Machine Learning Methods in Finance: Recent Applications and Prospects

Daniel Hoang, Kevin Wiegratz

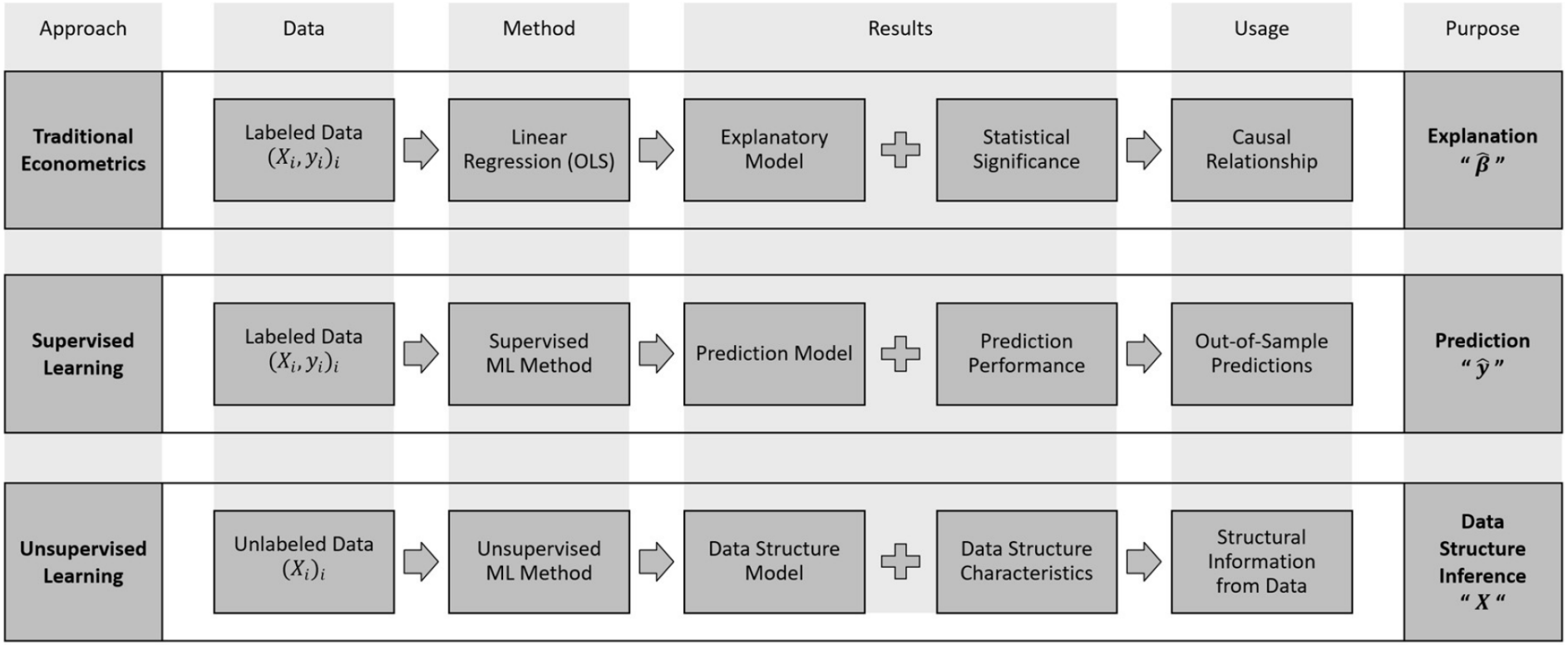
Long Zhen

Contents

- Introduction to ML
- Taxonomy of ML
- Summary

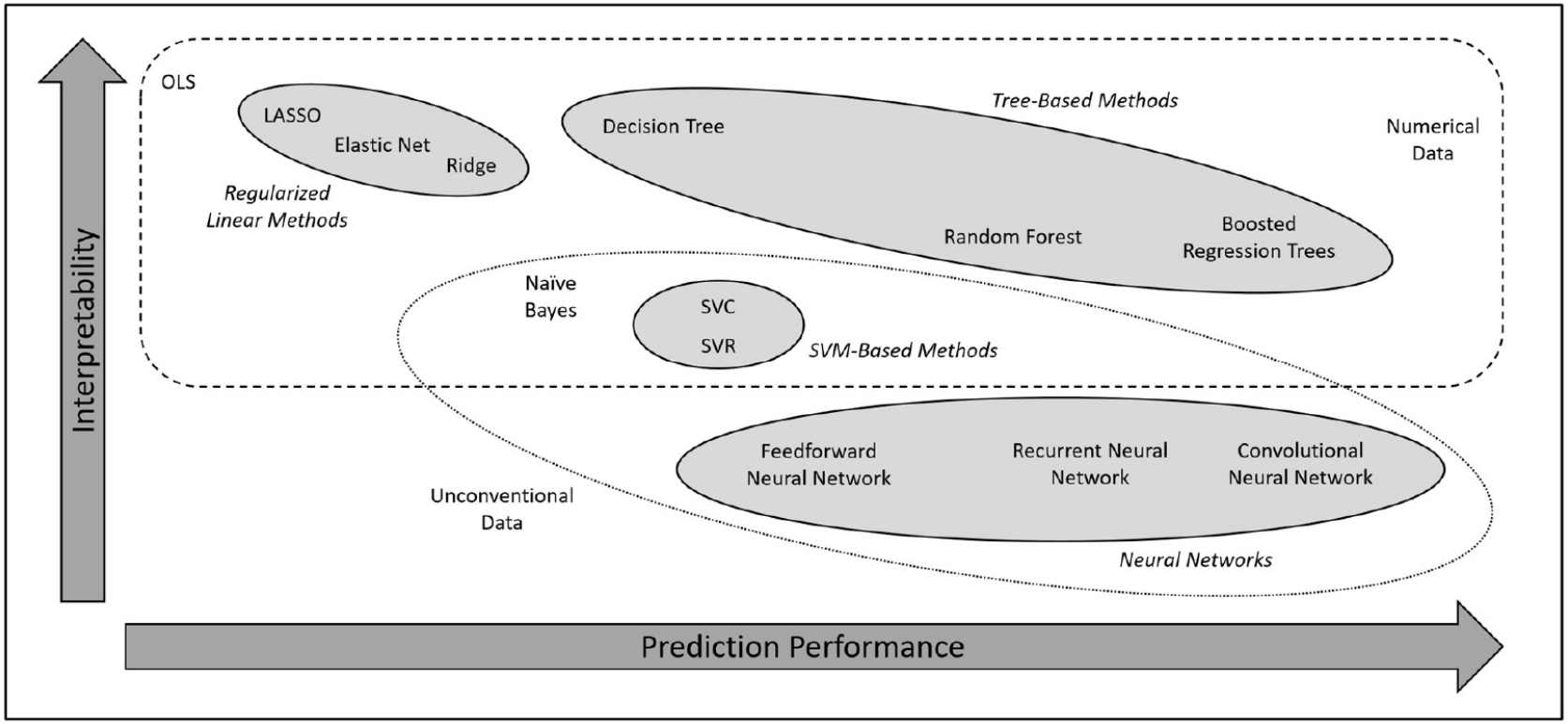
1. Introduction to ML

- Instead of causal explanations, ML serves for prediction (supervised) or data structure inference(unsupervised).



1.1 Supervised learning

- Purpose: prediction
- Training data & test data
