# SIMULATED POSTERIOR PREDICTIVE CHECKS
## FOR MIXTURE MODEL SELECTION

THOMAS SPROUL, UNIVERSITY OF RHODE ISLAND
JOSHUA WOODARD, CORNELL UNIVERSITY
APA CONFERENCE, UNIVERSITY OF WATERLOO
DECEMBER 2, 2017

# BACKGROUND

**Finite Gaussian Mixture Models are a data mining tool used to identify clustering in unlabeled data.**

**They are increasingly being used in applied econometrics:**

- Agricultural risk modeling/crop insurance pricing
- Behavioral economic models/insurance demand

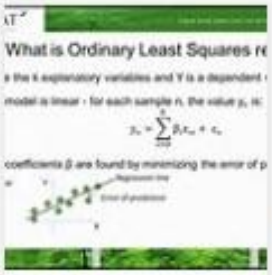
**But they have some drawbacks:**

- May not scale well to large data sets
- Picking the number of components (clusters) is hard

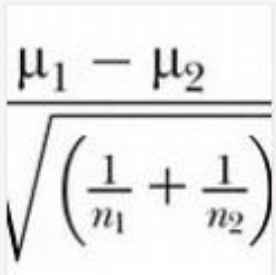
# MIXED MODELS

Some search results from **bing.com**…

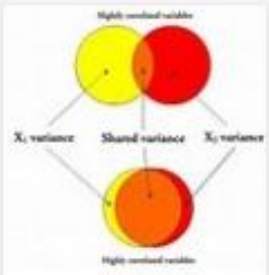
People interested in **linear regression** also searched for



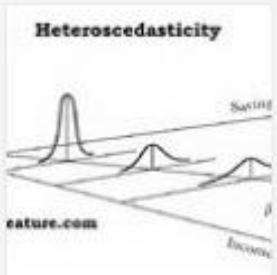

Ordinary Least Squares



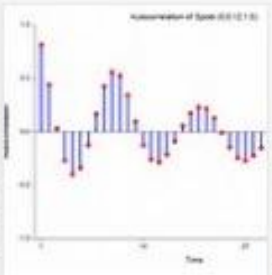

Student's T-Test




Multicollinearity




Heteroscedas…



Autocorrelation

Mixed Model

# BACKGROUND II

**Basic Model Selection Process:**

- Pick a number of mixture components, C (e.g., log N)
- Specify covariance structure of the (MVN) components.
- Fit using Expectation-Maximization:

$$\text{E-Step}: \ \pi_{ic} = \varphi_c\left(X_i\right) \Big/ \sum_{j=1}^{C} \varphi_j\left(X_i\right)$$

$$\text{M-Step}: \ \max_{\theta|\pi} LL\left(\theta \mid X\right) = \sum_i \ln\left(\sum_c \pi_{ic} \varphi_c\left(X_i \mid \theta\right)\right)$$

- Model selection via cross validation (LOO, K-fold) or information criteria (AIC, BIC, ICL).
- Alternatives:
  - DPGMM can pick C, but will require tuning.
  - *ad hoc*: "agreement" across procedures; $\pi_{ic}$ near 0 or 1.

# OUR APPROACH

**We simulate from fitted candidate models to estimate posterior probabilities that each model is the "true" one.**

**Goals:**

- robust inference about a population from a sample
- suitability of mixture models for a given domain

