Model Selection and Its Important Roles in Surveys

Jiming Jiang

University of California, Davis, USA

Nov. 14, 2023, Washington, D.C.

The 2023 Morris Hansen Lecture

Outline

1. Introduction

- The 21st Century has seen a flourish of statistical data science.

1. Introduction

- The 21st Century has seen a flourish of statistical data science.
- This is driven by an unprecedented capability of collecting data, or information.

1. Introduction

- The 21st Century has seen a flourish of statistical data science.
- This is driven by an unprecedented capability of collecting data, or information.
- Sometimes, there is too much "information";

1. Introduction

- The 21st Century has seen a flourish of statistical data science.
- This is driven by an unprecedented capability of collecting data, or information.
- Sometimes, there is too much "information";
- yet, in some other cases, there is (still) not enough.

1. Introduction

- The 21st Century has seen a flourish of statistical data science.
- This is driven by an unprecedented capability of collecting data, or information.
- Sometimes, there is too much "information";
- yet, in some other cases, there is (still) not enough.
- In any cases, especially in the latter, the idea of "borrowing strength" comes along.

- The term was originated from Tukey back in the 1960s.

- The term was originated from Tukey back in the 1960s.
- An ancient story from "Three kingdoms" (220–280 AD)

- The term was originated from Tukey back in the 1960s.
- An ancient story from "Three kingdoms" (220–280 AD)
- A "modern" story: IQ test (hypothetical; Mood *et al.* 1974, 370)

- The term was originated from Tukey back in the 1960s.
- An ancient story from “Three kingdoms” (220–280 AD)
- A “modern” story: IQ test (hypothetical; Mood *et al.* 1974, 370)
- Suppose the IQ of students in a particular age group are normally distributed with mean 100 and s.d. 15.

- The term was originated from Tukey back in the 1960s.
- An ancient story from "Three kingdoms" (220–280 AD)
- A "modern" story: IQ test (hypothetical; Mood *et al.* 1974, 370)
- Suppose the IQ of students in a particular age group are normally distributed with mean 100 and s.d. 15.
- It is also known that, for a given student, the test scores are normally distributed with mean equal to the student's IQ and s.d. equal to 5.

- Suppose that a student just took an IQ test and scored 130.

- Suppose that a student just took an IQ test and scored 130.
- What is the best prediction of the student's IQ?

- Suppose that a student just took an IQ test and scored 130.
- What is the best prediction of the student's IQ?
- The answer is not 130.

- Suppose that a student just took an IQ test and scored 130.
- What is the best prediction of the student's IQ?
- The answer is not 130.
- Why not?

- Suppose that a student just took an IQ test and scored 130.
- What is the best prediction of the student's IQ?
- The answer is not 130.
- Why not?
- If I had not told you the previous stories, the answer would be 130. But ...

- Suppose that a student just took an IQ test and scored 130.
- What is the best prediction of the student's IQ?
- The answer is not 130.
- Why not?
- If I had not told you the previous stories, the answer would be 130. But ...
- What stories?

- The stories provide additional information.

- The stories provide additional information.
- This, combined with the observed data, 130, leads to the best prediction of the person's IQ: 127.

- The stories provide additional information.
- This, combined with the observed data, 130, leads to the best prediction of the person's IQ: 127.
- Don't ask me how I get it (conditional expectation under Gaussian distribution).

- The stories provide additional information.
- This, combined with the observed data, 130, leads to the best prediction of the person's IQ: 127.
- Don't ask me how I get it (conditional expectation under Gaussian distribution).
- Bottom line: One can do better with more information — a simple, common-sense idea.

- The stories provide additional information.
- This, combined with the observed data, 130, leads to the best prediction of the person's IQ: 127.
- Don't ask me how I get it (conditional expectation under Gaussian distribution).
- Bottom line: One can do better with more information — a simple, common-sense idea.
- How to borrow strength?

- Some "common sense" lines:

- Some "common sense" lines:
- "If you know more, you can do better"

- Some "common sense" lines:
- "If you know more, you can do better"
- "Doing something is usually (not always) better than doing nothing"

- Some "common sense" lines:
- "If you know more, you can do better"
- "Doing something is usually (not always) better than doing nothing"
- "There's no free lunch"

- Some "common sense" lines:
- "If you know more, you can do better"
- "Doing something is usually (not always) better than doing nothing"
- "There's no free lunch"
- We can borrow strength by utilizing a statistical model.

- Some "common sense" lines:
- "If you know more, you can do better"
- "Doing something is usually (not always) better than doing nothing"
- "There's no free lunch"
- We can borrow strength by utilizing a statistical model.
- A statistical model allows you to, all of a sudden, known a lot more; but you don't know that for sure.

