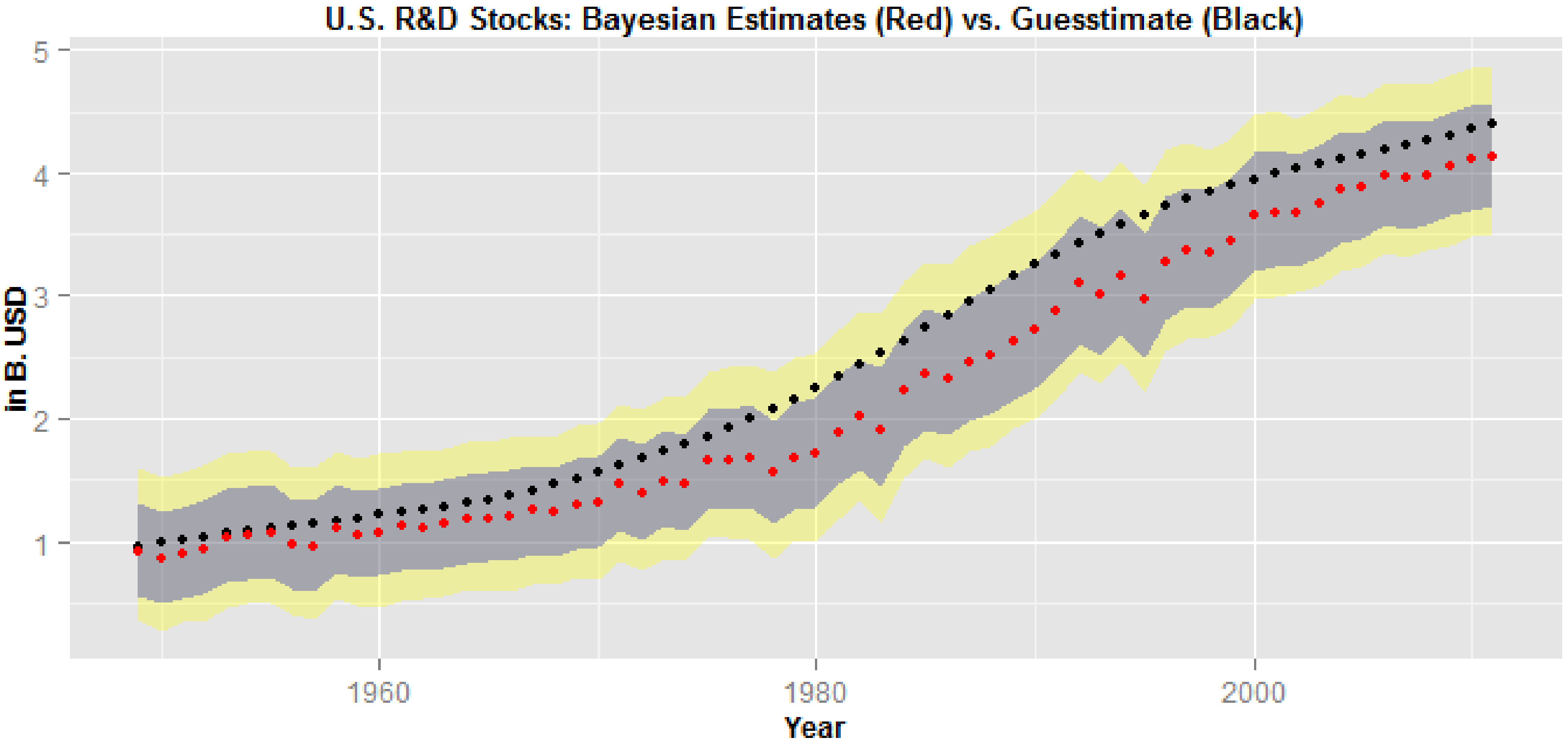
U.S. Public R&D experience : 1949-2011 using USDA data

- Estimated R&D lag parameters in color, Alston et al 2010 in black
- We use their expert opinion: prior $0.85 < \delta < 0.9$ and $0.7 < \lambda < 0.8$



