NBER WORKING PAPER SERIES

HOW DO AMERICANS REPAY THEIR DEBT? THE BALANCE-MATCHING HEURISTIC

John Gathergood
Neale Mahoney
Neil Stewart
Jörg Weber

Working Paper 25557
http://www.nber.org/papers/w25557

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2019, Revised June 2019

We thank Ruchi Mahadeshwar and Qi Zheng for excellent research assistance. The results in this paper were calculated (or derived) based on credit data provided by TransUnion, a global information solutions company, through a relationship with the Kilts Center for Marketing at The University of Chicago Booth School of Business. The work was supported by Economic and Social Research Council grants ES/K002201/1, ES/N018192/1, ES/P008976/1, Leverhulme grant RP2012-V- 022, and the Initiative on Global Markets at the University of Chicago. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Do Americans Repay Their Debt? The Balance-Matching Heuristic
John Gathergood, Neale Mahoney, Neil Stewart, and Jörg Weber
NBER Working Paper No. 25557
February 2019, Revised June 2019
JEL No. D12,D14,G02,G20

ABSTRACT

In Gathergood et al. (2019), we studied credit card repayments using linked data on multiple cards from the United Kingdom. We showed that individuals did not allocate payments to the higher interest rate card, which would minimize the cost of borrowing, but instead made repayments according to a balance-matching heuristic under which the share of repayments on each card is matched to the share of balances on each card. In this paper, we examine whether these results extend to the United States using a large sample of TransUnion credit bureau data. These data do not have interest rates, so we cannot examine the optimality of payments. However, we observe balances and repayments, so we can examine balance-matching behavior. We replicate our analysis and find that Americans also repay their debt in accordance with a balance-matching heuristic.

John Gathergood
School of Economics
Sir Clive Granger Building
University Park
Nottingham
NG8 1AA
United Kingdom
john.gathergood@nottingham.ac.uk

Neale Mahoney
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
Neale.Mahoney@chicagobooth.edu

Neil Stewart
Warwick Business School
University of Warwick
England
neil.stewart@warwick.ac.uk.

Jörg Weber
School of Economics
Sir Clive Granger Building
University Park
Nottingham
NG8 1AA
United Kingdom
joerg.weber@nottingham.ac.uk

1 Introduction

In Gathergood et al. (2019) we studied competing models of how individuals repay their debt across their portfolio of credit cards. The key to this analysis was a dataset with rich information on contract terms and utilization for multiple credit cards held by individuals in the United Kingdom. Using these data, we showed that individuals did not allocate payments to the higher interest rate card, which would minimize the cost of borrowing, but instead made payments according to a balance-matching heuristic under which the share of repayments on each card is matched to the share of balances on each card.[1]

In this paper, we examine whether these results extend to the United States using a large sample of TransUnion credit bureau data.[2] These data do not have interest rates, so we cannot examine the optimality of payments, but do include balances and repayments, so we can examine balance matching and other heuristic models. We are unaware of any U.S. dataset that has interest rates on multiple cards for a representative sample of individuals.

We evaluate balance matching and the other heuristics using the same methodology as Gathergood et al. (2019). As in the U.K., we find that balance matching outperforms the other heuristics in terms of goodness-of-fit (RMSE, MAE, Pearson's ρ) and also performs strongly in horse race analysis, where we determine the best fit model on an individual $\times$ month basis.[3] As before, we find that balance matching is persistent within individuals over time, suggesting it results from a stable feature of repayment behavior.

As we discussed in our prior research, balance matching could arise from the salient placement of balances on credit card statements and the well-documented tendency for humans (and other species) to engage in "matching behavior". Balance matching could also arise from individuals repaying a constant percentage of the balance on each card in a given month (e.g., 10% on each card this month), a rule-of-thumb that would lead to inefficient payments on both the allocative and extensive margins. While the precise underpinnings of balance matching are still an open question, the finding that balance matching also occurs in the U.S. indicates that it is a broad phenomenon.

[1] The first result builds on Ponce, Seira and Zamarripa (2017), who find in Mexican data that individuals carry a large fraction of their balances on their high interest rate card.

[2] Most Americans have two or more cards. Using 2015 data, we calculate that 71.5% of credit cards holders had two or more cards, and individuals with two or more cards accounted for 91.8% of balances and 91.7% of revolving balances.

[3] A $1/N$ rule performs better in the U.S. than the U.K. data.

2 Data

Our data is a panel of credit reports over 2000-2016 from TransUnion, a national credit reporting agency. The panel is based on a 10% random sample of individuals with a TransUnion report in 2000, with 10% of new entrants to the TransUnion database added to the panel each month. Our unit of analysis is the individual $\times$ month, which we refer to as observations. We construct separate samples based on the number of credit cards held by the individual in that month.

