Online Appendix

Big Loans to Small Businesses: Predicting Winners and Losers in an Entrepreneurial Lending Experiment

Authors - Gharad Bryan, Dean Karlan & Adam Osman

This online appendix includes (1) additional analysis tables, (2) a more detailed description of the machine learning strategy, (3) a discussion of the ethics of the experiment, and (4) a list of relevant survey questions.

1- Appendix Tables

Appendix Table 1. Balance on Psychometric Variables

	Control mean (1)	Treatment - Control Difference (OLS) (2)	Number of Obs (3)	Predictors of Take Up in Control (4)	Predictors of Take Up in Treatment (5)	p-value for test of (4)=(5) (6)
Baseline Value of Psychometric and Cognitive Variables						
Digit Span Recall	2.76	-0.06	981	-0.03	0.00	0.08
	{1.8}	(0.12)		(0.01)	(0.01)	
Raven's Matricies Score	1.72	-0.02	981	0.01	0.01	0.86
	{1.5}	(0.10)		(0.01)	(0.01)	
Hypothetical % allocated to risky investment	0.71	0.04	973	0.03	0.02	0.80
	{0.3}	(0.03)		(0.07)	(0.01)	
Financial Literacy Score	6.02	0.17	981	0.01	0.01	0.55
	{3.3}	(0.21)		(0.01)	(0.01)	
I tend to act first and worry about the consequences later	2.11	-0.01	972	0.01	-0.01	0.66
	{1.0}	(0.07)		(0.02)	(0.02)	
I can think of several solutions to any problem	1.76	-0.09	972	0.00	0.01	0.57
	{0.7}	(0.04)		(0.03)	(0.03)	
I prefer to have a flexible schedule - I don't like being tied down	1.98	0.01	970	-0.01	0.00	0.58
	{0.9}	(0.06)		(0.03)	(0.02)	
When I make a decision I usually go with my first, gut feeling.	2.03	-0.02	971	0.02	-0.01	0.77
	{1.0}	(0.07)		(0.02)	(0.02)	
I would work seven days a week if I could	1.67	0.06	972	-0.05	-0.01	0.80
	{0.7}	(0.05)		(0.03)	(0.02)	
I spend a lot of time planning for my future	2.09	0.02	972	0.01	0.01	0.76
	{1.0}	(0.07)		(0.02)	(0.02)	
I always get things done ahead of time	1.70	-0.04	972	-0.03	0.00	0.42
	{0.7}	(0.04)		(0.04)	(0.03)	
p-value for Joint Test of 50 psychometric & above questions		0.535	1004			

Notes: Column 1 reports the control group mean with standard deviations in braces. Column 2 reports the coefficient on a treatment indicator from a regression of each row on a treatment indicator and strata fixed effects. Columns 4 & 5 report coefficients from a single full interacted regression of a binary for take-up of the loan on each variable listed in the rows interacted with an indicator for control (coefficients in column 4) and an indicator for treamtnet (coefficients in column 5). Column 6 reports the p-value for testing the coefficients from column 4 to its relevant counterpart in column 5. Standard errors in parentheses.

Appendix Table 2. How Survey Attrition Differs by Treatment

	Coefficient (1)	Main Coefficient (2)	Interacted Coefficient (3)		Main Coefficient (4)	Interacted Coefficient (5)	
Treatment	0.007	0.029		Age	-0.004	-0.014	
	(0.008)	(0.066)			(0.006)	(0.024)	
Digit Span Recall		-0.007	0.016	Female (=1)	-0.004	0.006	
		(0.008)	(0.012)		(0.007)	(0.007)	
Raven's Matricies Score		0.005	-0.022**	Years of education	0.014*	-0.023	
		(0.006)	(0.009)		(0.008)	(0.014)	
Hypothetical % allocated to risky investment		-0.027**	0.030	Years of experience	0.003	-0.003	
		(0.014)	(0.017)		(0.008)	(0.009)	
Financial Literacy Score		-0.002	0.016	Monthly profits	0.002	-0.012	
		(0.008)	(0.014)		(0.007)	(0.010)	
In life, failure is not an option		-0.010	0.029*	Monthly expenditures	-0.025	-0.058	
		(0.008)	(0.017)		(0.023)	(0.063)	
When I make a decision I usually go with my first, gut feeling.		-0.009	-0.004	Monthly revenue	0.023	0.060	
		(0.007)	(0.014)		(0.023)	(0.067)	
I can think of several solutions to any problem		0.004	0.005	Monthly wagebill	0.014	-0.002	
		(0.008)	(0.014)		(0.015)	(0.016)	
I prefer to have a flexible schedule - I don't like being tied down		0.009	-0.026	Business is registered	-0.004	-0.003	
		(0.010)	(0.018)		(0.008)	(0.008)	
I tend to act first and worry about the consequences later		-0.002	0.008	Has employees (=1)	0.006	-0.018	*
		(0.006)	(0.013)		(0.009)	(0.010)	
I feel anxious outside my comfort zone		0.002	-0.002	Number of employees	-0.009	0.019	
		(0.007)	(0.012)		(0.008)	(0.016)	
Size of Previous ABA Loan		0.002	-0.015	Value of non-ABA Loans	-0.025*	0.026**	
		(0.012)	(0.014)		(0.014)	(0.013)	
				Have a loan other than ABA (=1)	0.042	-0.036	
					(0.030)	(0.030)	
p-value for joint test on treatment interactions for attrition				0.519			

Notes: Column 1 reports the coefficient from a regression of attrition in either survey on the treatment dummy while controling for survey round and strata (loan officer) fixed effects. Columns 2-5 report the coefficients from one regression. We regress a binary indicator for attrition in either survey on a treatment indicator as well as standardized values for all of the variables in each row (Columns 2 & 4) and an interaction of that variable with treatment (Columns 3 & 5), along with round and strata dummies. All non-binary control variables are standardized to aid in interpretation of the estimates. The last row reports the p-value of a joint test of significant for all the interacted variables in the regression. Overall response rate was 96%. Standard errors in parentheses. Statistical significance * 0.10; **0.05; ***0.01.