Preliminary Results:
U.S. Public R&D experience : 1949-2011 using USDA data

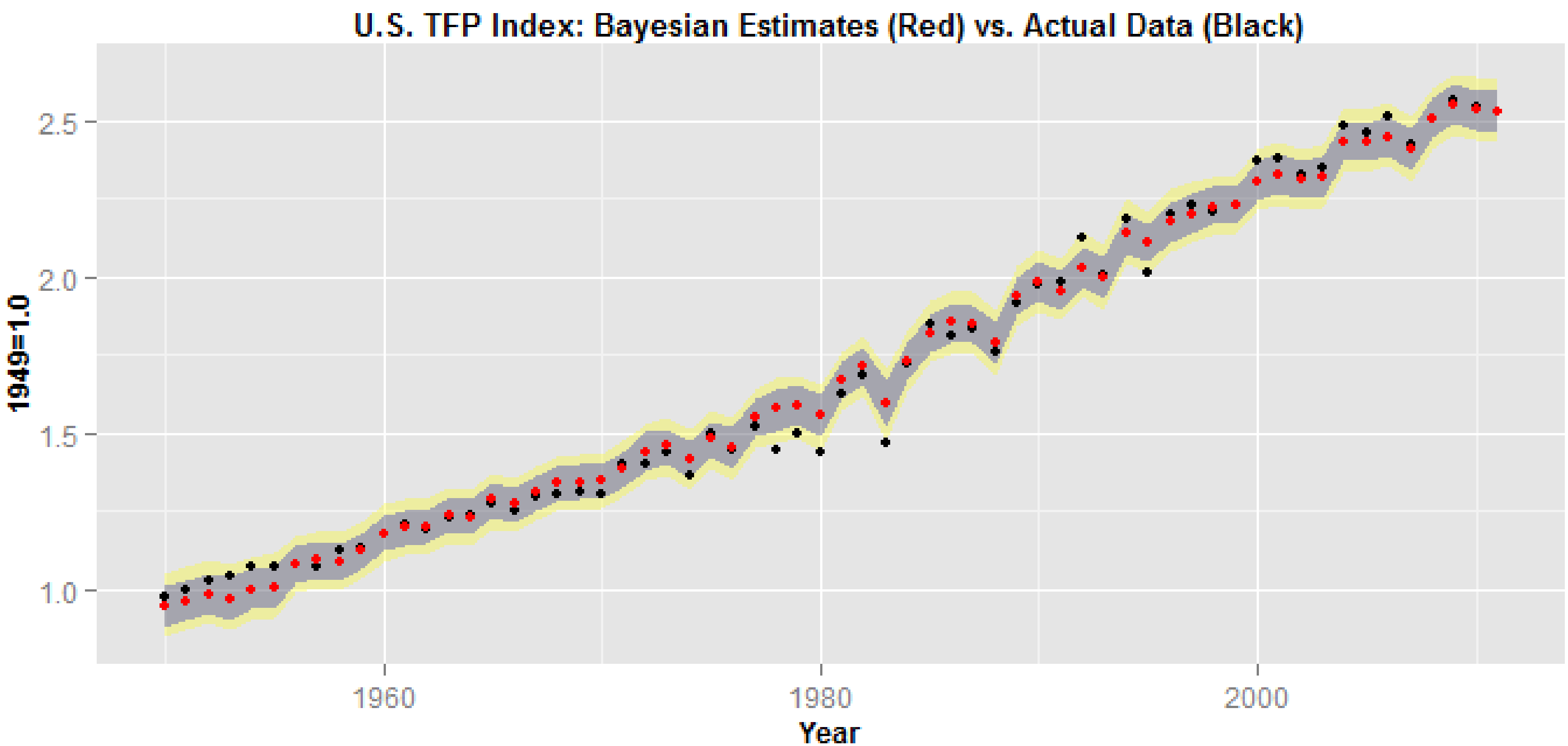
- Estimated R&D stocks (Alston et al 2010 in black vs. our results)



Preliminary Results:

U.S. Public R&D experience : 1949-2011 using USDA data

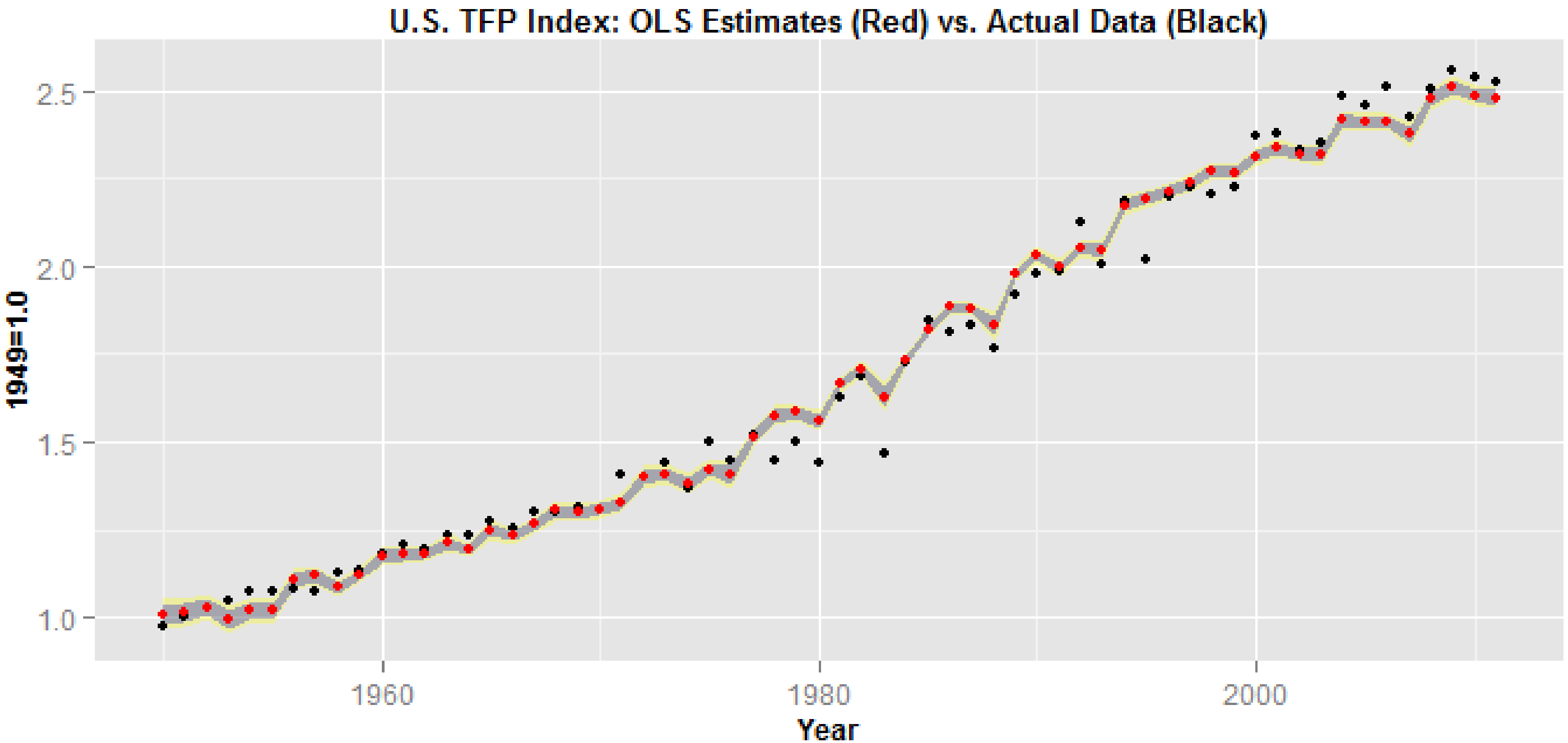
- Reconstructed TFP Index (USDA data: black vs. our results)
- A crucial “validation metric” : Empirical Coverage Probability
  - # data points inside yellow band must be roughly 95% of total # data points