- Some "common sense" lines:
- "If you know more, you can do better"
- "Doing something is usually (not always) better than doing nothing"
- "There's no free lunch"
- We can borrow strength by utilizing a statistical model.
- A statistical model allows you to, all of a sudden, known a lot more; but you don't know that for sure.
- Therefore, there is a little bit of "gamble", but it's an educated gamble.

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).

- Little (2008): Two approaches to weighting the survey data, $\bar{y}_w = \sum_{i=1}^{n} w_i y_i$.

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).

- Little (2008): Two approaches to weighting the survey data, $\bar{y}_w = \sum_{i=1}^{n} w_i y_i$.

- 1. Model-based (linear regression): $y_i = x_i' \beta + \epsilon_i$, $\text{var}(\epsilon_i) = \sigma^2 / a_i$, a_i known. Choose $w_i \propto a_i$.

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).
- Little (2008): Two approaches to weighting the survey data, $\bar{y}_w = \sum_{i=1}^{n} w_i y_i$.
- 1. Model-based (linear regression): $y_i = x_i'\beta + \epsilon_i$, $\mathrm{var}(\epsilon_i) = \sigma^2/a_i$, a_i known. Choose $w_i \propto a_i$.
- 2. Design-based: Choose $w_i \propto 1/\pi_i$, π_i = selection probability.

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).

- Little (2008): Two approaches to weighting the survey data, $\bar{y}_w = \sum_{i=1}^{n} w_i y_i$.

- 1. Model-based (linear regression): $y_i = x_i'\beta + \epsilon_i$, $\text{var}(\epsilon_i) = \sigma^2/a_i$, a_i known. Choose $w_i \propto a_i$.
- 2. Design-based: Choose $w_i \propto 1/\pi_i$, π_i = selection probability.

- Which one is correct?

2. Statistical modeling in surveys

- The use of statistical models in surveys is well established (e.g., Fuller 2009, ch. 5, Lohr 2022, ch. 11).

- Little (2008): Two approaches to weighting the survey data, $\bar{y}_w = \sum_{i=1}^{n} w_i y_i$.

- 1. Model-based (linear regression): $y_i = x_i'\beta + \epsilon_i$, $\mathrm{var}(\epsilon_i) = \sigma^2/a_i$, a_i known. Choose $w_i \propto a_i$.
- 2. Design-based: Choose $w_i \propto 1/\pi_i$, π_i = selection probability.

- Which one is correct?

- Does it matter?

- Does truth, or existence of truth, matter?

- Does truth, or existence of truth, matter?
- "You can't handle the truth" — Jack Nicholson, "A Few Good Men"

- Does truth, or existence of truth, matter?
- "You can't handle the truth" — Jack Nicholson, "A Few Good Men"
- "All models are wrong ... some are useful" — George Box

- Does truth, or existence of truth, matter?
- "You can't handle the truth" — Jack Nicholson, "A Few Good Men"
- "All models are wrong ... some are useful" — George Box
- "It doesn't matter whether a cat is white or black, as long as it catches mice" — Deng Xiaoping, former Chinese leader

- Does truth, or existence of truth, matter?
- "You can't handle the truth" — Jack Nicholson, "A Few Good Men"
- "All models are wrong ... some are useful" — George Box
- "It doesn't matter whether a cat is white or black, as long as it catches mice" — Deng Xiaoping, former Chinese leader
- It doesn't matter whether a statistical model is right or wrong, as long as it helps.

- How do we know a model is a good one (if there is no "right" model)?

- How do we know a model is a good one (if there is no "right" model)?
- Short-term and long-term answers.

- How do we know a model is a good one (if there is no "right" model)?
- Short-term and long-term answers.
- Short-term: The model has to be reasonable by the common sense.

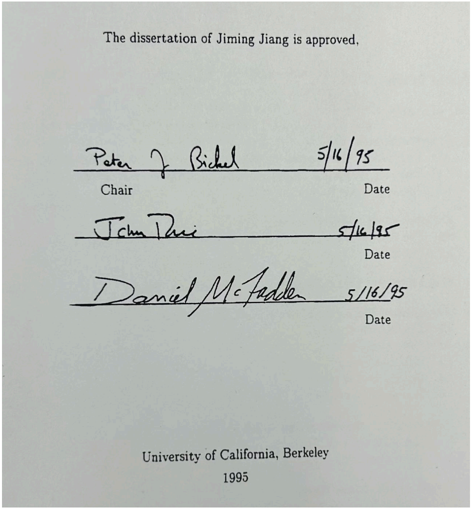
- How do we know a model is a good one (if there is no "right" model)?
- Short-term and long-term answers.
- Short-term: The model has to be reasonable by the common sense.
- Example. Discrete response model (e.g., McFadden & Kenneth 2000).

- How do we know a model is a good one (if there is no "right" model)?
- Short-term and long-term answers.
- Short-term: The model has to be reasonable by the common sense.
- Example. Discrete response model (e.g., McFadden & Kenneth 2000).
- Long-term: "Forever green is the tree of life" (Johann Wolfgang von Goethe, *Faust: Part I*).

