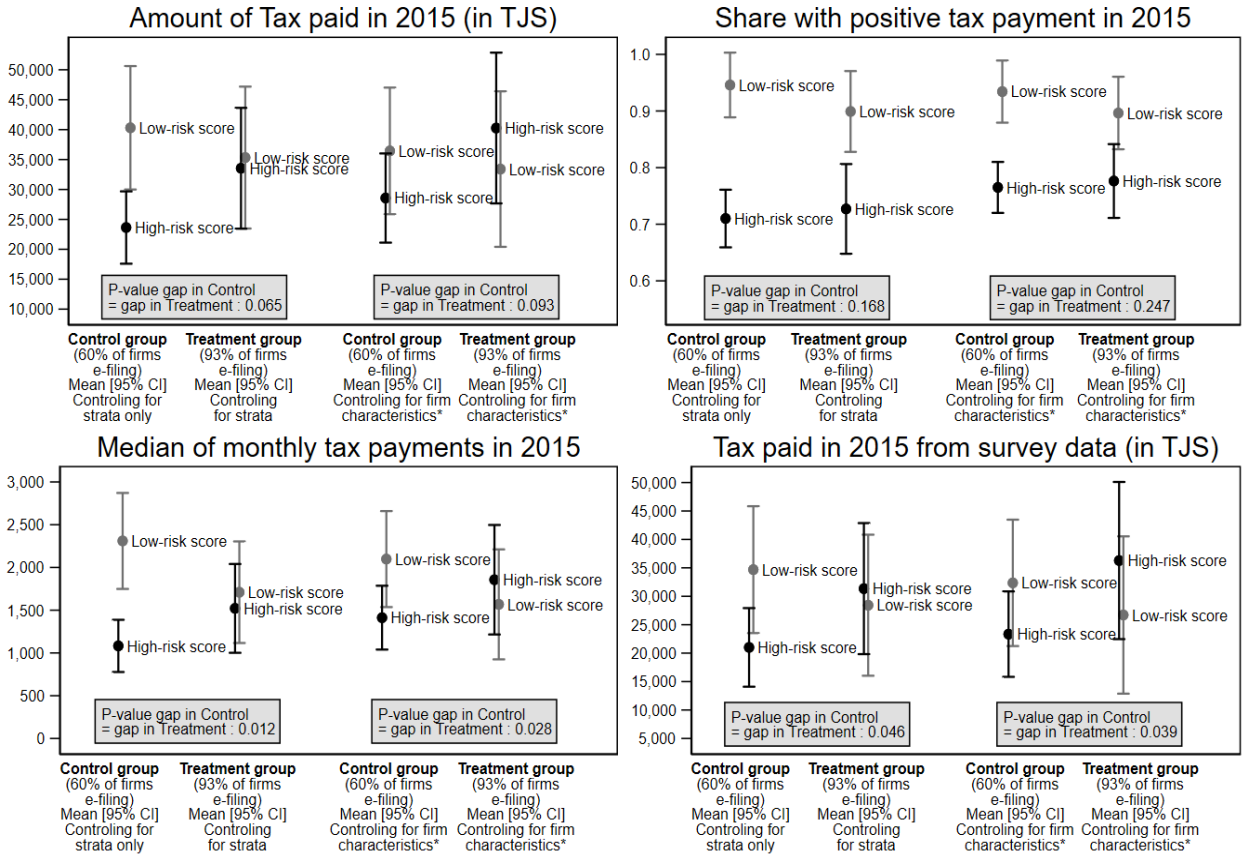
Technology, Taxation, and Corruption: Evidence from the Introduction of Electronic Tax Filing
ONLINE APPENDIX