Traditionally, credit bureaus provided data on credit card balances and credit limits, but did not provide data on payments or minimum payments due. However, during our time period, TransUnion was adding payments and minimum payments variables, allowing us to study balance matching behavior. We drop observations where payments or minimum payments data are missing or have not been updated since the previous month, since they are likely to be out of date. As in our earlier work, we also drop observations where the individual is delinquent, over their credit limit, or pays less than the minimum or more than the balance on at least one card.

Following Gathergood et al. (2019), we also implement "economic" sample restrictions to ensure that the resulting individuals have scope to reallocate their payments, holding total payments fixed. We drop observations with zero aggregate balances, since individuals have no balances to repay. We drop observations where the individual pays the minimum on all of their cards, since they could only reallocate payments by paying less than the minimum on at least one card, which would trigger a penalty. And we drop observations where the individual pays all of their cards in full, because any reallocation would result in the in paying more than the full balance on at least one card. See Gathergood et al. (2019) for more discussion of these sample restrictions.

Appendix Table AI shows the impact of these sample restrictions on individuals $\times$ months and aggregate balances in the two-card sample. The data coverage related restrictions, shown in Panel A, account for the majority of the reduction in sample size. The final two-card sample has 713,157 observations and $3.6 billion in balances.

3 Heuristics

With the exception of optimal repayments, we examine the same repayment models considered in Gathergood et al. (2019). Some of these heuristics are based on the capacity of a credit card, which we define as the difference between the credit limit and current balance. We describe these heuristics for the two-card sample, but they could be naturally extended to settings with three or more cards.

- **Balance Matching**: Match the share of repayments on each card to the share of balances on each card.
- **1/N Rule**: Make equal-sized repayments on each card. This is the debt repayment analogue to the $1/N$ rule for pension plan contributions (Benartzi and Thaler, 2001).
- **Heuristic 1**: Repay the card with the lowest capacity, subject to paying the minimum on each card. Once capacity is equalized across cards, allocate additional payments to both cards equally. This heuristic reduces the risk that an accidental purchase will put an individual over their credit limit.
- **Heuristic 2**: Repay the card with the highest capacity, subject to paying the minimum on each card. Once the highest capacity card is fully repaid, allocate remaining payments to the other card. This heuristic maximizes the "space" to make a large purchase on a single card.
- **Heuristic 3**: Repay the card with the highest balance, subject to paying the minimum on each card. Once balances are equalized across cards, allocate additional payments to both cards equally. This heuristic reduces the maximum balance across cards.
- **Heuristic 4**: Repay the card with the lowest balance ("debt snowball method"), subject to paying the minimum on each card. Once the balance on the lowest balance card is paid down to zero, allocate remaining payments to the other card. The debt snowball method is recommended by some financial advisors because paying off a card delivers a "win" that motivates further repayment behavior and simplifies an individual's debt portfolio.

4 Results

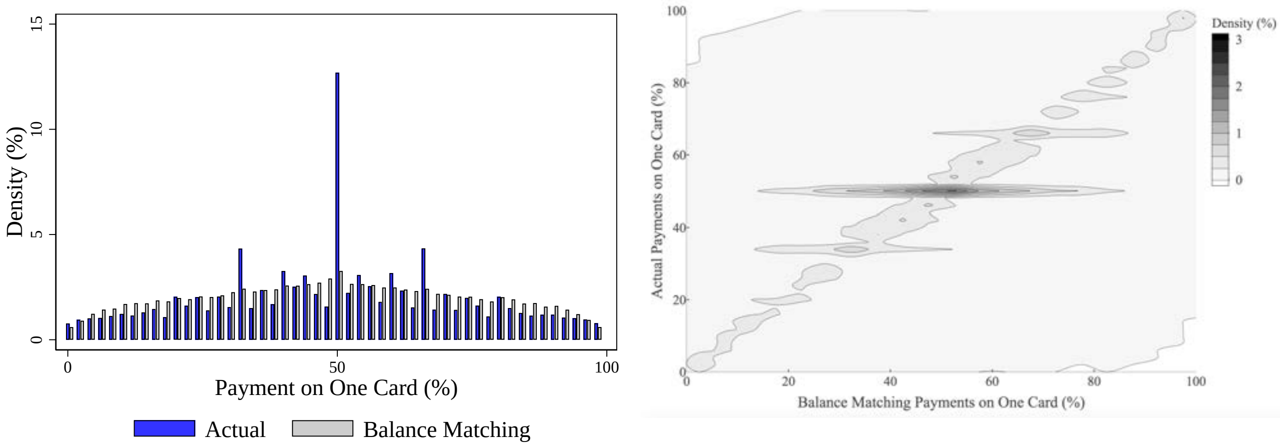
We evaluate balance matching and the other heuristics using the same methodology as Gathergood et al. (2019). For ease of comparison, we produce tables and figures with the same layout as our prior work.

We start by illustrating the distribution of actual and balance matching payments in the two-card sample. Panel A of Figure 1 shows the marginal distribution of actual and predicted payments on a randomly chosen card (of the two) under a balance matching rule. Panel B shows the joint distribution of actual and predicted payments.[4] Aside from the spike at 50%, the marginal distributions are similar.

