Advanced Regression Analysis

Predictive

Descriptive
Prescriptive

Strategic

Nick Mastronardi
PhD, CEO

POLCO.us

U.S. AIR FORCE

amazon

What Language You Speak Depends on Where You Come From

Undergraduate:
Math / Physics

→

Graduate / Academic:
Econ / OR

→

Private:
Econ / Data Science
Statistical Inference
Design of Experiment

Business Intelligence Engineer

Data Visualization

Database Engineering

Hosting Infrastructure

Predictive
Descriptive
Prescriptive
Strategic

Predictive Models

Care about Accuracy & Precision in Dependent / Outcome Variable

$$\vec{\beta} \min \sum_{i=1}^{N} (y_i - \hat{y})^2$$

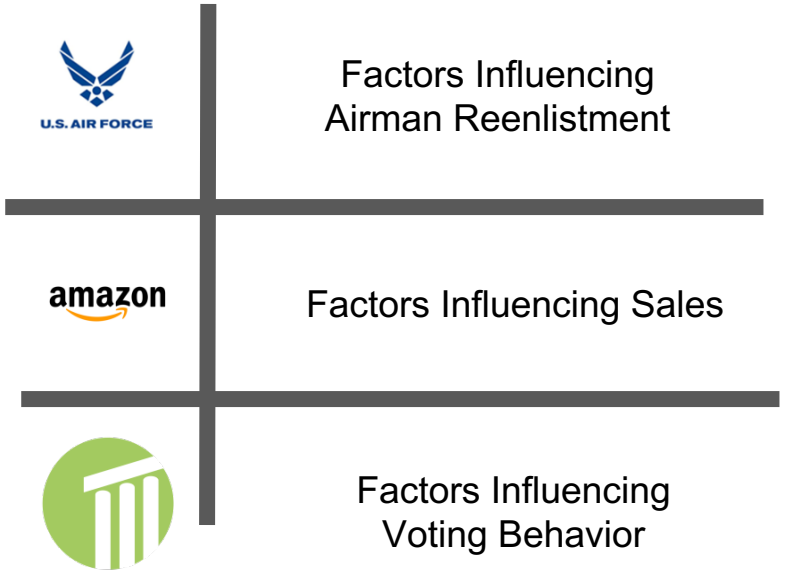
$$\hat{y} = \beta_0 + \beta_1 x_1 + \cdots \beta_K x_K + \varepsilon$$

Predictive

Descriptive

Prescriptive

Strategic



Descriptive Models

Causal Inference
Accuracy / Precision on inferred impact of how A causes B

Gauss Markov Thm:
(we'll come back to this)