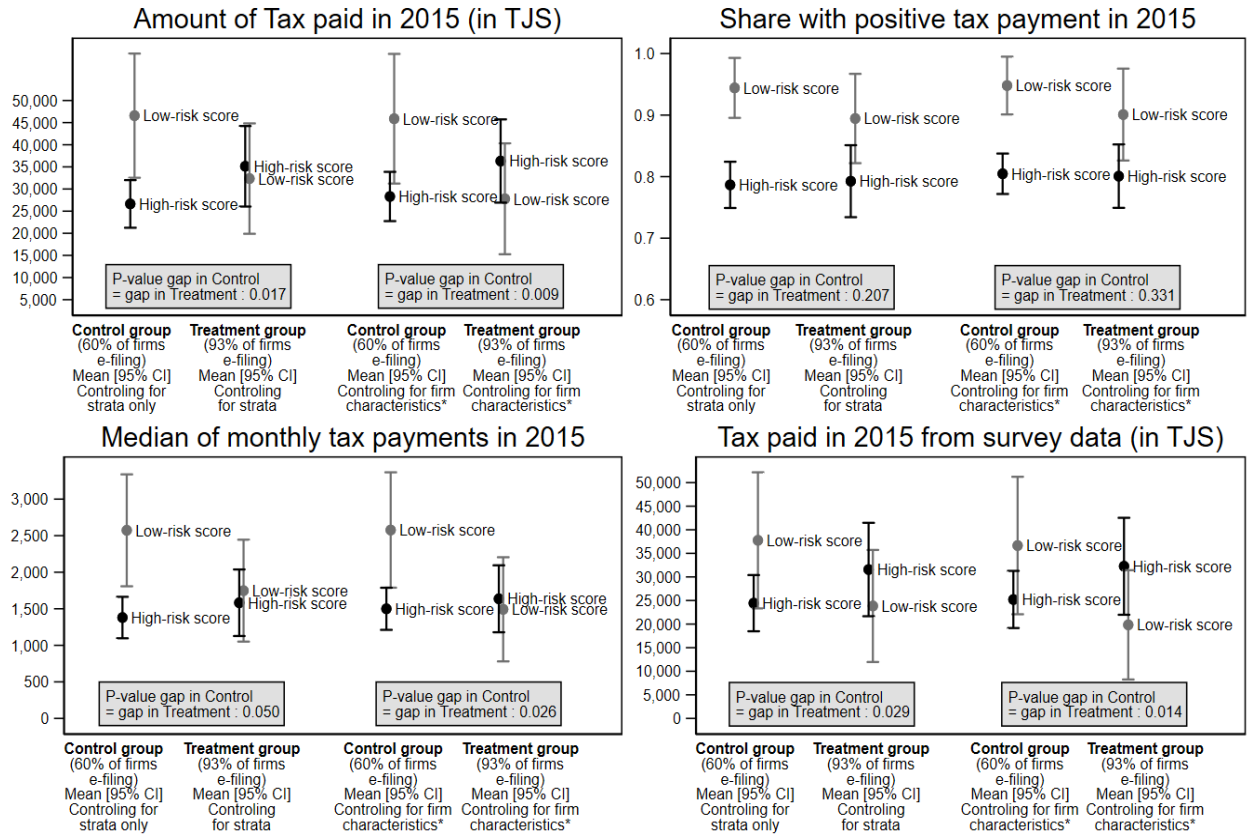
Oyebola Okunogbe and Victor Pouliquen

Appendix Figure A1 : Treatment Impact on Tax Outcomes by Risk Profile Score



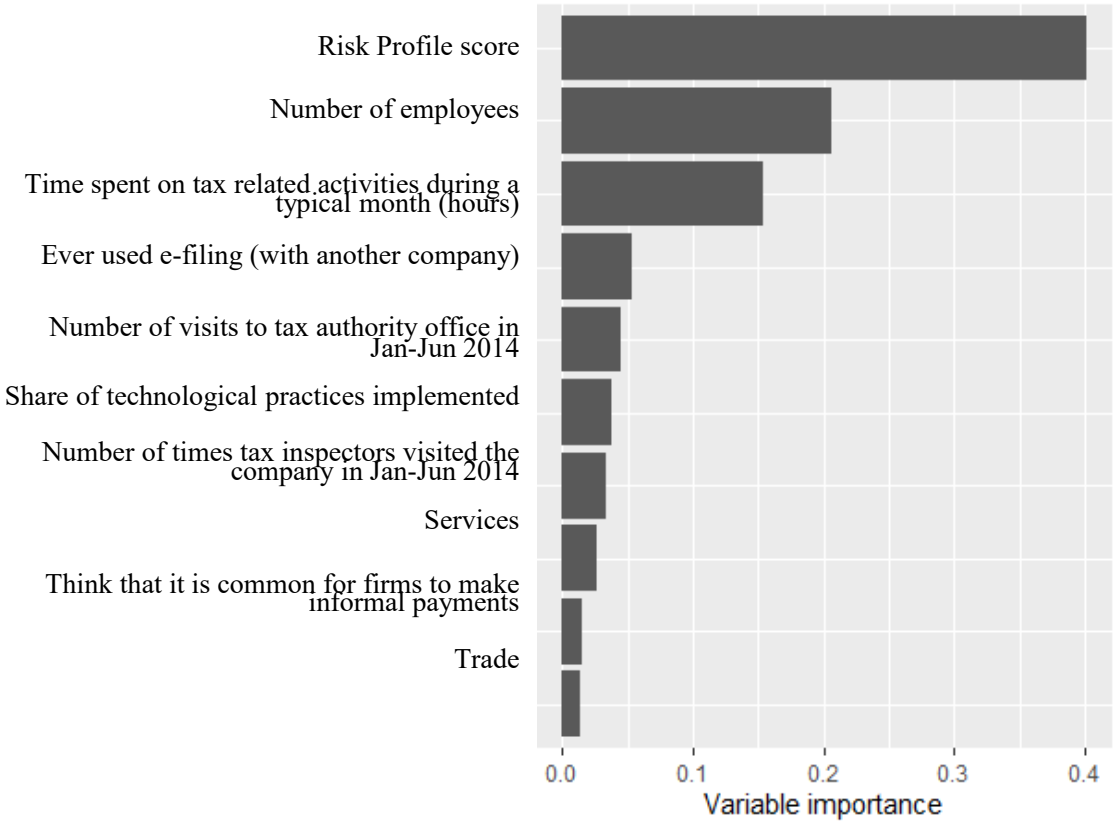
Notes: Administrative data from Tax Authority. Number of observations: 1,067. Low- (high-) risk score refers to firms with a baseline risk profile score below (above) the median. * Controling for sector, no employee, technological practices and time spent on tax activity. 1 TJS= US$ 0.14

Appendix Figure A2 : Treatment Impact on Tax Outcomes for First Quartile of Risk Profile Score and Firms in Quartiles 2, 3 and 4



Notes: Administrative data from Tax Authority. Number of observations: 1,067. Low- (high-) risk score refers to firms with a baseline risk profile score below (above) the 25th percentile. * Controling for sector, no employee, technological practices and time spent on tax activity.
1 TJS= US$ 0.14

Appendix Figure A3 : Variable Importance in Explaining Treatment Impact Heterogeneity



Note: This graph shows the frequency with which each variable is used as a splitting variable in the generalized random forest following Athey et al. (2019).