- How do we know a model is a good one (if there is no "right" model)?
- Short-term and long-term answers.
- Short-term: The model has to be reasonable by the common sense.
- Example. Discrete response model (e.g., McFadden & Kenneth 2000).
- Long-term: "Forever green is the tree of life" (Johann Wolfgang von Goethe, *Faust: Part I*).
- Example (cont.). Many real-life applications in Economics (leading to the 2000 Nobel Prize in Economical Sciences), and other fields.

- How do we know a model is a good one (if there is no “right” model)?
- Short-term and long-term answers.
- Short-term: The model has to be reasonable by the common sense.
- Example. Discrete response model (e.g., McFadden & Kenneth 2000).
- Long-term: “Forever green is the tree of life” (Johann Wolfgang von Goethe, *Faust: Part I*).
- Example (cont.). Many real-life applications in Economics (leading to the 2000 Nobel Prize in Economical Sciences), and other fields.
- Example 2. Small area estimation (later).

- Signature page of my Ph. D. thesis (UC Berkeley, 1995)



The dissertation of Jiming Jiang is approved,

Peter J. Bickel 5/16/95

Chair Date

John Rice 5/16/95

Date

Daniel McFadden 5/16/95

Date

University of California, Berkeley

1995

- Previous Morris Hansen Lectures: Use of models in surveys

- Previous Morris Hansen Lectures: Use of models in surveys
- Graham Kalton (2000) discussed what he called model-dependent methods; identified three areas in which models are particularly useful: SAE, missing data, and variance estimation.

- Previous Morris Hansen Lectures: Use of models in surveys
- Graham Kalton (2000) discussed what he called model-dependent methods; identified three areas in which models are particularly useful: SAE, missing data, and variance estimation.
- He discussed, in particular, the SAIPE (Small Area Income & Poverty Estimation) program and its evaluations.

- Previous Morris Hansen Lectures: Use of models in surveys
- Graham Kalton (2000) discussed what he called model-dependent methods; identified three areas in which models are particularly useful: SAE, missing data, and variance estimation.
- He discussed, in particular, the SAIPE (Small Area Income & Poverty Estimation) program and its evaluations.
- Concluded that "careful attention needs to be paid to the development of an appropriate model and its evaluation".

- Previous Morris Hansen Lectures: Use of models in surveys
- Graham Kalton (2000) discussed what he called model-dependent methods; identified three areas in which models are particularly useful: SAE, missing data, and variance estimation.
- He discussed, in particular, the SAIPE (Small Area Income & Poverty Estimation) program and its evaluations.
- Concluded that "careful attention needs to be paid to the development of an appropriate model and its evaluation".
- Rick Valliant (2022) discussed history of explicit models used in sample design and estimation, including an earlier paper of Hansen (1961), in which the author found purely design-based approach was inadequate for analyzing surveys subject to non-responses and other problems.

- Two major variants of statistical modelling (Little 2008)

- Two major variants of statistical modelling (Little 2008)
- 1. The super-population (SP) modelling

- Two major variants of statistical modelling (Little 2008)
- 1. The super-population (SP) modelling
- 2. Bayesian modelling

- Two major variants of statistical modelling (Little 2008)
- 1. The super-population (SP) modelling
- 2. Bayesian modelling
- Under an SP model, the finite (real) population is assumed to be a random sample from a "super-population".

- Two major variants of statistical modelling (Little 2008)
- 1. The super-population (SP) modelling
- 2. Bayesian modelling
- Under an SP model, the finite (real) population is assumed to be a random sample from a "super-population".
- A random sample, Y, from the super-population is assumed to have distribution $p(Y|\theta)$, where the probability distribution, $p(\cdot|\theta)$, is specified under an assumed model (e.g., regression) and θ is the vector of parameters under the model.

- Here, we focus on parametric models for simplicity.

- Here, we focus on parametric models for simplicity.
- Under a Bayesian model, there is additional prior information about θ, in terms of a prior distribution, $\pi(\theta)$.

- Here, we focus on parametric models for simplicity.
- Under a Bayesian model, there is additional prior information about θ, in terms of a prior distribution, $\pi(\theta)$.
- The prior distribution for Y is then expressed as

$$\pi(Y) = \int p(Y|\theta)\pi(\theta)d\theta.$$

- Here, we focus on parametric models for simplicity.
- Under a Bayesian model, there is additional prior information about θ, in terms of a prior distribution, $\pi(\theta)$.
- The prior distribution for Y is then expressed as

$$\pi(Y) = \int p(Y|\theta)\pi(\theta)d\theta.$$

- The posterior inference about the non-sampled part of the population is then carried out via the Bayes' Theorem.

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)
- "Selection" is the key word.

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)
- "Selection" is the key word.
- Example: Choice of weights.

