Machine Learning and Applications in Finance and Macroeconomics

Discussion by
Amit Seru
University of Chicago

Macro Financial Modeling
2016 Winter Meeting

Overview

- Machine Learning
  - Mortgage delinquency risk
  - Credit card delinquency risk
  - Matching datasets

- Comments
  - Where might we need this…
  - Predictions in a changing world…

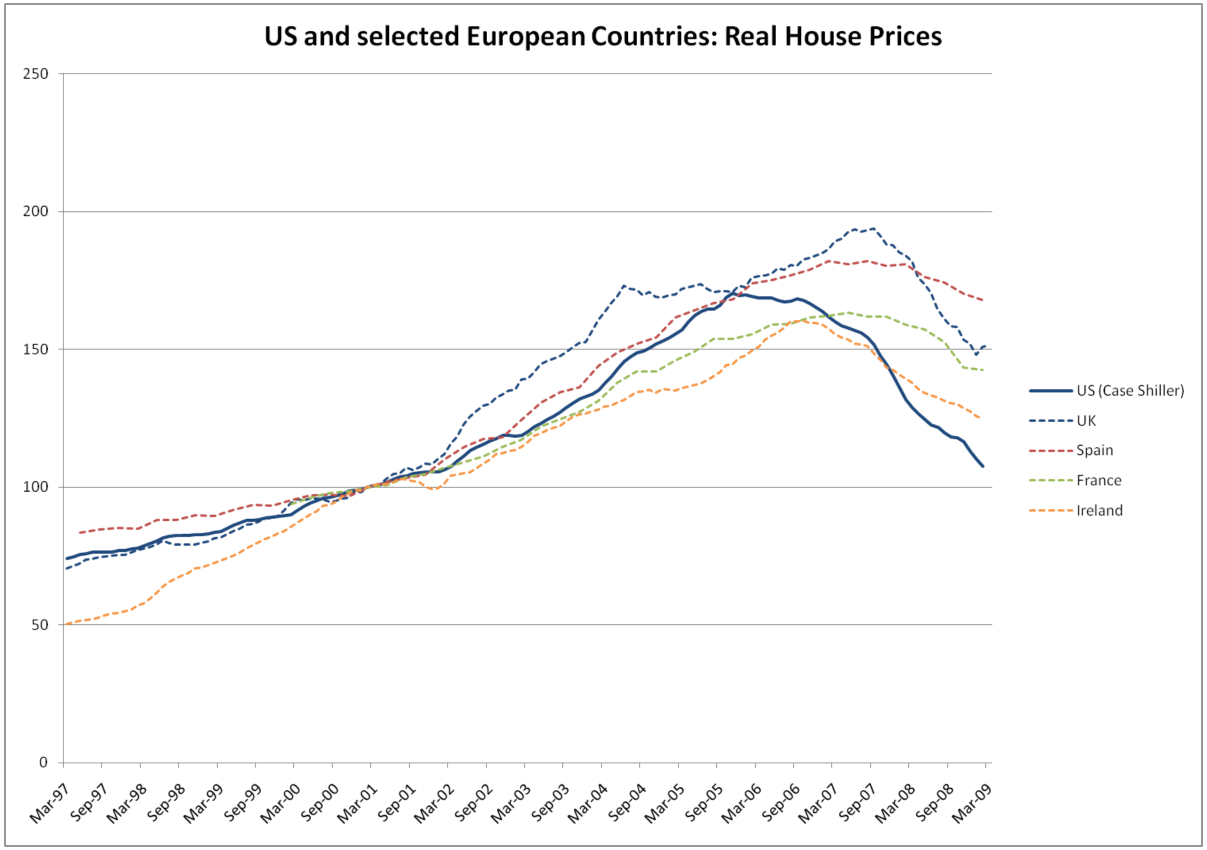
Where might we need this?

- Better risk assessment
  - Investments
  - Regulation/Capital charges

- Regulation/Supervision
  - At what frequency are we running these models?
    - ➢ Every supervision cycle?
    - ➢ Every stress test?
    - ➢ Does it matter if we can predict better over a short horizon at high frequencies?

  - Where did we need these predictions in supervision and regulation?

Mortgage Defaults Paper



- How well do we do when the market conditions change?
  - Train for 12 years (1999-2011) and test for 2012-2014
  - What if train for 2001-2006 and test for 2007-2009?