Table A1: Attrition at Endline Survey

	Mean [SD] Control group	Difference between control group and […]			P-values of the test:	
		Group A	Group B	N	Group A= Group B	Group A= Group B= 0
Survey completed and firm still operating	0.84	0.007	0	1498	0.786	0.932
	[0.366]	(0.021)	(0.026)			
Survey completed and firm liquidated	0.127	-0.008	-0.023	1498	0.506	0.586
	[0.333]	(0.019)	(0.023)			
Survey not completed: not available or not found	0.023	0.009	0.023	1498	0.307	0.226
	[0.15]	(0.009)	(0.014)			
Survey not completed: moved to another town	0.01	-0.008	0	1498	0.205	0.11
	[0.099]	(0.004)	(0.007)			

Notes: Endline survey data 2015. Column 1: Standard deviations presented in brackets. Columns 2-3: coefficients and standard errors (in parentheses) from an OLS regression of the firm owner/firm characteristic on treatment dummies, controlling for strata dummies.

Table A2: Balance Checks by Initial Treatment Groups

	Mean [SD] Control group	Difference between control group and […]		N	P-values of the test Group A= Group B
		Group A	Group B		
PANEL A: Administrative data from Tax Committee (2014)					
Legal entities$^{\beta}$	0.734	-	-	1,498	-
	[0.442]				
Sector of activity is Trade$^{\beta}$	0.413	-	-	1,498	-
	[0.493]				
Sector of activity is Services$^{\beta}$	0.419	-	-	1,498	-
	[0.494]				
Sector of activity is Manufacturing	0.11	0.02	0.025	1,498	0.699
	[0.313]	(0.012)	(0.014)		
Female owner	0.072	0.027	0.012	1,498	0.474
	[0.259]	(0.016)	(0.019)		
No employee	0.414	-0.019	-0.03	1,498	0.754
	[0.493]	(0.027)	(0.033)		
Number of employees	3.257	-0.089	0.119	1,498	0.376
	[3.431]	(0.189)	(0.24)		
Risk profile score in 2014$^{\lambda}$	62.1	1	2.8	1,067	0.433
	[28.7]	(1.8)	(2.3)		
Risk profile score in 2014 above median$^{\lambda}$	0.481	-0.004	0.034	1,067	0.351
	[0.5]	(0.033)	(0.04)		
PANEL B: Baseline Survey Data (2014)					
Firm has an accountant	0.72	0.044	0.047	1,498	0.903
	[0.449]	(0.025)	(0.03)		
Share of technological practices implemented	0.547	0.016	-0.023	1,498	0.159
	[0.432]	(0.023)	(0.028)		
Number of visits to tax authority office in Jan-Jun 2014	6.373	0.136	0.02	1,498	0.095
	[0.978]	(0.061)	(0.066)		
Time spent on tax related activities during a typical month (hours)	6.043	0.279	0.461	1,483	0.451
	[2.681]	(0.166)	(0.227)		
Number of times tax inspectors visited the company in Jan-Jun 2014	1.334	-0.032	-0.016	1,498	0.808
	[0.989]	(0.059)	(0.063)		
Ever used e-filing (with another company)	0.127	0.000	0.04	1,498	0.111
	[0.333]	(0.019)	(0.025)		
Think that it is common for firms to make informal payments	0.184	0.000	-0.014	1,498	0.591
	[0.388]	(0.022)	(0.026)		
P-values of joint orthogonality tests:		0.437	0.122		

Notes: Column 1: Standard deviations presented in brackets. Columns 2-3: coefficients and standard errors (in parentheses) from an OLS regression of the firm owner/firm characteristic on treatment dummies, controlling for strata dummies. β: variables used for stratification. λ: Risk profile scores are only calculated for legal entities. α: P-values of joint orthogonality tests obtained from regressions of the treatment dummy on all the variables in the right-hand column (not used for stratification) and testing for all coefficients are jointly equal to zero.

Table A3 : Balance Checks on Endline Survey Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All sample			**Below median of 2014 risk profile score**			**Above median of 2014 risk profile score**		
	Mean [SD] Control group (B&C)	Diff. in Group A	N	Mean [SD] Control group (B&C)	Diff. in Group A	N	Mean [SD] Control group (B&C)	Diff. in Group A	N
PANEL A: Administrative data from Tax Committee (2014)									
Legal entities$^{\beta}$	0.737	-	1,263	1	-	486	1	-	448
	[0.44]			[0]			[0]		
Sector of activity is Trade$^{\beta}$	0.401	-	1,263	0.346	-	486	0.252	-	448
	[0.49]			[0.477]			[0.435]		
Sector of activity is Services$^{\beta}$	0.425	-	1,263	0.512	-	486	0.567	-	448
	[0.495]			[0.501]			[0.496]		
Sector of activity is Manufacturing	0.124	0.008	1,263	0.104	0.004	486	0.119	0.011	448
	[0.33]	(0.012)		[0.306]	(0.018)		[0.324]	(0.023)	
Female owner	0.08	0.018	1,263	0.073	0.007	486	0.044	0.021	448
	[0.272]	(0.017)		[0.26]	(0.025)		[0.206]	(0.023)	
No employee	0.392	-0.015	1,263	0.197	0.003	486	0.422	-0.04	448
	[0.488]	(0.026)		[0.399]	(0.037)		[0.495]	(0.047)	
Number of employees	3.396	-0.157	1,263	4.294	-0.703	486	3.519	0.16	448
	[3.563]	(0.191)		[3.795]	(0.321)		[3.967]	(0.375)	
Risk profile score in 2014$^{\lambda}$	61.9	-0.7	934	41.7	1.7	486	83.6	-1.4	448
	[27.8]	(1.7)		[10.5]	(0.9)		[24.1]	(2.2)	
Risk profile score in 2014 above	0.483	-0.025	934	0	-	486	1	-	448
	[0.5]	(0.033)		[0]			[0]		
PANEL B: Baseline Survey Data (2014)									
Firm has an accountant	0.73	0.04	1,263	0.796	-0.026	486	0.752	0.074	448
	[0.444]	(0.025)		[0.404]	(0.039)		[0.433]	(0.038)	
Share of technological practices implemented	0.544	0.035	1,263	0.664	-0.026	486	0.651	0.108	448
	[0.438]	(0.023)		[0.417]	(0.04)		[0.408]	(0.039)	
Number of visits to tax authority office in Jan-Jun 2014	6.398	0.13	1,263	6.536	0.086	486	6.459	0.183	448
	[0.994]	(0.062)		[1.067]	(0.106)		[1.149]	(0.119)	
Time spent on tax related activities during a typical month (hours)	6.139	0.111	1,252	6.117	0.075	483	6.686	0.019	442
	[2.709]	(0.162)		[2.35]	(0.236)		[3.171]	(0.36)	
Number of times tax inspectors visited the company in Jan-Jun 2014	1.309	-0.038	1,263	1.26	-0.061	486	1.167	0.048	448
	[0.924]	(0.059)		[0.832]	(0.114)		[0.874]	(0.088)	
Ever used e-filing (with another company)	0.152	-0.01	1,263	0.208	-0.04	486	0.185	0.017	448
	[0.359]	(0.02)		[0.406]	(0.036)		[0.389]	(0.039)	
Think that it is common for firms to make informal payments	0.186	0.005	1,263	0.26	-0.053	486	0.178	0.009	448
	[0.389]	(0.022)		[0.439]	(0.037)		[0.383]	(0.037)	
P-values of joint orthogonality tests:	0.538			0.382			0.265		