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)
- "Selection" is the key word.
- Example: Choice of weights.
- Traditional approaches: Information criteria, shrinkage selection/estimation (e.g., Müller *et al.* 2013).

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)
- "Selection" is the key word.
- Example: Choice of weights.
- Traditional approaches: Information criteria, shrinkage selection/estimation (e.g., Müller *et al.* 2013).
- Big question: Is model selection a statistical problem?

3. Model selection

- Needs for model selection (e.g., Lahiri 2001, Pfeffermann 2013, Lumley and Scott 2017)
- "Selection" is the key word.
- Example: Choice of weights.
- Traditional approaches: Information criteria, shrinkage selection/estimation (e.g., Müller *et al.* 2013).
- Big question: Is model selection a statistical problem?
- Answer: No, it's not.

- If not, why is a criterion for model selection, such as the AIC, solely based on statistical consideration?

- If not, why is a criterion for model selection, such as the AIC, solely based on statistical consideration?
- Practical considerations, specific to the real-life problem we're dealing with, must be taken into consideration.

- If not, why is a criterion for model selection, such as the AIC, solely based on statistical consideration?
- Practical considerations, specific to the real-life problem we're dealing with, must be taken into consideration.
- How?

- If not, why is a criterion for model selection, such as the AIC, solely based on statistical consideration?
- Practical considerations, specific to the real-life problem we're dealing with, must be taken into consideration.
- How?
- The fence methods (Jiang *et al.* 2008); also see Müllet *et al.* 2013, Pfeffermann 2013, Rao & Molina 2015, Jiang & Nguyen 2016).

- If not, why is a criterion for model selection, such as the AIC, solely based on statistical consideration?
- Practical considerations, specific to the real-life problem we're dealing with, must be taken into consideration.
- How?
- The fence methods (Jiang *et al.* 2008); also see Müllet *et al.* 2013, Pfeffermann 2013, Rao & Molina 2015, Jiang & Nguyen 2016).
- Idea: 1. Build a "statistical fence" to satisfy statistical consideration of model fitting. The fence isolates a subset of candidate models that meet the model-fitting threshold.

- 2. Within the fence, incorporate other considerations to identify the optimal model.

- 2. Within the fence, incorporate other considerations to identify the optimal model.

- For example, parsimony is one “other” consideration that is often used;

- 2. Within the fence, incorporate other considerations to identify the optimal model.
- For example, parsimony is one “other” consideration that is often used;
- but (here is the key) practical considerations can also be incorporated in searching for the optimal model within the fence.

- 2. Within the fence, incorporate other considerations to identify the optimal model.
- For example, parsimony is one "other" consideration that is often used;
- but (here is the key) practical considerations can also be incorporated in searching for the optimal model within the fence.
- Such practical considerations can be scientific, economical, legal, or political (e.g., the model must not require privacy-protected information to "train").

- 2. Within the fence, incorporate other considerations to identify the optimal model.
- For example, parsimony is one "other" consideration that is often used;
- but (here is the key) practical considerations can also be incorporated in searching for the optimal model within the fence.
- Such practical considerations can be scientific, economical, legal, or political (e.g., the model must not require privacy-protected information to "train").
- 3. Finally, the threshold of the fence may be determined based on the principle of "letting the data speak".

- Building the fence:

$$Q(M) - Q(M_*) \leq c,$$

where $Q(\cdot)$ is a measure of lack-of-fit, M is a candidate model, and M_* is a (candidate) model that is optimal in terms of model fitting (i.e., one that minimizes Q).

- Building the fence:
$$Q(M) - Q(M_*) \leq c,$$
where $Q(\cdot)$ is a measure of lack-of-fit, M is a candidate model, and M_* is a (candidate) model that is optimal in terms of model fitting (i.e., one that minimizes Q).

- For example, in regression model selection, let $M_{\rm f}$ denotes the full model, that is, the model that includes all of the candidate predictors.

- Building the fence:

$$Q(M) - Q(M_*) \leq c,$$

where $Q(\cdot)$ is a measure of lack-of-fit, M is a candidate model, and M_* is a (candidate) model that is optimal in terms of model fitting (i.e., one that minimizes Q).

- For example, in regression model selection, let M_{f} denotes the full model, that is, the model that includes all of the candidate predictors.

- If Q is the SSR (sum of squares of residuals), then $M_* = M_{\mathrm{f}}$.

- Building the fence:

$$Q(M) - Q(M_*) \le c,$$

where $Q(\cdot)$ is a measure of lack-of-fit, M is a candidate model, and M_* is a (candidate) model that is optimal in terms of model fitting (i.e., one that minimizes Q).

- For example, in regression model selection, let $M_{\rm f}$ denotes the full model, that is, the model that includes all of the candidate predictors.

- If Q is the SSR (sum of squares of residuals), then $M_* = M_{\rm f}$.