Preliminary Results: non-Bayes comparison

U.S. Public R&D experience : 1949-2011 using USDA data

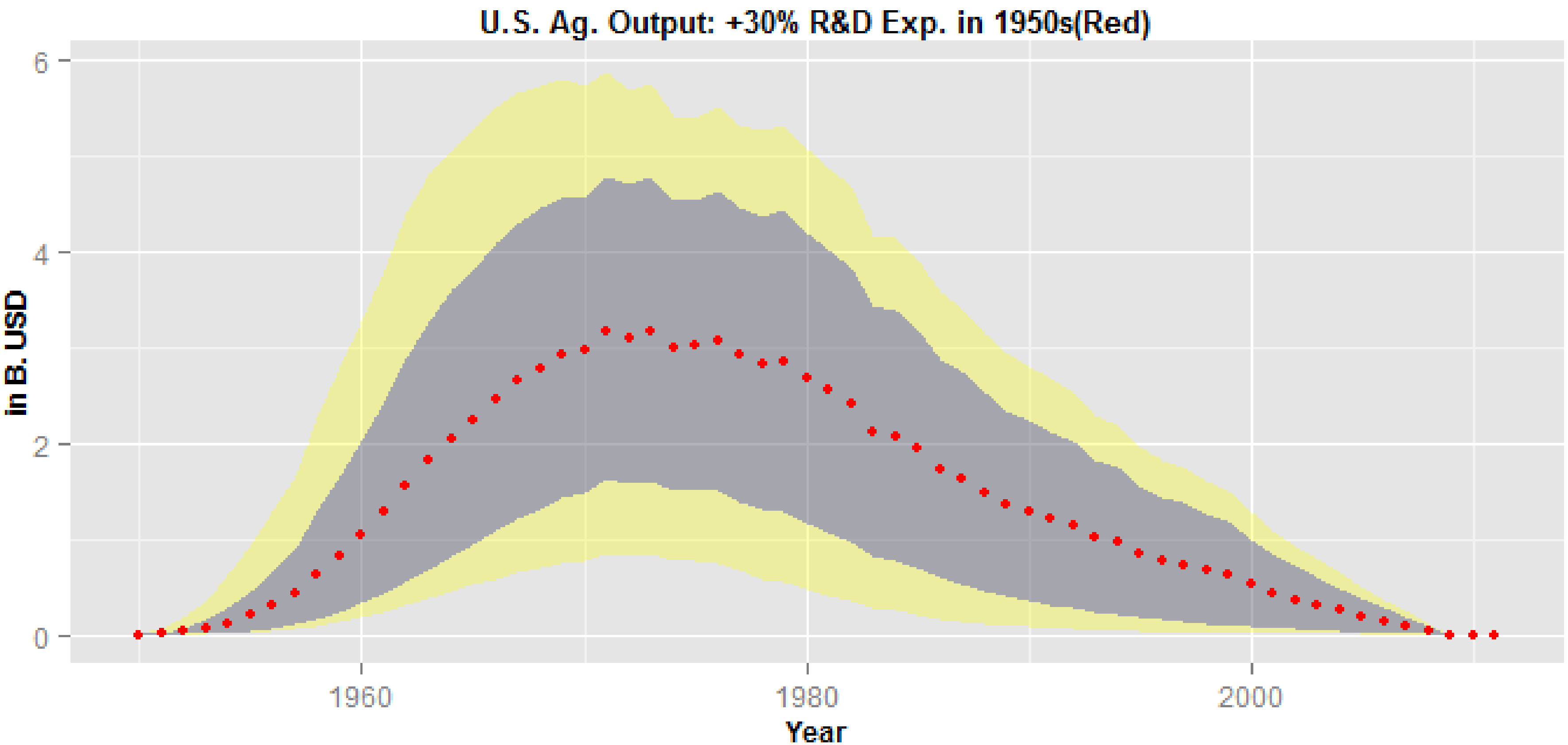
- Reconstructed TFP Index (USDA data: black vs. our OLS results)
- Uncertainty is very grossly underestimated:
  - ECP is ridiculously far from nominal value of 95%.



Counterfactual Analysis:

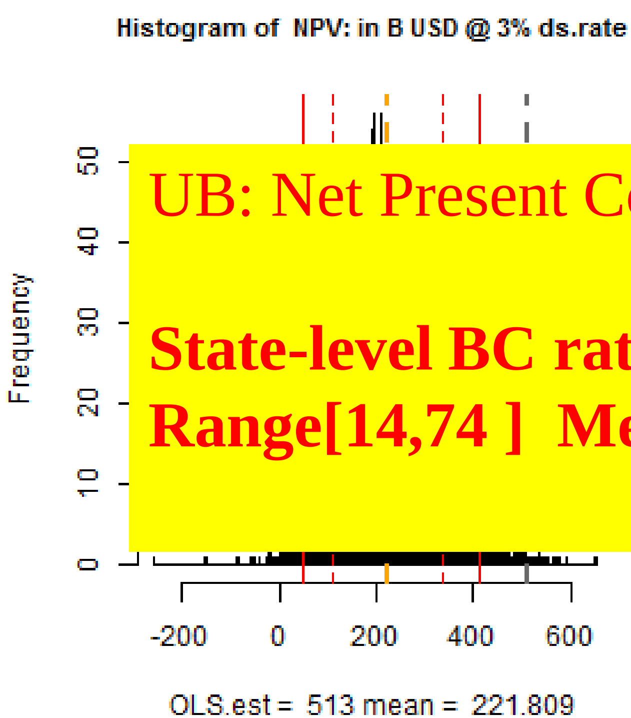
+30% increase in U.S. Public R&D in 1950-59 only

- Increase in U.S. Ag. Output from increased R&D investments
  - (how much would this investment have meant to U.S. ag. output?)