Notes: Columns 1, 4, 7: Standard deviations presented in brackets. Columns 2, 5, 8: coefficients and standard errors (in parentheses) from an OLS regression of the firm owner/firm characteristic on a treatment variable (group A), controlling for strata dummies. β: variables used for stratification. λ: Risk profile scores are only calculated for legal entities. α: P-values of joint orthogonality tests obtained from regressions of the treatment dummy on all the variables in the right-hand column (not used for stratification) and testing for all coefficients are jointly equal to zero.

Table A4: Correlation between 2014 Risk Profile Score and 2015 Tax Paid

	(1)	(2)	(3)	(4)
Dependent variables:	Tax paid in 2015		At least one positive tax payment in 2015	
Above median risk profile score in 2014	-5,167	-3,684	-0.147	-0.119
	(3,942)	(3,446)	(0.022)	(0.021)
Controlling for:				
- Trade, Services, gender, no employee, number of employees and tax office	yes	yes	yes	yes
- Index of profit and turnover in June and December 2015		yes		yes
Observations	1,067	934	1,067	934
R-squared	0.200	0.482	0.195	0.173
Mean Dependent Variable	33102	36245	0.825	0.866

Notes: OLS regresions with standard errors presented in parentheses.

Table A5: Impact of E-filing Adoption on Main Outcomes: Comparing Group A to Group C only

Dependent variables:	(1) # visits per month to tax authority office in 2015	(2) Time spent monthly on tax-related activities in 2015 (hours)	(3) Tax paid in 2015$^{\alpha}$ (admin. data) (in TJS)	(4) At least one positive tax payment in 2015	(5) Median of Tax paid in 2015$^{\alpha}$ (admin. data)	(6) Tax paid in 2015$^{\alpha}$ (Survey data) (in TJS)	(7) Think that it is common for firms to pay bribes	(8) Ever paid a bribe to a tax official (list exp.)
Overall Sample:								
(ITT) Impact of Assignment to Group A (compared to group C only)	-0.48	-1.731	1,990	0.006	-45	2,652	0.005	-0.065
	(0.03)	(0.366)	(3,193)	(0.021)	(168)	(6,219)	(0.031)	(0.062)
(IV) E-filing Impact (all firms)	-1.37	-4.92	5,891	0.019	-132	7,550	0.015	-0.191
	(0.06)	(1.04)	(9,383)	(0.061)	(499)	(17,673)	(0.087)	(0.172)
Observations	1,016	1,008	1,202	1,202	1,202	1,016	1,016	1,016
Mean outcome control group C	0.810	11.811	28,310	0.839	1,599	29,355	0.63	0.095
Mean outcome group A	0.33	10.083	30,152	0.843	1,551	30,806	0.632	0.02
E-filing Impact, Legal Entities Sample:								
(IV) E-filing Impact (Legal Entities)	-1.43	-5.17	7,477	0.015	-206	7,651	-0.055	-0.213
	(0.07)	(1.17)	(11,816)	(0.075)	(0.817)	(22,340)	(0.097)	(0.193)
Observations	756	750	883	883	883	756	756	756
Mean outcome control group C	0.84	11.842	31,149	0.794	1,665	33,669	0.627	0.129
Mean outcome group A	0.32	10.004	33,745	0.801	1,586	35,082	0.608	0.044
Impact by Baseline Risk Profile Score:								
(IV) E-filing Impact for below-median risk score	-1.55	-6.543	-9,433	-0.108	-1,941	-28,811	-0.237	-0.558
	(0.13)	(1.915)	(20,196)	(0.085)	(1,128)	(42,952)	(0.158)	(0.312)
(IV) E-filing Impact for above-median risk score	-1.34	-4.119	20,196	0.065	1,130	35,274	0.097	0.081
	(0.09)	(1.428)	(14,433)	(0.106)	(732)	(20,342)	(0.119)	(0.242)
P-val diff. (low vs. high risk score)	0.168	0.314	0.236	0.205	0.023	0.161	0.094	0.108
Observations	756	750	859	859	859	756	756	756
Mean Outcome in control group C for:								
… Firms below median risk score	0.820	12.08	38,292	0.938	2,305	43,718	0.638	0.129
… Firms above median risk score	0.88	11.59	25,554	0.692	1,087	23,023	0.616	0.131
Mean Outcome in group A for:								
… Firms below median risk score	0.33	10.03	36,314	0.908	1,785	37,140	0.569	-0.039
… Firms above median risk score	0.32	9.975	32,671	0.722	1,455	32,803	0.652	0.139

