How Do Individuals Repay Their Debt?  
The Balance-Matching Heuristic

John Gathergood,¹ Neale Mahoney,² Neil Stewart,³ & Joerg Weber¹

¹University of Nottingham  
²University of Chicago and NBER  
³University of Warwick

FDIC Consumer Research Symposium  
October 13, 2017
Motivation

Individual borrowing decisions underpin a broad set of economic behavior

- Consumption smoothing over the life-cycle
- Investment in human capital
- Purchases of durables

⇒ Understanding how individuals borrow is (i) an important input for many fields of economic research and (ii) directly relevant for consumer financial policy
Study how individuals allocate repayments across their portfolio of credit cards, for which optimal behavior can be clearly defined.

- Holding total repayments fixed, optimal to pay minimum on all cards, put any remaining payments on card with the highest interest rate.

Cleaner than studying other consumer financial products (e.g., mortgages, retirement savings, stock portfolio choices).

- Optimal decisions depend on unobserved preferences (e.g., risk preferences, discount factors).

Cleaner than studying credit card spending or balances.

- Spending depends on (subjective) value of rewards.
- Balances are the result of dynamic repayment and spending decisions.
Related Literature

- Misallocation on credit cards
  - Ponce et al. (2017) show misallocation in matched sample of Mexican card holders
  - Stango and Zinman (slides) find misallocation falls with high stakes in US opt-in consumer panel data

- Start by replicating the misallocation results from Ponce et al. (2017) in our UK data
  - Not main contribution, but necessary first step

- Main contribution is to evaluate “heuristics” that might better explain behavior
Data

- Argus Information and Advisory Services “Credit Card Payments Study” (CCPS)
  - Contract terms, spending and repayments from 5 UK issuers with combined 40% market share
  - 10% representative sample of CCPS covering Jan 2013 and Dec 2014
  - Anonymized individual-level identifiers
  - Analyze data at the individual × month level, focusing on 2 cards in main analysis but extend to up to 5 cards
Sample Restrictions

- Focus on months when individuals face *economically meaningful* decisions about how to allocate repayments
  - E.g., individuals who repay both cards in full do not face an allocative problem

- Restrict sample to individual × months where individual
  - Holds debt (revolving balance) on both cards
  - Makes at least minimum payment on both cards
  - Pays more than the minimum payment on at least one card
  - Does not pay both cards down in full

⇒ Resulting sample contains 68% of aggregate revolving balances in 2-card sample
## Summary Statistics, 2-Card Sample

<table>
<thead>
<tr>
<th></th>
<th>High APR Card</th>
<th>Low APR Card</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR</td>
<td>22.86</td>
<td>16.56</td>
<td>6.30</td>
</tr>
<tr>
<td><strong>Balances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balances</td>
<td>£3,018</td>
<td>£3,026</td>
<td>£8</td>
</tr>
<tr>
<td>Credit Limit</td>
<td>£6,385</td>
<td>£6,010</td>
<td>£375</td>
</tr>
<tr>
<td>Utilization</td>
<td>47.3%</td>
<td>50.3%</td>
<td>3.0%</td>
</tr>
<tr>
<td><strong>Payments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payments</td>
<td>£260</td>
<td>£230</td>
<td>£30</td>
</tr>
<tr>
<td>Minimum Payment</td>
<td>£63</td>
<td>£57</td>
<td>£6</td>
</tr>
</tbody>
</table>
Actual and Cost-Minimizing Repayments

- Refer to cost-minimizing allocation as the “optimal” allocation
- Holding total repayment fixed, it is optimal to:
  - Make minimum required payment on both cards
  - Repay as much as possible on card with high interest rate
  - Allocate any further payments to low interest rate card only when high interest rate card is repaid in full
Actual and Optimal Payments

![Graph showing Actual and Optimal Payments]

- **X-axis:** Payment on High APR Card (%)
- **Y-axis:** Density (%)
Summary of Misallocated Payments

Misallocated low-cost card debt

- Optimally, consumers should allocate 70.8% of payments to high interest card (97.1% of payments over the minimum)
- Actually, consumers allocate 51.2% of repayments to high interest card (51.5% of payments over and the minimum)
- Consumers misallocate 19.6% total monthly repayment to the low cost card (45.6% of payments over the minimum)
- 85% of card holders should put 100% of their excess repayment onto the high cost card, but only 10% do so
Actual and Optimal Payments with 3 Cards