Appendix Table 3. Heterogeneity Using Different Prediction Measures

Panel A: Heterogeniety based on Loan Officer (LO) Perceptions of Repayment				
LO thinks Default will:	Increase with Larger Loan	Stay same with Larger Loan	Decrease with Larger Loan	p-value for
	(1)	(2)	(3)	(1)=(2)=(3)
Treatment Effects on Profits	880	-3,127	-3,365	0.65
	(2932)	(3234)	(13984)	
Treatment Effects on Loan Penalties	-6	61	316	0.43
	(56)	(62)	(225)	
Panel B: Heterogeniety based on Loan Officer (LO) Perceptions of Revenue Changes				
LO thinks Revenue will:	Increase with Larger Loan	Stay same with Larger Loan	Decrease with Larger Loan	p-value for
	(1)	(2)	(3)	(1)=(2)=(3)
Treatment Effects on Profits	4,939	-7,574 *	3,116	0.08
	(3232)	(4140)	(4047)	
Treatment Effects on Loan Penalties	-8	96	40	0.42
	(54)	(71)	(73)	

Notes: Panels A & B utilize loan officer perceptions collected at baseline on a subset of the sample (N=559 from 288 individuals). Panel A looks at treatment effects split by how the loan officer graded repayment chances in the case of the large loan relative to the small loan. Panel B splits by loan officer grades on how revnues will increase with the large loan relative to the small loan. The regressions also include a survey round dummy and strata fixed effects. Monthly profits in the control group are 15649EGP. Standard errors clustered at the individual level. Statistical signficance *0.10; **0.05; ***0.01.

Appendix Table 4. Correlations between 6 Psychometric Variables

		1	2	3	4	5	6	7	8	9	10	11
1	I prefer to have a flexible schedule -I don't like being tied down	1.00										
2	I tend to act first and worry about the consequences later	0.25	1.00									
3	I can think of several solutions to any problem	0.30	0.19	1.00								
4	I feel anxious outside my comfort zone	0.13	0.18	0.10	1.00							
5	In life, failure is not an option	0.21	0.17	0.35	0.16	1.00						
6	When I make a decision I usually go with my first, gut feeling	0.14	0.17	0.19	0.11	0.16	1.00					
7	Digit Span Recall	0.12	0.15	0.03	0.13	0.00	0.08	1.00				
8	Ravens Matrices	0.12	0.18	-0.05	0.08	0.00	0.10	0.32	1.00			
9	Financial Literacy Score	0.10	0.16	0.00	0.13	0.02	0.08	0.37	0.35	1.00		
10	Hypothetical % in risky investment	-0.10	-0.16	-0.09	-0.01	-0.08	-0.09	0.02	-0.09	-0.10	1.00	
11	Willingness to Take Risks	0.06	-0.05	0.06	0.08	0.04	-0.03	0.10	0.01	0.00	0.42	1.00

Appendix Table 5. What Predicts Performance? Correlates of Individual Treatment Effects

	7 Variable ITE Prediction (1)	Individual Treatment Effect (2)	Individual Treatment Effect (3)
Age	-0.044	-275	-129
	(0.041)	(196)	(176)
Female (=1)	-0.058	-428	-277
	(0.089)	(432)	(388)
Years of Education	0.000	231	231
	(0.038)	(183)	(165)
Years of Experimence	0.054	133	-44
	(0.038)	(187)	(168)
Monthly Profits	0.064	-210	-327
	(0.055)	(271)	(244)
Business is Registered	0.018	27	-4
	(0.036)	(175)	(158)
Has employees (=1)	0.040	178	234
	(0.040)	(194)	(174)
Value of other non-ABA loans	-0.001	-430***	-421***
	(0.034)	(164)	(150)
Size of Previous ABA Loan	-0.010	512***	537***
	(0.044)	(192)	(173)
Monthly Expenses	0.323	919	227
	(0.258)	(1264)	(1134)
Monthly Revenue	-0.280	-848	-153
	(0.275)	(1348)	(1209)
Monthly Wagebill	-0.134***	-355 *	-163
	(0.043)	(210)	(190)
6 Variable ITE Prediction			0.994***
			(0.064)
Number of Observations	992	995	991
p-value on joint test	0.255	0.007	0.008

Notes: Outcomes and non-binary dependent variables standardized. 7 variable ITE prediction utilizes the questions outlined in Table 6. Statistical significance *0.10; **0.05; ***0.01.