Note: The first two parts of the table present results from 2SLS regressions measuring the impact of e-filing adoption (instrumented by assignment to group A) on outcomes presented in colums. The last part of the table presents results from 2SLS regressions measuring the heterogeneous impact of e-filing adoption (instrumented by assignment to group A) by risk profile scores. For the heterogeneous analysis, samples are restricted to legal entities because risk scores are not calculated for individual enterprises. Column (8) presents results from a list experiment measuring bribe payments (see main text for a description). Robust standard errors in parentheses. All regressions include control dummies for strata. α: winsorized at the 99th percentile. Sharpened two-stage q-values corrected for multiple hypothesis testing as described in Anderson (2008).

Table A6: Impact on E-filing Adoption by Baseline Risk Profile Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All sample			**Below median of 2014 risk profile score**			**Above median of 2014 risk profile score**		
	Mean [SD] Control group (B&C)	Diff. in Group A	N	Mean [SD] Control group (B&C)	Diff. in Group A	N	Mean [SD] Control group (B&C)	Diff. in Group A	N
PANEL A: Administrative data from TC (Aug 2014-Dec 2015)									
Used E-filing	0.604	0.325	1,498	0.626	0.321	547	0.547	0.39	520
	[0.489]	(0.02)		[0.485]	(0.032)		[0.499]	(0.035)	
Used E-filing conditional on survival	0.658	0.326	1,275	0.681	0.32	488	0.592	0.365	474
	[0.475]	(0.019)		[0.467]	(0.029)		[0.492]	(0.035)	
Still using e-filing	0.563	0.224	1,498	0.614	0.248	547	0.537	0.209	520
	[0.496]	(0.024)		[0.488]	(0.037)		[0.499]	(0.042)	
PANEL B: Endline survey data (Feb 2016)									
Firm used electronic filing to submit tax reports in 2015	0.579	0.427	1,263	0.526	0.475	486	0.53	0.488	448
	[0.494]	(0.019)		[0.5]	(0.031)		[0.5]	(0.032)	
Found out about e-filing during intervention training	0.815	0.185	1,263	0.792	0.202	486	0.796	0.211	448
	[0.388]	(0.015)		[0.406]	(0.025)		[0.403]	(0.026)	
Found out about e-filing from business network	0.154	-0.154	1,263	0.176	-0.172	486	0.152	-0.158	448
	[0.362]	(0.014)		[0.382]	(0.024)		[0.36]	(0.024)	
Found out from another source (other training, tax Committee publication...)	0.03	-0.031	1,263	0.031	-0.03	486	0.052	-0.053	448
	[0.172]	(0.006)		[0.174]	(0.01)		[0.222]	(0.014)	

Notes: Columns 1, 4, 7: Standard deviations presented in brackets. Columns 2, 5, 8: coefficients and standard errors (in parentheses) from an OLS regression of the firm outcome on a treatment variable (group A), controlling for strata dummies. Colomn 4 to 9: Sample restricted to legal entities because risk profile scores are only calculated for legal entities.

Table A7: Cost-Effectiveness Analysis (only Accounting for Reduced Compliance Costs)

	Control Group	Group A	Group B
Number of firms	608	594	296
Number of training conducted	12	36	16
Program costs (in USD):			
Training organization (specific by group)	7,573	10,167	4,931
Logistical help to register to e-filing (group A only)	0	4,696	0
Total program costs	**7,573**	**14,863**	**4,931**
Cost effectiveness analysis :			
Cost per firm included in treatment (in USD)	12	25	17
Additional cost with respect to control Group (in USD)		13	
Program impact on e-filing adoption (from table 3)		0.34	
Cost per additional e-filing adoption (in USD)		**37**	
Program impact on compliance costs (in hours saved) (from Table 5)		-4.7	
Amount of money saved monthly by firms[α] (in USD)		5.5	
Number of months for private benefits in term of time saved to exceded program costs		**7**	

Note: Training costs include direct costs of organizating trainings (trainers salaries, equipement rental, calls to invite participants) and costs to develop training materials. Costs related to the logistical help to register include costs to call and visit firms for the software instalation and costs related to firms registration. Exchange rate from Oanda.com on January 1st 2016: USD 1 = TJS 6.99. **α**: assuming the wage of the person in charge of tax declaration is in average USD 178 per month or USD 1.11 per hour.

Table A8: Impact on Bribe Payments Using a List Experiment

	(1)	(2)	(3)	(4)	(5)	(6)
	All Sample		**Below median of 2014 risk profile score**		**Above median of 2014 risk profile score**	
	Group B&C	Group A	Group B&C	Group A	Group B&C	Group A
Short list (4 questions without the question on bribe payments):						
Mean	1.119	1.146	1.109	1.182	1.088	1.075
SE	0.02	0.027	0.03	0.046	0.027	0.035
N	361	246	138	99	136	80
Long list (5 questions, including the question on bribe payments):						
Mean	1.194	1.166	1.252	1.143	1.187	1.214
SE	0.029	0.033	0.05	0.05	0.051	0.06
N	397	259	151	98	134	98
Difference short list - long list:						
Difference	0.075	0.02	0.143	-0.039	0.098	0.139
SE Difference	0.036	0.042	0.06	0.068	0.057	0.074
P-value T-test Difference	0.038	0.643	0.017	0.569	0.087	0.06
Difference in Difference with Control group:						
Difference in Difference		-0.055		-0.182		0.041
SE Difference		0.056		0.092		0.093
P-value T-test Difference		0.326		0.048		0.658

Notes: Endline survey data, Feb 2016. Colomn 3 to 6: samples restricted to legal entities because risk scores are not calculated for individual enterprises.