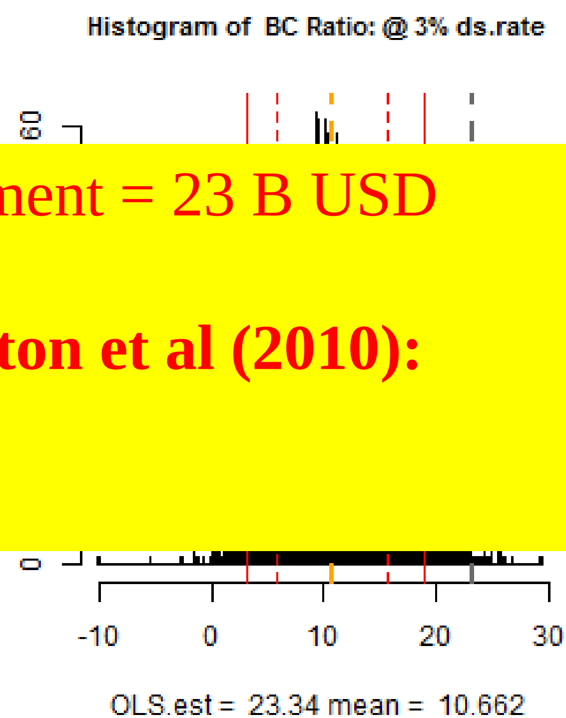
Counterfactual Analysis:
+30% increase in U.S. Public R&D in 1950-59 only

- Net Present Values & Benefit Cost Ratios (our OLS results in grey and Bayesian results in yellow): uncertainty accuracy matters !



UB: Net Present Cost of Investment = 23 B USD

State-level BC ratios from Alston et al (2010):
Range[14,74] Mean: 39.7



Methodological conclusions

- Bayes allows full evaluation of uncertainty on all parameters.
- No other method can estimate lag parameters from data.
- We use expert opinion on lags (from Alston et al. 2010) and sharpen the conclusions, lowering uncertainty.
- Hierarchical structure exploits all relevant data (TFP and public spending), to construct unobserved R&D stocks.
- Any question, realized or counterfactual, can be answered in precise probabilistic terms.
- Model selection can be based on validation metrics (ECP) :
- Bayes appears far superior to least squares regression
    - ➢ linear models are more appropriate than log-linear ones
    - ➢ uncertainty quantification is adequately conservative
    - ➢ lag uncertainty is high
    - ➢ R&D uncertainty is comfortable, and robust to lag structure