- Prediction performance vs. interpretation



- **OLS**: BLUE; best interpretation, weak performance
  - Nonlinear transformation and interactions?
- →**LASSO/Ridge/Elastic Net**: introduce bias by adding penalty term
- **Regression tree**
- → random forest: bootstrap samples and build separate trees
- → GBDT: build trees iteratively

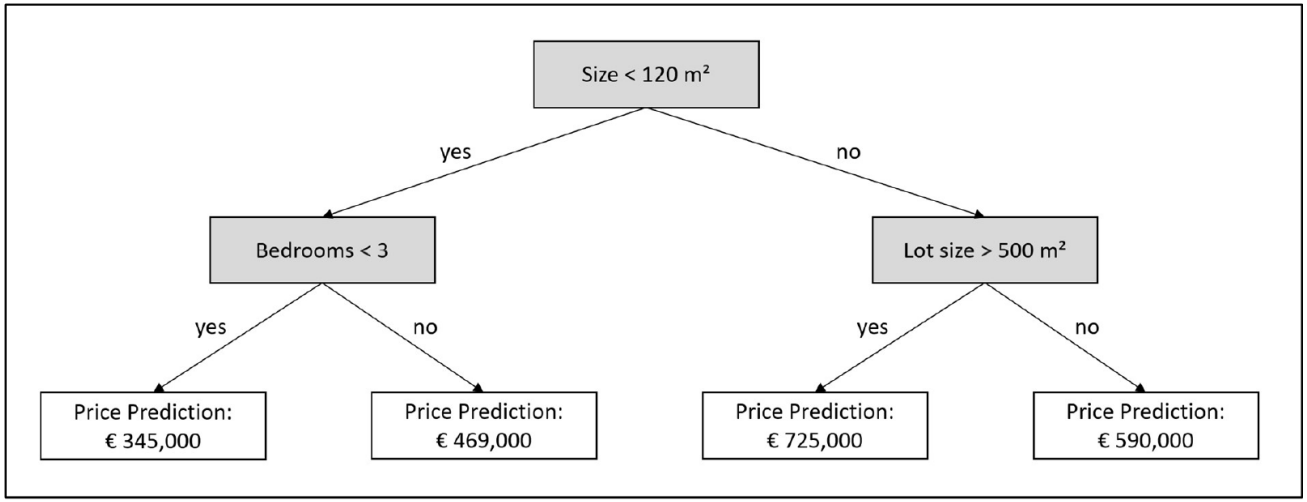
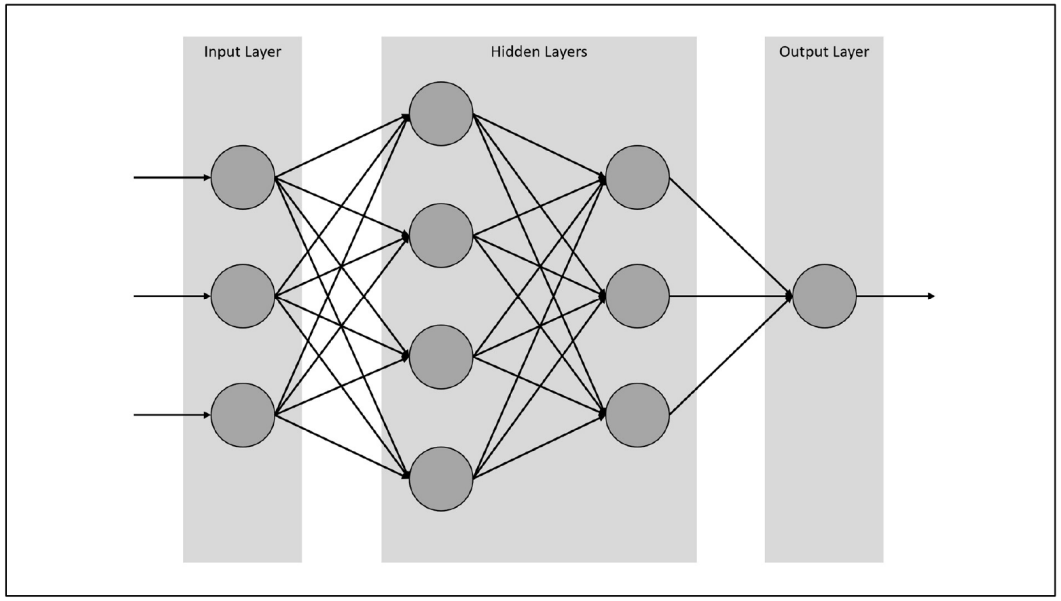


Figure 4. Illustrative depiction of a decision tree trained for house price prediction

