How Does a Firm Adapt in a Changing World? The Case of Prosper Marketplace

Xinlong Li* and Andrew Ching+

*Nanyang Technological University, NTU Business School
+Johns Hopkins University, Carey Business School

Research Questions

How does a firm use past data in a fast-changing world?

- The world is changing fast (so-called *concept drift*).
- If we rely on old data to predict what will happen next, we could be way off, e.g., #COVID cases (omicorn surge), Afghanistan's situation, Ukraine's situation, etc.
- Not all past data are equally informative about the current environment.
- With high frequency streaming data, tech firms can select more relevant data to use.
- Research Question: How do they select the past data?

Our Approach

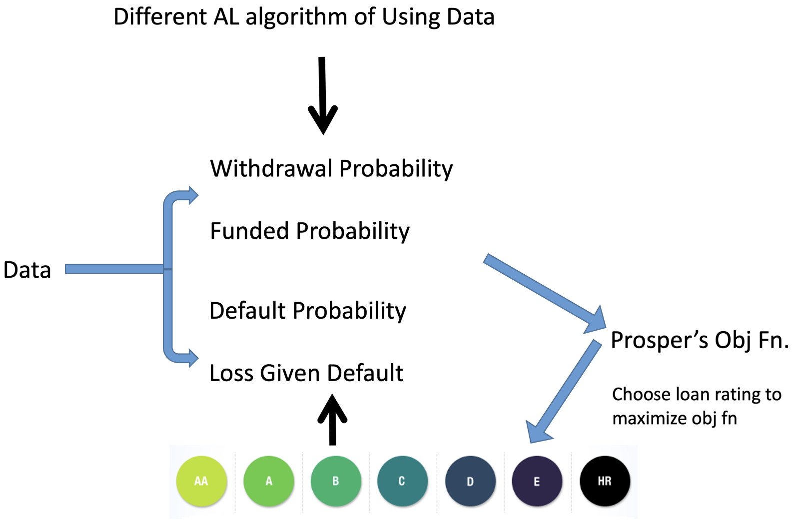
We propose a Generalized Revealed Preference Approach.

- Typically, Revealed Preference allows us to uncover structural parameters.
- We argue that firm's choice is also a function of its Adaptive Learning/Data Selection algorithm.
- By observing their choices, together with a structural model of its decision, we can back out their adaptive learning/data selection (AL) algorithm.
- A firm (or any agent) may need to set price, choose rating, anticipate upcoming inflation, economy outlook, etc.
- Helps improve our fundamental understanding about decision making.
- Useful for market intelligence – a firm may want to learn about how its rivals use the past data to set its market strategies.

Prosper.com

- A P2P lending platform.
- Two years data (12/19/2010 – 12/31/2012) from Prosper.com, right after they have introduced posted price business model.
- Two-sided market for lenders and borrowers.
- Prosper assesses the risk of each borrower and assigns rating to each loan.
- Prosper only earns a fee when a loan is successfully funded.
- When assigning a rating, we allow Prosper to care about both (i) current profits, and (ii) reflecting the true evaluated risk of a loan (ethic & reputation concern).
- We estimate Prosper's objective function to infer their relative importance.

Model Framework



Prosper Ratings (Observed Choice)

Summary of Results

- Among five AL/data selection methods we consider, Ensemble-Recession Probability Method (E-RPM) explains Prosper's decisions (loan ratings) the best.
- Counterfactual Experiments:
  - Only focusing on reporting evaluated loan risks (revenue dec by 4.25%, risk misreporting dec by 5.58%).
  - Only focusing on current revenue (rev. inc by 6.47%, risk misreporting inc by 7.26%).
  - No Adaptive Learning (revenue decreases by 3.06%, risk misreporting inc by 4.69%).
  - 'Know' the true DGP (rev. inc by 8.04%, risk misreporting dec by 9.61%).
- Main Contributions:
  - Show that Revealed Preference allows us to infer how a firm adaptively learn.
  - Also recover the structural parameters of Prosper's objective function – allowing us to see how it does trade-off between current revenue and reporting its evaluated loan risks.

Outline

- Related Literature.
- Institutional Background of Prosper.com.
- Data and Reduced Form Evidence for Concept Drift
- Empirical Framework to Evaluate Adaptive Learning (AL) algorithms
- Five AL algorithms under consideration
- Results

Related Literature

- Firms' Learning.
  - Adaptive Learning as Econometrician: Doraszelski, Lewis and Pakes (2018), Evans and Honkapohja (2001, 2013), Sargent (1993), etc.
  - Bayesian Learning: Hitsch (2006), Ching (2010), Huang et al. (2019).
  - Agnostic about learning mechanism: Huang, Lovett and Ellickson (2022).
  - But these papers assume data generating process remains unchanged (no concept drift).

- Consumer's Changing Preference Over Time.
  - Dew, Ansari and Li (2020), Sriram, Chintagunta and Neelamegham (2006), Liechty, Fong and DeSarbo (2005), etc.

- Peer-to-Peer Lending (Crowdfunding).
  - Lin et al. (2009), Netzer et al. (2016). Kawai, Onishi and Uetake (2020), Zhang and Liu (2012), Liu, Wei and Xiao (2020)
  - None of them deals with concept drift and adaptive learning.

Prosper Marketplace

Table: Interest Rates, (Posted) Estimated Loss Rate, and Commission Rates

	AA	A	B	C	D	E	HR
Interest Rate	7.74%	10.80%	15.88%	19.92%	24.85%	30.43%	31.78%
Commission Fee Rate	0.5%	3%	3%	4.5%	4.5%	4.5%	4.5%
Posted Est. Loss Rate	1.42%	3.03%	5.56%	7.94%	10.83%	14.41%	17.08%
Service Fee Rate	1%	1%	1%	1%	1%	1%	1%

- 1-1 mapping between rating and interest rate and posted est. loss rate [or "posted loss rate"]
- A loan application:
  - is Funded in 14 days, or Expired.
  - can be Withdrawn by a borrower before it is expired.
- A borrower may Default.
- Price (interest rate) setting: Prior to Dec 2010, Prosper uses auction to set interest rate. After that, Prosper sets interest rates on its own (as a function of rating).