Appendix Table 6. Differences Between Groups on All Other Psychometric Measures

	Bottom GATES Group Mean {Std. Dev.} (1)	Difference in Top Group Coeff (s.e.) (2)			Bottom GATES Group Mean {Std. Dev.} (3)	Difference in Top Group Coeff (s.e.) (4)	
At work, I need to be in control	4.70	-0.51	***	I am critical of myself	4.29	-0.55	***
	{0.67}	(0.06)			{1.03}	(0.08)	
I always say what I think	4.79	-0.54	***	I feel anxious outside my comfort zone	4.26	-0.95	***
	{0.48}	(0.05)			{1.02}	(0.10)	
I don't follow, I lead	4.07	-0.42	***	Success is never down to luck	4.29	-0.47	***
	{1.20}	(0.10)			{1.14}	(0.09)	
I will do anything to get what I want	4.66	-0.62	***	Whenever I cross something off my 'to-do list',	4.55	-0.79	***
	{0.76}	(0.07)		I add something new straight away	{0.65}	(0.07)	
Deep inside, I know I am better than most	3.70	-0.35	***	You should never take shortcuts in life	3.72	-0.15	
people	{1.49}	(0.12)			{1.39}	(0.11)	
I don't get the recognition I deserve	3.48	-0.35	***	I stay calm even during a crisis	4.44	-0.65	***
	{1.38}	(0.11)			{0.93}	(0.08)	
People often tell me how great I am at	4.30	-0.45	***	When I need to, I act quickly without thinking	4.32	-0.62	***
what I do	{0.92}	(0.08)		too much	{1.05}	(0.08)	
I've got a great sense of humour	4.08	-0.46	***	I plan everything	4.62	-0.70	***
	{1.09}	(0.09)			{0.68}	(0.06)	
It is always better to be in the background	3.34	-0.27	**	I am more concerned about getting the job	3.91	-0.70	***
than in the centre of attention	{1.59}	(0.13)		done than following office rules	{1.33}	(0.11)	
Modesty gets you nowhere	2.22	0.45	***	I can concentrate well even when my office is	3.50	-0.17	
	{1.57}	(0.13)		messy	{1.54}	(0.12)	
Without risk there is no reward	4.61	-0.77	***	I always check and double check my work	4.50	-0.53	***
	{0.80}	(0.07)			{0.92}	(0.08)	
It's always good to question authority	4.63	-0.46	***	I can leave work unfinished and move on	2.47	0.38	***
	{0.72}	(0.06)			{1.54}	(0.12)	
I find it difficult to take orders from other	4.03	-0.52	***	I am more concerned about the big picture	3.24	-0.40	***
people	{1.21}	(0.10)		than the details	{1.56}	(0.12)	
I only trust myself	3.93	-0.55	***	I keep my promises	4.71	-0.53	***
	{1.31}	(0.11)			{0.54}	(0.05)	
People often struggle to understand my	3.93	-0.75	***	I see business opportunities where others	4.23	-0.43	***
ideas	{1.30}	(0.11)		don't	{1.01}	(0.08)	
I see patterns and connections where				I have a strong desire to be successful in life			
other's don't	4.09	-0.48	***		4.72	-0.61	***
	{1.01}	(0.09)			{0.55}	(0.05)	
I always know when to give up, and move	4.57	-0.64	***	I am a results oriented person	3.98	-0.47	***
on to something else	{0.73}	(0.07)			{1.29}	(0.10)	
I can't wait to get out of bed in the morning	4.65	-0.59	***	I am a very competitive person	4.21	-0.26	***
- there is always so much to do	{0.72}	(0.06)			{1.19}	(0.09)	
It is not that I don't see profitable business	3.98	-0.56	***	Some people think I am lazy	1.88	0.31	***
opportunities, I just don't have the	{1.08}	(0.09)			{1.33}	(0.11)	
motivation to do anything about them				I am always trying to improve my	4.78	-0.58	***
				performance, whatever I am doing	{0.42}	(0.05)	

Notes: Columns 1 & 3 report the baseline averages of characteristics for the individuals who are in the bottom group of estimated impacts on profits based on the Chernozhukov (2023) method utilizing psychometric, cog & risk data, while Columns 2 & 4 report the coefficient on an indicator for being in the top group from a regression of the variable in each row. Regressions only include people in the top or bottom groups. Psychometric questions are on a 1-5 scale, with 5=strongly agree & 1=strongly disagree. We report all the psychometric measures collected at baseline and not included in Table 6. Standard errors in parentheses. Statistical significance *0.10; **0.05; ***0.01.

Appendix Table 7. Inclusion Rounds for Lasso ITE Predictions

Variables	Inclusion Round in the Lasso Regression	Variables	Inclusion Round in the Lasso Regression
I can think of several solutions to any problem	1	I can't wait to get out of bed in the morning - there is always so much to do	17
I prefer to have a flexible schedule - I don't like being tied down	1	I find it difficult to take orders from other people	17
When I make decisions I usually go with my first,	2	I stay calm even during a crisis	18
I would work seven days a week if I could	2	Deep inside, I know I am better than most people	18
I spend a lot of time planning for my future	3	I see patterns and connections where other's don't	18
I tend to act first and worry about the consequences later	4	I can concentrate well even when my office is messy	18
I always get things done ahead of time	4	I am always trying to improve my performance, whatever I am doing	19
I feel anxious outside my comfort zone.	5	Success is never down to luck	20
In life, failure is not an option	6	You should never take shortcuts in life.	20
I have always believed I am going to be successful	7	Without risk there is no reward	21
Modesty gets you nowhere	7	Ravens Matricies	21
Whenever I cross something off my ‘to-do list’, I add something new straight away	7	It is always better to be in the background than in the centre of attention	22
I plan everything	8	I always know when to give up, and move on to something else	22
Digit Span	8	I am critical of myself	23
I can leave work unfinished and move on	9	When I need to, I act quickly without thinking too much	24
It's always good to question authority	10	It is not that I don't see profitable business opportunities, I just don’t ha	25
I have a strong desire to be successful in life	10	I am a results oriented person	26
People often struggle to understand my ideas	11	I am a very competitive person	26
I keep my promises	11	I will do anything to get what I want	26
When I make a business decision it is almost always the right decision	12	I see business opportunities where others don’t	27
I prefer to focus on opportunities rather than risks	12	I am more concerned about getting the job done than following office rules	28
Some people think I am lazy	12	I am more concerned about the big picture than the details	29
At work, I need to be in control	13	Risk Attitude	29
I always say what I think	13	I only trust myself	30
I always check and double check my work	13	I don't get the recognition I deserve	31
Financial Literacy	14		