**Our results suggest:**

- mixture model selection is hard
- some published results may be in question

# SAMPLE DATA

**USDA/NASS Corn and Soybeans from IL, Wheat from KS.**

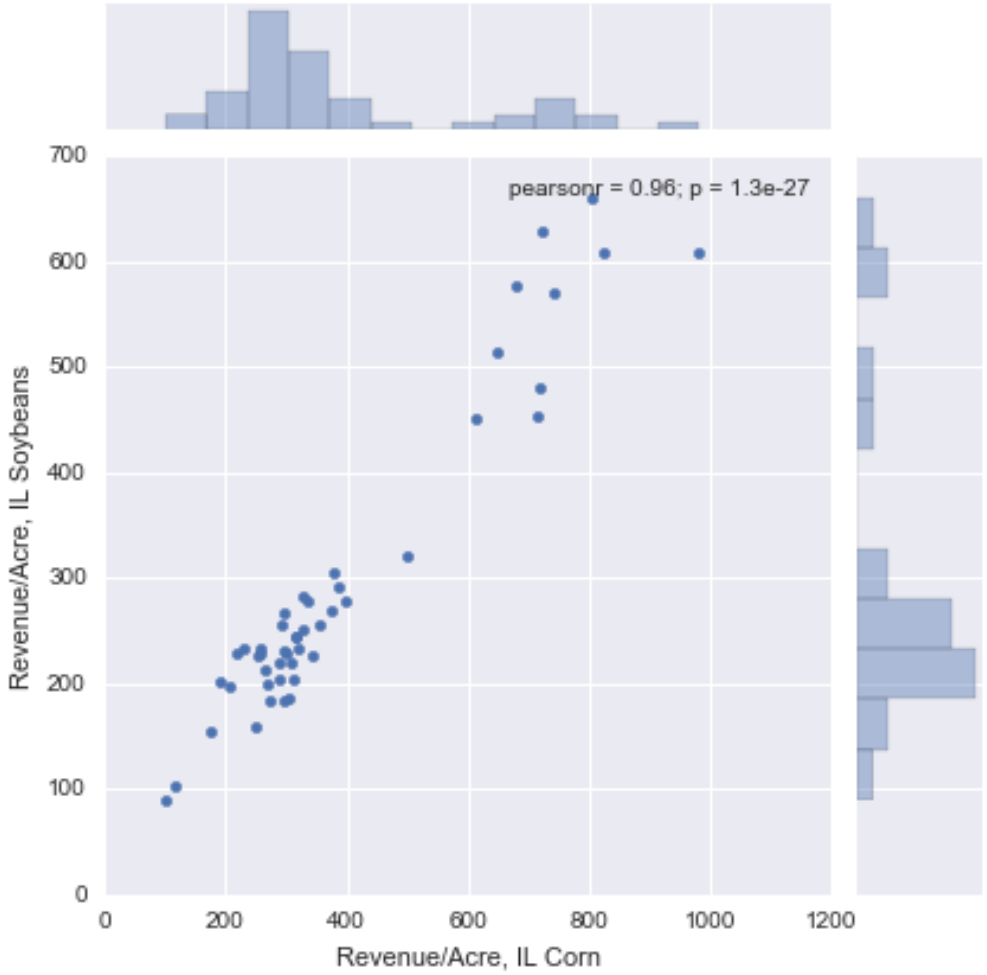
- State-level averages for Yield/Acre and Revenue/Acre.
- 47 annual observations, 1970-2016.