Prosper's Decision Process

- Prosper needs to assign a rating to each loan application.
  - Default Probability and Loss Rate Given Default.
  - Two-sided market model: Funding Probability (lender side) and Withdrawal Probability (borrower side).

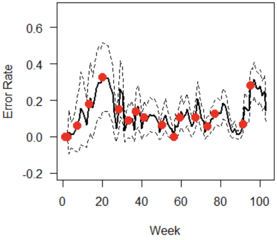
Table: Interest Rates Across Ratings

	AA	A	B	C	D	E	HR
Interest Rate	7.74%	10.80%	15.88%	19.92%	24.85%	30.43%	31.78%

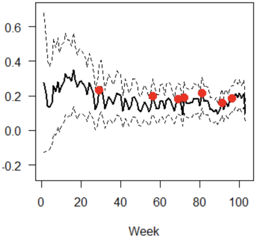
Reduced Form Evidence of Concept Drift

- Data range: 12/19/2010 to 12/31/2012.
- 31,807 loan applications. 50+ borrower characteristics.
- For each loan, we observe (i) whether a borrower withdraws, (ii) whether it is funded, (iii) whether default, (iv) if default, % loss.
- We use the test by Gama, et al., (2004). Take lenders' investment behavior for example.
  - We build a model to predict lender's investment behavior based on the past data.
  - Lenders' investment behavior is stationary over time $\Rightarrow$ The model's prediction performance should remain stable.
  - Concept drift: lenders' investment behavior changes $\Rightarrow$ The model's prediction performance will drop.

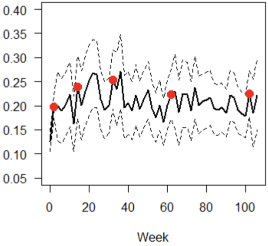
Reduced Form Evidence of Concept Drift



(a) Prediction Error Rate in Funded Probabilities

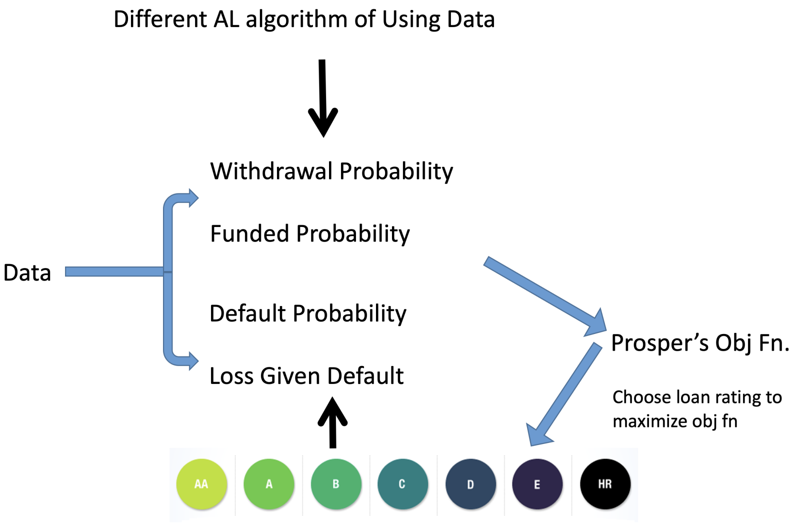


(b) Prediction Error Rate in Withdrawal Probabilities



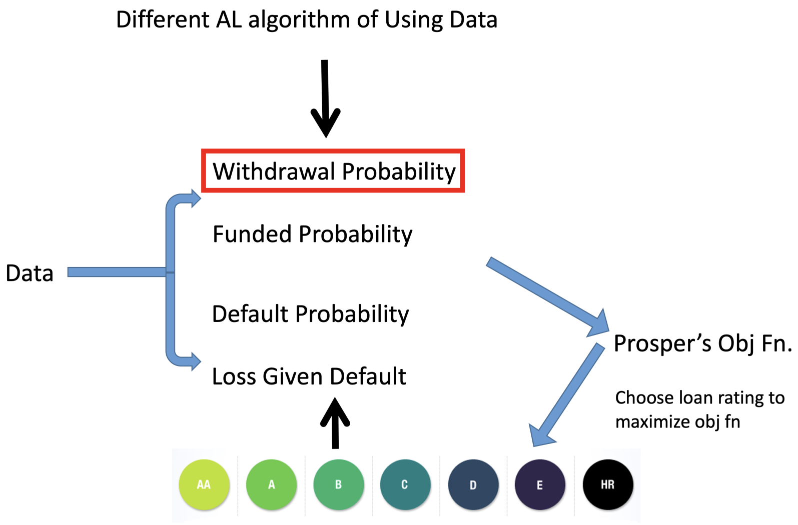
(c) Prediction Error Rate in Default Probabilities

Model Framework



Prosper Ratings (Observed Choice)

Model Framework



Prosper Ratings (Observed Choice)

Borrower's Utility

Borrower i's utility of accepting a rating l loan:

$$U_{ilt}(X_i, Z_{il}; \boldsymbol{\gamma_t}) = \gamma_{lt} + \gamma_{1t} \cdot O_l \cdot M_i + \gamma_{2t} \cdot MP_{il} + \gamma_{3t} \cdot M_i + \epsilon_{i1t}, \tag{1}$$

- O_l: Commission (Origination) fee rate at rating l.
- M_i: Amount requested by loan application i.
- MP_{il}: Monthly payment.