Notes: This table reports the results from a lasso regression of the individual treatment effect estimated using the method from Chernozhukov (2023) on our psychometric, risk and cognitive variables. We first implement the Puffer transformation (Jia and Rohe, 2015) to generate a stable ranking and then report which variables are including at different levels of inclusion restrictions. Those including in the first inclusion round are most predictive of the treatment effect, and each subsequent round improves the prediction conditional on those variables that have come before it.

Appendix Table 8. Feature Importance Scores for Random Forest ITE Predictions

Variable	Feature importance score	Variable	Feature importance score
I prefer to have a flexible schedule - I don't like	736,654	Some people think I am lazy	65,205
I can think of several solutions to any problem	602,199	I see business opportunities where others don't	65,115
When I make decisions I usually go with my first, gut feeling	426,525	When I need to, I act quickly without thinking too much	63,896
I would work seven days a week if I could	420,520	Deep inside, I know I am better than most people	62,576
I spend a lot of time planning for my future	342,674	I find it difficult to take orders from other people	61,885
I always get things done ahead of time	275,521	I can concentrate well even when my office is messy	61,350
I tend to act first and worry about the consequences later	274,575	I always know when to give up, and move on to something else	59,500
I feel anxious outside my comfort zone.	264,970	You should never take shortcuts in life.	59,255
Modesty gets you nowhere	185,872	Success is never down to luck	58,203
I have always believed I am going to be successful	170,854	Without risk there is no reward	57,669
In life, failure is not an option	150,927	I only trust myself	56,632
I plan everything	148,293	I am critical of myself	55,837
It's always good to question authority	142,466	I don't follow, I lead	55,741
Whenever I cross something off my 'to-do list', I add something new straight away	112,994	It is always better to be in the background than in the centre of attention	55,575
I can leave work unfinished and move on	101,918	People often tell me how great I am at what I do	54,112
I have a strong desire to be successful in life	92,855	At work, I need to be in control	53,095
Digit Span Recall	91,913	Risk Attitude	52,719
People often struggle to understand my ideas	88,019	Financial Literacy Score	51,800
I prefer to focus on opportunities rather than risks	84,302	I am always trying to improve my performance, whatever I am doing	51,556
I keep my promises	77,425	I will do anything to get what I want	51,085
I can't wait to get out of bed in the morning - there is always so much to do	76,817	I am more concerned about the big picture than the details	50,179
When I make a business decision it is almost	74,656	I don't get the recognition I deserve	49,929
I've got a great sense of humour	73,787	Hypothetical Investment Choice	49,457
I see patterns and connections where other's don't	73,150	I am more concerned about getting the job done than following office rules	49,109
I am a very competitive person	70,582	It is not that I don't see profitable business opportunities, I just don't ha	48,976
I always check and double check my work	70,099	I am a results oriented person	47,494
I always say what I think	67,435	Ravens Matrices	46,841
I stay calm even during a crisis	66,861		

Notes: This table reports the results from a random forest that was trained to estimate the individual treatment effect estimated using the method from Chernozhukov (2023) using our psychometric, risk and cognitive variables. We first implement the Puffer transformation (Jia and Rohe, 2015) to stay consistent with the lasso regressions. We then extract values for feature importance, where a larger number represents the variable being more predictive of the treatment effect.

2- Machine Learning Appendix

This appendix details the machine learning procedures for estimating treatment effect heterogeneity. We mostly follow the generic machine learning techniques of Chernozhukov et al. (2023), with a couple modifications explained below.

To build some intuition for the method, suppose that, based on prior knowledge or theory, we were able to predict a treatment and a control outcome for each of our firms, based only on baseline data. We would then be able to subtract these two outcomes to form a predicted individual treatment effect (ITE) for each participant in our study. We could then use our experimental data to test whether the predicted ITE accurately captures heterogeneity in experimental treatment effects. For example, we might run a linear regression including the predicted ITE, interacted with treatment, and test whether the resulting coefficient is equal to zero. If the coefficient was greater than zero, this would imply that the predicted ITE captures some of the actual treatment effect heterogeneity, and if it was equal to 1 then it would perfectly capture it. Because the experimental data has not been used to generate the predicted ITE, there are no concerns about overfitting, and standard hypothesis testing methods can be used.