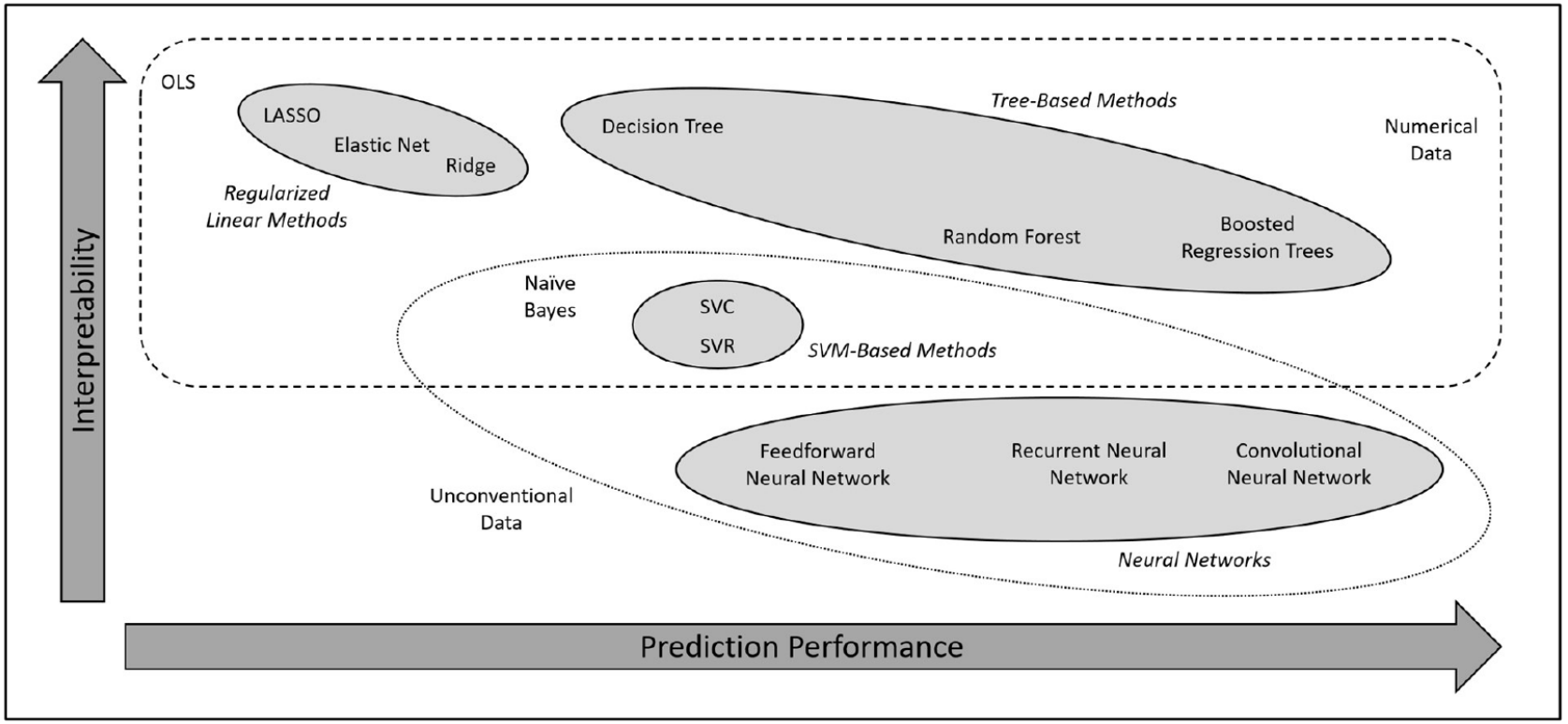
Hard to interpret:

- Feed-forward NN



- → recurrent NN: for sequential data
  - Eg. Gated recurrent units/LSTM
- → convolutional NN: for visual data

- Older methods:
  - Naïve Bayes
  - SVM/SVR



1.2 Unsupervised learning

- Purpose: data structure inference

Clustering

Goal	Methods
Grouping of observations: • High within-group similarity • Low cross-group similarity	• Centroid-based methods • Density-based methods • Distribution-based methods • Hierarchical methods • ...

Dimensionality Reduction

Goal	Methods
Increase of information density: • Decrease dimensionality • Retain most information	• PCA-based methods • NN-based methods • ...

Other

Association Rule Mining	Outlier Detection	Synthetic Data Generation	...

Clustering

- Centroid-based methods: eg. K-means (重心法)
  - After the initial positioning of the centroids, they iteratively update their position to arrive at suitable clusters
- Density-based methods: eg. DBSCAN
  - Group observations with many similar observations nearby into clusters
- Distribution-based methods: eg. Gaussian mixture models
  - Based on whether observations belong to the same statistical distribution
- Hierarchical methods: eg. BIRCH
  - Iteratively combine smaller clusters into larger clusters

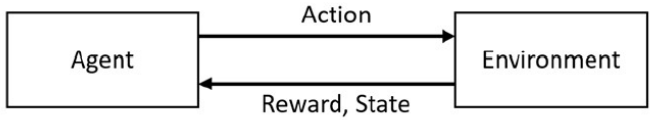
Dimensionality reduction

- Decrease the dimensionality while retain most of the inherent information
- PCA-based: cover as much of the data's variance as possible
- NN-based: autoencoder
  - Consist of an encoder network that creates a condensed representation of the input data and a subsequent decoder network that reconstructs the original data. A special bottleneck layer connect the encoder and decoder to train

- Others:
- Association rule mining: identify relations between variables
- Outlier detection methods

1.3 Reinforcement learning and other types

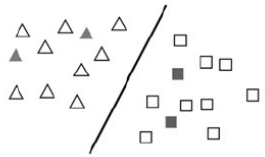
Reinforcement Learning



Description

- Markov decision process model with environment and agent whose actions bring rewards and change the environment
- Goal: policy that maximizes the expected total reward
- Data: description of the Markov decision process
- Idea: the algorithm trades short-term vs. long-term rewards

Semi-Supervised Learning



Description

- Combination of supervised and unsupervised learning
- Goal: prediction (as in supervised learning)
- Data: few labeled examples and many unlabeled examples
- Idea: the additional unlabeled data adds information about the probability distribution of the input data

Other

演绎学习 Deductive Learning

联邦学习 Federated Learning

遗传算法 Genetic Algorithms

...