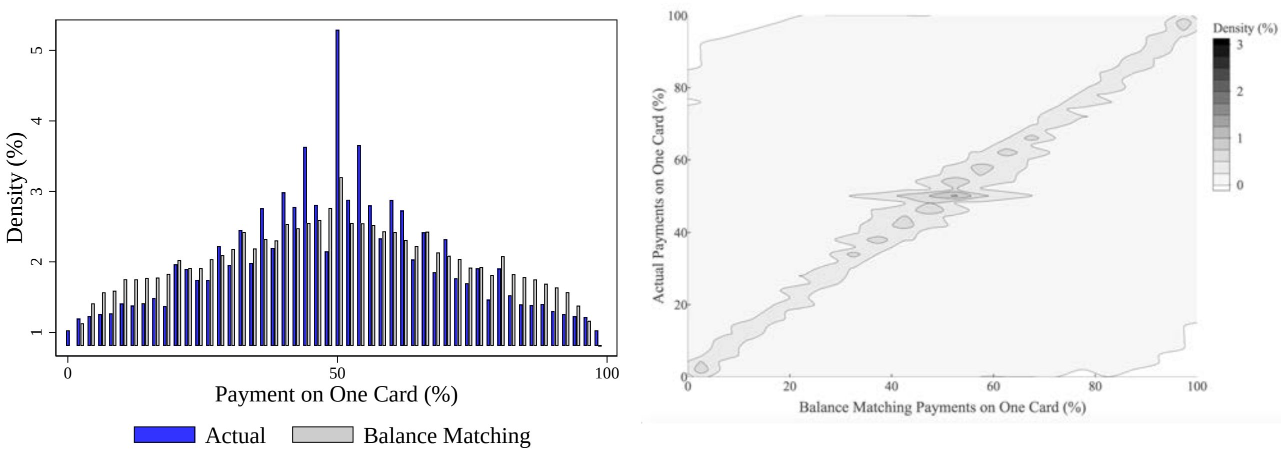
[4]In Gathergood et al. (2019), we showed results for the high APR card. Since we do not observe interest rates, in this paper we randomly choose one of the two cards to focus on. Because our goodness-of-fit metrics are invariant to the card which is chosen, this distinction has no bearing on our results.

Figure 1: Balance Matching

(A) Baseline Sample



(B) Non-Round Number Payment Sample

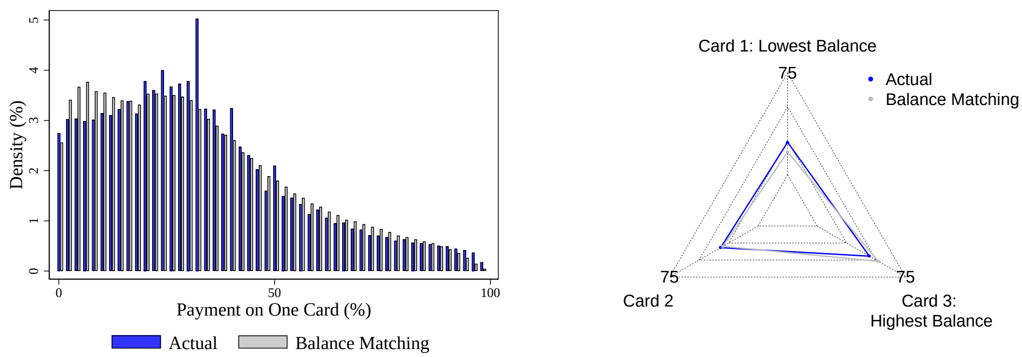


Note: Left panels show the distribution of actual and balance-matching payments on one card. Right panels show the joint density of actual and balance-matching payments. Panel A shows the baseline sample two-card sample; Panel B restricts the sample to non-round payment amounts (not multiples of $50).

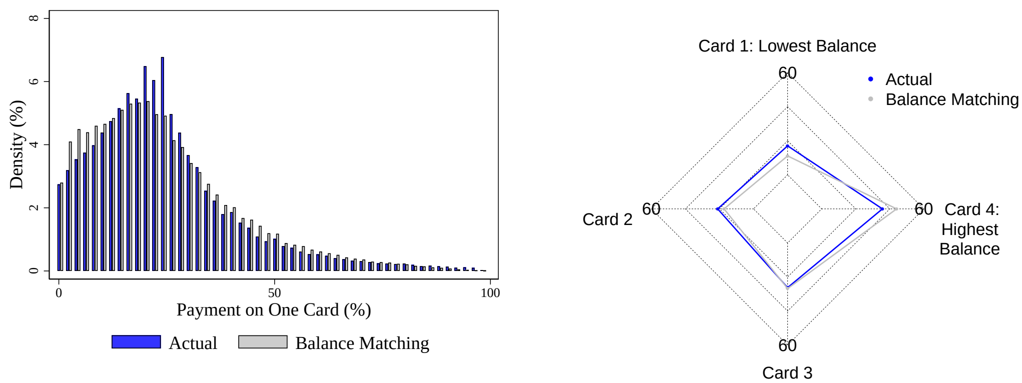
The joint distribution indicates a strong positive correlation ($\rho = 0.61$).