In practice, we do not have sufficient prior knowledge to undertake this task, but we do know that, with sufficiently rich baseline data, machine learning methods are good at producing well-performing predictive models. This suggests using a subset of our experimental data (a training set) to train a predictive model, and then testing for heterogeneity using the remaining part of the data (a testing set). Again, because the testing set has not been used to train the model, overfitting is not a concern. Of course, the results of this method would depend on the exact split of the data used, which suggests trying multiple splits and then taking steps to correct for the number of splits used. We would also like to account for the fact that the researcher may have used multiple different machine learning approaches and chosen the one with the "best" performance.

An important implication of this discussion is that the specific machine learning algorithm used is not relevant, it only matters that the model is successful at pre-

dicting outcomes using baseline data. This is the logic, that we follow, put forward in Chernozhukov et al. (2023). We use four different and common machine learning algorithms, but we could use any algorithm with good predictive performance. A second implication is that, while it may not be possible to do correct inference for the entire conditional average treatment effect (CATE) function, it may be relatively straightforward to do correct inference for specific features of the CATE, as suggested by the linear heterogeneous treatment effect example above. A major contribution of Chernozhukov et al. (2023) is to provide a procedure that accounts for multiple sample splits and multiple machine learning methods and provides correct inference around parameters or features that matter for the analysis of heterogeneous treatment effects. In our case we care about two primary parameters, (1) whether the data show evidence of important heterogeneity in treatment effects (BLP), and (2) how the treatment effect differs for different groups in the sample (GATES).

For the first parameter the procedure calls for estimating a model for the individual treatment effect using a random half of the data (training set), and then regressing that value interacted with treatment on the actual outcomes in the other half of the data (testing set). If the models are predictive then when we run regression 1 we would find that the coefficient on the predicted individual treatment effect, interacted with treatment, will be positive and statistically significant. This would mean that the individual treatment effect estimated for the people in the testing test (using the model generated from the data for the people in the training set) is indeed predictive of the actual outcomes from the experiment.

Part of the inference procedure includes rerunning the analysis 100 times with different splits of the data into training & testing sets and taking the median values of the coefficients to protect against the uncertainty of the data splitting. To account for this splitting and rerunning confidence intervals are considered to be at the 1-2 $*$ α instead of 1-α level. To account for the use of multiple machine learning models, a Bonferroni adjustment is used.

For the second parameter, the predicted ITE is sorted for everyone in the testing set,

and then they are split into quartiles. This is followed by a fully interacted regression of treatment and the set of estimated group quartile bins, so that we can estimate what the treatment effect is for each group. Again, confidence intervals are at the 1-2 $*$ α level to account for sample splitting and because the method initially generates predictive models using four different machine learning methods it multiplies all p-values by four, i.e. a Bonferroni correction to conservatively control for multiple hypothesis testing. We report results from the method that performs best (i.e. has the lowest p-value on the coefficient from regression 1), and with this Bonferroni adjustment we should not need to worry about false positives. Nonetheless, we also find heterogeneity using other methods. But looking for heterogeneity using less predictive algorithms is like estimating a causal effect using an instrument with a less powerful first stage - you could do it but a researcher would naturally prefer the stronger option, for both statistical and theoretical reasons.

Chernozhukov et al. (2023) estimates the GATES parameters using only the testing half of the sample. In each of the 100 splits of the data, they generate ITEs for half of the sample, and then run their regressions on that half of the sample. They do this 100 times and then take the median coefficient and median standard error from the 100 regressions. We follow this procedure for our GATES estimates on profits (panel A in Table 4). In our CGATES estimates (panels B & C in Table 4) we deviate slightly by taking the average group allocation for individuals across all 100 splits of the data for each person and then allocate them final groups based on this value.[57] We then run the fully interacted regression of these "final groups" and treatment to recover the group average treatment effects. This allows us to utilize the full sample, while ensuring that the estimates remain "honest", since no person's data went into estimating their own ITE or Group.

Research degrees of freedom: While the method is relatively mechanical, researchers have some degree of freedom which could allow them to cherry-pick results. Re-

[57] Alternatively we could take the median ITE from all 100 splits and allocate people to groups based off of that. When we do so we get largely the same results. We prefer to do it based on average group allocation because average ITE estimate could differ across splits, and so allocating based on groups provides a method that is immune to these types of "level" effects.

searchers can choose which data to include in the analysis, which machine learning algorithms to try out, what proportion of the data to put into training vs testing sets, and the number of groups in the GATES procedure. In our setting, we show how we change the baseline data by including only psychometric and cognitive data in one case, standard data in another, and all data in a third. We use only the four machine learning algorithms that are chosen by Chernozhukov et al. (2023), and we correct for the multiple methods as noted above. We split testing and training data in half just like the initial authors, and believe that an alternative approach would seem manufactured. We do diverge by using four groups instead of five in the GATES analysis. This does not change the estimate of whether heterogeneity exists (BLP), but provides us more power when considering impacts on each group. Estimates are economically similar, but less precise when we use five groups.

Who are in the different groups? A limitation to these methods is that the models produced by the algorithms are not easy interpreted. While the models can do a good job predicting the outcomes for people outside the sample, the model itself is not informative about why some people are predicted to do well and others are predicted to do badly. To address this Chernozhukov et al. (2023) suggests using "CLAN Analysis" which compares the baseline characteristics of people in the top group and those in the bottom group. But it's not clear which characteristics are worth comparing, which is a difficult problem when there are many baseline covariates that the algorithms could be combining non-parametrically.

To address this concern we take the median value of the estimated individual treatment effect across the 100 splits of the data, and then attempt to see which baselines variables are most predictive of the estimated effect. We do this in two ways, once using a lasso, and another time using a random forest algorithm. We choose the top 5 most predictive variables, in line with other recent work utilizing machine learning techniques to generate summary measures (Jayachandran et al., 2021).