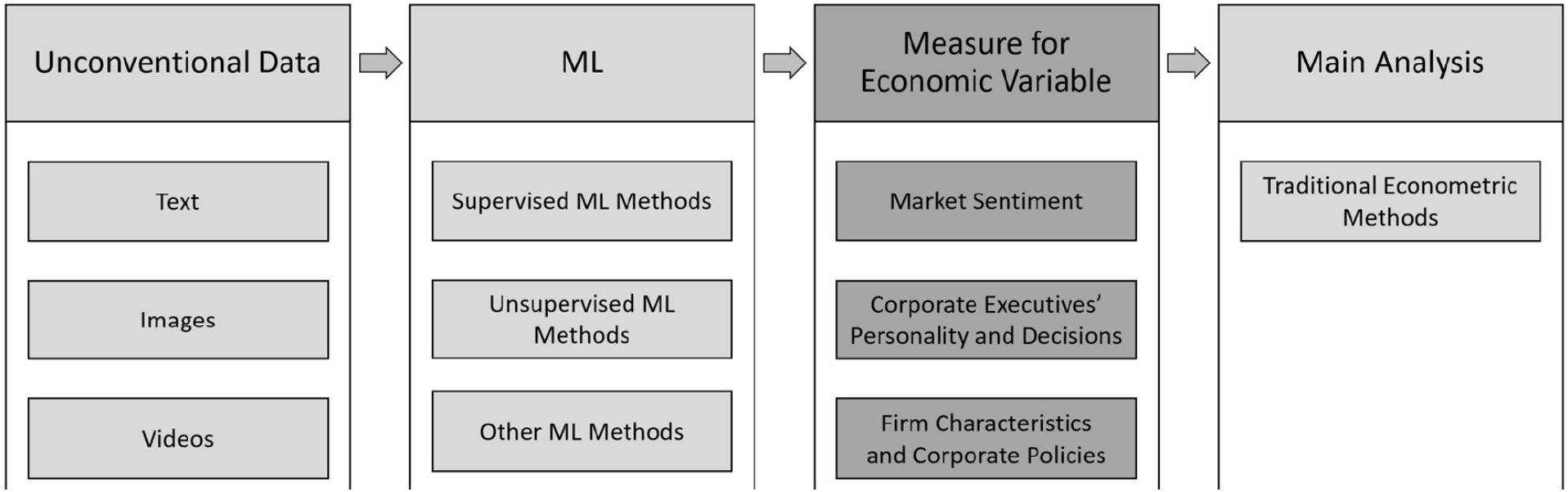
- Semi-supervised: eg. Active learning (closely related variant of semi-supervised learning): train the model with labeled examples → calculate importance score for the unlabeled → label and reduce costs

2. Taxonomy of ML Applications in Financial Economics

- ML solves different problems compared to OLS
- Three typical applications:

1	Construction of Superior and Novel Measures	$y = \beta \boldsymbol{X} + \varepsilon$
2	Reduction of Prediction Error in Economic Prediction Problems	$\widehat{\boldsymbol{y}} = f(X)$
3	Extension of the Existing Econometric Toolset	$y = \beta X + \varepsilon$ & $\boldsymbol{ML}$

2.1 Construction of superior and novel measures



Measures of Market Sentiment

Stock Market Sentiment

	Methods	Data
Antweiler and Frank (2004)	Naïve Bayes/SVM	Yahoo Finance message board
Renault (2017)		StockTwits
Bartov et al. (2017)		Twitter
Barbon et al. (2019)	Naïve Bayes	Firm-specific news
Ke et al. (2019)	Customized ML	Dow Jones Newswire articles
Huang et al. (2014)	Naïve Bayes	Analyst reports
Manela and Moreira (2017)		Wall Street Journal front-page articles
Vamossy (2020)	Deep learning	StockTwits posts
Liew and Wang (2016)	Commercial ML	Twitter posts

Product Market Sentiment

	Methods	Data
Tang (2018)	Commercial ML	Twitter posts

Measures of Corporate Executives' Personality and Decisions

Executives' Personality

	Methods	Data	Measures
Gow et al. (2016)	Naïve Bayes/SVM	Yahoo Finance message board	Naïve Bayes/SVM
Hrazdil et al. (2020)	IBM service	Conference calls	CEO and CFO personality
Hsieh et al. (2020)	Face detection	Executives' business headshot images	trustworthiness
Du et al. (2019)	Textual analysis	Mutual fund managers' letters to shareholders	Managers' level of confidence

Executives' Decisions

	Methods	Data	Measures
Bandiera et al. (2020)			Whether perform low/high-level tasks
Barth et al. (2020)		Earnings conference call	How they withhold information
Hu and Ma (2020)		Youtube videos	How founders speak? What they say? How they present visually