Table A9: Share of Firms that Stopped Using E-filing at Some Point after Registration

	(1) **Among those who registered for e-filing by June 2015:** Stopped using e-filing at some point after registration
Group A x Below median risk profile score in 2014	0.060
	(0.030)
Group A x Above median risk profile score in 2014	0.172
	(0.032)
Administrative data from Tax Committee (2014)	
Above median risk profile score in 2014	-0.029
	(0.027)
Female owner	-0.052
	(0.040)
No employee	0.126
	(0.032)
Number of employees	-0.001
	(0.002)
Survey data (baseline)	
Firm has an accountant	0.041
	(0.032)
Share of technological practices implemented	-0.032
	(0.030)
Number of visits to tax authority office in Jan-Jun 2014	-0.011
	(0.008)
Time spent on tax related activities during a typical month (hours)	0.004
	(0.004)
Number of times tax inspectors visited the company in Jan-Jun 2014	0.005
	(0.010)
Ever used e-filing (with another company)	-0.063
	(0.025)
Think that it is common for firms to make informal payments	0.014
	(0.031)
P-val difference (low vs. high risk score)	0.012
Observations	623
Mean Dependent variable in control group (B&C) for:	
…all firms (with a risk profile score)	0.028
…firms below median risk profile score in 2014	0.031
…firms above median risk profile score in 2014	0.024
Mean Dependent variable in group A for:	
…all firms (with a risk profile score)	0.151
…firms below median risk profile score in 2014	0.089
…firms above median risk profile score in 2014	0.215

Notes: OLS regresions with standard errors presented in parentheses controling for gender of the owner, no employees, number of employees, strata dummies and all baseline survey variables presented in Table 2. Risk profile scores are only calculated for legal entities.

Table A10: Bribe Behavior of Firms that Dropped out of E-filing at Some Point after Registration

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep variable: Number of answers in the list experiment					
	All firms		Below median of 2014 risk profile score		Above median of 2014 risk profile score	
Ever paid a bribe to a tax official (Long List)	0.078	0.070	0.059	0.060	0.099	0.086
	(0.035)	(0.035)	(0.049)	(0.049)	(0.049)	(0.049)
Stopped using e-filing at some point after registration	0.032	0.048	0.115	0.130	-0.005	0.001
	(0.080)	(0.079)	(0.175)	(0.160)	(0.059)	(0.069)
Ever paid a bribe to a tax official X Stopped using e-filing at some point after registration	0.148	0.159	-0.109	-0.126	0.259	0.289
	(0.146)	(0.145)	(0.220)	(0.214)	(0.168)	(0.177)
Observations	741	734	396	394	345	340
Share of firms that stopped using e-filing at some point after registration	0.093	0.093	0.056	0.056	0.136	0.136
Controling for baseline variables: firm sector, number of employees, technological practises, time spent on tax related activities and location	No	Yes	No	Yes	No	Yes

Notes: Endline survey data, Feb 2016. Samples restricted firms that registered to e-filing and to legal entities because risk scores are not calculated for individual enterprises.

Table A11: E-filing Impact by Quartiles of Baseline Risk Profile Score

Dependent variables:	(1) # visits per month to tax authority office in 2015		(2) Time spent monthly on tax-related activities in 2015 (hours)		(3) Tax paid in 2015[α] (admin. data) (in TJS)		(4) At least one positive tax payment in 2015		(5) Median of Tax paid in 2015 (admin. data)		(6) Tax paid in 2015[α] (Survey data) (in TJS)		(7) Think that it is common for firms to make informal payments	
E-filing (IV) (impact for firms with risk score in Q1)	-1.6	-1.6	-8.802	-9.142	-50,630	-56,537	-0.186	-0.164	-2,943	-3,298	-163,879	-177,632	-0.368	-0.460
	(0.2)	(0.2)	(2.787)	(2.943)	(30,552)	(33,849)	(0.124)	(0.144)	(1,755)	(1,945)	(92,909)	(98,104)	(0.220)	(0.237)
E-filing (IV) x [P25-P50[(Additional impact for firms in Q2)	-0.0	-0.1	4.332	4.559	65,322	71,739	0.077	0.064	2,072	2,508	196,854	202,269	0.276	0.350
	(0.2)	(0.3)	(3.580)	(3.770)	(39,902)	(41,365)	(0.158)	(0.169)	(2,178)	(2,268)	(106,445)	(111,789)	(0.295)	(0.310)
E-filing (IV) x [P50-P75[(Additional impact for firms in Q3)	0.1	0.1	1.062	1.130	92,374	101,829	0.254	0.223	4,935	5,333	226,827	245,065	0.811	0.949
	(0.3)	(0.3)	(4.143)	(4.401)	(42,001)	(47,435)	(0.196)	(0.210)	(2,308)	(2,582)	(105,340)	(112,029)	(0.345)	(0.368)
E-filing (IV) x [P75-P100] (Additional impact for firms in Q4)	0.4	0.3	6.754	6.539	68,842	81,186	0.221	0.197	3,680	4,063	186,265	202,687	0.388	0.542
	(0.2)	(0.2)	(3.212)	(3.587)	(33,465)	(38,855)	(0.175)	(0.197)	(1,891)	(2,150)	(96,874)	(100,053)	(0.255)	(0.290)
Controling for heterogeneity of E-filing impact with variables: firm sector, no employees, technological practices, time spent on tax related activities	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	934	(0.1)	925	(1.064)	1,067	(13,218)	1,067	(0.072)	1,067	(683)	934	(22,606)	934	(0.089)

Note: The table presents results from 2SLS regressions measuring the heterogeneous impact of e-filing adoption (instrumented by assignment to group A) by quartile of risk profile scores. The samples are restricted to legal entities because risk scores are not calculated for individual enterprises. Robust standard errors in parentheses. All regressions include control dummies for strata. α: winsorized at the 99th percentile.