- Another popular choice of Q is the negative log-likelihood, which applies beyond the linear models. See Jiang & Nguyen (2016) for other examples of Q.

- A challenging problem: How to choose c, the threshold of the fence?

- A challenging problem: How to choose c, the threshold of the fence?

- Example (Adaptive fence; Jiang *et al.* 2008, 2009): To be specific, let parsimony be the criterion of selecting the optimal model within the fence.

- A challenging problem: How to choose c, the threshold of the fence?

- Example (Adaptive fence; Jiang *et al.* 2008, 2009): To be specific, let parsimony be the criterion of selecting the optimal model within the fence.

- Also, for simplicity, consider a "classical" setting, in which the space of candidate models contains a true model, that is, a model under which the data are samples generated.

- A challenging problem: How to choose c, the threshold of the fence?

- Example (Adaptive fence; Jiang *et al.* 2008, 2009): To be specific, let parsimony be the criterion of selecting the optimal model within the fence.

- Also, for simplicity, consider a "classical" setting, in which the space of candidate models contains a true model, that is, a model under which the data are samples generated.

- Under such a setting, an "over-fitting model", that is, a model that includes the true model as a special case, is also a correct model; it may not be optimal, though, because it can be simplified.

- A challenging problem: How to choose c, the threshold of the fence?
- Example (Adaptive fence; Jiang *et al.* 2008, 2009): To be specific, let parsimony be the criterion of selecting the optimal model within the fence.
- Also, for simplicity, consider a "classical" setting, in which the space of candidate models contains a true model, that is, a model under which the data are samples generated.
- Under such a setting, an "over-fitting model", that is, a model that includes the true model as a special case, is also a correct model; it may not be optimal, though, because it can be simplified.
- In particular, M_{f} is a correct model.

- Let M_c denotes the model selected using the fence procedure with the given threshold c.

- Let M_c denotes the model selected using the fence procedure with the given threshold c.

- Ideally, one would like to choose c such that

$$P = \mathrm{P}(M_c = M_{\mathrm{opt}}) \tag{1}$$

is maximized, where M_{opt} denotes the optimal model.

- Let M_c denotes the model selected using the fence procedure with the given threshold c.

- Ideally, one would like to choose c such that

$$P = \mathrm{P}(M_c = M_{\text{opt}}) \tag{1}$$

is maximized, where M_{opt} denotes the optimal model.

- However, two things are unknown in (1).

- Let M_c denotes the model selected using the fence procedure with the given threshold c.

- Ideally, one would like to choose c such that

$$P = \mathrm{P}(M_c = M_{\mathrm{opt}}) \tag{1}$$

is maximized, where M_{opt} denotes the optimal model.

- However, two things are unknown in (1).

- 1. How to compute the probability, P, which depends on the underlying model?

- Let M_c denotes the model selected using the fence procedure with the given threshold c.
- Ideally, one would like to choose c such that

$$P = \mathrm{P}(M_c = M_{\mathrm{opt}}) \tag{1}$$

is maximized, where M_{opt} denotes the optimal model.

- However, two things are unknown in (1).
- 1. How to compute the probability, P, which depends on the underlying model?
- 2. What is M_{opt}.

- Idea 1: Because we know $M_{\rm f}$ is, at least, a correct model, we can bootstrap under $M_{\rm f}$.

- Idea 1: Because we know $M_{\rm f}$ is, at least, a correct model, we can bootstrap under $M_{\rm f}$.
- Idea 2: "Maximum likelihood". For a given candidate model, M, let $p^*(M) = {\rm P}^*(M_c = M)$, that is, the empirical (bootstrap) probability that M is selected.

- Idea 1: Because we know $M_{\rm f}$ is, at least, a correct model, we can bootstrap under $M_{\rm f}$.

- Idea 2: "Maximum likelihood". For a given candidate model, M, let $p^*(M) = {\rm P}^*(M_c = M)$, that is, the empirical (bootstrap) probability that M is selected.

- Let p^* be the maximum of $p^*(M)$ overall candidate models M. Note that p^* depends on c, that is, $p^* = p^*_c$.

- Idea 1: Because we know $M_{\rm f}$ is, at least, a correct model, we can bootstrap under $M_{\rm f}$.
- Idea 2: "Maximum likelihood". For a given candidate model, M, let $p^*(M) = {\rm P}^*(M_c = M)$, that is, the empirical (bootstrap) probability that M is selected.
- Let p^* be the maximum of $p^*(M)$ overall candidate models M. Note that p^* depends on c, that is, $p^* = p^*_c$.
- That is exactly the point!

- Idea 1: Because we know $M_{\rm f}$ is, at least, a correct model, we can bootstrap under $M_{\rm f}$.

- Idea 2: "Maximum likelihood". For a given candidate model, M, let $p^*(M) = {\rm P}^*(M_c = M)$, that is, the empirical (bootstrap) probability that M is selected.