Measures of Firm Characteristics and Corporate Policies

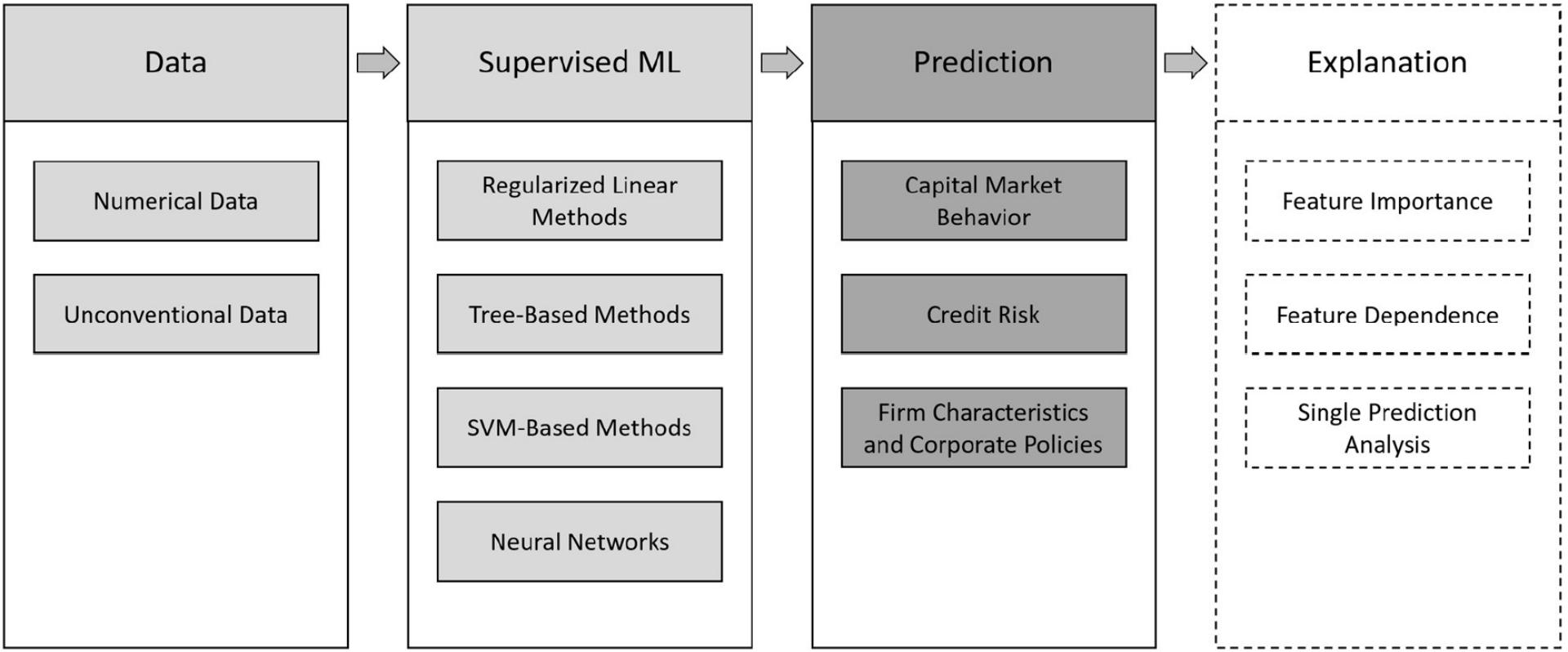
Corporates

Methods	Data	Measures
Li et al. (2020)	Conference call transcripts	Corporate cultures
Buehlmaier and Whited (2018)	Annual report	Financial constraints
Lowry et al. (2020)	SEC letters	Regulatory IPO concern