3- Ethics Appendix

We have developed this appendix in an effort to describe the ethical considerations of this experiment, and clarify the nature of the collaboration between the researchers and ABA. We follow the framework put forth in Asiedu et al. (2021), for the sake of comparability within economics.

1. **Equipoise** As we describe in the paper, there was significant hesitancy from bank management in providing these larger loans to clients. We used the evaluation to provide the lender an opportunity to test whether or not these loans would be beneficial to their clients as well as to the bank's bottom line.

2. **Role of Researchers with Respect to Implementation:** Bryan, Karlan and Osman are not active researchers in the project. Bryan, Karlan and Osman designed the randomization and evaluation protocols and managed the data collection activities and all of the data analysis. ABA, and not the researchers, designed and implemented the treatments (i.e., the loans).

3. **Potential Harms to Research Participants from the Interventions:** We designed the study to limit the potential harms to participants. In particular participants voluntarily applied for a larger loan, and were approved by their loan officer and the bank's credit committee. All participants expected to benefit from the intervention, and all the bank staff also expected participants to benefit. Our results show that some did not benefit, leading to decreases in business profits and related outcomes.

4. **Potential Harms to Research Participants from Data Collection or Research Protocols** We do not believe participants were subjected to any harms from data collection. Participants were able to refuse to answer any questions they wanted to and were told they could stop the interview at any time.

5. **Financial and Reputational Conflicts of Interest:** Bryan, Karlan and Osman did not receive any form of financial compensation as part of this study (nor did any

assistants or staff associated with the research team). No employee of ABA was named as a PI or participant in any research grant that provided funding for this project.

The research questions pursued in this study and the results described in this study are novel and different form of prior work conducted by the authors. We perceive no reputational conflicts of interest.

6. **Intellectual Freedom:** This study was conducted through a collaboration between PIs and the Alexandria Business Association. The study was conceived and designed by the PIs, who maintained full intellectual freedom throughout all stages of the project. At no point did the partner have undue influence on the analysis or interpretation of results.

7. **Feedback to Participants or Communities:** We intend to share our results with participants via email after our work is subject to peer-review.

8. **Foreseeable Misuse of Research Results:** The authors recognize that the results are relevant for public policy and regulatory activities in credit markets. While misinterpretation or deliberate mischaracterization of the results could have implications for individuals, communities and firms, we are unaware of any set of outcomes or analyses that ex-ante would be predicted to favor advantaged parties. In fact, our search for heterogeneity is partly motivated by this kind of concern regarding studying merely the "average" impact of credit for firms, since average results may mask the presence of both winners and losers.

4- Survey Questions

Risk Questions

R1: How do you see yourself? Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please select one number on the scale, where 0 equals "unwilling to take risks" and 10 equals: "fully prepared to take risks"

R2: Imagine you have EGP 10,000 to invest. You may choose to "invest" all or a portion of your money. There is a 50% chance to double the amount of money you choose to "invest". It is equally possible that you could lose half the investment. For example, if you choose to invest EGP 1,000, there is a 50-50 chance of getting EGP 2,000 or EGP 500. How much of the EGP 10,000 do you invest?

Financial Literacy Questions

FL1: As the demand for a good or service rises, all other things being equal, its price is likely to:

Fall; Rise; Stay the same; There is not enough information to know; I don't know; Refused to answer

FL2: The amount by which revenue from sales exceeds costs in a business is known as:

The profit margin; The cost of capital ; Owner's equity; Gross sales; I don't know; Refused to answer

FL3: As the price of raw materials used to produce a product fall, all other things being equal, the profit margin on sales of that product will:

Fall; Rise; Stay the same; Remain unchanged ; I don't know; Refused to answer

FL4: Monies that go in and out of business over a period of time are known as:

Cash flow; Sales; Assets; Liabilities ; I don't know; Refused to answer

FL5: A sum of money that is owed or due is known as:

A credit; Equity; A debt; A grant; I don't know; Refused to answer

FL6: Annual sales net of all discounts and sales taxes is known as:

Profit; Turnover; Gross sales ; Cash flow; I don't know; Refused to answer

FL7: A sum of money paid regularly (typically annually) by a company to its shareholders out of its profits (or reserves) is known as:

Interest; Profit; A loan; A dividend; I don't know; Refused to answer

FL8: You export the majority of the goods you sell. All other things being equal, if the value of your own currency falls (depreciates) relative to the currency of the market you export to, the goods you sell to that market will be:

More expensive; Cheaper; The same price as before; There is not enough information to know; I don't know; Refused to answer

FL9: You import raw materials to produce the goods you sell. All other things being equal, if the value of your own currency rises (appreciates) relative to the currency you import raw materials from, the cost of producing the goods you sell will:

Rise; Fall; Stay the same; There is not enough information to know; I don't know; Refused to answer

FL10: Debts or financial obligations incurred during business operations are known as:

Equity; Credit; Assets; Liabilities ; I don't know; Refused to answer

FL11: A financial report summarizing a company's assets (what it owns), liabilities (what it owes) and owner's equity at a given time is known as:

The Bought Ledger; The Balance Sheet; The Cash flow Statement; The Profit & Loss Statement; I don't know; Refused to answer

FL12: A financial statement summarizing a company's performance and financial po-

sition by reviewing revenues, costs and expenses during a specific period of time is known as:

The Bought Ledger; The Balance Sheet; The Cash flow Statement; The Profit & Loss Statement; I don't know; Refused to answer

FL13: The total assets of a company, minus its total liabilities represent the capital available for distribution to its shareholders. This is known as:

Profit; Owner's equity; Loss; Debt; I don't know; Refused to answer

FL14: The risk of a customer not paying for goods they have received is an example of:

Financial risk; Operational risk; Strategic risk; Compliance risk; I don't know; Refused to answer

FL15: Which of the following is not included in a cash flow control statement?