Individuals
Actual and Optimal Payments with 4 Cards

Individual:
Actual and Optimal Payments with 5 Cards

Individual

Card 1 = Lowest APR
Card 2
Card 3
Card 4
Card 5 = Highest APR
Optimization Frictions

- Theories of optimization frictions (switching costs, rational inattention) predict less misallocation
  - When stakes are high
  - Over longer time horizons

- We show misallocation invariant to:
  - Difference in interest rates
  - Level of total payment
  - Age of high cost card
  - Days between payment due dates
Misallocated vs. Difference in Interest Rates

Excess Misallocated vs Interest Rates
Misallocated vs. Total Repayments

Excess Misallocated vs Total Payment
Interest Savings from Optimizing

- Calculate savings from optimizing repayments in two-card sample
  - Shift as much of the balance as possible to the low APR card
  - Calculate reduction in annualized interest payments
  - Think of this as “steady state” under optimal repayments

- In steady state
  - 44% of individuals reduce their high APR card balance to zero

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>p75</th>
<th>p90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saving in £</td>
<td>64.2</td>
<td>111.0</td>
<td>68.7</td>
<td>167.0</td>
</tr>
<tr>
<td>% Annualized Interest (%)</td>
<td>11.7</td>
<td>23.4</td>
<td>10.2</td>
<td>24.3</td>
</tr>
</tbody>
</table>

- Total savings larger if individuals “re-invest” savings or if we consider 3+ cards
Two Potential Explanations

1. Individuals use $1/N$ rule

2. Individuals use “some other rule” and then round repayment amounts
   - Suppose that under other rule, individual would repay £220 on high APR card and £180 on the low APR card
   - However, with rounding, ends up repaying £200 on each
   - Then will observe behavior than looks like $1/N$
Density of Payments in £

Payments tend to be made in prominent round value amounts.

- 19.2% take on multiples of £100 and 33% of take on multiples of £50.
Rounding and 1/N Type Behavior

1/N splitting much more common among round value amount payments (left panel) compared with other payments (right panel)

(a) Round Payments (£50s)  (b) Non-Round Payments
Summary

- Cannot reject hypothesis that nearly all splitting due to rounding
  - Restricting to non-rounders, excess mass at $1/N$ is less than 2%
- However, cannot reject hypothesis that individuals who round are different from individuals who do not round, and that these individuals would split even if they did not repay round number amounts
Balance Matching

Definition: Match the ratio of repayments to the ratio of balances

- Let $q_k$ indicate balances and $p_k$ indicate payments, balance-matching payments are given by

\[
\frac{p_A}{p_B} = \frac{q_A}{q_B}
\]

subject to the constraint that the individual pays at least the minimum on both cards and no more than the full balance on either card

- Constraints rarely bind and dropping these observations does not affect results
Example Card Card Statement

02 March 2017

At a glance
Your new balance: £562.13
Minimum payment: £12.64
Please pay by: 27 March 2017

Your activity
Your previous balance: £639.63
Payments towards your account: £484.21
Your new activity: £406.71
Interest charged: £0.00
Other charges: £0.00
Your new balance: £562.13
Available to spend: £4,437.87
Your current credit limit: £5,000.00

Your current interest rates
Simple standard rate p.a.: 17.18% (18.6% compound equivalent)
Simple cash rate p.a.: 24.60% (27.6% compound equivalent)
Estimated interest next month: £0.00

This month's updates
Balance Transfer
Your Balance Transfer and Money Transfer rates are available this month. You can easily and securely make a transfer anytime.

Card issuer contact details
Matching behavior has been observed across domains (and species)

- Herrnstein (1961) “Matching Law”: Pigeons peck keys for food in proportion to the time it takes the keys to rearm (instead of concentrating their effort on the key that rearms most quickly)
- Economic example: Rubinstein (2002): Subjects diversify across independent 60%-40% gambles even though betting on the gamble with a 60% probability of payout is a strictly dominant strategy
Other Heuristics

We also consider 4 alternative heuristics which individuals might adopt in their repayment decisions:

- **Heuristic 1:** Pay down card with lowest capacity (e.g., to avoid going over the credit limit)
- **Heuristic 2:** Pay down card with highest capacity (e.g., to free up space for large purchase)
- **Heuristic 3:** Pay down card with highest balance (e.g., because of aversion to debt on a card-by-card basis)
- **Heuristic 4:** Pay down card with lowest balance (e.g., “debt snowball” method)
Testing Between Models