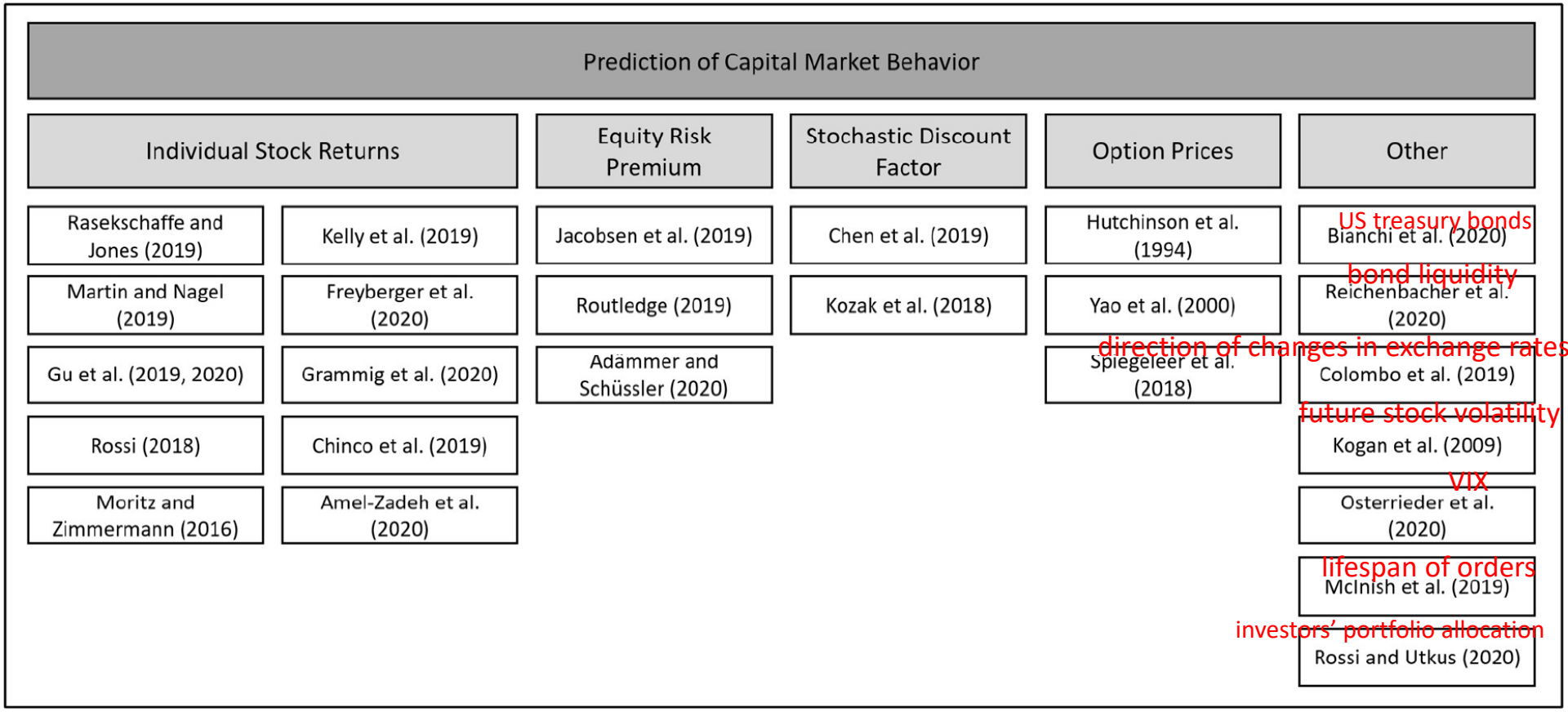
Financials

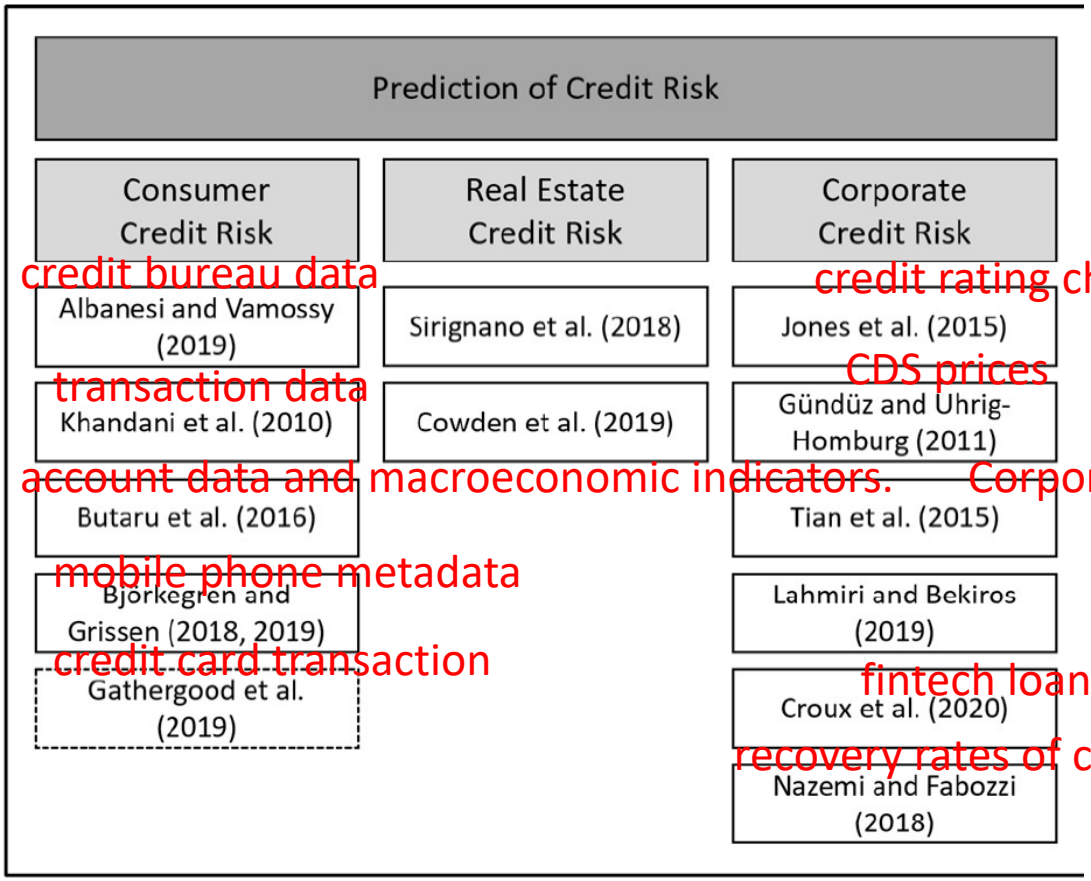
Methods	Data	Measures
Hanley and Hoberg (2019)	Commercial ML	Banks' annual report	Risk exposure
Bubna et al. (2020)	Cluster		Venture capital relatedness
Bertsch et al. (2020)		Customer complaint texts	Bank misconduct

2.2 Reduction of Prediction Error in Economic Prediction



- Explanation
  - Feature importance: importance score
    - eg. Permutation importance (permute X_i and calculate the score)
  - Feature dependence: relations between predictor and target
    - Partial dependence plots
  - Single prediction analysis: disentangle the contribution of every predictor to a specific prediction value
    - Sharpley Additive Explanation(SHAP)


Prediction of Capital Market Behavior
Individual Stock Returns
Equity Risk Premium
Stochastic Discount Factor
Option Prices
Other
Rasekschaffe and Jones (2019)
Kelly et al. (2019)
Jacobsen et al. (2019)
Chen et al. (2019)
Hutchinson et al. (1994)
Bianchi et al. (2020)
US treasury bonds
Martin and Nagel (2019)
Freyberger et al. (2020)
Routledge (2019)
Kozak et al. (2018)
Yao et al. (2000)
Reichenbacher et al. (2020)
bond liquidity
Gu et al. (2019, 2020)
Grammig et al. (2020)
Adämmer and Schüssler (2020)
Spiegeleer et al. (2018)
Colombo et al. (2019)
direction of changes in exchange rates
Rossi (2018)
Chinco et al. (2019)
Kogan et al. (2009)
future stock volatility
Moritz and Zimmermann (2016)
Amel-Zadeh et al. (2020)
Osterrieder et al. (2020)
vix
McInish et al. (2019)
lifespan of orders
Rossi and Utkus (2020)
investors' portfolio allocation



credit bureau data

credit rating changes

transaction data

CDS prices

account data and macroeconomic indicators.