- Let p^* be the maximum of $p^*(M)$ overall candidate models M. Note that p^* depends on c, that is, $p^* = p_c^*$.

- That is exactly the point!

- Choose c that maximizes p_c^*.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?
- Recall $Q(M) - Q(M_{\rm f}) \le c$.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?
- Recall $Q(M) - Q(M_\mathrm{f}) \leq c$.
- $M = M_\mathrm{f}$ is the choice, every single time; hence $p_0^* = 1$.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?
- Recall $Q(M) - Q(M_{\rm f}) \le c$.
- $M = M_{\rm f}$ is the choice, every single time; hence $p_0^* = 1$.
- Similarly, when c is sufficiently large, everybody is in the fence; as a result, M_c is the model that has the minimum dimension (assumed unique), denoted by $M_{\min}$, every single time.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?
- Recall $Q(M) - Q(M_{\rm f}) \leq c$.
- $M = M_{\rm f}$ is the choice, every single time; hence $p_0^* = 1$.
- Similarly, when c is sufficiently large, everybody is in the fence; as a result, M_c is the model that has the minimum dimension (assumed unique), denoted by $M_{\rm min}$, every single time.
- Thus, once again, $p_c^* = 1$, if c is sufficiently large.

- It should be noted that such a plot of p_c^* vs c is typically W shaped, rather than Λ shaped.
- Why?
- What happens when $c = 0$?
- Recall $Q(M) - Q(M_{\rm f}) \le c$.
- $M = M_{\rm f}$ is the choice, every single time; hence $p_0^* = 1$.
- Similarly, when c is sufficiently large, everybody is in the fence; as a result, M_c is the model that has the minimum dimension (assumed unique), denoted by $M_{\min}$, every single time.
- Thus, once again, $p_c^* = 1$, if c is sufficiently large.
- One is typically looking for the “peak in the middle”.

4. Example

- Battese, Harter & Fuller (1988; BHF) presented data for 12 Iowa counties obtained from the 1978 June Enumerative Survey of the U.S. Dept. of Agriculture.

4. Example

- Battese, Harter & Fuller (1988; BHF) presented data for 12 Iowa counties obtained from the 1978 June Enumerative Survey of the U.S. Dept. of Agriculture.

- Also available were data from land observatory satellites on crop areas involving corn and soybeans.

4. Example

- Battese, Harter & Fuller (1988; BHF) presented data for 12 Iowa counties obtained from the 1978 June Enumerative Survey of the U.S. Dept. of Agriculture.
- Also available were data from land observatory satellites on crop areas involving corn and soybeans.
- This is a classical example of borrowing strength via a statistical model.

4. Example

- Battese, Harter & Fuller (1988; BHF) presented data for 12 Iowa counties obtained from the 1978 June Enumerative Survey of the U.S. Dept. of Agriculture.
- Also available were data from land observatory satellites on crop areas involving corn and soybeans.
- This is a classical example of borrowing strength via a statistical model.
- The latter is a linear mixed model (LMM) in the form of

$$y_{ij} = x_{ij}'\beta + v_i + e_{ij},$$

$i = 1, \dots, m$, $j = 1, \dots, n_i$, where ...

- The authors discussed the choice of $x'_{ij}\beta$ in the LMM, which is viewed as a model selection problem.

- The authors discussed the choice of $x'_{ij}\beta$ in the LMM, which is viewed as a model selection problem.
- They suggested $x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}$, where x_{ij1} and x_{ij2} are the numbers of pixels classified as corn and soybeans, respectively, for both the corn observation and the soybeans observation.

- The authors discussed the choice of $x'_{ij}\beta$ in the LMM, which is viewed as a model selection problem.
- They suggested $x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}$, where x_{ij1} and x_{ij2} are the numbers of pixels classified as corn and soybeans, respectively, for both the corn observation and the soybeans observation.
- Why?

- The authors discussed the choice of $x'_{ij}\beta$ in the LMM, which is viewed as a model selection problem.
- They suggested $x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}$, where x_{ij1} and x_{ij2} are the numbers of pixels classified as corn and soybeans, respectively, for both the corn observation and the soybeans observation.
- Why?
- This is what Fuller said: ...

- The authors discussed the choice of $x'_{ij}\beta$ in the LMM, which is viewed as a model selection problem.
- They suggested $x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}$, where x_{ij1} and x_{ij2} are the numbers of pixels classified as corn and soybeans, respectively, for both the corn observation and the soybeans observation.
- Why?
- This is what Fuller said: ...
- Nevertheless, the authors also discussed other choices of $x'_{ij}\beta$, such as including the squares and cross product of $x_{ijr}, r = 1, 2$.