Cash coming in; Assets; Cash going out; Cash balance; I don't know; Refused to answer

FL16: Which of the following is not an operating expense?

Payroll; Taxes; Supplies; Dividend payments ; I don't know; Refused to answer

Psychometric Questions

Baseline

At work, I need to be in control

I always say what I think

I don't follow, I lead

I will do anything to get what I want

When I make a business decision it is almost always the right decision

I have always believed I am going to be successful

Deep inside, I know I am better than most people

I don't get the recognition I deserve

People often tell me how great I am at what I do

I've got a great sense of humour

It is always better to be in the background than in the centre of attention

Modesty gets you nowhere

When I make decisions I usually go with my first, gut feeling

I prefer to focus on opportunities rather than risks

Without risk there is no reward

I tend to act first and worry about the consequences later

It's always good to question authority

I find it difficult to take orders from other people

I only trust myself

People often struggle to understand my ideas

I see patterns and connections where other's don't

I always know when to give up, and move on to something else

I would work seven days a week if I could

I can't wait to get out of bed in the morning - there is always so much to do

In life, failure is not an option

I am critical of myself

I feel anxious outside my comfort zone

Success is never down to luck

Whenever I cross something off my "to-do list", I add something new straight away

You should never take shortcuts in life

I stay calm even during a crisis

I can think of several solutions to any problem

I prefer to have a flexible schedule - I don't like being tied down

I always get things done ahead of time

When I need to, I act quickly without thinking too much

I plan everything

I am more concerned about getting the job done than following office rules

I can concentrate well even when my office is messy

I always check and double check my work

I can leave work unfinished and move on

I am more concerned about the big picture than the details

I keep my promises

I see business opportunities where others don't

It is not that I don't see profitable business opportunities, I just don't have the motivation to do anything about them

I spend a lot of time planning for my future

I have a strong desire to be successful in life

I am a results oriented person

I am a very competitive person

Some people think I am lazy

I am always trying to improve my performance, whatever I am doing

Follow Up 1

Discussion is fine, as long as I make the final decision

I don't like being told what I can and cannot do

People think twice before confronting me

In years time, I want people to still be talking about the great work I have done

As a child I always kept my parents on their toes

I am a very easy person to get along with

I am a dreamer

I often have too many ideas in my head at once

I get annoyed when people do not take their work seriously

I like it when people are straight with me and give honest feedback

I feel in complete control of my life

I hate letting people down

I have practical skills that others don't

People often ask for my opinion on business matters

I know that others may be better at a task than I am

I create my own path in life

I choose my words carefully

I like being spontanious

I make decisions quickly and move on

I do my best work when I leave things to the last minute

If I disagree with someone, I tell them what I think

I am good at making last minute changes to plans

When working in teams, I tend to come up with more ideas than others

I am always trying to find new ways of doing things

People who believe in a better future are just naive

Great business ideas change the world

When it comes to exploiting business opportunities, I am often too cautious

Even if I know how to do something, I would always try to do it in a different way

Every decision I make is part of my long-term career plan

I don't have any big plans for the future

I need to do my work exactly right

I get annoyed when plans fall through

I have a brain for business

I always know when there is a gap in the market for a new product

In my group of friends I am the most creative person

I am always willing to take financial risks

When investing my money, I would rather be safe than sorry

I am someone who likes to win at whatever I do

Follow Up 2

Status is the most important thing in life

I would love to be famous

People are jealous of me

I am naturally impatient

I rarely recognize valuable business openings unless they are really obvious

I am a natural risk taker

I was in trouble a lot at school

I was in trouble a lot when I was a kid

I do things my way

I often hesitate to act on valuable business opportunities

I don't like following rules

I sometimes feel overwhelmed by my own creativity

I am an 'ideas person' details bore me

Being different is necessary for success

Before I go to sleep I think, "what could I have done better today?"

