Comments slides for Thursday, Oct 15: Measuring predictability

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".

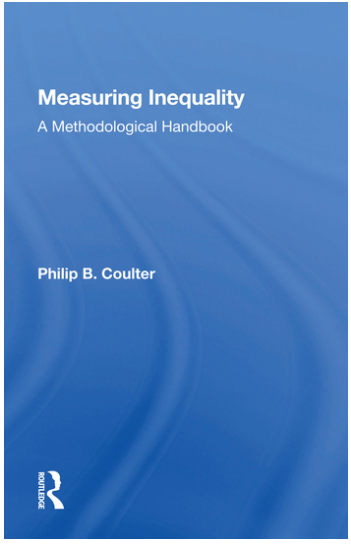
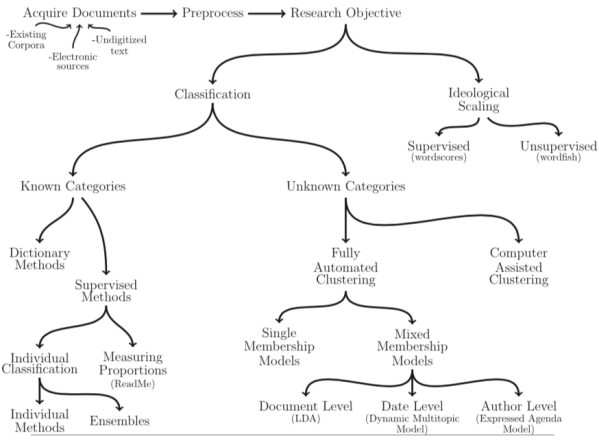

Measuring Inequality
A Methodological Handbook
Philip B. Coulter
ROUTLEDGE

Table 3-2: Primary Properties of Indexes Based on the Deviations Model

	Polarity	Type of index	Concentration	Comparative standard	Constant additions	Transfers	Scale invariance	Population symmetry	Lorenz dominance	Scale	Definition	Information use	Upper and lower limits	Simplicity
Wilcox's deviation from the mode (DM)	equality	relative, ANONC	no	mode	increase	yes	yes	no	no	nominal	yes	good	1, 0	good
Dahl's polyarchy (P)	equality	absolute, type A	yes	mean	increase	yes	yes	yes	yes	interval	yes	good	1-(1/K),0	fair
Schutz's inequality (S)	inequality	absolute, type A	no	mean	decrease	no	yes	yes	no	nominal	yes	poor	1-(1/K),0	good
Mayer's uniformity (M)	equality	absolute, type A	yes	mean	increase	yes	yes	yes	yes	nominal	yes	good	K-1, 0	good
Nagel's equality (E)	equality	relative, ANONC	yes	mean	decrease	yes	yes	yes	yes	nominal	yes	good	1, 0	good
Gini's mean relative difference (MRD)	equality	relative, ANONC	yes	all other components	increase	yes	yes	yes	yes	interval	yes	good	1, 0	fair
Gini coefficient (G)	inequality	absolute, type A	yes	all other components	decrease	yes	yes	yes	yes	interval	yes	good	1-(1/K), 0	poor
Przeworski's instability (D_t)	inequality	absolute, type A	no	adjacent component	decrease	yes	yes	yes	yes	nominal	yes	good	t_n-1, 0	poor


Acquire Documents
Preprocess
Research Objective
-Existing Corpora
-Electronic sources
-Undigitized text
Classification
Ideological Scaling
Supervised (wordscores)
Unsupervised (wordfish)
Known Categories
Unknown Categories
Dictionary Methods
Supervised Methods
Fully Automated Clustering
Computer Assisted Clustering
Individual Classification
Measuring Proportions (ReadMe)
Single Membership Models
Mixed Membership Models
Individual Methods
Ensembles
Document Level (LDA)
Date Level (Dynamic Multitopic Model)
Author Level (Expressed Agenda Model)

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity
- $R^2_{Holdout}$ and $MSE_{Holdout}$ are very different

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity
- $R^2_{Holdout}$ and $MSE_{Holdout}$ are very different
- picking a metric because it is "famous" can be dangerous, but picking a metric that is not standard can be weird too

Observations/comments/questions/provocations:

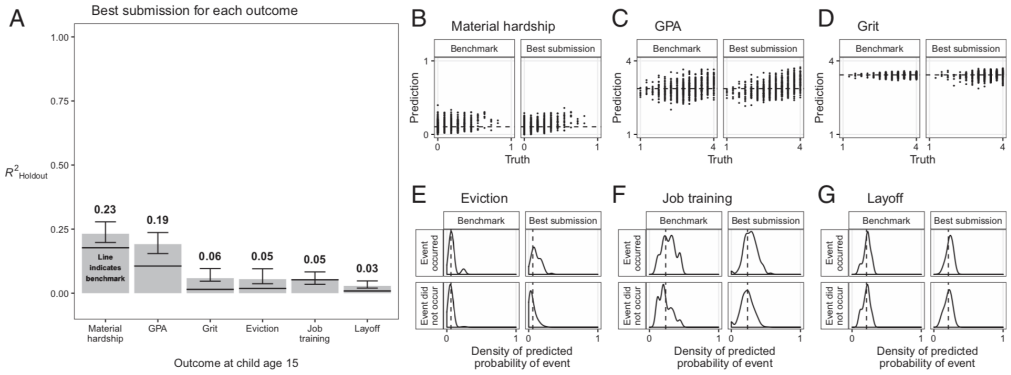
- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity
- $R^2_{Holdout}$ and $MSE_{Holdout}$ are very different
- picking a metric because it is "famous" can be dangerous, but picking a metric that is not standard can be weird too
- For policy we care about ranks and for science we care about values! Maybe?

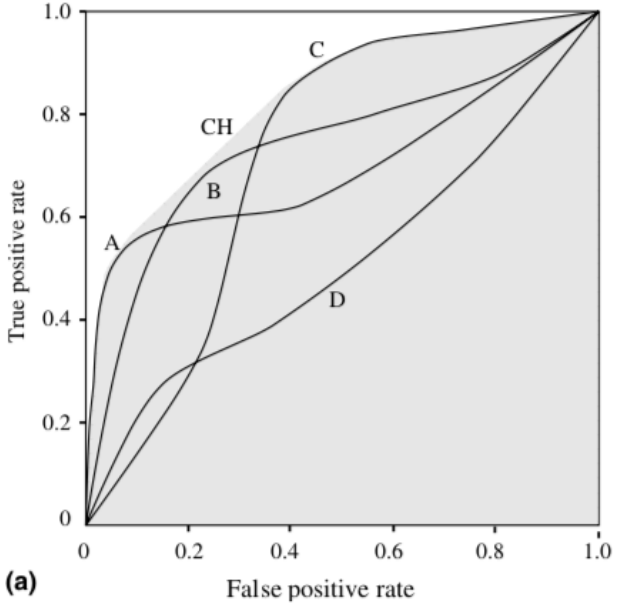
Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity
- $R^2_{Holdout}$ and $MSE_{Holdout}$ are very different
- picking a metric because it is "famous" can be dangerous, but picking a metric that is not standard can be weird too
- For policy we care about ranks and for science we care about values! Maybe?
- Classification problems are rare in sociology because we are rarely trying to do anything

Observations/comments/questions/provocations:

- I'm surprised at how non-systematized this is. Measure of "predictability" seem more complex than measure of "inequality".
- We didn't talk any about measure of "unpredictability"
- Single number summaries lose information and hide heterogeneity
- $R^2_{Holdout}$ and $MSE_{Holdout}$ are very different
- picking a metric because it is "famous" can be dangerous, but picking a metric that is not standard can be weird too
- For policy we care about ranks and for science we care about values! Maybe?
- Classification problems are rare in sociology because we are rarely trying to do anything
- Imagine statistical learning produces and uncalibrated score $\tilde{y}$. Higher $\tilde{y}$ means more likely to get evicted. We want $\hat{y} = f(\tilde{y})$. We can learn $f()$. https://machinelearningmastery.com/probability-calibration-for-imbalanced-classification/. Platt scaling aka Platt calibration. Could this work for continuous outcomes too? How could it?


A
Best submission for each outcome
1.00
0.75
0.50
0.25
0.00
R^2_{Holdout}
0.23
0.19
0.06
0.05
0.05
0.03
Line indicates benchmark
Material hardship
GPA
Grit
Eviction
Job training
Layoff
Outcome at child age 15
B
Material hardship
C
GPA
D
Grit
Benchmark
Best submission
Prediction
Truth
E
Eviction
F
Job training
G
Layoff
Event occurred
Event did not occur
Density of predicted probability of event


1.0
0.8
0.6
0.4
0.2
0
True positive rate
0
0.2
0.4
0.6
0.8
1.0
False positive rate
A
B
CH
C
D

(a)

Comments slides for Thursday, Oct 15: Measuring predictability

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University