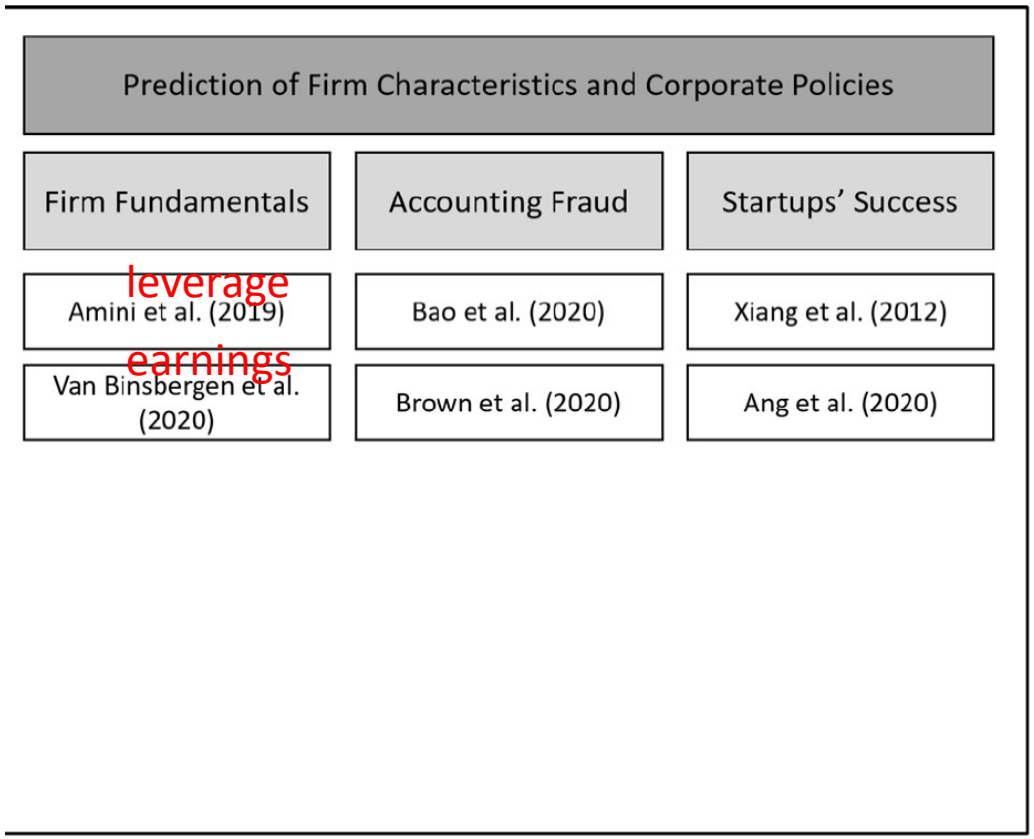
Corporate bankrupt

mobile phone metadata

credit card transaction

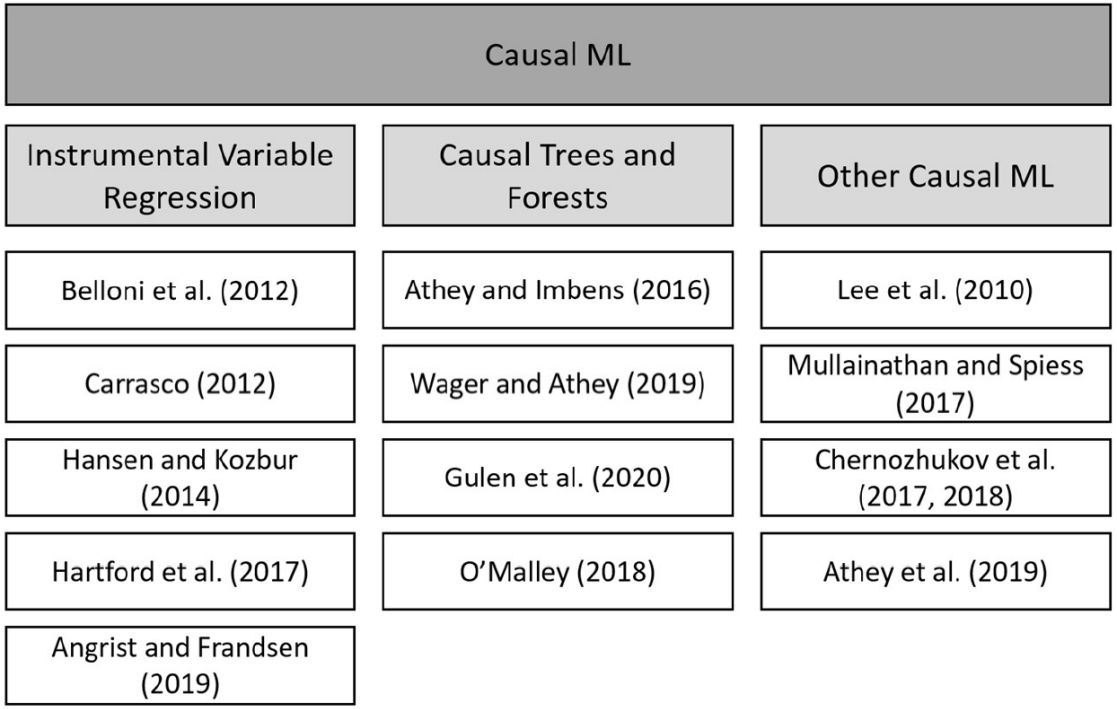
fintech loan default

recovery rates of corporate bonds


Prediction of Firm Characteristics and Corporate Policies
Firm Fundamentals
Accounting Fraud
Startups' Success
leverage
Amini et al. (2019)
Bao et al. (2020)
Xiang et al. (2012)
earnings
Van Binsbergen et al. (2020)
Brown et al. (2020)
Ang et al. (2020)

2.3 Extension of the Existing Econometric Toolset

- ML-enhanced instrumental variable regression
- Causal trees and forests



- Recall 2SLS:

$$y_i = \alpha + \beta x_i + \epsilon_i (i = 1, \dots, n)$$
$$cov(x_i, \epsilon_i) \neq 0$$

  - IV: $cov(z_i, \epsilon_i) = 0; cov(z_i, x_i) \neq 0$
  - First stage:

$$x_i = \gamma + \delta z_i + u_i$$
$$\widehat{x_i} = \hat{\gamma} + \hat{\delta} z_{\mathrm{i}}$$