- Evaluate models using two statistical approaches
  1. Assess explanatory power using standard measures of goodness of fit (RMSE, MAE, $\rho$)
  2. Horse-race analysis where we determined best fit model on an individual $\times$ month basis
- Both tests are useful for determining the “best” model
  - E.g., a model could be closest to actual behavior on average, while ranking second-best to another model in fitting most observations
**Benchmarks**

Useful to compare goodness-of-fit to upper and lower benchmarks

1. Lower benchmark: “Even a broken clock is right twice a day”
   - Uniform (0,100) percentage on the high APR card

2. Upper benchmark: Max predictive power with our data
   - Machine Learning (Decision Tree, Random Forest, XGBoost)
   - Potentially “unfair” as ML models also predict rounding. Planning to redo analysis dropping individuals who pay round amounts
## Goodness of Fit

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Draw (0,100)</td>
<td>36.52</td>
<td>29.99</td>
<td>0.00</td>
</tr>
<tr>
<td>Optimal Balance Matching</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Heuristics*
- Heuristic 1
- Heuristic 2
- Heuristic 3
- Heuristic 4

*Machine Learning*
- Decision Tree
- Random Forest
- XGBoost
## Goodness of Fit

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Draw (0, 100)</td>
<td>36.52</td>
<td>29.99</td>
<td>0.00</td>
</tr>
<tr>
<td>Optimal</td>
<td>35.19</td>
<td>25.48</td>
<td>0.31</td>
</tr>
<tr>
<td>Balance Matching</td>
<td>23.83</td>
<td>17.02</td>
<td>0.47</td>
</tr>
</tbody>
</table>

*Heuristics*

- Heuristic 1
- Heuristic 2
- Heuristic 3
- Heuristic 4

*Machine Learning*

- Decision Tree
- Random Forest
- XGBoost
Balance Matching and Actual Repayments
## Goodness of Fit

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Draw (0, 100)</td>
<td>36.52</td>
<td>29.99</td>
<td>0.00</td>
</tr>
<tr>
<td>Optimal</td>
<td>35.19</td>
<td>25.48</td>
<td>0.31</td>
</tr>
<tr>
<td>Balance Matching</td>
<td>23.83</td>
<td>17.02</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>Heuristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic 1</td>
<td>36.43</td>
<td>27.27</td>
<td>0.08</td>
</tr>
<tr>
<td>Heuristic 2</td>
<td>33.47</td>
<td>23.86</td>
<td>0.29</td>
</tr>
<tr>
<td>Heuristic 3</td>
<td>35.23</td>
<td>25.91</td>
<td>0.27</td>
</tr>
<tr>
<td>Heuristic 4</td>
<td>34.09</td>
<td>24.59</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>Machine Learning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Tree</td>
<td>19.35</td>
<td>14.97</td>
<td>0.53</td>
</tr>
<tr>
<td>Random Forest</td>
<td>16.22</td>
<td>11.56</td>
<td>0.71</td>
</tr>
<tr>
<td>XGBoost</td>
<td>17.15</td>
<td>12.90</td>
<td>0.66</td>
</tr>
</tbody>
</table>

More Detail on ML Models
Interpreting Machine Learning Models

- ML models often criticized for being “black boxes”
- Useful for learning about relative importance of different variables for predicting behavior
  - How important are balances?
  - How important are interest rates?
- Do not want to interpret variable importance when variables are highly correlated
  - Not a concern in our setting

[Correlation Matrix]
Random Forest and Extreme Gradient Boosting

• Variable importance measures the proportion of the total reduction of sum of squared errors in prediction of outcome variable that results from the split(s) of each input variable across all nodes and trees.

<table>
<thead>
<tr>
<th>Random Forest</th>
<th>Gradient Boosting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Importance</td>
</tr>
<tr>
<td>High Card Balance</td>
<td>0.20</td>
</tr>
<tr>
<td>Low Card Balance</td>
<td>0.18</td>
</tr>
<tr>
<td>Low Card CL</td>
<td>0.13</td>
</tr>
<tr>
<td>High Card CL</td>
<td>0.13</td>
</tr>
<tr>
<td>High Card Purchases</td>
<td>0.10</td>
</tr>
<tr>
<td>Low Card Purchases</td>
<td>0.10</td>
</tr>
<tr>
<td>High Card APR</td>
<td>0.08</td>
</tr>
<tr>
<td>Low Card APR</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Horse Race Between Models

- We estimate model fit on observation-by-observation basis
  - For each observation we can observe counterfactual behavior under each rule
  - Comparing rules with one another, we calculate which rule is closest to actual behavior
  - The rule which is closest “wins” the horse race, we report the % win rate for each rule