Predictive
Descriptive
Prescriptive
Strategic

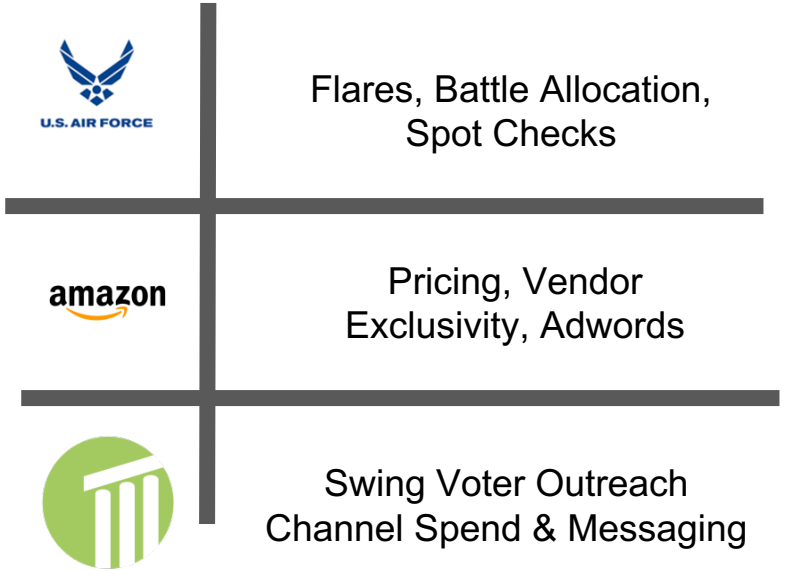
Prescriptive Models

Use Descriptive Models and Choice Constraints to Optimize Performance

$\vec{x}$ max $\Pi(\vec{x}|\vec{\beta})$ s.t. constraints

Predictive
Descriptive
Prescriptive

Strategic



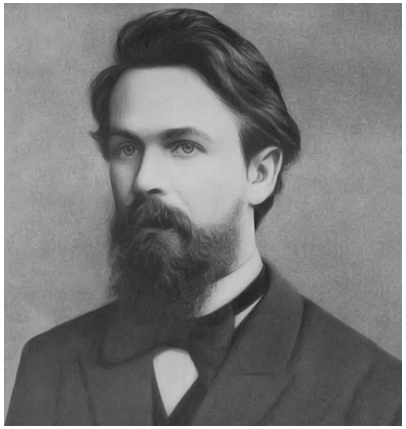
Strategic Models

Interdependent Payoffs
Game Theoretic, eg Nash Equilibrium

The Future:

But let's not get ahead of where the biggest need is

Gauss Markov Theorem



Conditions under which, if satisfied, you can trust accuracy and precision of causal impact parameters in regression models.

And, if conditions aren’t satisfied, gives insights to how you can adjust the regression model to recover accuracy and precision.

Gauss Markov Theorem

If none of the following things are true, then coefficients are:
Unbiased / Accurate
Efficient / Precise / Confident

Multicollinearity	Correlation in factors
Heteroskedasticity	Errors vary systematically with causal factors
Autocorrelation	Observations' errors are not independent across time
Improper Variables	Omitted variables, Extra variables
Model Mis-specification	Not the correct posited relationship between factors

Recovery High Level Description

Multicollinearity	Correlation in factors
	Ooooh!
Heteroskedasticity	Error varies systematically with factors
	Reweight emphases of observations
Autocorrelation	Observation errors are not independent
	Remove correlated component of errors
Improper Variables	Omitted variables, Extra variables
	Add in the right variables, Remove extraneous
Model Mis-specification	Not the correct posited relationship between factors
	Think hard, Be an industry expert, Be careful

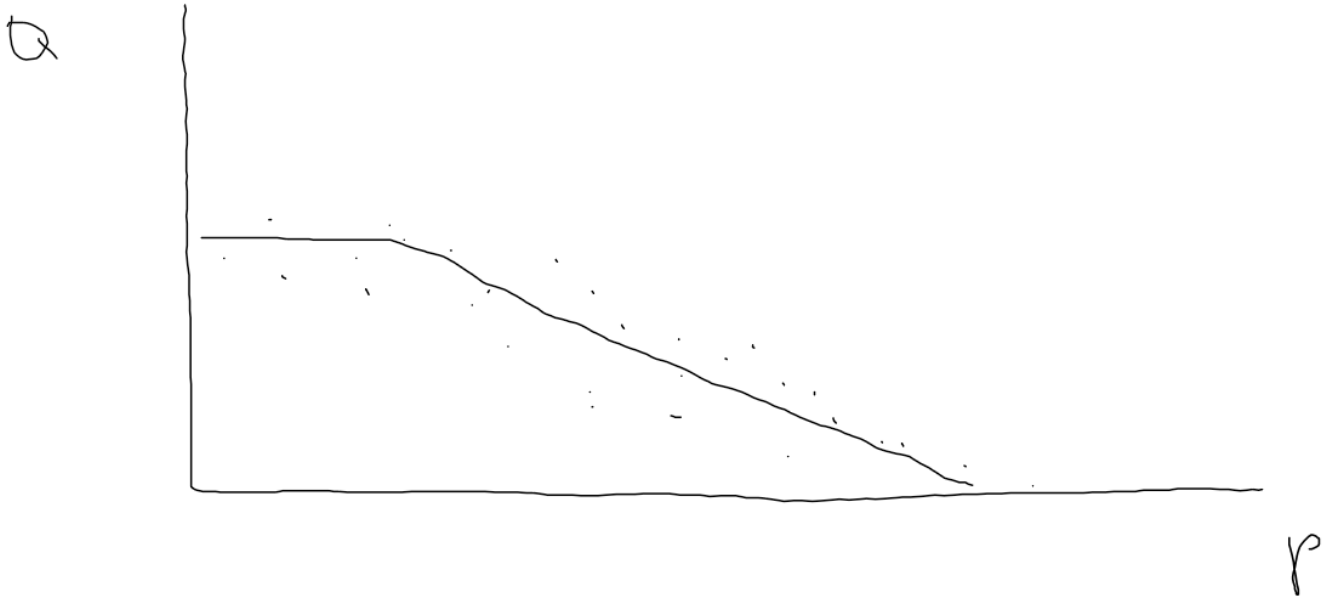
Air Force Reenlistment
Labor supply elasticities