Figure 2: Actual and Balance-Matching Payments on Multiple Cards

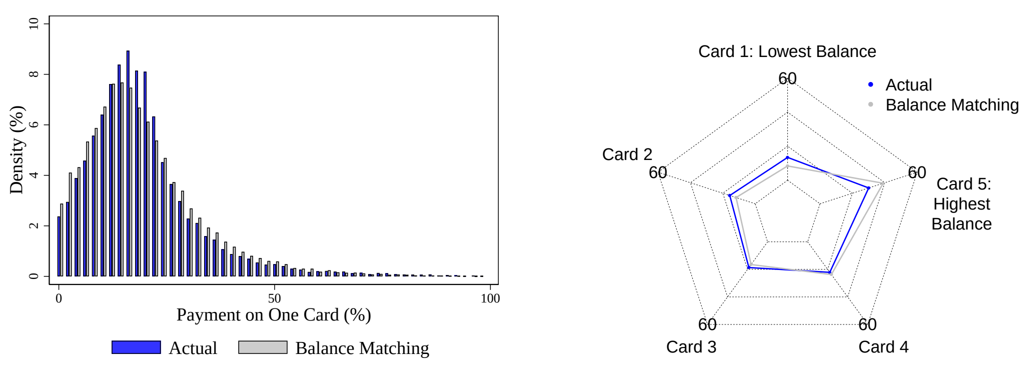
(A) Three Cards



(B) Four Cards



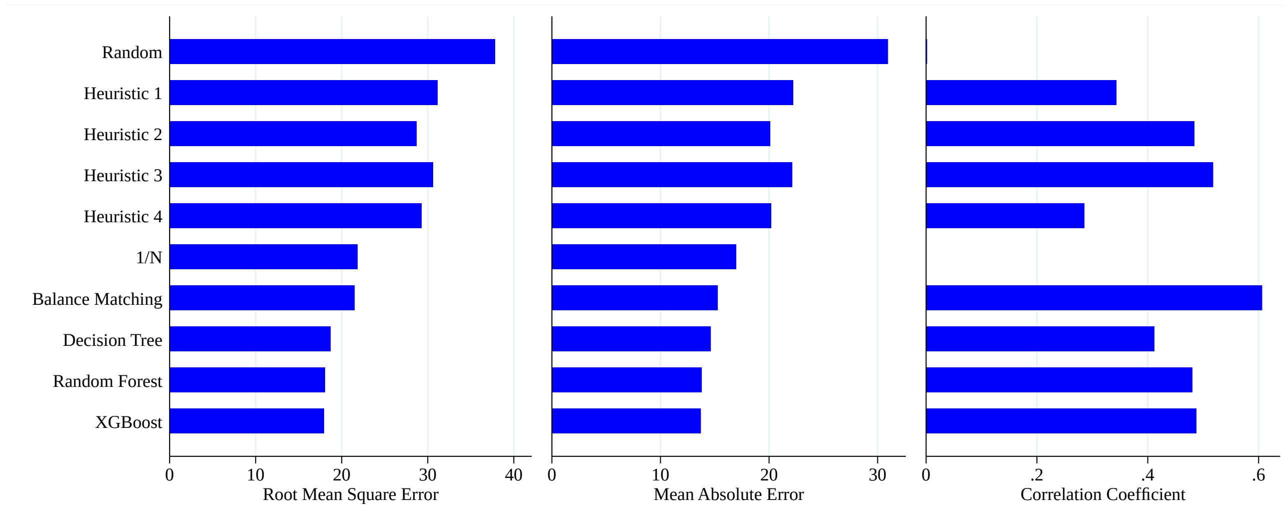
(C) Five Cards



Note: Left column shows the marginal distributions of actual and balance-matching payments on one card. Right column shows radar plots of the mean percentage of actual payments and payments under the balance-matching rule allocated to each card. In the radar plots, cards are ordered clockwise from the highest to the lowest balance (starting at the first node clockwise from noon).

Since credit card payments bunch at round numbers, we follow our prior work and also conduct our analysis separately for observations with round and non-round payments, where we define round number payments as multiples of \$50. The correlation between actual and balance-matching payments is higher in the non-round number sample (Figure 1 Panel B) than in the round number sample (Appendix Figure A1). Also, the spike at 50% is much more pronounced in the round number sample, suggesting that $1/N$ allocations might arise due to rounding, a possibility we discuss in more detail in Gathergood et al. (2019).

Figure 3: Goodness-of-Fit for Different Models



Note: Goodness-of-fit for different models of the percentage of payments on one card in the baseline two-card sample.