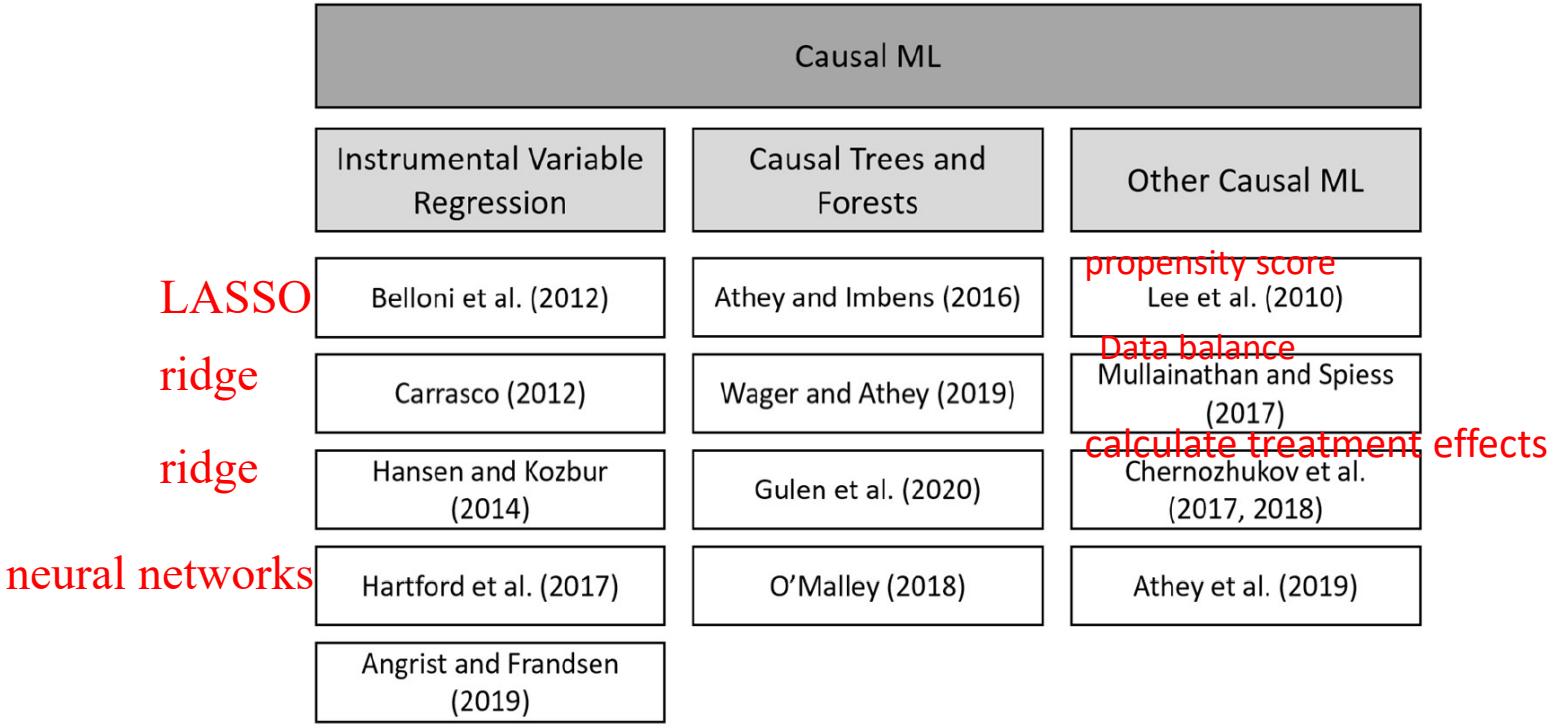
  - Second stage:

$$y_i = \alpha + \beta \widehat{x_i} + (\epsilon_i + \beta \widehat{u_i})$$

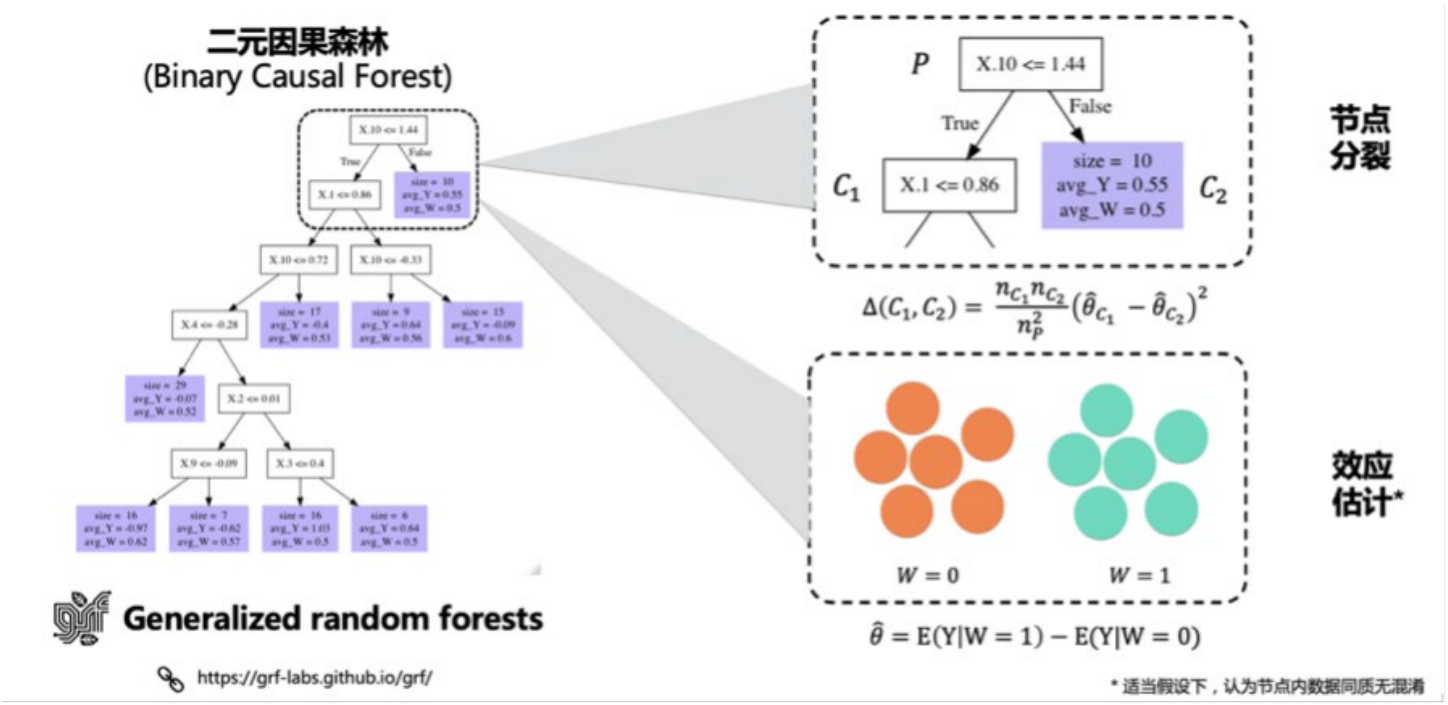
- Causal ML & IV
  - Better predictions for the IV results in more precise estimates in the second stage.

Extension of the Existing Econometric Toolset

- ML-enhanced instrumental variable regression
- Causal trees and forests



- Causal trees and forests
  - 从协变量中，找到一个最优分裂节点，**最大化子节点间处理效应差异**



Summary

- ML methods:
  - Supervised learning
  - Unsupervised learning
  - Others (reinforcement learning…)
- Taxonomy of ML
  - Measure construction
  - Prediction
  - Econometric extension
- Summary