Table A12 : Comparison of Machine Learning Methods

Dep variable: Total Tax Paid in 2015	(1) Elastic Net	(2) Boosting	(3) Neutral Network	(4) Random Forest
Best BLP	122,759,798	202,261,385	149,314,308	220,851,637
Best GATES	4,854	11,322	5,448	14,395

Notes: Medians over 100 splits. BLP: Best Linear Predictors of the conditional average treatment effect. GATES: Group average treatement effect. Best BLP and GATES are computed following the methodology describe in Chernozhukov et al. (2018)

Table A13 : Best Linear Predictors of the Conditional Average Treatment Effect using Random Forest

Dep variable: Total Tax Paid in 2015	(1) Average Treatment Effect (ATE)	(2) Heterogeneity loading parameter (HET)
Coefficient	3,600	0.417
95% Confidence interval	(-6883,13800)	(0.111,0.724)
P-value	[1.000]	[0.013]

Notes: Medians over 100 splits. Best Linear Predictors of the conditional average treatment effect computed with Random Forest, following the methodology describe in Chernozhukov et al. (2018)

Table A14: Classification Analysis using Random Forest: Baseline Characteristics of the Most and Least Affected

Dep variable: Total Tax Paid in 2015	(1) Bottom 50% of predicted effect	(2) Top 50% of predicted effect	(3) Difference (95% CI)	(4) P-values difference
Risk profile score in 2014	59	67	9	0.001
	(55.90,62.62)	(63.92,70.48)	(3.713,13.24)	
Risk profile score in 2014 above median	0.43	0.54	0.12	0.024
	(0.370,0.490)	(0.481,0.601)	(0.038,0.208)	
Sector of activity is Trade	0.36	0.28	-0.09	0.064
	(0.305,0.416)	(0.218,0.331)	(-0.169,-0.007)	
Number of employees	3.47	3.951	0.40	0.415
	(3.050,3.907)	(3.510,4.386)	(-0.221,1.045)	
Share of technological practices implemented	0.65	0.67	0.01	1
	(0.602,0.700)	(0.622,0.721)	(-0.060,0.081)	
Time spent on tax related activities monthly (hrs)	6.457	6.404	-0.01	1
	(6.080,6.832)	(6.044,6.804)	(-0.533,0.514)	

Note: Baseline average characteristics of the 50% most and least affected firms according to Random Forest predictions. Medians over 100 splits. 95% confidence intervals computed using double/debiased machine learning as described in Chernozhukov et al. (2018).

Appendices

A Script Used to Invite Firms to Training

Good day. My name is "OPERATOR NAME". We are calling you from company "IMPLEMENTING PARTNER NAME" LLC. Our company in cooperation with the Tax Committee and the International Finance Corporation is holding a seminar. Your company - LLC "COMPANY NAME" is in the short-list of taxpayers we are inviting to participate in our training. Training will take place on "DATE AND TIME" at "PLACE". All participants will be given training and guidance materials on tax and taxation. Some firms will be selected to be registered for e-filing system free of charge. Due to limited resources, we cannot register all firms now, so the firms selected will be based on chance as chosen by a computer program. In this regard, please prepare and bring the following documents:

- Copy of the registration certificate
- Extract from the Unified State Registry
- Passport copy of Director
- Company Seal

This training will provide important information but it is optional and there will be no penalty for your firm if you do not attend. Thanks in advance for your participation. Telephone number for inquiries and information: xxx-xx-xx

B Measurement of Key Variables and Outcomes

Baseline administrative data:

- Legal entities: equal 1 if the firm is a legal entity (75% of the sample) and 0 if it is an individual enterprise.
- Sector of activity: Sector of activity reported by the firm when registering
- Female owner: the firm owner is a woman
- Number of employees: Number of employees reported by the firm to the tax authority
- The firm was audited in 2014: equal 1 if the tax administration reported that the firm was audited in 2014.
- Amount of fine following an audit in 2014 (in TJS/1000): amount of fine in 2014 in TJS/1000, as reported in administrative data. Equal zero if the firm was not audited in 2014.

- Risk score in 2014: Risk score calculated by the tax authority in 2014 (see section on Data for a detailed description of the risk score)

Baseline survey data: (self-administered survey)

- Firm has an accountant: the firm reported that an accountant (full-time or part-time) is responsible for tax accounting (as opposed to the firm owner himself).
- Share of technological practices implemented: Share of technological practices implemented by the firm from the following list: "the firm has high speed internet on premises," "the firm uses emails for business communication," and "the firm maintains accounting and tax records in your organization electronically using a specialized program (such as Excel)."
- Number of visits to tax authority office in Jan-Jun 2014: Number of times any employees of the firm visited a tax authority office between January and June 2014.
- Time spent on tax-related activities during a typical month (hours): This variable is the sum of the total amounts of time spent during a typical month (in January-June 2014) on tax report preparation and on visiting a tax office (including travel time).
- Number of times tax inspectors visited the company in Jan-Jun 2014: number reported by the firm owner
- Ever used e-filing (with another company): Equals 1 if the respondent reported that he/she had ever used e-filing to submit a tax document with another company.
- Think that it is common for firms to make informal payments: answered that the following statement is always or often true: "It is common for firms in my line of business to have to pay some irregular 'additional payments/gifts' to tax officials."