- If we consider this as a variable selection problem, the space of candidate predictors may be chosen as

$$\left\{1,\ x_{ij1},\ x_{ij2},\ x_{ij2}^2,\ x_{ij2}^2,\ x_{ij1}x_{ij2}\right\}.$$

- If we consider this as a variable selection problem, the space of candidate predictors may be chosen as

$$\left\{1,\ x_{ij1},\ x_{ij2},\ x_{ij2}^2,\ x_{ij2}^2,\ x_{ij1}x_{ij2}\right\}.$$

- If we apply the fence methods, a standard measure of lack-of-fit, Q, is the negative log-likelihood.

- If we consider this as a variable selection problem, the space of candidate predictors may be chosen as

$$\left\{1,\ x_{ij1},\ x_{ij2},\ x_{ij2}^2,\ x_{ij2}^2,\ x_{ij1}x_{ij2}\right\}.$$

- If we apply the fence methods, a standard measure of lack-of-fit, Q, is the negative log-likelihood.
- For simplicity, let us use the parsimony criterion for selecting the model within the fence.

- If we consider this as a variable selection problem, the space of candidate predictors may be chosen as

$$\left\{1,\ x_{ij1},\ x_{ij2},\ x_{ij2}^2,\ x_{ij2}^2,\ x_{ij1}x_{ij2}\right\}.$$

- If we apply the fence methods, a standard measure of lack-of-fit, Q, is the negative log-likelihood.

- For simplicity, let us use the parsimony criterion for selecting the model within the fence.

- The threshold, c, is chosen adaptively as described above.

- The selection results, compared with the BHF models, are presented in the table below:

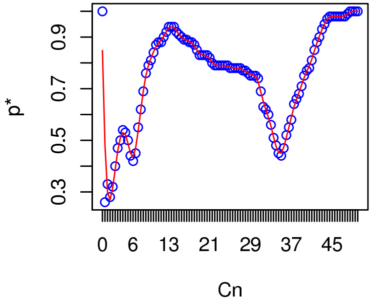
Outcome Variable	Predictors	
	BHF Model	Adaptive Fence
Corn	Corn & Soybeans	Corn
Soybeans	Corn & Soybeans	Soybeans

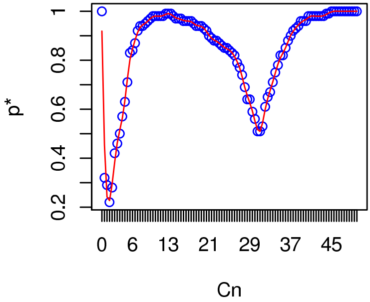
- The selection results, compared with the BHF models, are presented in the table below:

Outcome Variable	Predictors	
	BHF Model	Adaptive Fence
Corn	Corn & Soybeans	Corn
Soybeans	Corn & Soybeans	Soybeans

- The plots of p_c^* vs c?

- AF Selection for the Crops Data. Left: p^* vs $c = c_n$ for selecting the corn model. Right: p^* vs $c = c_n$ for selecting the soybeans model.





5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.

5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.

- Question: Do you have a practical example of doing the fence this way?

5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.
- Question: Do you have a practical example of doing the fence this way?
- Answer: I don't.

5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.
- Question: Do you have a practical example of doing the fence this way?
- Answer: I don't.
- What?

5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.
- Question: Do you have a practical example of doing the fence this way?
- Answer: I don't.
- What?
- That's why I'm here.

5. Discussion

- It was said that a potential advantage of the fence methods is incorporating consideration of practical interest.
- Question: Do you have a practical example of doing the fence this way?
- Answer: I don't.
- What?
- That's why I'm here.
- It would've been nicer to come here with all questions and answers, but it is just as important to have some questions but no answers.

- And I believe the answers are on their way.

- And I believe the answers are on their way.
- Biggest challenges in implementing the fence methods: Computation, and artificial intelligence (AI).

- And I believe the answers are on their way.
- Biggest challenges in implementing the fence methods: Computation, and artificial intelligence (AI).
- Why computation?

- And I believe the answers are on their way.
- Biggest challenges in implementing the fence methods: Computation, and artificial intelligence (AI).
- Why computation?
- For example, the adaptive fence requires bootstrapping under the full model.

- And I believe the answers are on their way.
- Biggest challenges in implementing the fence methods: Computation, and artificial intelligence (AI).
- Why computation?
- For example, the adaptive fence requires bootstrapping under the full model.
- Bootstrap for high-dimensional model selection?

- Why AI?

- Why AI?
- For example, finding the "peak in the middle" in AF is something human eyes can arguably do better than a computer.

- Why AI?
- For example, finding the "peak in the middle" in AF is something human eyes can arguably do better than a computer.
- Really?

- Why AI?
- For example, finding the "peak in the middle" in AF is something human eyes can arguably do better than a computer.
- Really?
- Many years ago, people also said that a computer could NOT defeat a chess (or go) grand master.

- Why AI?
- For example, finding the "peak in the middle" in AF is something human eyes can arguably do better than a computer.
- Really?
- Many years ago, people also said that a computer could NOT defeat a chess (or go) grand master.
- We need some genius young people with good computer science training.