Machine Learning: Interpreting the “Black Box”

- Combine several observable factors to “improve” predictions
  - How do we interpret the results from the black box?
    - Would matter for what policy intervention is designed.

Credit Card Defaults Paper

- Why is risk management different across banks?
  - Selection of consumers different. (Citi v/s Capital One)
  - Selection across regions different (Citi urban v/s BofA rural)
  - Rewards program different (gas, airlines, stores, etc etc)
  - Risk management practices different -- some estimate at the account level; others at the portfolio level.
  - Attrition rates could be different

Machine Learning: Interpreting the "Black Box"

- Combine several observable factors to "improve" predictions
  - How do we interpret the black box?
- Not modeling primitives ➔ could potentially miss key quantitatively important aspects. Would this matter?
  - Incentives of agents that influence data generating process
    - ➢ Meaning of observables can change over time

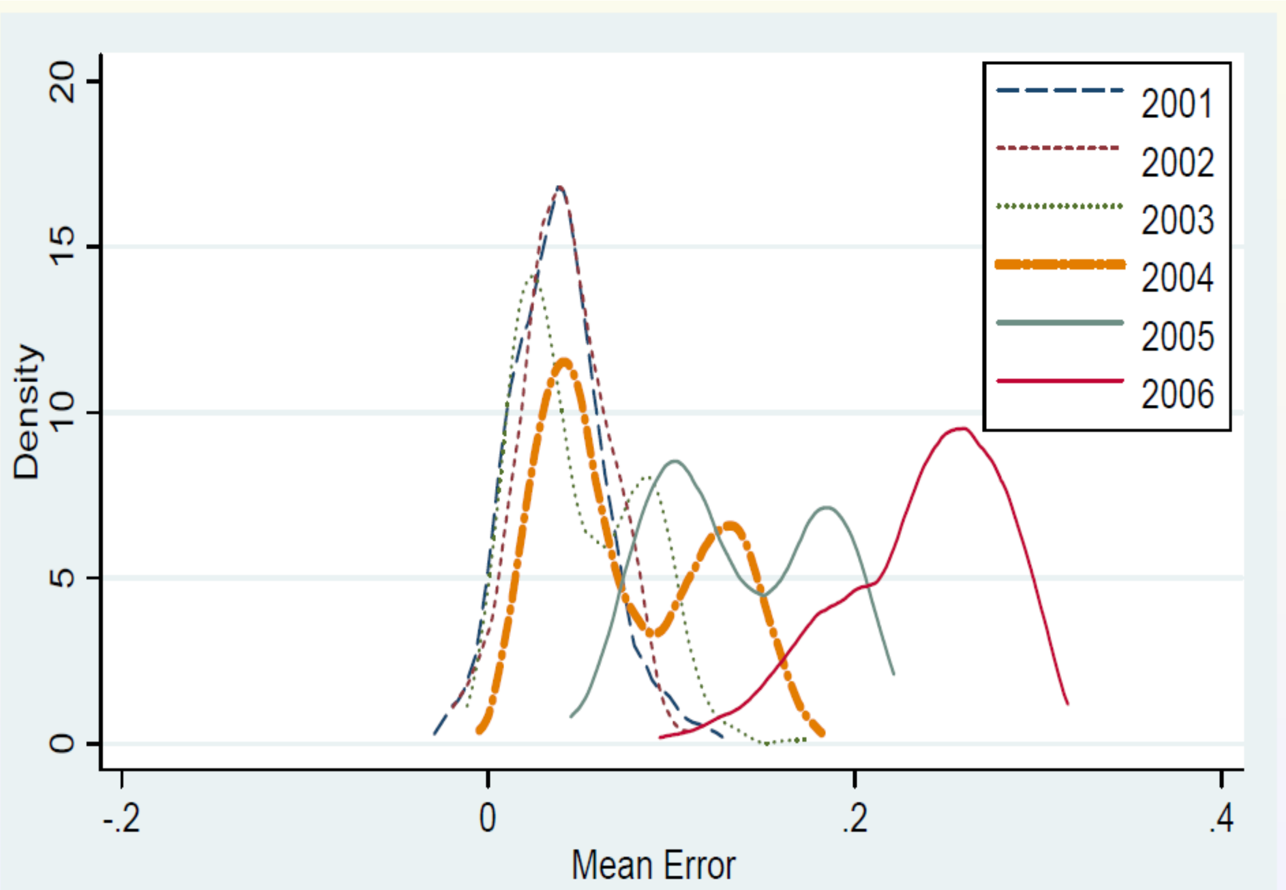
Incentives: Reliance on Hard Information

$$r_i = \alpha + \beta_{FICO} \times FICO_i + \beta_{LTV} \times LTV_i + \epsilon_i.$$