**Two sample applications:**

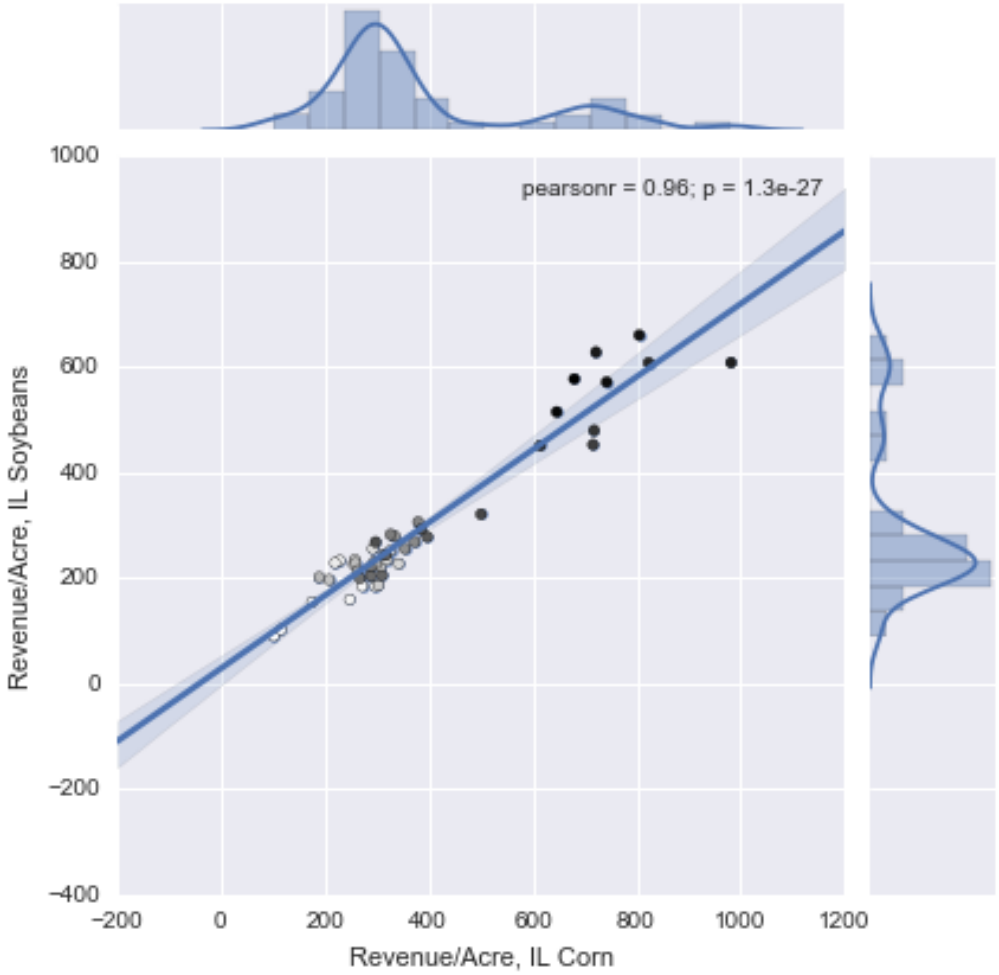
- IL Corn and IL Soybeans, Revenue/Acre
  - 0.96 linear correlation, but has apparent clustering
  - possibility of structural break due to 2006 RFS
- IL Corn and KS Wheat, Yield/Acre
  - 0.35 linear correlation, clustering not obvious if it exists

Proof of concept only - nominal prices, no trend adjustment, etc.