Borrower's Utility

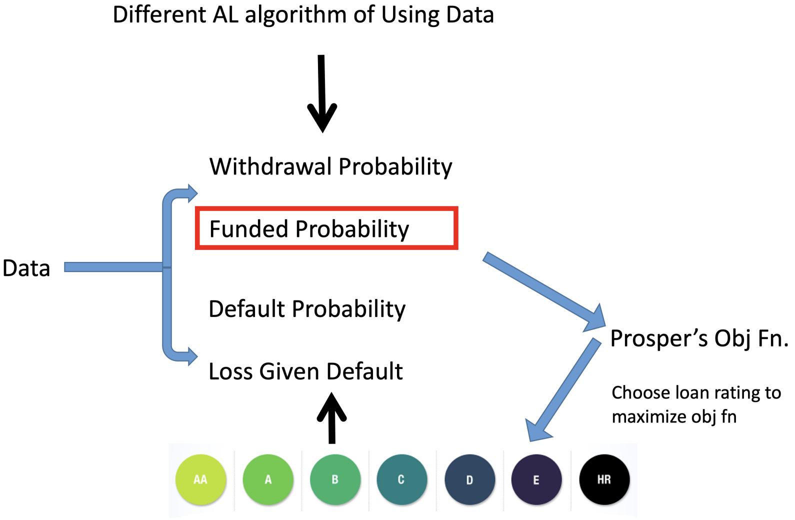
Utility of outside option:

$$U_{i0t} = f_0(X_i, E_t; \boldsymbol{\gamma_t}) + \epsilon_{i0t}, \tag{2}$$

- X_i: Borrower specific characteristics like income, etc
- E_t: Mortgage rate, S & P 500 index, etc.
- ϵ_{i1t}, ϵ_{i0t} follow type I extreme value distribution.

$$W_{ilt} = Pr(Withdrawal = 1|\gamma_t) = \frac{\exp(U_{i0t})}{\exp(U_{i0t}) + \exp(U_{i1t})} \tag{3}$$

Model Framework



Prosper Ratings (Observed Choice)

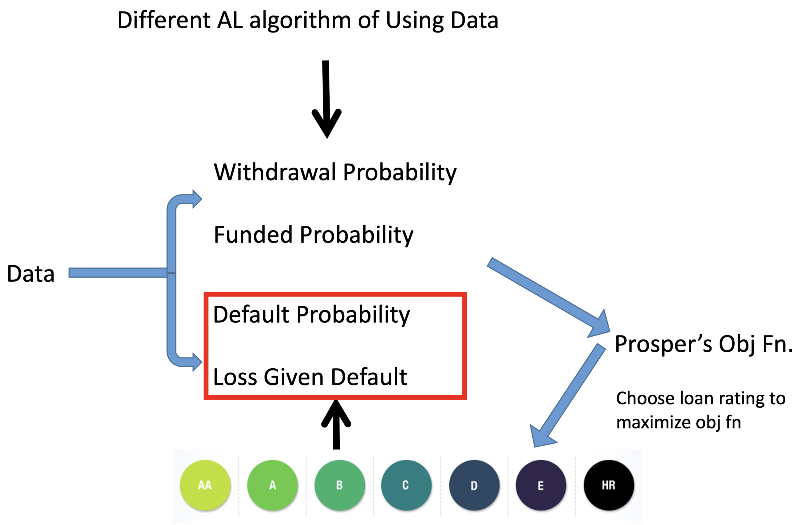
Funding Probability (Lender Side)

Loan application i's funding probability at rating l is given by:

$$F_{ilt} = Pr(\text{Funded} = 1 | X_i, Z_{il}, E_t; \boldsymbol{\beta_t}) = \frac{\exp(\beta_{lt} + X_i\beta_{1t} + Z_{il}\beta_{2t} + E_t\beta_{3t})}{1 + \exp(\beta_{lt} + X_i\beta_{1t} + Z_{il}\beta_{2t} + E_t\beta_{3t})}, \tag{4}$$

- X_i: Borrower specific characteristics like income, etc
- Z_{il}: Rating specific characteristics like interest rate, etc
- E_t: Macroeconomic factors like mortgage rate, S&P 500 index, etc.

Model Framework



Prosper Ratings (Observed Choice)

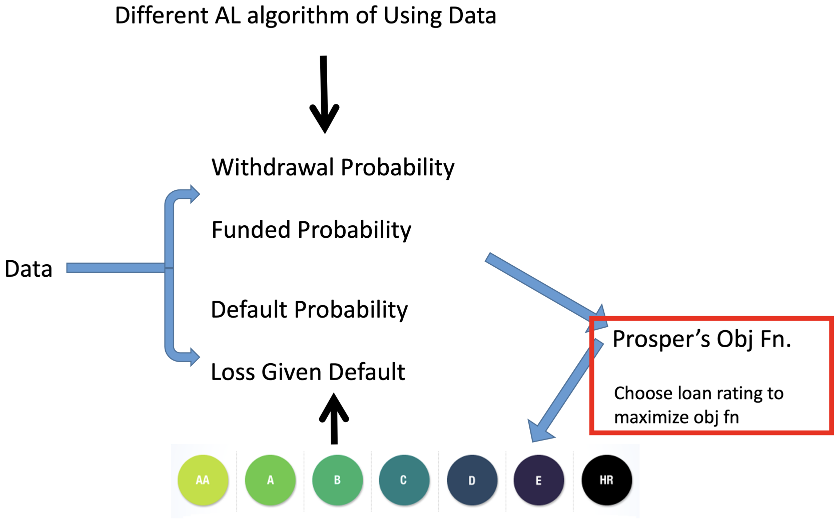
Borrower Risk: Naive Bayes Model

- Default Probability: ED_{ilt}.
- Loss Rate Given Default: $ELGD_{ilt}$.