	β_{FICO}	β_{LTV}	R^2 (in %)	Observations
1997	-0.004***	0.030***	3	24067
	(.0002)	(.0013)		
1998	-0.007***	0.035***	7	60094
	(.0001)	(.0008)		
1999	-0.007***	0.020***	8	104847
	(.0001)	(.0005)		
2000	-0.010***	0.035***	14	116778
	(.0001)	(.0004)		
2001	-0.012***	0.038***	20	136483
	(.0001)	(.0004)		
2002	-0.011***	0.071***	18	162501
	(.0001)	(.0001)		
2003	-0.012***	0.079***	32	318866
	(.0001)	(.0001)		
2004	-0.010***	0.097***	40	610753
	(.0001)	(.0001)		
2005	-0.009***	0.110***	48	793725
	(.0001)	(.0001)		
2006	-0.011***	0.115***	50	614820
	(.0001)	(.0001)		

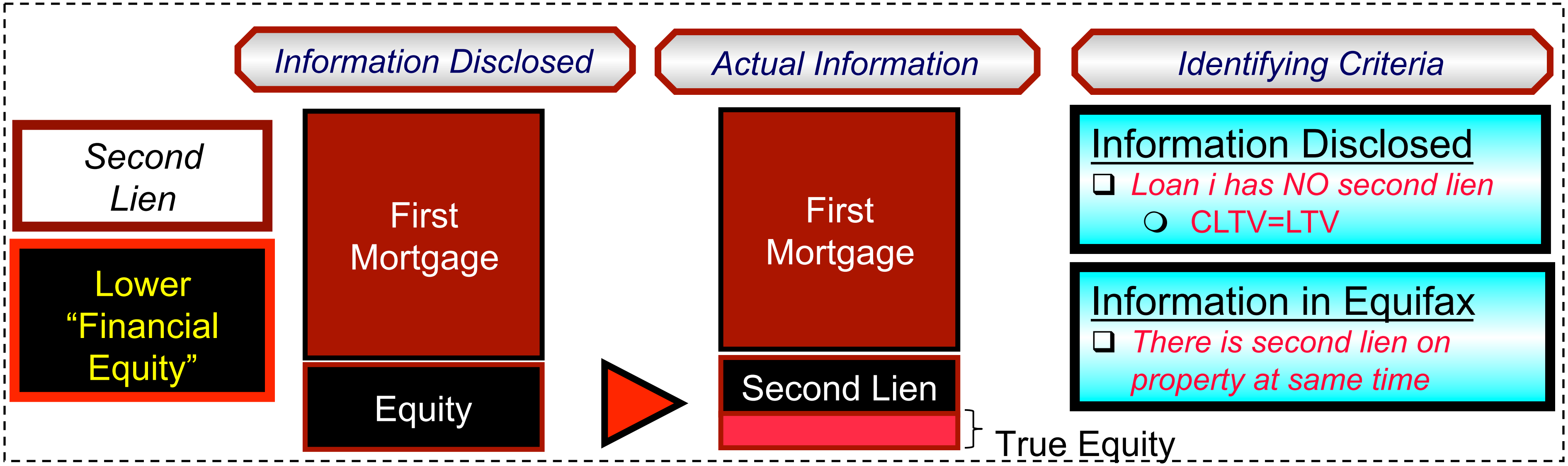
- Dramatic increase in the R^2: about 3% in 1997, goes upto almost 50%
  - $\beta_{FICO} < 0$, $\beta_{LTV} > 0$.
- In the low securitization regime, the hard information variables explain very little variation in interest rates
  - Omitted variables are particularly important
  - Soft information, by its very nature, is one of the omitted variables