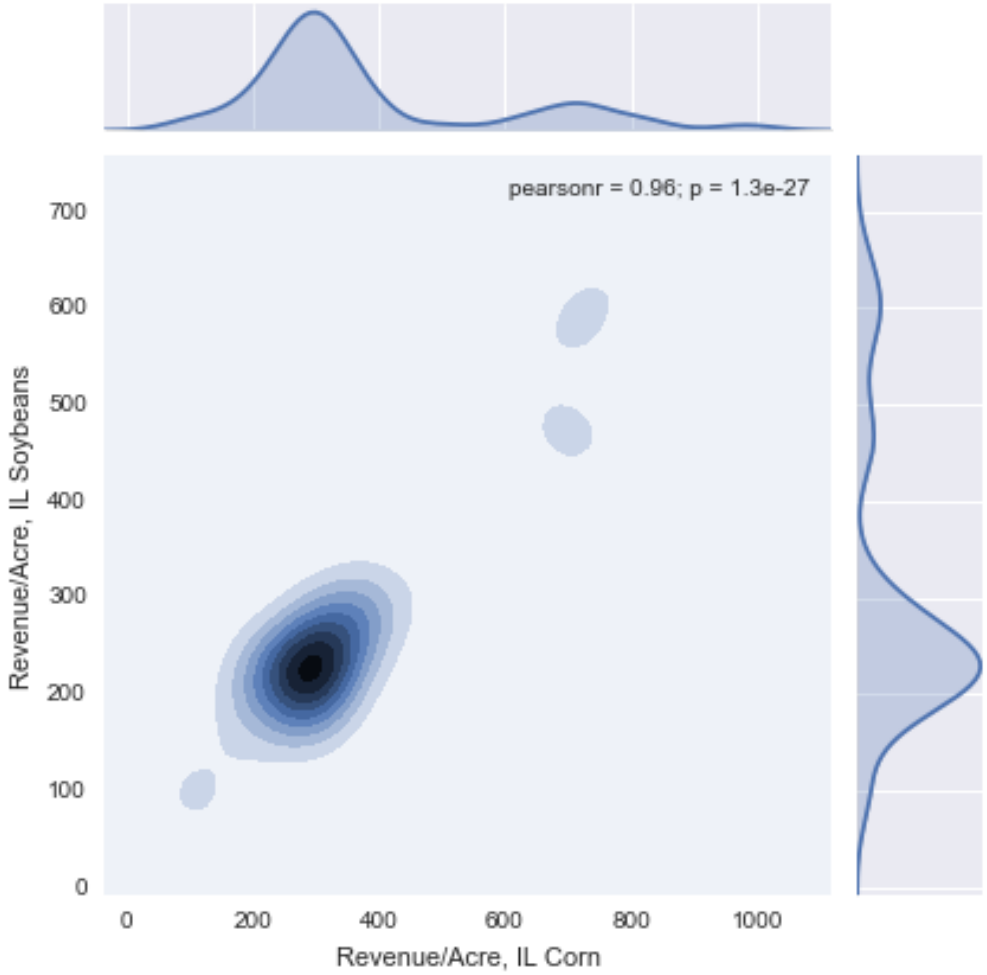
# CORN AND BEANS



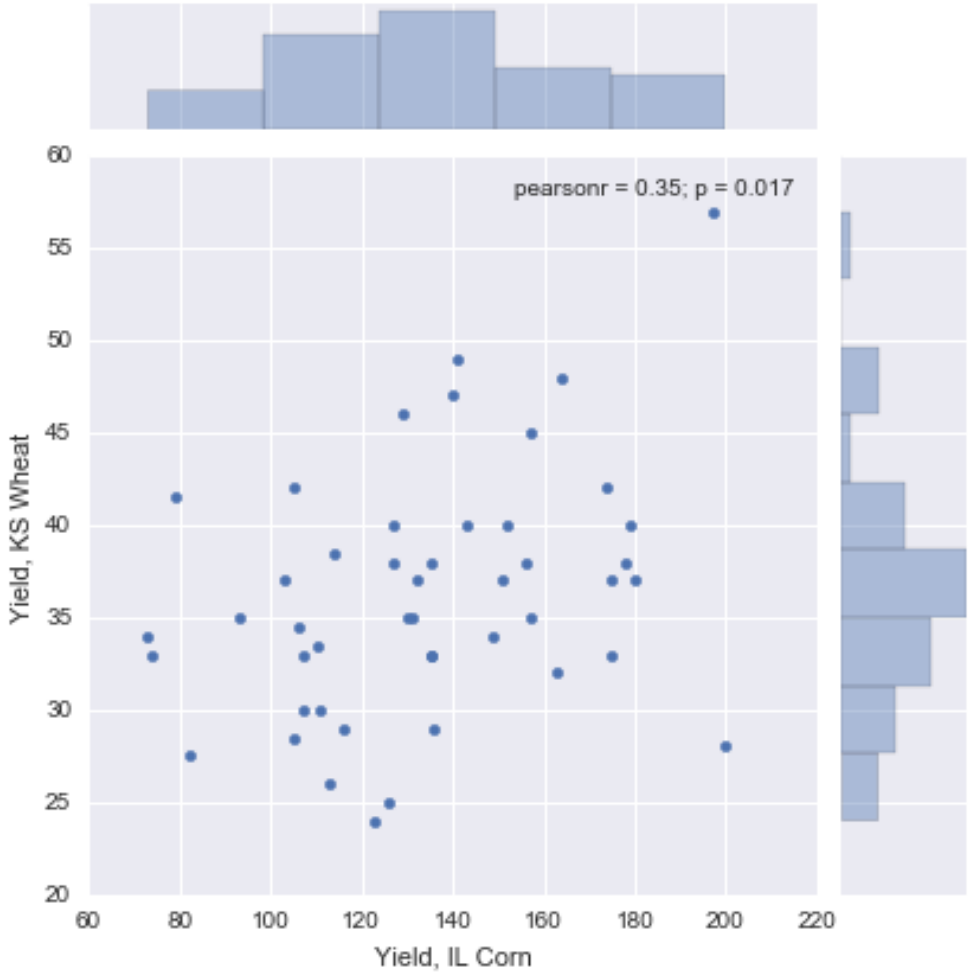
# CORN AND BEANS



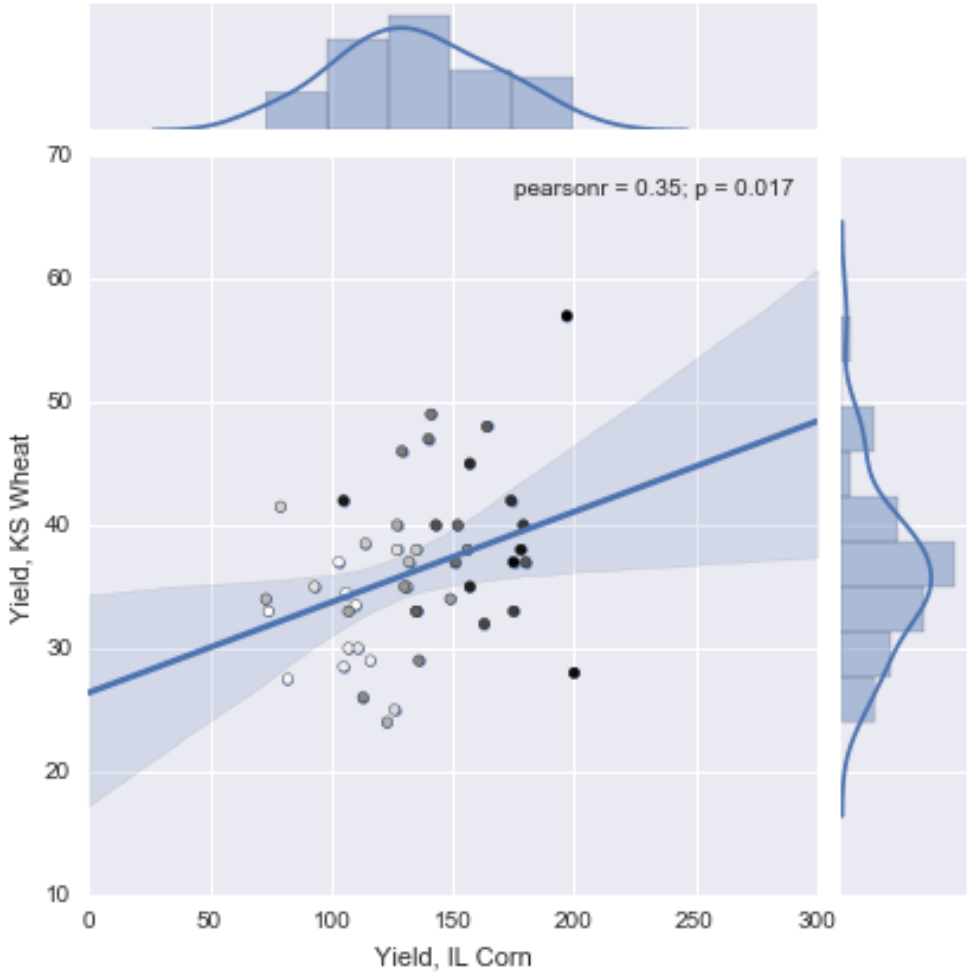
# CORN AND BEANS



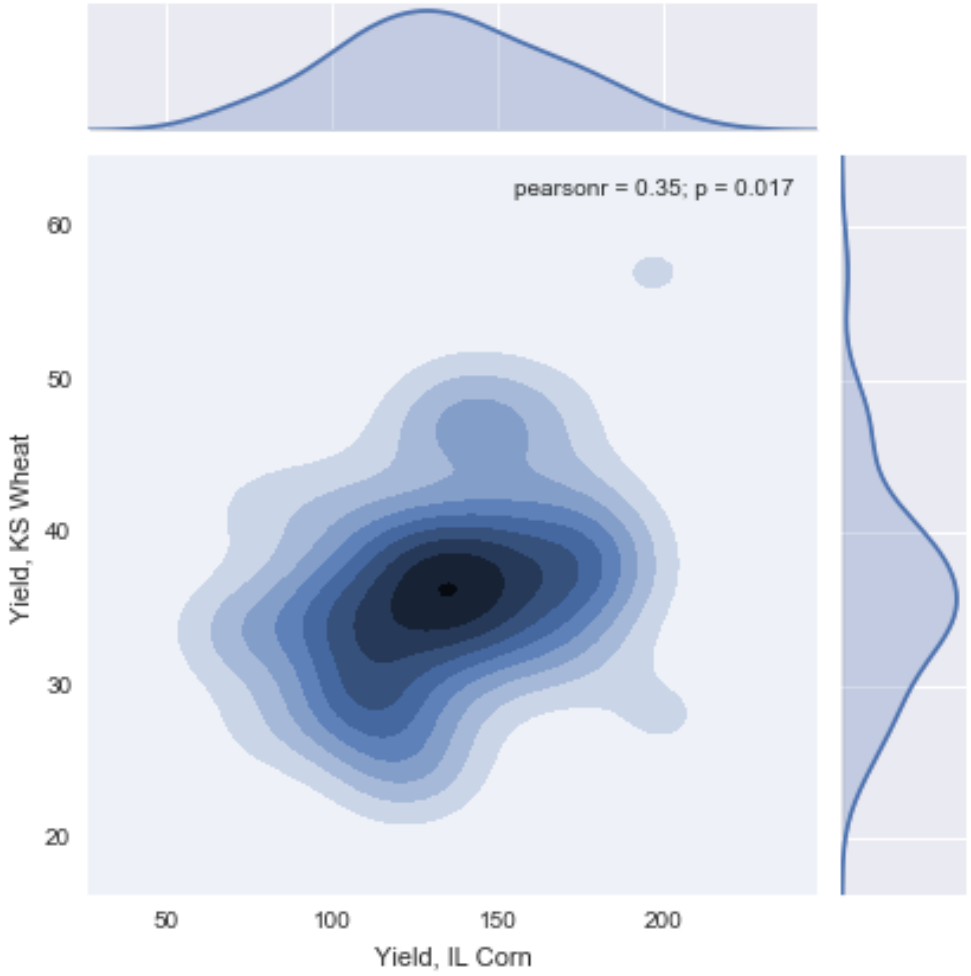
# CORN AND WHEAT



# CORN AND WHEAT



# CORN AND WHEAT



# A FIRST EXAMPLE

Take the data, **X**, and the set of candidate models, **M**, as given.

Take also a model selection procedure as given which selects only one model, *s*.

For two models, *m* and *n* in M, we want to know the probability that *n* is the "true model" *t*, given *m* chosen:

$$\Pr(t = n \mid s = m) = \frac{\Pr(s = m \mid t = n) \cdot \Pr(t = n)}{\Pr(s = m)}$$

$$\propto \Pr(s = m \mid t = n) \cdot \Pr(t = n)$$

If all candidate models have equal prior probability, then this is proportional to the probability of *m* being chosen given *n* is true.