The reason to use Naive Bayes Model

- Computational advantages.
- Good at detecting credit card fraud. (Albashrawi and Lowell 2016, Zareapoor and Shamsolmoali 2015, Ngai et al. 2011).

Model Framework



Prosper Ratings (Observed Choice)

Firm's Objective Function

Prosper maximizes the following indirect utility:

$$U_{ilt}(X_i, Z_{il}|S_t, AL, \alpha, \delta_1, \delta_2) = \alpha_l + \delta_1 \cdot \overbrace{(O_l + L_{ilt}) \cdot M_i \cdot F_{ilt} \cdot (1 - W_{ilt})}^{\text{Profits}} + \delta_2 \cdot |PLR_l - ED_{ilt} \cdot ELGD_{ilt}| + \epsilon_{ilt},$$

- AL: Adaptive Learning algorithm.
- O_l: Commission rate for rating l.
- L_{ilt}: Service fee rate.
- M_i: Amount requested by loan application i.
- F_{il}: Funding probability of loan application i with rating l conditional on no withdrawal.
- W_{il}: Withdrawal probability of loan application i with rating l.

Firm's Objective Function

Prosper maximizes the following indirect utility:

$$U_{ilt}(X_i, Z_{il}|S_t, AL, \alpha, \delta_1, \delta_2) = \alpha_l + \delta_1 \cdot (O_l + L_{ilt}) \cdot M_i \cdot F_{ilt} \cdot (1 - W_{ilt}) +$$

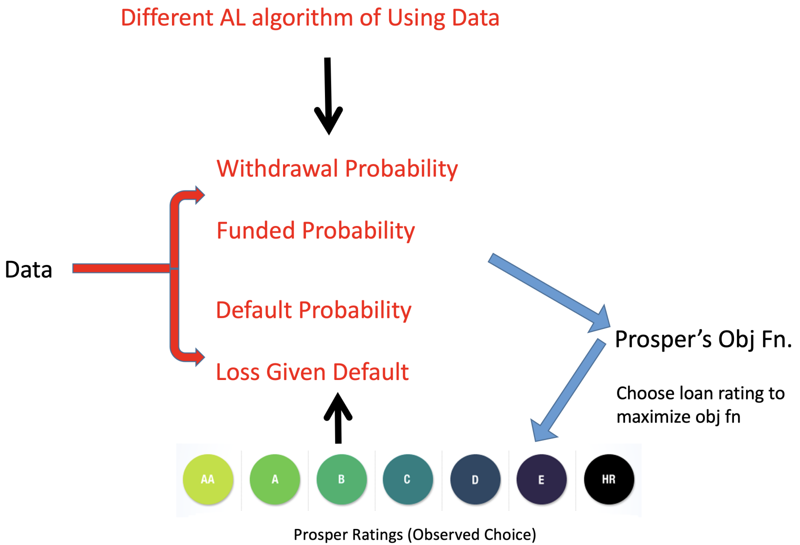
$$\delta_2 \cdot \overbrace{|PLR_l - ED_{ilt} \cdot ELGD_{ilt}|}^{\text{Reputation}} + \epsilon_{ilt},$$

- PLR_l: Posted loss rate at rating l reported by Prosper.
- ED_{ilt}: Estimated default probability of loan application i with rating l.
- $ELGD_{ilt}$: Estimated loss given default of loan application i with rating l.

Table: Posted Loss Rate, PLR_l

	AA	A	B	C	D	E	HR
Estimated Loss	1.42%	3.03%	5.56%	7.94%	10.83%	14.41%	17.08%

Model Framework



Adaptive Learning/Data Selection Methods

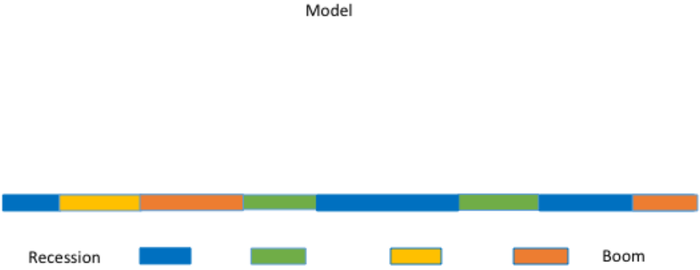
- 1. Equal Weight Method (EWM)
    - Assumption: World is stationary (Dorazelski et al. 2018; Sargent, 1993; Evans and Honkapohja, 2001).

- 2. Gradual Forgetting Method (GFM)
    - Assumption: More recent data carries more information about the current market.

- 3. Moving Window Method (MWM)
    - Assumption: Only data within a certain window size is relevant to the current market.

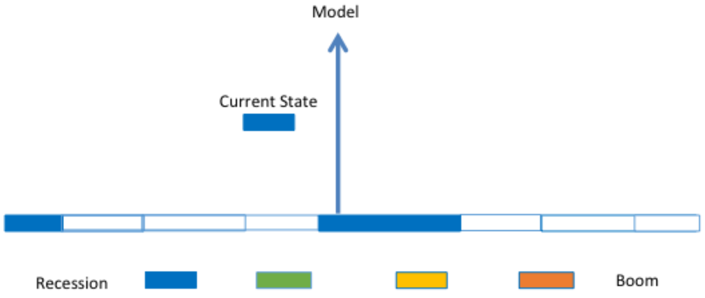
RPM

- 4. Recession Probability Method (RPM)
  - Assumption: Consumers tend to behave closely in similar economic environments.