I like to live dangerously

As a child, I was never interested in school

It's my way or the high way

I do not let other people's opinions affect me

I am excellent at what I do

I handle difficult tasks well

There is no challenge I feel I can't overcome

I can easily deal with unexpected events

I am more optimistic than most people I know

I prefer to do things in ways that are well established

My judgements can be wrong

I always stick to the rules

I am so busy I sometimes forget things

I need everything to be just right

I change and update my plans constantly

I act exactly according to plan

I will risk making mistakes to get things done

I am always looking out for new business opportunities

I think my decisions through carefully

I feel great after organising my workspace

I always follow directions

I see commercial opportunities in everything

I don't like rushing into things

Some people think I am risk averse

I like to be prepared at all times

I am quick to spot ways of making money

Other people think I am highly creative

If there is a profitable business opportunity, I rarely miss it

I prefer to solve problems in novel ways

It is always best to follow rules and social norms

I am always trying to make things better

I find it easy to apply my creativity in everyday life

I rarely see good business opportunities, even in my area of expertise

My aim in life is to find new ways to make money

Even when I spot a profitable business opportunity, I rarely act on it

I am not afraid to take business risks

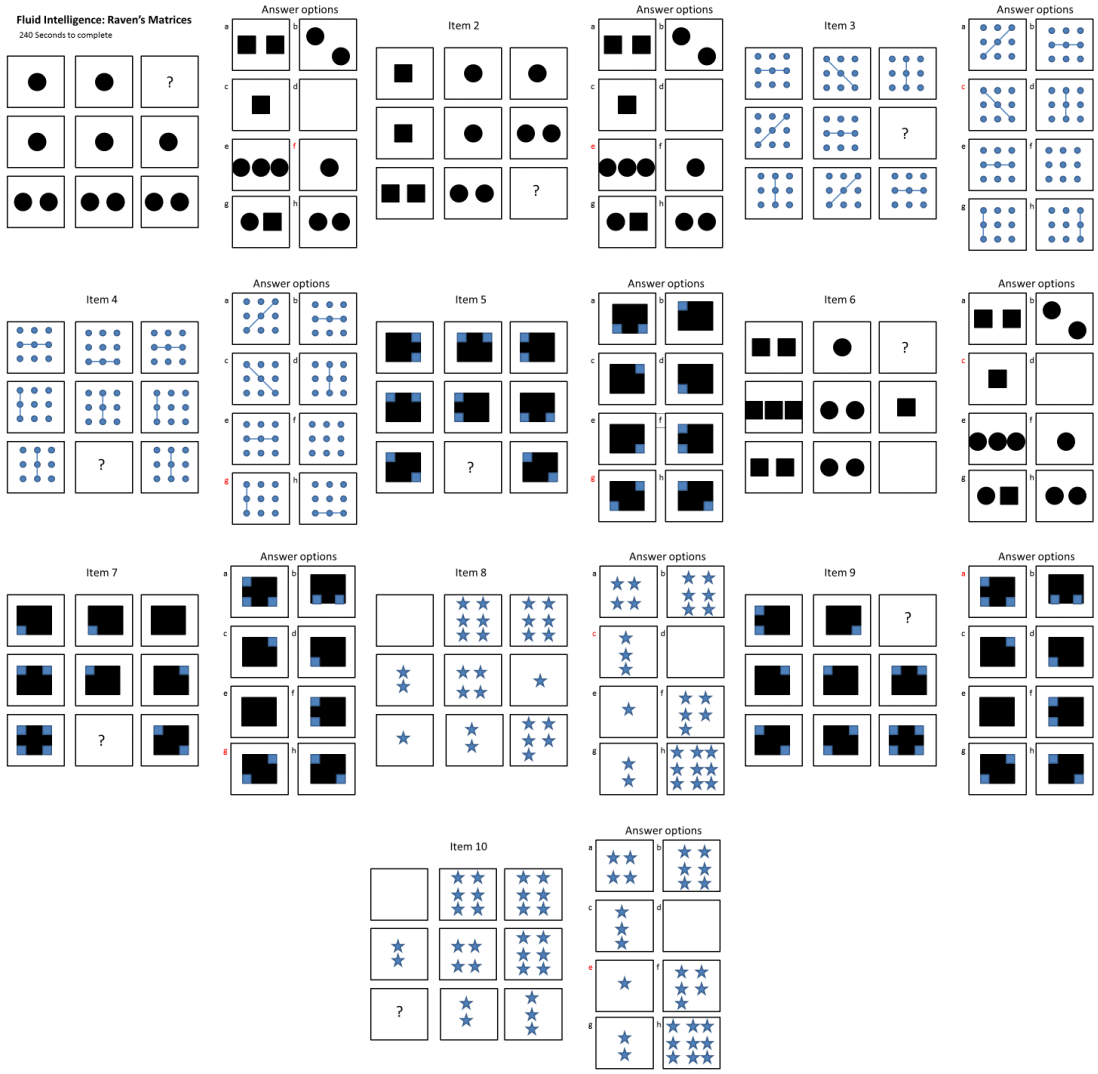
I often solve problems in unusual ways

Cognitive Questions

Digit Span Recall

Now we're going to play a game. A series of digits will appear for a few seconds. Try to remember the digits and repeat them back in order. After each successful turn an extra digit will be added. [Max at 10].

Raven's Matrices



Variables Included in "Standard Data"

Gender; Years of education; 14 Educational Attainment dummies; Birth Year; 9 dummies for different types of income sources; 5 dummies for business sector; number of business partners; 7 dummies for where the business is located; business registration status; 4 dummies for the ownership status of the business premises; the year the business was started; number of branches; expenses; revenues; profits; 5 dummies on how the owner pays themselves from the business; income from other businesses; 3 binaries for the trend of sales over last year; the number of suppliers; if they import inputs; highest and lowest profits over the past 6 months; 8 different dummies about who makes pricing decisions; 3 continuous variables about how much of revenues come from cash, consignment or credit; the number of new products introduced in the previous year; 8 different continuous variables about the number of different types of employees (e.g. full time, part time, etc); 4 binaries on different advertising practices; digit span recall; 4 binaries on bookkeeping; 7 binaries on usage of financial services; 14 continuous variables about different types of expenses for the business; 4 continuous variables about total amount of different types of borrowing at baseline, and 4 corresponding variables about total fees fir each type of borrowing from the different sources (banks, MFI's, friends & family, ROSCAs).