# SIMPLE MODEL SELECTION

1. Set upper bound B = 4 ≈ log(47), where C ≤ B.
2. Pick model selection procedure.
3. Fit models and compare C*:

| Procedure | Corn/Beans | Corn/Wheat |
|---|---|---|
| AIC | 4 | 4 |
| BIC | 4 | 2 |
| ICL | 2 | 2 |
| LOOCV | 3 | 4 |
| 10FCV | 2 | 2 |

4. Now what?

# AD HOC APPROACHES

Corn and Beans

| Procedures | C | π > 0.90 | π > 0.95 | π > 0.99 | Weights |
|---|---|---|---|---|---|
| AIC, BIC | 4 | 0.96 | 0.94 | 0.94 | .75, .11, .10, .04 |
| LOOCV | 3 | 0.96 | 0.96 | 0.94 | .79, .11, .11 |
| ICL, 10FCV | 2 | 1.00 | 1.00 | 0.98 | .79, .21 |

Corn and Wheat

| Procedures | C | π > 0.90 | π > 0.95 | π > 0.99 | Weights |
|---|---|---|---|---|---|
| AIC, LOOCV | 4 | 0.66 | 0.57 | 0.40 | .41, .22, .20, .18 |
| – | 3 | 0.60 | 0.45 | 0.19 | .45, .34, .22 |
| BIC, ICL, 10FCV | 2 | 0.64 | 0.51 | 0.17 | .57, .43 |

# CONDITIONAL PROBS

1. For each candidate model and procedure, pick C in {1,…,B}
2. Simulate the distribution of *s*, given *t* is true.

**Pr(*s*|*t* ) using AIC, Corn and Beans Data**

| *t* / *s* | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.81 | 0.14 | 0.04 | 0.01 |
| 2 | 0.05 | 0.93 | 0.02 | 0.00 |
| 3 | 0.06 | 0.75 | 0.17 | 0.03 |
| 4 | 0.07 | 0.64 | 0.21 | 0.09 |

Thus, Pr(t = 4|s = 4) = 0.70 (0.09 normalized by sum of column 4).

However, note that *s* = *4* is itself an unlikely event.

# CONDITIONAL PROBS II

**Pr(*s*|*t* ) using 10FCV, Corn and Beans Data**

| *t* / *s* | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.97 | 0.03 | 0.00 | 0.00 |
| 2 | 0.03 | 0.96 | 0.012 | 0.00 |
| 3 | 0.03 | 0.51 | 0.35 | 0.11 |
| 4 | 0.04 | 0.40 | 0.31 | 0.26 |

**Pr(*s*|*t* ) using BIC, Corn and Wheat Data**

| *t* / *s* | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.99 | 0.01 | 0.00 | 0.00 |
| 3 | 0.99 | 0.01 | 0.00 | 0.00 |
| 4 | 0.98 | 0.02 | 0.00 | 0.00 |

# CHECKING ACROSS PROCEDURES

Pr(t|s) by selection procedure, Corn and Beans

| | | *t* | | | |
|---|---|---|---|---|---|
| Procedure | *s* | 1 | 2 | 3 | 4 |
| AIC | 4 | 0.09 | 0.00 | 0.21 | 0.70 |
| BIC | 4 | 0.00 | 0.00 | 0.18 | 0.81 |
| ICL | 2 | 0.00 | 0.35 | 0.34 | 0.32 |
| LOOCV | 3 | 0.01 | 0.07 | 0.50 | 0.43 |
| 10FCV | 2 | 0.01 | 0.51 | 0.27 | 0.21 |

Under AIC, the theoretical relative likelihood is $L_3/L_4 = \exp(-10.46)$ but our posterior value here is $L_3/L_4 = 0.3 = \exp(-1.20)$.

# CHECKING ACROSS PROCEDURES

Pr(t|s) by selection procedure, Corn and Wheat

| | | *t* | | | |
|---|---|---|---|---|---|
| Procedure | *s* | 1 | 2 | 3 | 4 |
| AIC | 4 | 0.04 | 0.04 | 0.46 | 0.46 |
| BIC | 2 | 0.00 | 0.22 | 0.27 | 0.51 |
| ICL | 2 | 0.00 | 0.18 | 0.25 | 0.57 |
| LOOCV | 4 | 0.00 | 0.02 | 0.22 | 0.76 |
| 10FCV | 2 | 0.02 | 0.45 | 0.28 | 0.25 |

# ANOTHER EXAMPLE

For robust inference, we must address the potential for sampling variability/outliers to influence *s*, as well as uncertainty in the estimation routines.