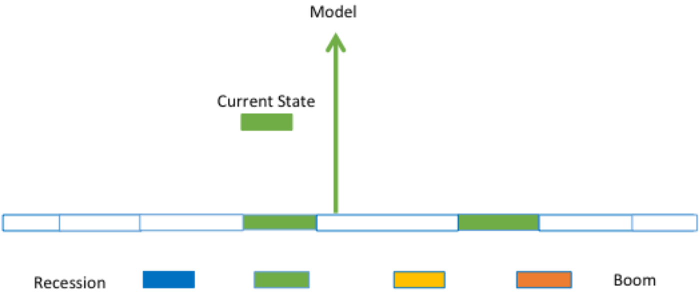


RPM

Recession Probability Method

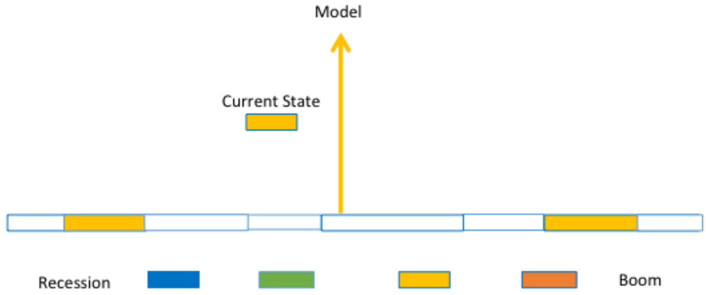


Recession Probability Method



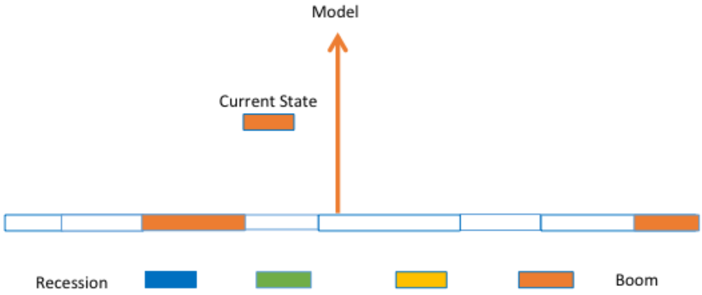
RPM

Recession Probability Method



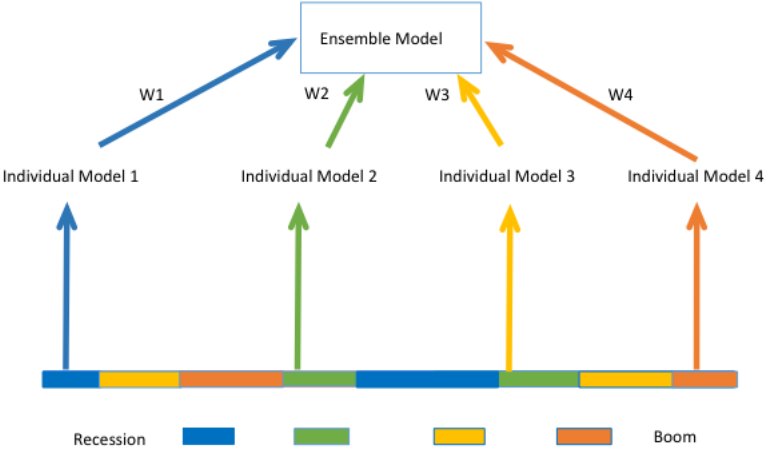
RPM

Recession Probability Method



E-RPM

- 5. Ensemble Recession Probability Method (E-RPM)
  - Assumption: Each individual model represents a certain economic environment.
  - Take weighted average of those individual models.



E-RPM

- 5. Ensemble Recession Probability Method (E-RPM)
  - Weight on each individual model is determined by the model's prediction performance in the last period.
  - Wang et al. (2003) proposes setting weight for poorly performed model to zero.

$$w'_{jt} = \begin{cases} 0 & \text{if model } j \text{ has the largest or} \\ & \text{second largest MSE in period t-1;} \\ [MSE_{jt-1}]^{-1} & \text{otherwise.} \end{cases}$$

Identification

- β_t (Lender Side) can be identified from lenders' investment decisions.
- γ_t (Borrower Side) can be identified from borrowers' withdrawal decisions.
- Parameters in the Naive Bayes Model can be identified from borrowers' default behavior.
- δ_1 and δ_2 can be identified from the trade-off between profits and reputation when assigning the rating.

Identification

- At the very beginning of 2022:
  - If Prosper uses data from Jan 2020 to Dec 2021, it would predict that a restaurant worker is very risky, and assign a low rating to the borrower.
  - if Prosper only uses the most recent 6 months data to calibrate its predictive models, restaurant workers would receive much better ratings.

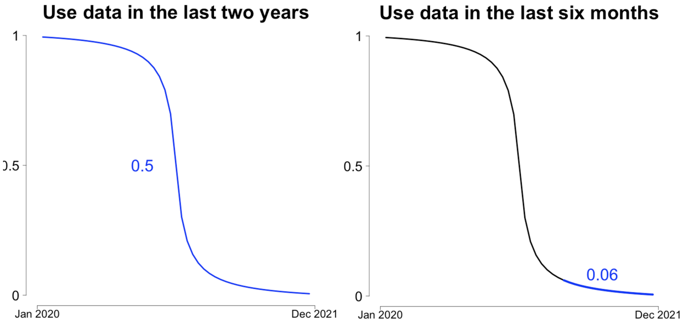


Figure: Average Default Rate of Restaurant Workers

Identification

- We have run simulation analysis to demonstrate we can successfully identify the parameters, as well as the learning algorithms.

Estimation Strategy

Two-step estimation: For any given AL/Data Selection Algorithm,

- Step 1: W_{ilt}, F_{ilt}, ED_{ilt} and $ELGD_{ilt}$ are estimated from consumers' investment, withdrawal and default outcomes.

$$\alpha_l + \delta_1 \cdot (O_l + L_{ilt}) \cdot M_i \cdot (1 - W_{ilt}) \cdot F_{ilt} + \\ \delta_2 \cdot |PLR_l - ED_{ilt} \cdot ELGD_{ilt}| + \epsilon_{ilt}$$

- Step 2: Maximum likelihood approach to estimate α_l, δ_1 and δ_2.