- Idea & Execution

- Idea & Execution

- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.

- Idea & Execution
- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.
- Which one is more important?

- Idea & Execution
- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.
- Which one is more important?
- Well, if you aim to publish a *JASA* paper, then Idea probably is.

- Idea & Execution
- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.
- Which one is more important?
- Well, if you aim to publish a *JASA* paper, then Idea probably is.
- But, look at Physics — for one thing, many Nobel prizes have been awarded to work that executed an existing idea.

- Idea & Execution
- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.
- Which one is more important?
- Well, if you aim to publish a *JASA* paper, then Idea probably is.
- But, look at Physics — for one thing, many Nobel prizes have been awarded to work that executed an existing idea.
- And, there are much more, if one goes beyond academics.

- Idea & Execution
- It may take a second to come up with a beautiful idea, but many years of hard work to execute the idea.
- Which one is more important?
- Well, if you aim to publish a *JASA* paper, then Idea probably is.
- But, look at Physics — for one thing, many Nobel prizes have been awarded to work that executed an existing idea.
- And, there are much more, if one goes beyond academics.
- Real-life surveys is one of them.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.
- Fuller, W. A. (2009), *Sampling Statistics*, Wiley

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.
- Fuller, W. A. (2009), *Sampling Statistics*, Wiley
- Lohr, S. L. (2022), *Sampling: Design and Analysis*, 3rd ed., Chapman & Hall.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.
- Fuller, W. A. (2009), *Sampling Statistics*, Wiley
- Lohr, S. L. (2022), *Sampling: Design and Analysis*, 3rd ed., Chapman & Hall.
- Little, R. J. (2008), Weighting and prediction in sample surveys (with discussion), *Calcutta Statist. Assoc. Bull.* 60, 147–194.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.
- Fuller, W. A. (2009), *Sampling Statistics*, Wiley
- Lohr, S. L. (2022), *Sampling: Design and Analysis*, 3rd ed., Chapman & Hall.
- Little, R. J. (2008), Weighting and prediction in sample surveys (with discussion), *Calcutta Statist. Assoc. Bull.* 60, 147–194.
- Lahiri, P. (2001), *Model Selection — IMS Lec. Ser.* 38, Institute of Mathematical Statistics.

References

- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- McFadden, D. and Kenneth, T. (2000), Mixed MNL model for discrete responses, *J. Appl. Econ.* 15, 447–470.
- Fuller, W. A. (2009), *Sampling Statistics*, Wiley
- Lohr, S. L. (2022), *Sampling: Design and Analysis*, 3rd ed., Chapman & Hall.
- Little, R. J. (2008), Weighting and prediction in sample surveys (with discussion), *Calcutta Statist. Assoc. Bull.* 60, 147–194.
- Lahiri, P. (2001), *Model Selection — IMS Lec. Ser.* 38, Institute of Mathematical Statistics.
- Pfeffermann, D. (2013), New important developments in small area estimation, *Statist. Sci.* 28, 40–68.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.
- Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008), Fence method for mixed model selection, *Ann. Statist.* 36, 1669–1692.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.
- Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008), Fence method for mixed model selection, *Ann. Statist.* 36, 1669–1692.
- Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation*, 2nd ed., Wiley.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.
- Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008), Fence method for mixed model selection, *Ann. Statist.* 36, 1669–1692.
- Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation*, 2nd ed., Wiley.
- Jiang, J. and Nguyen, T. (2016), *The Fence Methods*, World Scientific, Singapore.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.
- Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008), Fence method for mixed model selection, *Ann. Statist.* 36, 1669–1692.
- Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation*, 2nd ed., Wiley.
- Jiang, J. and Nguyen, T. (2016), *The Fence Methods*, World Scientific, Singapore.
- Jiang, J., Nguyen, T. and Rao, J. S. (2009), A simplified adaptive fence procedure, *Statist. Probab. Letters* 79, 625–629.

- Lumley, T, and Scott, A. (2017), Fitting regression models to survey data, *Statist. Sci.* 32, 265–278.
- Müller, S., Scealy, J. L., and Welsh, A. H. (2013), Model selection in linear mixed models, *Statist. Sci.* 28, 135-167.
- Jiang, J., Rao, J. S., Gu, Z., and Nguyen, T. (2008), Fence method for mixed model selection, *Ann. Statist.* 36, 1669–1692.
- Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation*, 2nd ed., Wiley.
- Jiang, J. and Nguyen, T. (2016), *The Fence Methods*, World Scientific, Singapore.
- Jiang, J., Nguyen, T. and Rao, J. S. (2009), A simplified adaptive fence procedure, *Statist. Probab. Letters* 79, 625–629.
- Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), An error-components model for prediction of county crop areas using survey and satellite data, *J. Amer. Statist. Assoc.* 80, 28–36.