# Big Data, Model Selection, Aggregation-Indexing, Oracles

Esfandiar Maasoumi[1]

[1]Emory University

August 1, 2018

# What Does Big Data Mean?

# What Does Big Data Mean?

1. Big Data and Big Models
   - Big Data-Micro; Big Data Macro/Time series!
   - Mostly variable selection: Multiple Indicators, **Latent Objects**.

# What Does Big Data Mean?

1. Big Data and Big Models
   - Big Data-Micro; Big Data Macro/Time series!
   - Mostly variable selection: Multiple Indicators, **Latent Objects**.
2. Latent Objects:
   - Well-being; Happiness, Permanent Income, Expectations.
   - What about Data Generation Process (DGP) as a "Latent Object"?

# What Does Big Data Mean?

1. Big Data and Big Models
   - Big Data-Micro; Big Data Macro/Time series!
   - Mostly variable selection: Multiple Indicators, **Latent Objects**.
2. Latent Objects:
   - Well-being; Happiness, Permanent Income, Expectations.
   - What about Data Generation Process (DGP) as a "Latent Object"?
3. Common Themes
   - More Variables than Observations $p >> n$
   - RELATED: Shrinkage, Penalization, Averaging, Model Selection, model uncertainty, Misspecification

# The ProtoType Case: Variable Selection

# The ProtoType Case: Variable Selection

1. Outcome Y, Target T, Variable Set X of p vars
   - WRITE THE model. formulae here with definitions.
   - This is typically a linear "MODEL"!
   - Estimation and Big Data strategies: Shrinkage, LASSO, Other Penalization methods.

# Inference

# Inference

1. Best Practice Inference Theory
   - Exemplified by Victor Chernozhukov and coworkers.
   - Asymptotic Inference, especially for $p > n$.
   - I will advertize my old work in this area Below!!:)
   - Ridge Regression; Stein; shrinkage estimation and forecasting in systems of equations; LASSO

# Inference

1. Best Practice Inference Theory
   - Exemplified by Victor Chernozhukov and coworkers.
   - Asymptotic Inference, especially for $p > n$.
   - I will advertize my old work in this area Below!!:)
   - Ridge Regression; Stein; shrinkage estimation and forecasting in systems of equations; LASSO
2. "Bayesian" Interpretation; Extraneous Statistical "Information" (important);
   - Model selection and uncertainty.
   - What do we want to learn or do? (KEY question)
   - "Causal"? Policy analysis/decision making needs this.
   - Trearment effect and Program/policy evaluation

# Mechanism vs. Blackbox Prediction vs. Indexing

# Mechanism vs. Blackbox Prediction vs. Indexing

1. Aggregation-Indexing Multiple Indicators
   - Aggregate-"average" all Xs. Not "causal"
   - What is "average"? An "INDEX"
   - Classic Index Number Problem
   - Aggregation-Averaging of Models is related, but not the same

# Mechanism vs. Blackbox Prediction vs. Indexing

1. Aggregation-Indexing Multiple Indicators
   - Aggregate-"average" all Xs. Not "causal"
   - What is "average"? An "INDEX"
   - Classic Index Number Problem
   - Aggregation-Averaging of Models is related, but not the same
2. Models for Mechanism Learning
   - Many Moments Paradigm (GMM)
   - Empirical Likelihood-Information Theory
   - Hopeless!? All models are Misspecified
   - The Map allegory!

1. Machine learning, Deep Learning, Ridge, Adaptive LASSO..... (Brief)

1. Machine learning, Deep Learning, Ridge, Adaptive LASSO..... (Brief)
2. Habibnia and Maasoumi (Neural Nets), Later! Many moments case and wrong moments (School 2016)

1. Machine learning, Deep Learning, Ridge, Adaptive LASSO..... (Brief)
2. Habibnia and Maasoumi (Neural Nets), Later! Many moments case and wrong moments (School 2016)

Aggrgegation-Indexing

- Multiple Indicators of Well-Being
- Maasoumi (1986. Econometrica)
- TODAY: Model Averaging as Indexing, when all are misspecified
- Gospodinov-Maasoumi (2018)