Results: Prosper is Using Ensemble Recession Probability Method

Table: Estimation Results of Prosper's Objective Function

	EWM	MWM	RPM	GFM	E-RPM
llh	-51999.34	-53343.86	-51121.85	-52219.56	-49314.87
BIC	104081.62	106770.66	102326.64	104522.06	98712.68
$\hat{\delta}_1$	2.21 (0.378)	1.58 (0.132)	4.41 (0.169)	2.09 (0.401)	4.26 (0.185)
$\hat{\delta}_2$	-11.0 (0.161)	-6.41 (0.120)	-10.44 (0.145)	-10.94 (0.165)	-13.98 (0.154)
$\hat{\alpha}_1$	-2.33 (0.035)	-1.44 (0.033)	-1.70 (0.037)	-2.38 (0.035)	-1.41 (0.037)
$\hat{\alpha}_2$	-1.41 (0.024)	-0.69 (0.021)	-1.27 (0.023)	-1.44 (0.024)	-0.99 (0.023)
$\hat{\alpha}_3$	-1.30 (0.023)	-0.69 (0.021)	-1.18 (0.022)	-1.30 (0.023)	-0.96 (0.023)
$\hat{\alpha}_4$	-1.30 (0.024)	-0.84 (0.022)	-1.27 (0.023)	-1.30 (0.024)	-1.06 (0.023)
$\hat{\alpha}_5$	-0.61 (0.020)	-0.32 (0.018)	-0.57 (0.019)	-0.59 (0.020)	-0.42 (0.019)
$\hat{\alpha}_6$	-0.74 (0.020)	-0.67 (0.020)	-0.75 (0.020)	-0.74 (0.020)	-0.68 (0.020)