35/58
# Horse Race Between Models

<table>
<thead>
<tr>
<th>Win %</th>
<th>Uniform</th>
<th>33.6</th>
<th>45.3</th>
<th>50.0</th>
<th>44.7</th>
<th>47.0</th>
<th>46.5</th>
<th>38.1</th>
<th>31.3</th>
<th>34.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance Matching</td>
<td>66.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td></td>
<td>54.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic 1</td>
<td></td>
<td></td>
<td>50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic 2</td>
<td></td>
<td></td>
<td></td>
<td>55.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XGB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65.6</td>
</tr>
</tbody>
</table>
# Horse Race Between Models

<table>
<thead>
<tr>
<th>Win %</th>
<th>Balance Matching</th>
<th>Optimal</th>
<th>Heuristic 1</th>
<th>Heuristic 2</th>
<th>Heuristic 3</th>
<th>Heuristic 4</th>
<th>Decision Tree</th>
<th>Random Forest</th>
<th>XGB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>67.5</td>
<td>32.5</td>
<td>27.8</td>
<td>34.1</td>
<td>25.6</td>
<td>36.0</td>
<td>44.4</td>
<td>53.0</td>
<td>49.2</td>
</tr>
</tbody>
</table>
Heuristic Behavior Within Individuals

- If balance matching is a model of individual behavior, we naturally expect balance matching behavior to be correlated within individuals over time
  - For each individual × month our horse race analysis identifies which model best fits the data
  - Calculate a transition matrix between best-fit models at the individual level
  - Use uniform model as a lower-benchmark for persistence of individual behavior
## Probability of Transition Between Rules

<table>
<thead>
<tr>
<th></th>
<th>Uniform</th>
<th>Bal Mat</th>
<th>Optimal</th>
<th>H 1</th>
<th>H 2</th>
<th>H 3</th>
<th>H 4</th>
<th>1/n Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>22%</td>
<td>59%</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>13%</td>
</tr>
<tr>
<td>Bal Mat</td>
<td>8%</td>
<td>82%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>Optimal</td>
<td>8%</td>
<td>69%</td>
<td>12%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>H 1</td>
<td>12%</td>
<td>50%</td>
<td>0%</td>
<td>23%</td>
<td>1%</td>
<td>3%</td>
<td>4%</td>
<td>7%</td>
</tr>
<tr>
<td>H 2</td>
<td>6%</td>
<td>62%</td>
<td>0%</td>
<td>1%</td>
<td>25%</td>
<td>3%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>H 3</td>
<td>12%</td>
<td>50%</td>
<td>1%</td>
<td>4%</td>
<td>4%</td>
<td>18%</td>
<td>1%</td>
<td>10%</td>
</tr>
<tr>
<td>H 4</td>
<td>10%</td>
<td>56%</td>
<td>0%</td>
<td>5%</td>
<td>0%</td>
<td>1%</td>
<td>23%</td>
<td>4%</td>
</tr>
<tr>
<td>1/n Rule</td>
<td>16%</td>
<td>31%</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>49%</td>
</tr>
</tbody>
</table>
Balance Matching and Minimum Payments

Balance matching could arise due to anchoring onto *minimum payments*

- Min pay amounts are prominent displayed on credit card statements
- Min pay amounts are sometimes a proportion of the balance

⇒ Balance matching could stem from “minimum payment matching”
Balance Matching and Minimum Payments

- Test for importance of minimum payment matching using non-linearities in minimum payment rules
- Typical minimum payment formula
  \[ \text{Minimum Payment} = \max\{25, 2\% \times \text{Balance}\} \]
- Separately identify balance matching from minimum payment matching by comparing “slope sample” (where balance matching and minimum payments are strongly correlated) and “floor sample” (where correlation is much weaker)
Same Slopes Sample

- Balance matching and minimum payment matching amounts are virtually identical ($\rho = 0.96$)

(c) $\rho = 0.63$

(d) $\rho = 0.62$
Floor Sample

- Balance matching and minimum payment matching amounts are different \((\rho = 0.56)\)

\[(e) \quad \rho = 0.57 \quad (f) \quad \rho = 0.22\]
Conclusion

- Allocation of credit card repayments is highly non-optimal and not explained by fixed costs of adjustment.
- Repayment behavior is better explained by a balance matching heuristic:
  - Captures more than half of the predictable variation in repayments.
  - Performs substantially better than other models.
  - Highly persistent within-individuals over time.
  - Consistent with machine learning models, which place high weight on balances and very low weight on prices in explaining repayment decisions.
Backup Slides
## Sample Selection