Name	Date	Decision	Bonus	other factors

$Y_i \in \{0,1\}$

Probit Regression

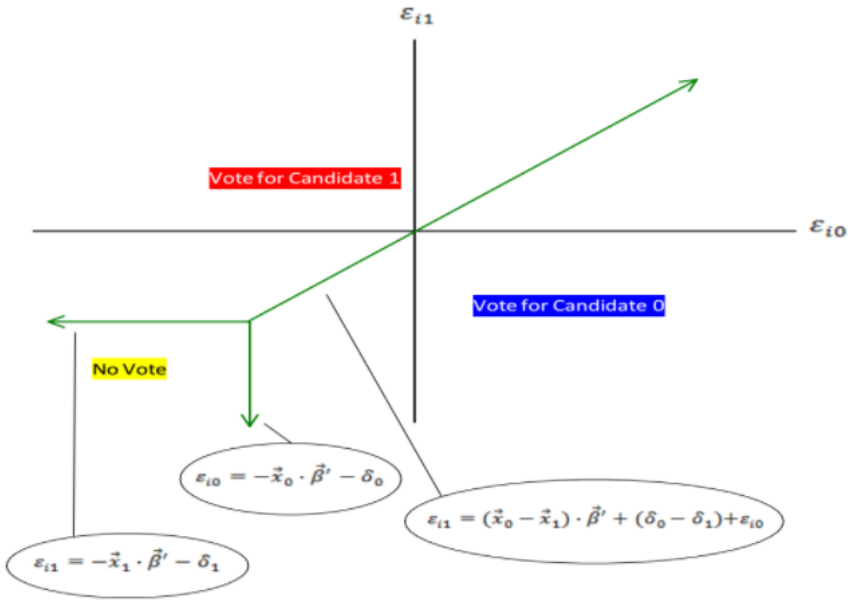
Amazon pricing



Voter behavior

$$max_{x_j \epsilon \{x_0, x_1, x_n\}} U_i(x_j) = \vec{x}_j \cdot \vec{\beta}' + \delta_j + \varepsilon_{ij}$$

$$\vec{x}_0 \cdot \vec{\beta}' + \delta_0 + \varepsilon_{i0} > \vec{x}_1 \cdot \vec{\beta}' + \delta_1 + \varepsilon_{i1}$$



How Big Data and Increased Computational Capabilities are Changing Things

1. Ordinary Least Squares - (Generalized Method of Moments) - Maximum Likelihood Estimation

2. Model Iteration: Random Forests, Genetic Algorithms, Taylor/Fourier/Bessel Functional Bases

3. Oaxaca Blinder Regression Machine-Learning Hybrids to Identify Treatment Effects

4. Now that you can expect competitors to be employing these techniques - Strategic Effects