A Closer Look at E-RPM

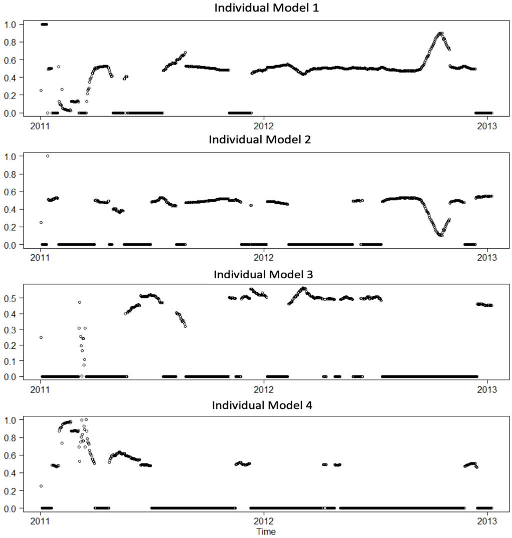


Figure: Weights on Individual Funding Model in E-RPM

Prediction of Customer Behavior

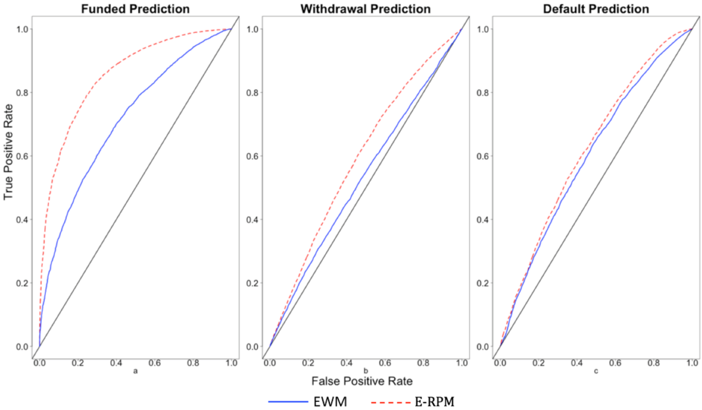


Figure: EWM vs ERPM

Prediction of Customer Behavior

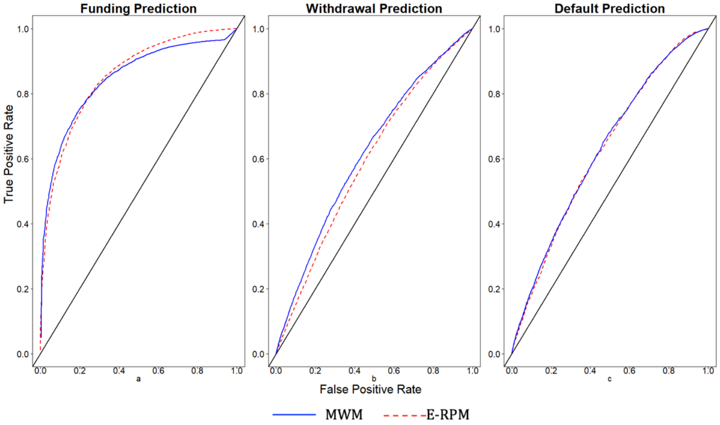


Figure: MWM vs ERPM

Prediction of Customer Behavior

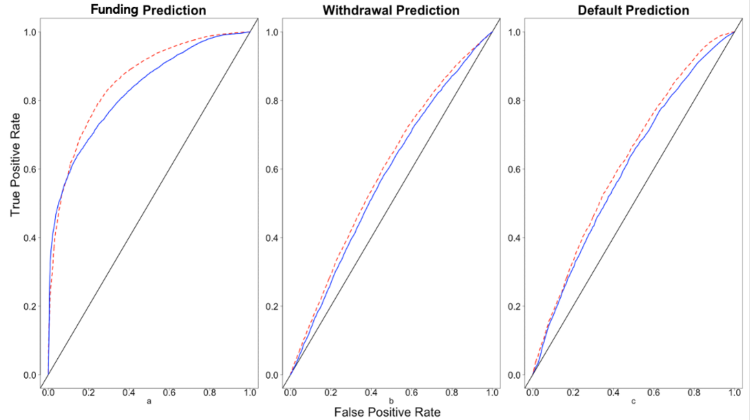


Figure: RPM vs ERPM

Prediction of Customer Behavior

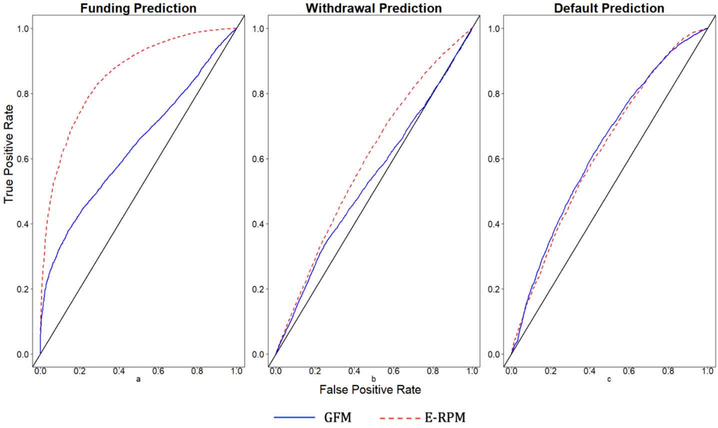


Figure: GFM vs ERPM

Prediction of Customer Behavior

Table: Comparison of Prediction Results

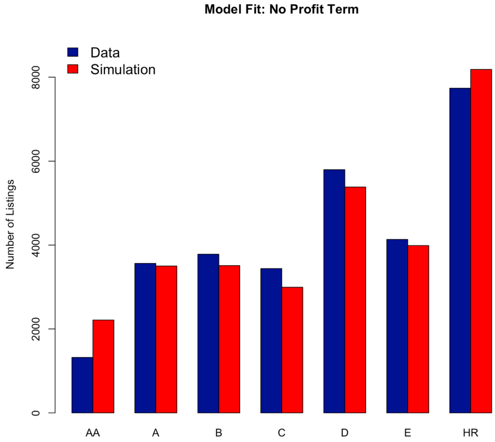
	EWM	MWM	RPM	GFM	E-RPM
Funding AUC	0.712	0.847	0.796	0.638	0.854
Withdrawal AUC	0.534	0.618	0.587	0.535	0.595
Default AUC	0.597	0.609	0.603	0.615	0.617
Loss Given Default MSE	0.045	0.092	0.110	0.095	0.035

Notes: AUC represents area under ROC. The larger the AUC is, the better prediction performance the corresponding method has.

Counterfactuals

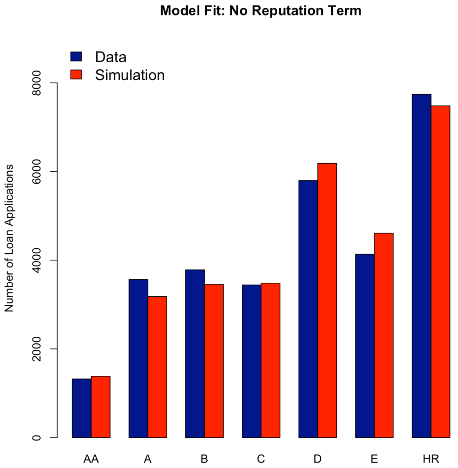
- No profits term ($\delta_1 = 0$).
- No long-term reputation term ($\delta_2 = 0$).
- No Adaptive Learning.
- Know the true DGP (use both historical and future data to make decisions in the current period).

NO Profit Term Expected Rating Distribution



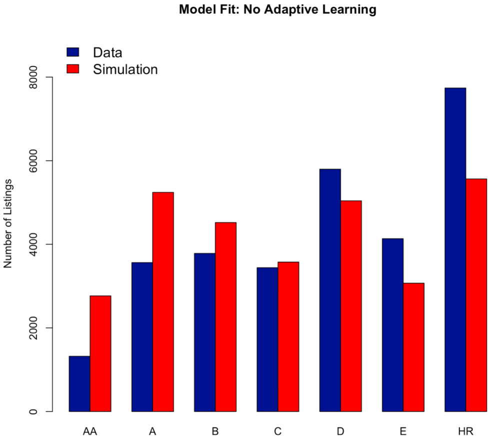
Prosper assigns more AA and HR loans. It is sensible to assign more AA loans because they are less profitable due to the lower origination fee. It also makes sense that we see more HR loans because they are less likely to get funded and generate revenue.

NO Reputation Term Expected Rating Distribution



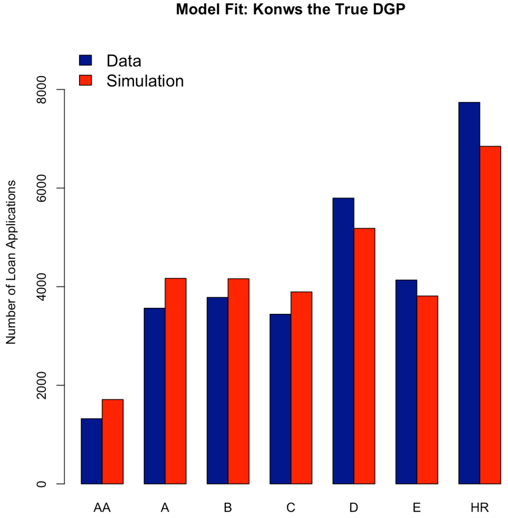
More loan application were assigned D, E ratings and fewer loan applications were assigned A and B ratings. These changes make sense because the origination fee rate of ratings D and E is 50% higher than that of ratings A and B.