Incentives: Actual versus Predicted Defaults

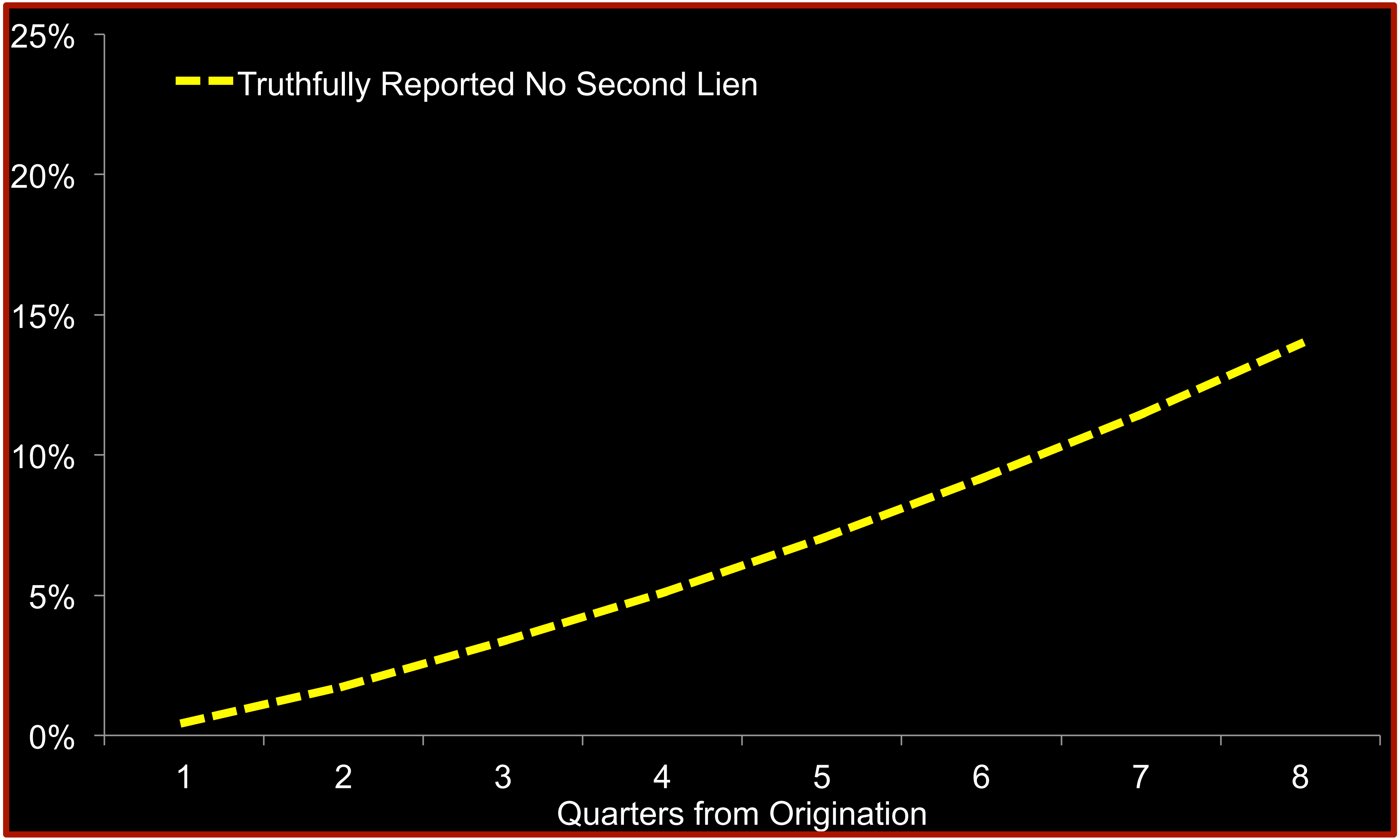


Machine Learning: Interpreting the "Black Box"