Figure 2 shows actual and balance matching payments for the samples with 3-5 cards. The left column shows the marginal distributions of actual and balance-matching payments on a randomly chosen card, and the right column shows radar plots with the mean percentage of repayments allocated to each card ordered clockwise by balance. The plots indicate that actual payments are close to what is predicted by the balance matching rule.

We formally measure the performance of the balance matching and alternative models using three standard measures of goodness-of-fit: the square root of the mean square error (RMSE), the mean absolute error (MAE), and the correlation between actual and predicted payments (Pearson's ρ).

To help interpret the goodness-of-fit values, we also establish lower and upper benchmarks. For a lower benchmark, we calculate goodness-of-fit under the assumption that the percentage of repayments allocated to the card is randomly drawn from a uniform distribution with support on the 0-100% interval.

To provide an upper benchmark, we use machine learning techniques to construct a set of purely statistical models of repayment behavior. Specifically, we estimate decision tree, random forest, and extreme gradient boosting models for the percentage of payments on the high balance card. We use the same set of variables which enter into our heuristics (balances and credit limits on both cards) as input variables and "tune" the models to maximize out-of-sample power.[5] Technical details are provided in the Online Appendix accompanying Gathergood et al. (2019).[6]

Figure 3 reports the goodness of fit from this analysis. Appendix Table AII shows the numerical values. The lower benchmark of random repayment has the worst fit. Balance matching performs close to the upper benchmark, determined by the machine learning models, as measured by RMSE and MAE, and better than this benchmark, as measured by Pearson's ρ.[7] Heuristics 1-4 do not perform much better than the lower benchmark. The $1/N$ rule performs similarly well to the balance matching rule as measured by RMSE and MAE, but has zero correlation with actual repayments, by construction.[8]

To complement the goodness-of-fit analysis, we also evaluate the models using "horse races" where we determine the best fit model on an observation-by-observation basis. A model that fits a smaller number of observations very poorly, but a larger number quite well, would perform poorly under the goodness-of-fit analysis but well under this approach.

Panel A of Table 1 compares each model against the lower benchmark of randomly distributed payments. Balance matching is the best fit model for 67.4% of observations, twice the percentage of the random benchmark. This is much better than Heuristics 1-4, slightly better than the $1/N$ heuristic, and slightly worse than the upper benchmarks provided by the machine learning models. Panel B compares each model against balance matching, excluding the comparison with random benchmark shown in Panel A. Balance matching has the best fit for a substantially higher percentage of observations than Heuristics 1-4, a slightly lower percentage than $1/N$, and a slightly lower percentage than the machine learning models.

As we discussed in our prior work, it is not surprising that the machine learning models sometimes fit the data better than balance matching. These models use balances as an input and could,

[5]In Gathergood et al. (2019) we also included APRs and spending amounts, which are not available in the TransUnion data.

[6]Appendix Figure A2 displays the estimated decision tree.

[7]This result should be interpreted with the caveat that the machine learning models, by design, minimize RMSE.

[8]Appendix Figure A3 and Appendix Table AIII show goodness-of-fit separately for the round and non-round number samples. The results are similar.

Table 1: Horse Races Between Alternative Models

Panel (A): Random vs. Other Rules

	Horse Race								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Win Percent									
Random	32.63	34.47	49.32	46.04	48.08	47.29	30.95	29.59	29.22
Balance Matching	67.37								
1/N		65.53							
Heuristic 1			50.68						
Heuristic 2				53.96					
Heuristic 3					51.92				
Heuristic 4						52.71			
Decision Tree							69.05		
Random Forest								70.41	
XGBoost									70.78

Panel (B): Balance Matching vs. Other Rules

	Horse Race							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Win Percent								
Balance Matching	49.03	59.81	57.25	61.23	55.82	49.76	47.15	46.06
1/N	50.97							
Heuristic 1		40.19						
Heuristic 2			42.75					
Heuristic 3				38.77				
Heuristic 4					44.18			
Decision Tree						50.24		
Random Forest							52.85	
XGBoost								53.94

Note: Table shows percentage of individual $\times$ month observations that are best fit by different models of repayment behavior using the baseline two-card sample.

Table 2: Transition Matrix for Best-Fit Model

	Current Period						
	Random	Balance Matching	H 1	H 2	H 3	H 4	1/N
Previous Period							
Random	32.70%	17.98%	5.74%	6.93%	7.31%	6.50%	22.84%
Balance Matching	7.74%	44.48%	5.84%	6.96%	7.84%	5.05%	22.09%
Heuristic 1 (Pay Down Lowest Capacity)	5.86 %	12.98%	24.12%	9.90%	15.16%	17.54%	14.44%
Heuristic 2 (Pay Down Highest Capacity)	5.64 %	11.76%	8.71%	25.97%	17.97%	16.83%	13.13%
Heuristic 3 (Pay Down Highest Balance)	5.64%	13.68%	14.65%	18.24%	26.37%	7.81%	13.60%
Heuristic 4 (Pay Down Lowest Balance)	5.04%	10.11%	15.76%	18.64%	8.96%	25.65%	15.84%
1/N	7.39%	16.14%	5.40%	5.78%	5.52%	6.15%	53.63%

Note: Table shows transition matrix for the best-fit payment model using the baseline two-card sample.

with large enough sample size, replicate the balance-matching heuristic. The advantage of balance matching is that it is easy to understand, has psychological underpinnings (e.g., probability matching, Herrnstein's matching law), and might provide intuition in yet-to-be-studied environments.