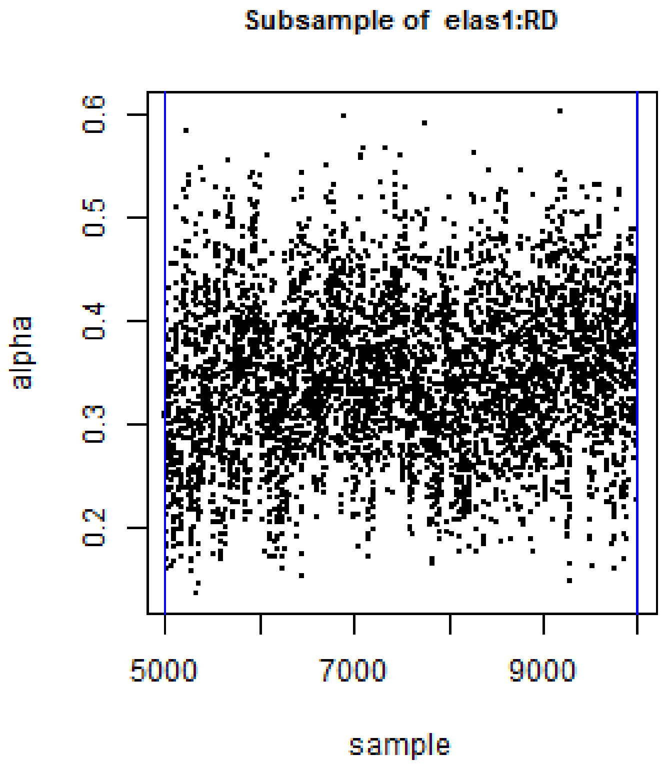
Quantitative policy implications

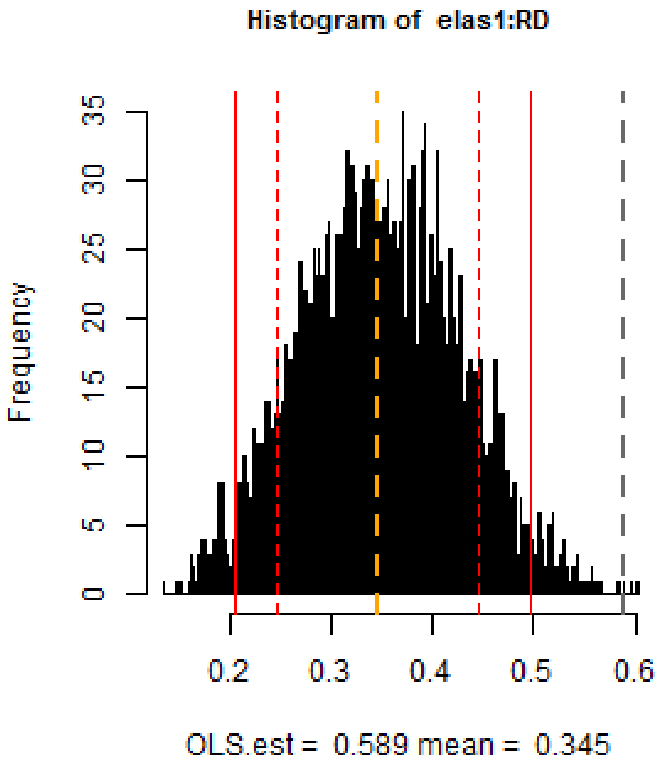
- Lag parameter estimates and uncertainty suggest :
  - ➢ peak of spending impact probably occurs later than guessed
  - ➢ impact is likely to be sustained for longer
  - ➢ non-negligible possibility that impact might occur more quickly, but betting on this would be unwise .
- Given lag uncertainty, a public investment policy should favor steady investments spread out and sustained over time :
  - ➢ we will check by running various future spending scenarios
  - ➢ then doing forecasting via future TFP Bayes reconstruction
- Previous R&D estimates are on the high side, suggesting :
  - ➢ policy-makers may need more conservative expectations of spending effectiveness
  - ➢ counterfactual analysis with historical data tells same story

Preliminary Results: change to Log model

U.S. Public R&D experience : 1949-2011 using USDA data

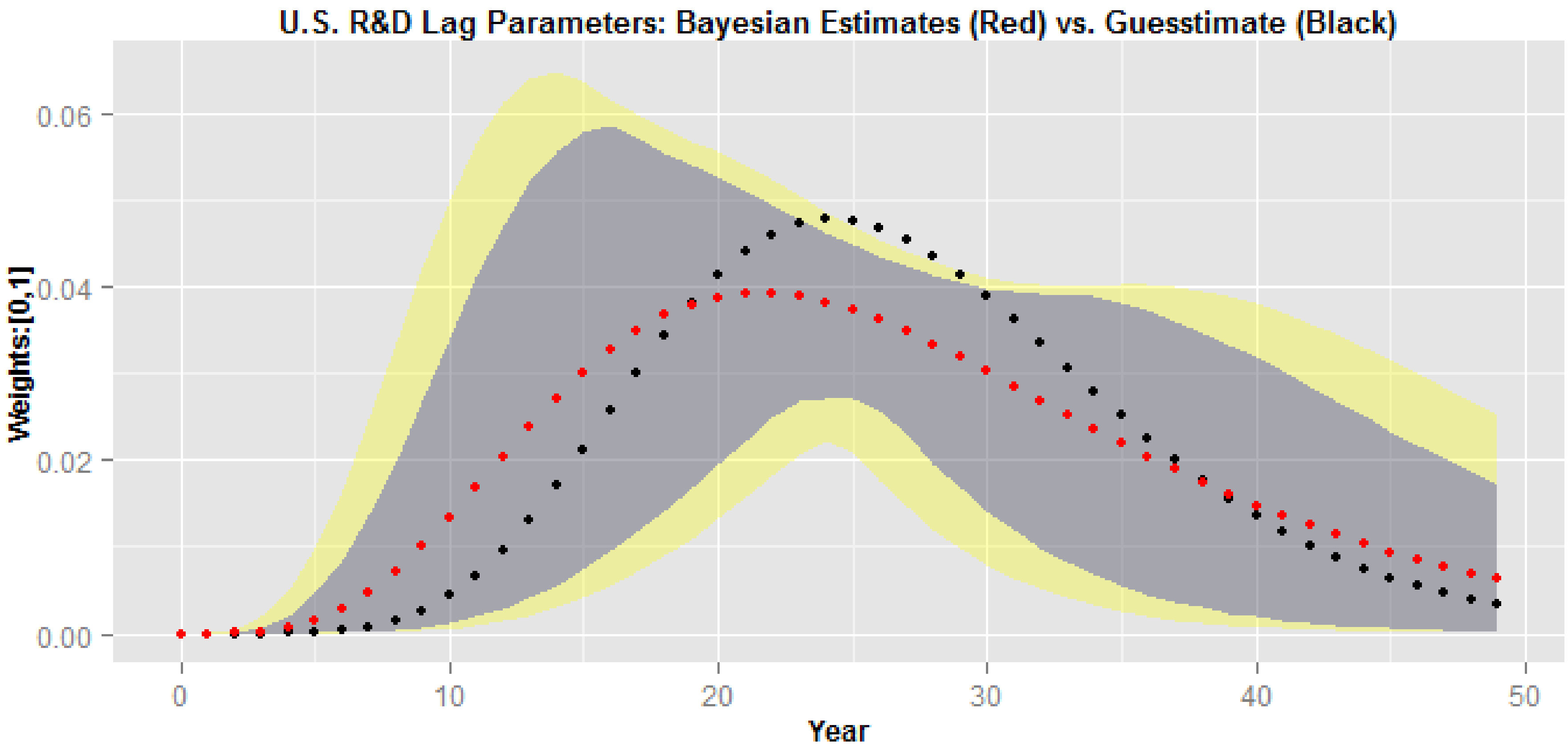
- Estimated point elasticity of U.S. Ag. TFP with respect to R&D stocks (Alston et al 2010 in grey vs. mean in yellow) **log model**