No AL Expected Rating Distribution



Much more AA, A and B loans.

True DGP Expected Rating Distribution



The deviation from the data in experiment (iv) falls somewhere between those in experiment (i-ii) and (iii).

Counterfactual Revenue

Table: Simulated Revenue in Counterfactual Exercises (Objective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Ave. Revenue per loan	239.70	229.51	255.21	232.36	258.98
ΔRevenue ($)	0	-10.19	15.50	-7.34	19.27
ΔRevenue (%)	0	-4.25%	6.47%	-3.06%	8.04%

Notes: We use E-RPM as the comparison baseline.

- If Prosper ignores the short-term profits (setting $\delta_1 = 0$), the expected revenue drops by about 4.25%.
- If Prosper does not care about long-term reputation, its revenue increases by 6.47%.

Counterfactual Revenue

Table: Simulated Revenue in Counterfactual Exercises (Objective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Ave. Revenue per loan	239.70	229.51	255.21	232.36	258.98
ΔRevenue ($)	0	-10.19	15.50	-7.34	19.27
ΔRevenue (%)	0	-4.25%	6.47%	-3.06%	8.04%

Notes: We use E-RPM as the comparison baseline.

- If Prosper does not adaptively learn at all, its expected revenue drops by about 3.06%.
- If Prosper knows the "true" parameters of the predictive models, the expected revenue would increase by 8.04%.

Counterfactual Risk Misreporting

Table: Simulated Risk Misreporting in Counterfactual Exercises (Objective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
ΔAverage Risk Misreporting	0	-5.58%	7.26%	4.69%	-9.61%

Notes: We use E-RPM as the comparison baseline.

- If Prosper ignores the current profits (setting $\delta_1 = 0$), the risk misreporting term drops by about 5.58%.
- If Prosper does not care reporting her evaluated risk (setting $\delta_2 = 0$), its risk misreporting term increases by 7.26%.

Counterfactual Risk Misreporting

Table: Simulated Risk Misreporting in Counterfactual Exercises (Objective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
ΔAve. Risk Misreporting	0	-5.58%	7.26%	4.69%	-9.61%

Notes: We use E-RPM as the comparison baseline.

- If Prosper does not adaptively learn at all, its risk misreporting increases by about 4.69%.
- If Prosper knows the "true" parameters of the predictive models, the risk misreporting would decrease by 9.61%.

Conclusion

1. We use Prosper as an example to illustrate how to infer a firm's AL/Data Selection algorithm.
   - Generalized Revealed Preference approach.
   - Improve our understanding about decision making.
   - Understand your rivals' decisions.
2. Find evidence that Prosper cares about both current profits and reporting the evaluated risks.
3. The approach proposed here can be applied to other settings where the environment is changing over time.

Thank you!

[Skip]Counterfactual Revenue

Table: Simulated Revenue in Counterfactual Exercises (Subjective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Revenue (\$)	7,626,098	7,320,768	7,611,017	4,628,401	7,974,155
ΔRevenue (\$)	0	-305,330	-15,081	-2,997,697	348,057
ΔRevenue (%)	0	-4.00%	-0.20%	-39.30%	4.56%

Notes: This table shows the simulated expected revenue (subjective) in different scenarios. We use E-RPM as the comparison baseline.

- If Prosper ignores the current profits (setting $\delta_1 = 0$), the expected revenue drops by about 4.00%.
- If Prosper does not care about the “reputation” term, its revenue drops by 0.20%.

[Skip]Counterfactual Revenue

Table: Simulated Revenue in Counterfactual Exercises (Subjective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Revenue ($)	7,626,098	7,320,768	7,611,017	4,628,401	7,974,155
ΔRevenue ($)	0	-305,330	-15,081	-2,997,697	348,057
ΔRevenue (%)	0	-4.00%	-0.20%	-39.30%	4.56%

Notes: This table shows the simulated expected revenue (subjective) in different scenarios. We use E-RPM as the comparison baseline.

- If Prosper does not adaptively learn at all, its expected revenue drops by about 39.30%.
- If Prosper knows the "true" parameters of the predictive models, the expected revenue would increase by 4.56%.

[Skip]Counterfactual Reputation Loss

Table: Simulated Reputation Loss in Counterfactual Exercises (Subjective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Average Reputation Loss	8.23%	8.16%	10.97%	9.46%	15.40%
ΔAverage Reputation Loss	0	-0.07%	2.74%	1.23%	7.17%
ΔAverage Reputation Loss (%)	0	-0.85%	33.29%	14.95%	87.12%

Notes: This table shows the simulated average reputation loss (subjective) in different scenarios. We use E-RPM as the comparison baseline.

- If Prosper ignores the current profits (setting $\delta_1 = 0$), the long-term reputation loss drops by 0.85%.
- If Prosper does not care about long-term reputation (setting $\delta_2 = 0$), its long-term reputation loss increases by 33.29%.

[Skip]Counterfactual Reputation Loss

Table: Simulated Reputation Loss in Counterfactual Exercises (Subjective)

	E-RPM	$\delta_1 = 0$	$\delta_2 = 0$	No AL	True DGP
Average Reputation Loss	8.23%	8.16%	10.97%	9.46%	15.40%
ΔAverage Reputation Loss	0	-0.07%	2.74%	1.23%	7.17%
ΔAverage Reputation Loss (%)	0	-0.85%	33.29%	14.95%	87.12%

Notes: This table shows the simulated average reputation loss (subjective) in different scenarios. We use E-RPM as the comparison baseline.

- If Prosper does not adaptively learn at all, its long-term reputation loss increases by 14,95%.
- If Prosper knows the "true" parameters of the predictive models, the long-term reputation loss would increase by 87.12%.