To the extent that we think of the competing models as actually representing different models of individual decision-making, we would naturally expect the best-fit model to be persistent within individuals over time. Table 2 shows the within-person transition matrix for the best-fit model. The sample is restricted to individual $\times$ months where we observe repayment behavior for at least two months in a row. For this exercise, we include all of the candidate models in the horse race, and we fix the uniformly distributed repayment to be constant within individuals over time. Consistent with Gathergood et al. (2019), balance matching and 1/N exhibit high degrees of persistence, suggesting they result from a stable feature of repayment behavior.[9]

5 Conclusion

In Gathergood et al. (2019), we examined linked data on multiple credit cards from the United Kingdom. We showed that individuals did not allocate payments to the higher interest rate card, which

[9] Among individuals whose repayments are best fit by the uniform model in a given month, 33% make repayments that are closest to the uniform model in the next month. This persistence likely reflects the fact that balances and repayments are sticky over time: If the uniform model happens to be accurate in a given month, and balances and payments are sticky, then the uniform model, which is fixed to be constant within an individual over time, will mechanically be accurate in the next month as well.

would minimize the cost of borrowing, but instead made repayments according to a balance-matching heuristic under which the share of repayments on each card is matched to the share of balances on each card. In this paper, we replicated our analysis using a large sample of TransUnion credit bureau data, and found that Americans also make payments in accordance with a balance-matching heuristic, indicating that it is a broad phenomenon.

References

Benartzi, Shlomo, and Richard H Thaler. 2001. "Naive Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review*, 91(1): 79–98.

Gathergood, John, Neale Mahoney, Neil Stewart, and Joerg Weber. 2019. "How Do Individuals Repay Their Debt? The Balance-Matching Heuristic." *American Economic Review*, 109(3): 844–75.

Ponce, Alejandro, Enrique Seira, and Guillermo Zamarripa. 2017. "Borrowing on the Wrong Credit Card? Evidence from Mexico." *American Economic Review*, 107(4): 1335–61.

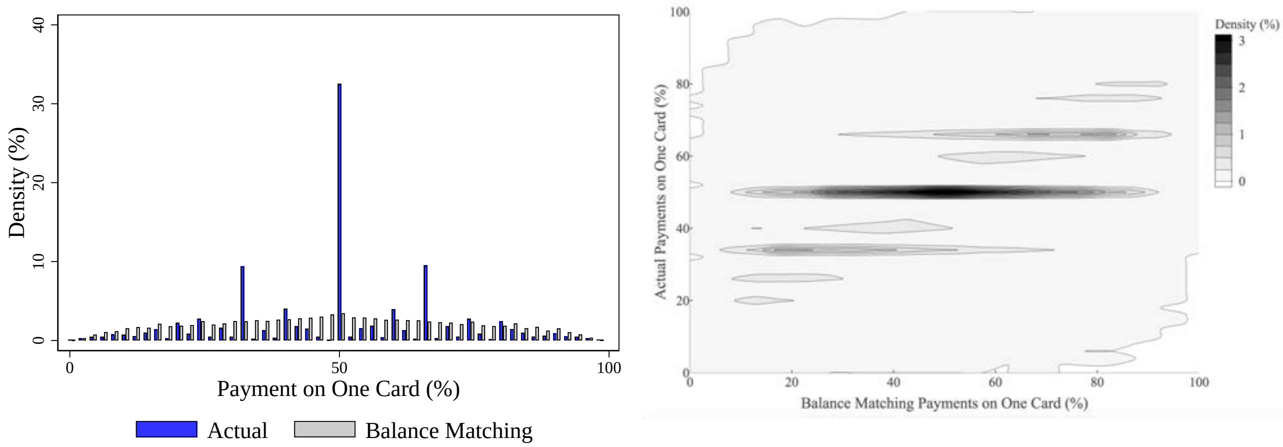
How Do *Americans* Repay Their Debt? The Balance-Matching Heuristic