We do so by bootstrapping the model selection process to get an IID sample, *S*, over values of *s,* conditional on data.

$$\Pr(t=n \mid S) = \frac{\Pr(S \mid t=n)\cdot \Pr(t=n)}{\Pr(S)}$$

$$\propto \Pr(S \mid t=n)\cdot \Pr(t=n)$$

Here, Pr(S|t=n) is the likelihood of our sample, given knowledge of P(s=m|t=n) from simulation.

# BOOTSTRAPPED SELECTION

Pr(s=m) by selection procedure, Corn and Beans (N=1,000)

| | *s* | | | |
|---|---|---|---|---|
| Procedure | 1 | 2 | 3 | 4 |
| AIC | 0.00 | 0.00 | 0.11 | 0.89 |
| BIC | 0.00 | 0.05 | 0.26 | 0.69 |
| ICL | 0.00 | 0.50 | 0.25 | 0.25 |
| LOOCV | 0.01 | 0.25 | 0.33 | 0.41 |
| 10FCV | 0.01 | 0.37 | 0.23 | 0.39 |

# BOOTSTRAPPED SELECTION II

Pr(s=m) by selection procedure, Corn and Wheat (N=1,000)

| | *s* | | | |
|---|---|---|---|---|
| Procedure | 1 | 2 | 3 | 4 |
| AIC | 0.07 | 0.13 | 0.22 | 0.59 |
| BIC | 0.51 | 0.16 | 0.15 | 0.19 |
| ICL | 0.93 | 0.03 | 0.03 | 0.01 |
| LOOCV | 0.27 | 0.30 | 0.18 | 0.25 |
| 10FCV | 0.38 | 0.31 | 0.13 | 0.18 |

# POSTERIORS

Pr(t|S) by selection procedure, Corn and Beans

| | *t* | | | |
|---|---|---|---|---|
| Procedure | 1 | 2 | 3 | 4 |
| AIC | 0.00 | 0.00 | 0.00 | 1.00 |
| BIC | 0.00 | 0.00 | 0.00 | 1.00 |
| ICL | 0.00 | 1.00 | 0.00 | 0.00 |
| LOOCV | 0.00 | 0.00 | 1.00 | 0.00 |
| 10FCV | 0.00 | 1.00 | 0.00 | 0.00 |

# POSTERIORS II

Pr(t|S) by selection procedure (N=100), Corn and Wheat

| | *t* | | | |
|---|---|---|---|---|
| Procedure | 1 | 2 | 3 | 4 |
| AIC | 0.00 | 0.68 | 0.00 | 0.32 |
| BIC | 0.59 | 0.00 | 0.00 | 0.41 |
| ICL | 0.49 | 0.00 | 0.26 | 0.26 |
| LOOCV | 1.00 | 0.00 | 0.00 | 0.00 |
| 10FCV | 1.00 | 0.00 | 0.00 | 0.00 |

# POSTERIORS IIA

Pr(t|S) by selection procedure (N=500), Corn and Wheat

| | *t* | | | |
|---|---|---|---|---|
| Procedure | 1 | 2 | 3 | 4 |
| AIC | 0.00 | 0.98 | 0.00 | 0.02 |
| BIC | 0.85 | 0.00 | 0.00 | 0.15 |
| ICL | 0.93 | 0.00 | 0.04 | 0.04 |
| LOOCV | 1.00 | 0.00 | 0.00 | 0.00 |
| 10FCV | 1.00 | 0.00 | 0.00 | 0.00 |

# DISCUSSION

**We demonstrate a simulation/bootstrap-based procedure to generate posterior model probabilities for model selection in Finite Gaussian Mixture Models.**

- We show that simulated posteriors can converge to models not selected by EM or bootstrapping alone, as well as models not suggested by use of a naïve posterior.
- We don't solve the problem that selection procedures often do not agree on the final model.

**Future work:**

- Out of sample robustness checks
- Simulated data set with known DGP