E-filing use and registration:

- Used E-filing: equal 1 if the firm used e-filing at any time during the study period.

Firm outcomes from endline survey data:

- Number of visits per month to tax authority office in 2015: sum of all visits reported by the firm owner to any tax office and for any reason.
- Total time spent on tax-related activities by month in 2015 (hours): This variable is the sum of the total amounts of time spent during a typical month in 2015 on the following activities:
  - Collate records
  - Submit tax returns (including travel time)

  - Get the reconciliation act
  - Prepare primary documents used for tax purposes

- Think that it is common for firms to make informal payments: answered that the following statement is always or often true: "It is common for other companies to have to pay some irregular 'additional payments/gifts' to tax officials."

- Tax paid in 2015: Total amount of tax paid in 2015, calculated using the average (declared) amount of tax paid across two focal months of 2015 (June and December) multiplied by 12.

Firm outcomes from administrative data:

- Tax paid in 2015: Sum of all taxes paid in 2015 using monthly administrative data.

- At least one positive tax payment in 2015

- Median of Tax paid in 2015

- Number of positive tax payments in 2015

List Experiment:
Respondents were asked the following question:
"Now, I will read a list of various actions that a company can take to solve or prevent problems with the Tax Administration. After I read the entire list, I would like you to tell me ***how many of these actions your company took in 2015"****.*

50 percent of the sample had to choose among a short list of answers (which did not include unofficial payments) and 50 percent among a longer list (which included unofficial payments). The selection was random and was stratified on firm legal and treatment status.

Short list of answers (50% of sample)	Long list of answers (50% of sample)
1. Received help from trade associations	1. Received help from trade associations
	2. Made unofficial payments or provided free services/goods
2. Had detailed discussions with tax officials	3. Had detailed discussions with tax officials
3. Provided additional documents	4. Provided additional documents
4. Pursued court action	5. Pursued court action

C Cost Effectiveness Analysis Using Time Saved by Firms

Administering the program (organizing the trainings, inviting firms, and providing logistical support for e-filing registration to Group A firms) cost $25 per firm in Group A, compared to $17 per firm in Group B and $12 per firm in Group C (Table A7). Given the 34 percentage point difference in take-up between Group A and the control group, and the $13 per firm difference in program costs, the cost per additional e-filing adoption in Group A relative to the control group is $37. The lack of any significant difference between adoption in Group B and the control group indicates that the relevant aspect of Group A treatment was the logistical help with registration.

The difference in program costs per firm between Group A and Group B (cost of logistical support for registration) is $8 per firm. Given the 30 percentage point difference between Group A and Group B, the cost per additional e-filing adoption in Group A relative to Group B is $27.

We can compare the program costs to the benefits accruing to firms from the reduction in compliance costs. Data limitations preclude our calculating other potential benefits of the program such as savings in tax administration costs. In addition, from the government's perspective, we detect no significant average effects on tax revenue (although any revenue impact would be a transfer from firms to the government). Table 5 estimates that firms save 4.7 hours each month they would otherwise have spent on tax-related activities. From survey data, the average wage of the person in charge of tax declaration in firms is $178 per month (or $1.11 per hour), creating an estimated $5.5 savings per firm each month. As such, it would take five to seven months for private benefits in terms of time saved to exceed program costs. Although firms may not necessarily be willing to pay the full costs of the program,[1] these results provide guidance for a social planner on types of interventions that may be considered in promoting e-filing adoption. Appendix Table A7 provides details on the cost effectiveness analysis.

D Machine Learning Methodology

We follow the methodology described in Chernozhukov et al. (2017) which is based on the double/ debiased machine learning approach developed by Chernozhukov et al. (2018).

A key principle of this methodology is to focus on few important features of the heterogeneity instead of trying to estimate all the possible heterogeneity. This method relies on repeated data splitting between training and test samples. Inference is obtained by taking the adjusted medians of p-values and confidence intervals obtained from multiple data splits which allows taking into account the uncertainty coming from parameter estimation and data splitting. We use this method to answer two questions: first, is there any heterogeneity? And, second, what are the characteristics of the most and least affected

[1] Take-up remained quite low when firms had to pay $40 to register and obtain a token.

groups?

The key steps implemented (based on Chernozhukov et al. (2017)) are as follows:

- We determine which Machine learning method is the most appropriate by computing the best BLP (Best linear predictor) and the best GATE (group average treatment effect). The results are available in Appendix Table A12 and show that the Random Forest method outperform the other methods (Elastic net, Boosting and Neutral Network). For the rest of the analysis, we focus on the Random Forest method which gives more precise results (results are consistent with the other methods but imprecise).

- Using Random Forest, we compute the BLP of impact heterogeneity. The heterogeneity term presented in column (2) of Appendix Table A13 shows that there is significant heterogeneity of impact on total tax paid in 2015.

- We perform classification analysis and compute the average baseline characteristics of the 50 percent most affected and least affected firms defined in terms of the random forest predicted effect. The samples were split into two, with 50 percent each allocated to training and test samples. P-values were computed using medians across 100 splits. The results are available in Appendix Table A14 and show that firms among the 50 percent most affected firms have significantly higher risk score and are more likely to have a risk score above the median.