Appendix

John Gathergood Neale Mahoney Neil Stewart Jörg Weber

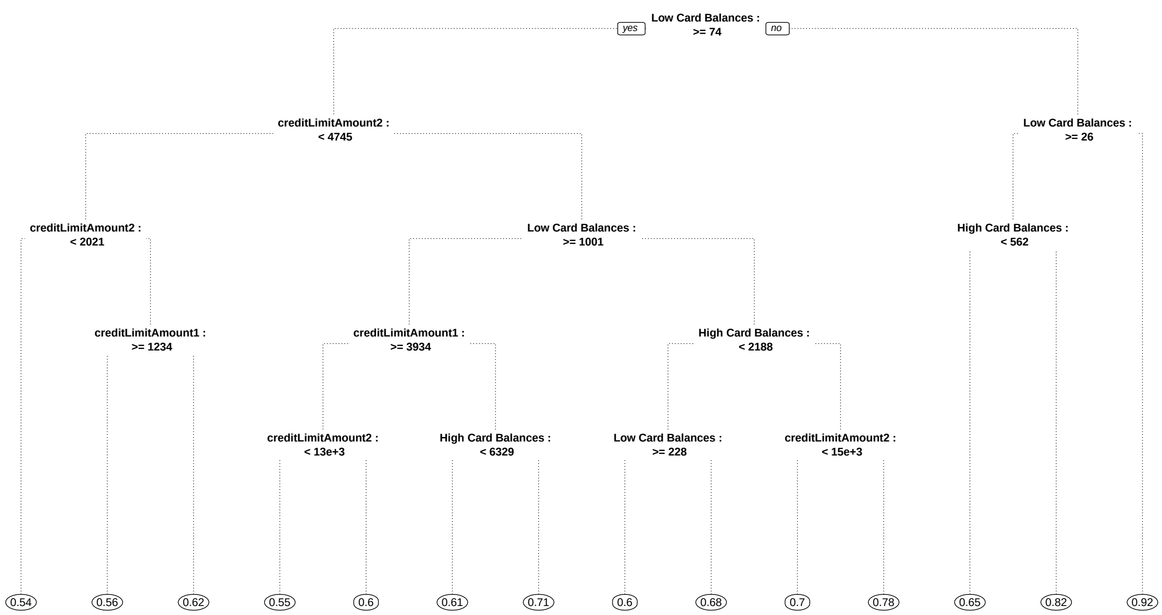
Figure A1: Balance Matching

(A) Round Number Payment Sample



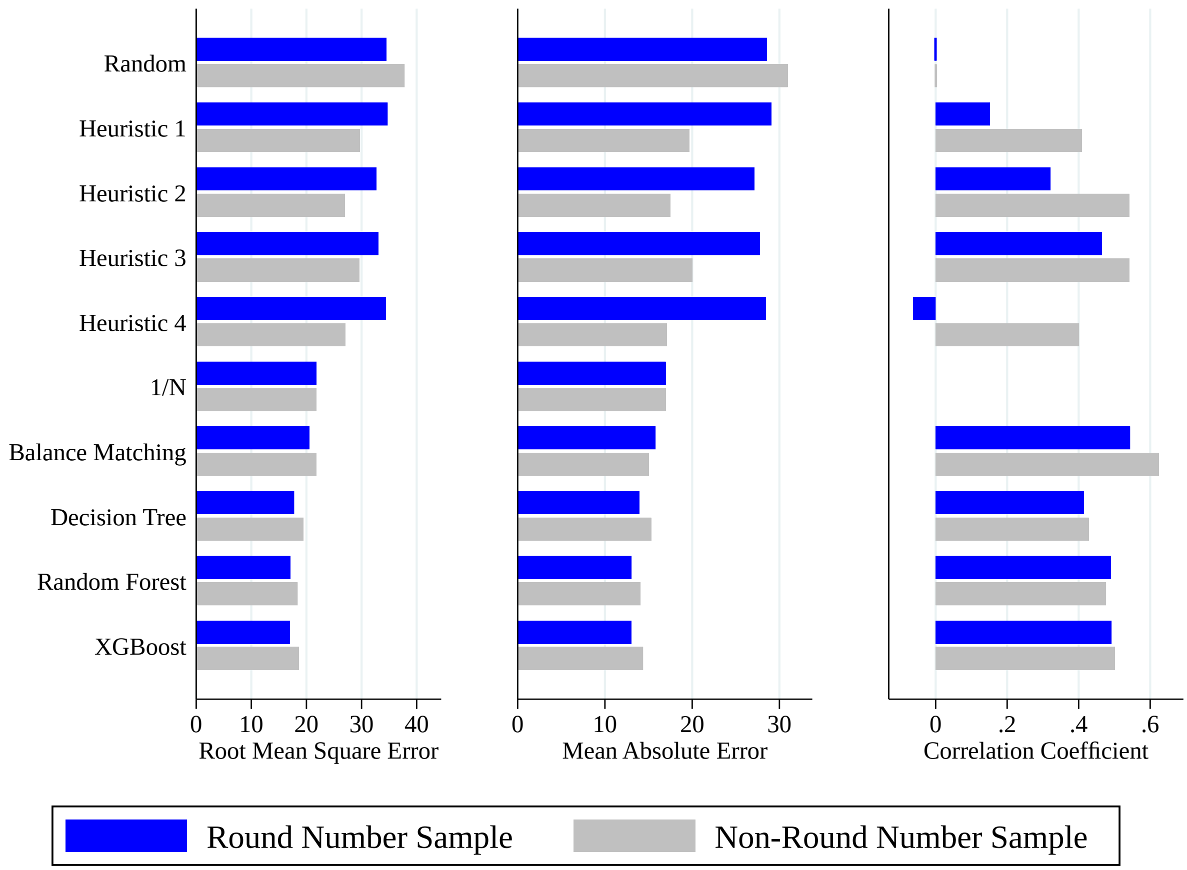
Note: Left panel shows the distribution of actual and balance-matching payments one card in the two-card sample with round number payments (multiples of $50).