Preliminary Results:
U.S. Public R&D experience : 1949-2011 using USDA data

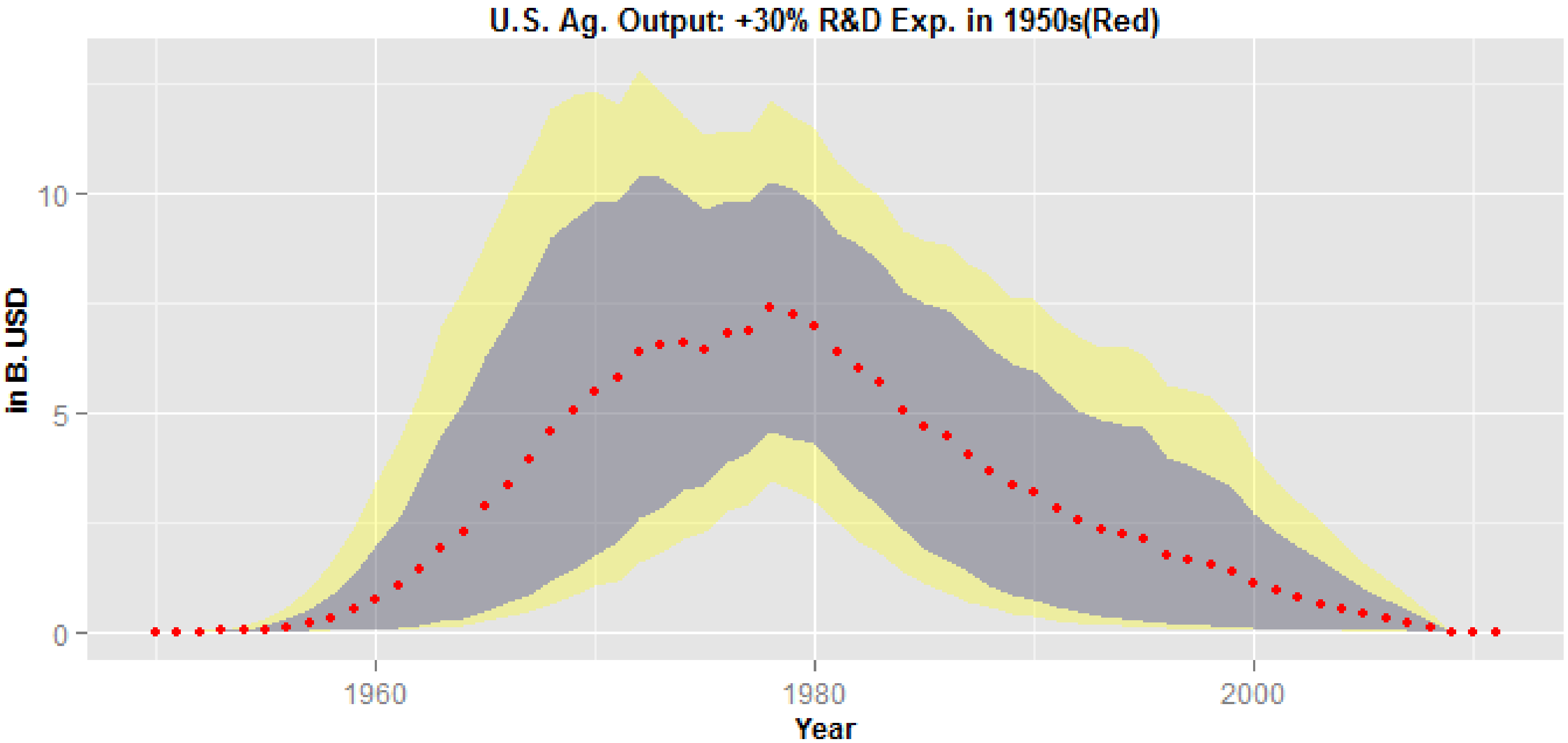
- Estimated R&D lag parameters (Alston et al 2010 in black vs. our results) **log model**