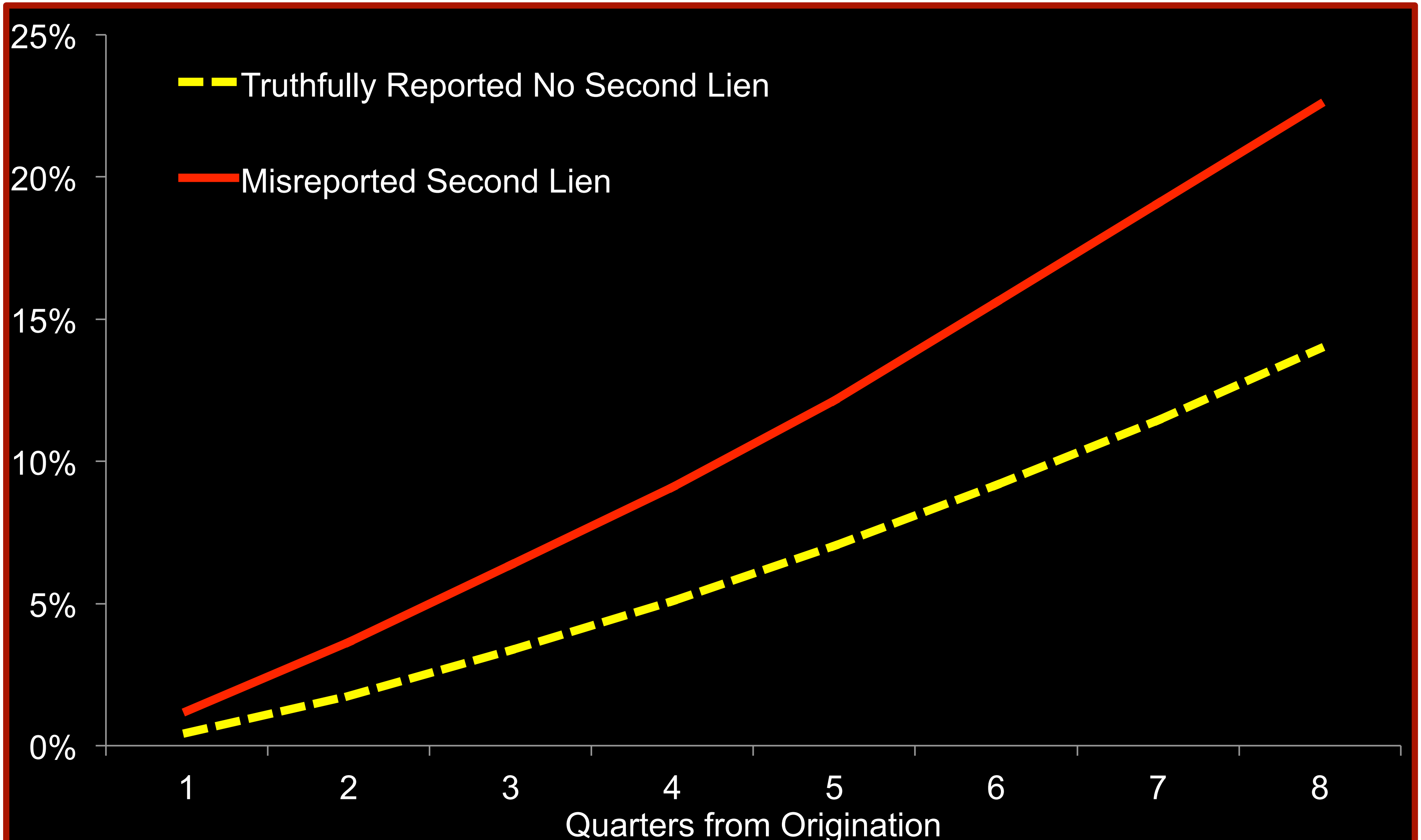
- Combine several observable factors to "improve" predictions
  - How do we interpret the black box?
- Not modeling primitives ➔ could potentially miss key quantitatively important aspects. Would this matter?
  - Incentives of agents that influence data generating process
    - ➢ Meaning of observables can change over time