Figure A2: Decision Tree



Note: Figure shows the decision tree for percentage of repayments on one card. Top row is tree root. Nodes show the variable and split value at each branch. Bottom rows show predicted values at the end of each branch.

Figure A3: Goodness-of-Fit for Different Models, Round and Non-Round Number Samples



Note: Goodness-of-fit for different models of the percentage of payments. Round and non-round samples are defined by whether repayments on the high APR card are multiples $50.

Table AI: Two-Card Sample Restrictions

		Individual × Months		Aggregate Balance	
	Step	Count	Dropped (%)	Amount ($, Millions)	Dropped (%)
Panel A: Data Coverage					
All credit cards	0	503,036,516		2,379,064	
Two-card sample (both open)	1	92,510,310	81.6%	364,093	84.7%
Drop if either lacks payment data	2	2,527,248	97.3%	9,550	97.4%
Drop if either has no minimum payment	3	1,765,050	30.2%	7,345	23.1%
Drop if payment not updated since last month	4	1,391,563	21.2%	5,987	18.5%
Drop if unique credit card ID is duplicated	5	1,391,379	0.00%	5,986	0.00%
Panel B: Economic Sample					
Drop if either has negative capacity	6	1,223,815	12.0%	5,185	13.4%
Drop if either is delinquent	7	1,222,553	0.1%	5,178	0.1%
Drop if payment less than minimum or more than balance	8	914,056	25.2%	4,253	17.9%
Drop if both cards pay only minimum payment	9	840,588	8.0%	3,883	8.7%
Drop if both cards pay full balance	10	713,157	15.2%	3,622	6.7%

Note: Table shows the sample restrictions.

Table AII: Goodness-of-Fit for Different Models

	(1) RMSE	(2) MAE	(3) Corr
i) Main Models			
Random	37.81	30.96	0.00
1/N	21.83	16.96	0.00
Balance Matching	21.47	15.25	0.61
ii) Alternative Heuristics			
Heuristic 1 (Pay Down Lowest Capacity)	31.13	22.22	0.34
Heuristic 2 (Pay Down Highest Capacity)	28.66	20.11	0.48
Heuristic 3 (Pay Down Highest Balance)	30.59	22.14	0.52
Heuristic 4 (Pay Down Lowest Balance)	29.24	20.18	0.29
iii) Machine Learning Models			
Decision Tree	18.69	14.62	0.41
Random Forest	18.04	13.78	0.48
XGBoost	17.91	13.72	0.49

Note: Goodness-of-fit for different models of the percentage of repayments on one card. Column 1 shows root mean square error (RMSE), column 2 shows mean absolute error (MAE) and column 3 shows Pearson Correlation Coefficient.

Table AIII: Goodness-of-Fit for Different Models, Round Number and Non-Round Number Payment Samples

	Round Number Sample			Non-Round Number Sample		
	(1) RMSE	(2) MAE	(3) Corr	(4) RMSE	(5) MAE	(6) Corr
i) Main Models						
Random	34.53	28.58	0.00	37.81	30.96	0.00
1/N	21.83	16.96	0.00	21.86	17.00	0.00
Balance Matching	20.53	15.80	0.54	21.81	15.04	0.62
ii) Alternative Heuristics						
Heuristic 1	34.74	29.08	0.15	29.67	19.66	0.41
Heuristic 2	32.71	27.11	0.32	26.99	17.50	0.54
Heuristic 3	33.05	27.55	0.47	29.61	20.04	0.55
Heuristic 4	34.41	28.43	-0.06	27.06	17.11	0.40
iii) Machine Learning Models						
Decision Tree	17.78	13.95	0.41	19.47	15.31	0.43
Random Forest	17.11	13.05	0.49	18.41	14.08	0.48
XGBoost	17.01	13.03	0.49	18.64	14.38	0.50

Note: Goodness-of-fit for different models of the percentage of payments on one card. Round and non-round samples are defined by whether repayments on the high APR card are multiples $50. Column 1 shows root mean square error (RMSE), column 2 shows mean absolute error (MAE) and column 3 shows Pearson Correlation Coefficient.

Table AIV: Heterogeneous Types from 3-Way Horse Race Model

Win Percent	(1)
1/N	39.76
Balance Matching	41.01
Random	19.23

Note: Table shows percentage of individual $\times$ month observations that are best fit by different models of repayment behavior.