Counterfactual Analysis:
+30% increase in U.S. Public R&D in 1950-59 only

- Increase in U.S. Ag. Output from increased R&D investments

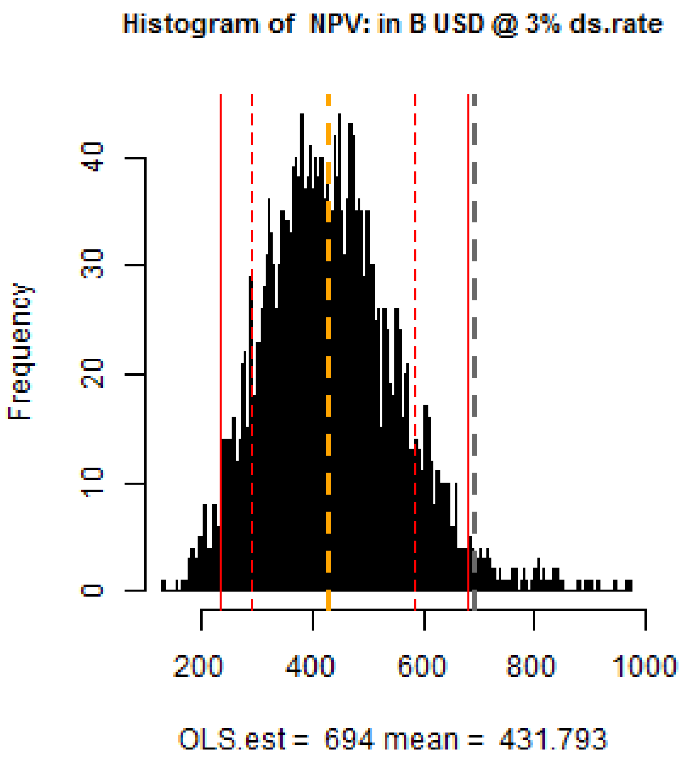
log model

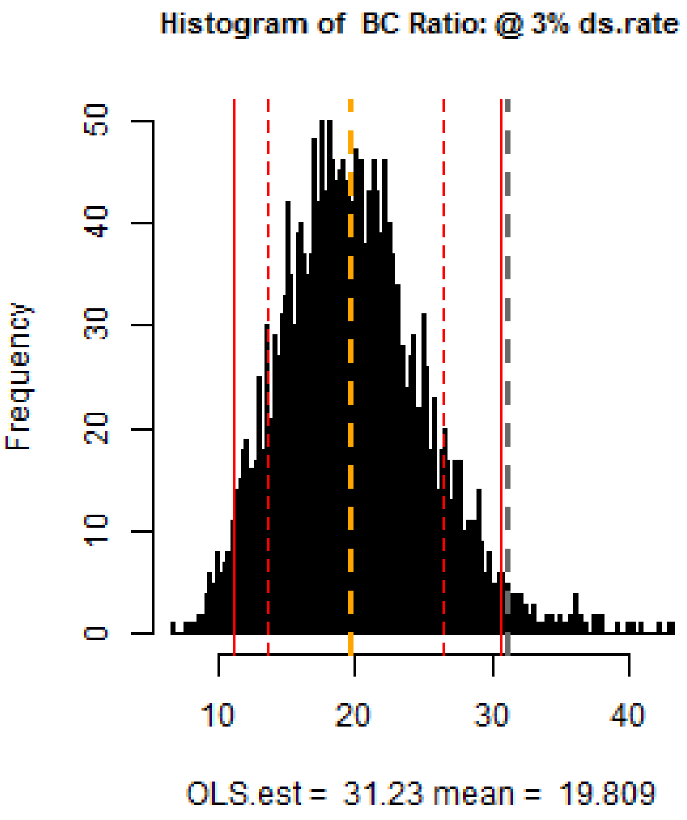


Counterfactual Analysis:
+30% increase in U.S. Public R&D in 1950-59 only

- Net Present Values and Benefit Cost Ratios (our OLS results in grey and Bayesian results in yellow)

log model





Preliminary Results:

log model

U.S. Public R&D experience : 1949-2011 using USDA data

- Reconstructed TFP Index (USDA data: black vs. our results)
- A crucial "validation metric" : Empirical Coverage Probability
  - # data points inside yellow band must be roughly 95% of total # data points