DGP

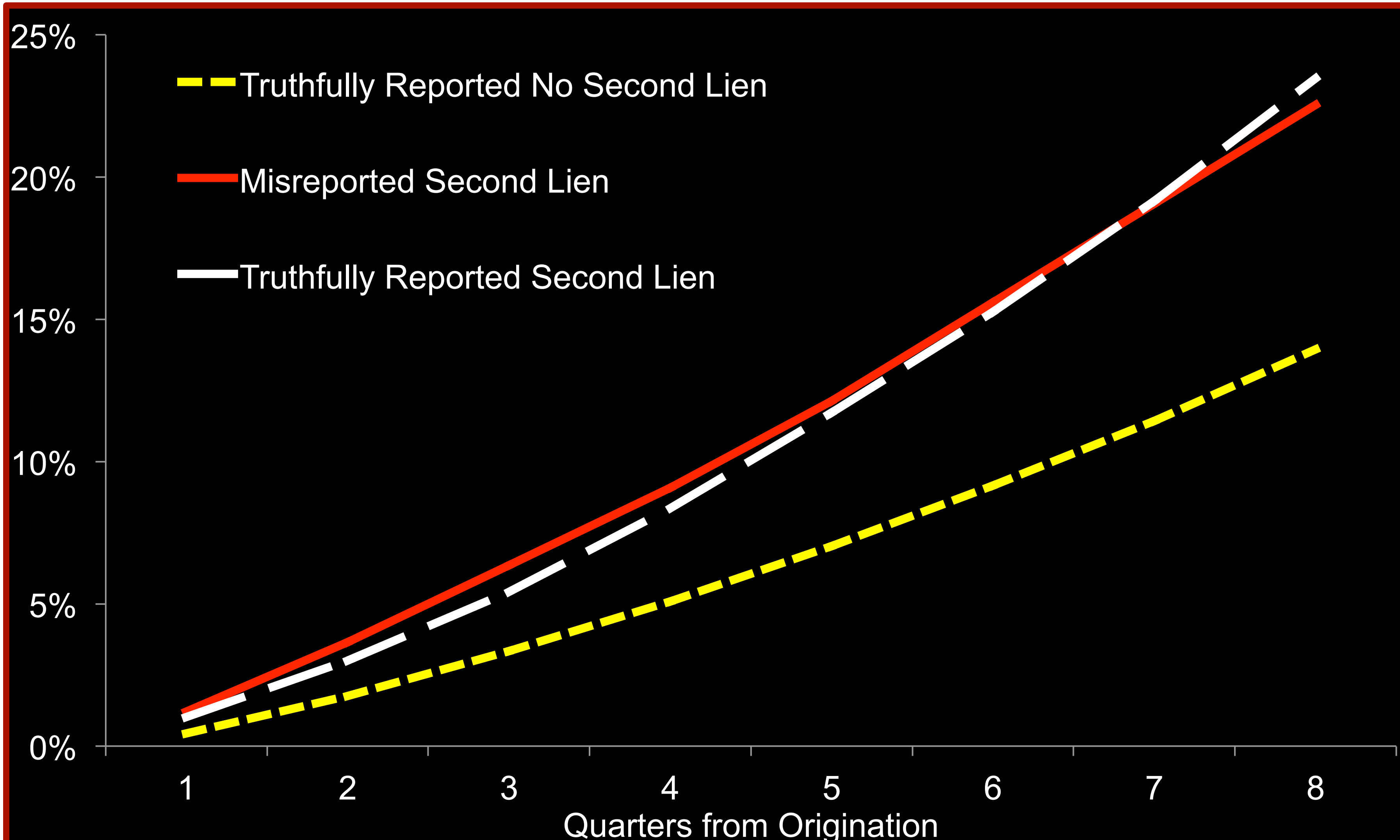

25%
20%
15%
10%
5%
0%
Truthfully Reported No Second Lien
1
2
3
4
5
6
7
8
Quarters from Origination

DGP



25%
20%
15%
10%
5%
0%
Truthfully Reported No Second Lien
Misreported Second Lien
1
2
3
4
5
6
7
8
Quarters from Origination

DGP


25%
20%
15%
10%
5%
0%
Truthfully Reported No Second Lien
Misreported Second Lien
Truthfully Reported Second Lien
1
2
3
4
5
6
7
8
Quarters from Origination

Machine Learning: Interpreting the "Black Box"

- Combine several observable factors to "improve" predictions
- Not modeling primitives ➔ could potentially miss key quantitatively important aspects
  - Policy interventions at times when "value of prediction" the most
    - ➢ Foreclosure behavior different before and after HAMP
    - ➢ Foreclosure behavior different before and after HARP

  ➔ accounting for timing and eligibility of some borrowers important

  - Institutional factors could change data generation
    - ➢ Different incentives to foreclose depending on ownership status
    - ➢ Organizational ability of servicers to pass through government subsidies

Conclusion

- Nice and interesting set of papers
  - Machine learning can help improve predictions
  - Better data would be better
- Going forward
  - What are we using these models for?
    - ➢ Regulation/Supervision/Investments?
  - Do these models do better on the Lucas critique?
    - ➢ Incentives/Institutional Factors/Government Interventions
  - How would one interpret the model with…
    - ➢ …Other agencies also fitting same data (FICO, Zillow, TransUnion)
    - ➢ …Human capital/political constraints in supervision