<table>
<thead>
<tr>
<th></th>
<th>Account X (months)</th>
<th>Customers</th>
<th>Aggregate debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td>Unrestricted Sample</td>
<td>7,876,760</td>
<td>229,260</td>
<td>330,495,722</td>
</tr>
<tr>
<td><strong>Drop if:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal Interest Rates</td>
<td>315,070</td>
<td>2,293</td>
<td>6,609,914</td>
</tr>
<tr>
<td>No Debt on Either Card</td>
<td>3,544,542</td>
<td>71,071</td>
<td>0</td>
</tr>
<tr>
<td>Debt on Only One Card</td>
<td>1,260,282</td>
<td>16,048</td>
<td>49,574,358</td>
</tr>
<tr>
<td><strong>of which:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debt on High Card Only</td>
<td>708,908</td>
<td>9,170</td>
<td>29,744,615</td>
</tr>
<tr>
<td>Debt on Low Card Only</td>
<td>551,373</td>
<td>6,878</td>
<td>19,829,743</td>
</tr>
<tr>
<td>Pays Min Only</td>
<td>1,654,120</td>
<td>25,219</td>
<td>39,659,487</td>
</tr>
<tr>
<td>Pays Both in Full</td>
<td>315,070</td>
<td>2,293</td>
<td>9,914,872</td>
</tr>
<tr>
<td>Restricted Sample</td>
<td>788,122</td>
<td>112,796</td>
<td>226,059,074</td>
</tr>
</tbody>
</table>
Actual and Optimal Excess Payments

Back to Actual and Optimal Payments
Excess Misallocated vs. Difference in Interest Rates

Back to Misallocated Payments and Interest Rates
Excess Misallocated vs. Total Repayments
Excess Misallocated vs. Age of High Cost Card

Back to Misallocated Payments and Card Age
Excess Misallocated vs. Days Between Payment Due Dates

Back to Misallocated Payments and Days Between Due Dates
Days Between Payment Due Dates

![Histogram showing the distribution of days between payment due dates. The x-axis represents the number of days between payment due dates, and the y-axis represents the density (%). The histogram shows a decreasing trend as the days increase.]

Back to Misallocated Payments and Days Between Due Dates
Heuristic 1: Pay Down Lowest Capacity Card
Heuristic 2: Pay Down Highest Capacity Card

Back to Goodness of Fit

Density (%) vs Actual Heuristic 2

Back to Goodness of Fit

Density (%) vs Actual Payments on High APR Card (%)
Heuristic 3: Pay Down Highest Balance Card

Back to Goodness of Fit
Heuristic 4: Pay Down Lowest Balance Card

![Graph showing density and actual heuristic 4 payments on high APR card percentage]

Density (%)

Heuristic 4 Payments on High APR Card (%)
Machine Learning

- **Decision (Regression) Tree**
  - Set of branches from root to leaves that classifies observations, tree is grown on training data and cross-validated with hold-out data to avoid over-fitting. Form of localised regression with recursive partitioning.

- **Random Forest**
  - Use a large number of randomly bootstrapped regression trees to reduce variance of predictions, averaging results across all trees to avoid over-fitting

- **XGBoost**
  - ‘Extreme gradient boosting’ method grows a large number of trees classifier - by - classifier to improve the ensemble of trees, cross-validating with hold-out data to avoid over-fitting
Variable importance interpretation is spurious if omitted variables (i.e. card APRs) are highly correlated with included variables. In our case they are not.

<table>
<thead>
<tr>
<th></th>
<th>APR(H)</th>
<th>APR(L)</th>
<th>Bal(H)</th>
<th>Bal(L)</th>
<th>Pur(H)</th>
<th>Pur(L)</th>
<th>Lim(H)</th>
<th>Lim(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR(H)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APR(L)</td>
<td>0.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal(H)</td>
<td>0.11</td>
<td>0.13</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal(L)</td>
<td>0.10</td>
<td>0.09</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pur(H)</td>
<td>-0.06</td>
<td>-0.06</td>
<td>0.04</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pur(L)</td>
<td>-0.05</td>
<td>-0.03</td>
<td>0.09</td>
<td>0.05</td>
<td>0.01</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lim(H)</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.65</td>
<td>0.24</td>
<td>0.15</td>
<td>0.10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Lim(L)</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.27</td>
<td>0.69</td>
<td>0.07</td>
<td>0.12</td>
<td>0.37</td>
<td>1.00</td>
</tr>
</tbody>
</table>