

Do audits deter future non-compliance? Evidence on Self-Employed Taxpayers

Sebastian Beer, Matthias Kasper, Erich Kirchler,
and Brian Erard

November 11, 2016

Do audits deter future non-compliance?

What happens when you experience an audit?

- ▶ affect subjective probability of future audits?
- ▶ impact on how you perceive the IRS: trustworthy, efficient, service-oriented?
- ▶ inform about tax law (legitimacy of deductions) or indicate the need for tax consultants?

This paper: measures impact of audits on future compliance

Contribution based on two pillars

- ▶ Distinguish between compliant and non-compliant taxpayers, using audit outcome (adjustment)
- ▶ Rely on operational audit data for Schedule C filers

⇒ Identify average treatment effect on the treated (ATT), conditional on audit outcome

Data and baseline definition of experimental groups

Combination of two panel datasets (years 2005 to 2012)

- ▶ operational audit data (type of audit, audit duration, audit outcome, IRS-internal risk scores)
- ▶ granular tax return information on Schedule C filers

Definition of experimental groups

- ▶ No audit ongoing in 2005, 2006, 2007, and 2009
- ▶ Treatment: Audit of TY 2007 started before taxpayer filed return for TY 2008
- ▶ Controls: $N \approx 7000$, Treatment: $N \approx 2000$

Empirical strategy: propensity score matching

ΔY denotes the change in reported income; D treatment assignment. Average treatment effect on the treated is

$$\tau^{ATT} = \underbrace{E[\Delta Y^1 | D = 1]}_{\text{observed}} - \underbrace{E[\Delta Y^0 | D = 1]}_{\text{not observed}}$$

Non-parametric matching estimators use

$$\hat{\tau}^{ATT} = \frac{1}{N^1} \sum_{i:D_i=1} [\Delta Y_i^1 - \hat{m}(\rho_i)]$$

- ▶ where $\hat{m}(\rho_i)$, is a weighted average of control group outcomes
- ▶ with weights determined by similarity between individual i and control group members

1st stage: what triggers audits?

.. and might affect reported income..

Capture difference between control and treatment, using three sets:

Set I: DIF score, total taxable income, profitability in 2007

Set II: Set I + three distance measures (12 income variables, 8 business structure variables, 6 other variables) in 2007

Set III: Set II for 2007 + Set II for 2006 + interaction

Estimate probability of treatment assignment (for both treatments) conditional on controls:

$$\rho_i = P(\widehat{D} | X_{Set,i}) \quad (1)$$

2nd stage: create averages based on similarity

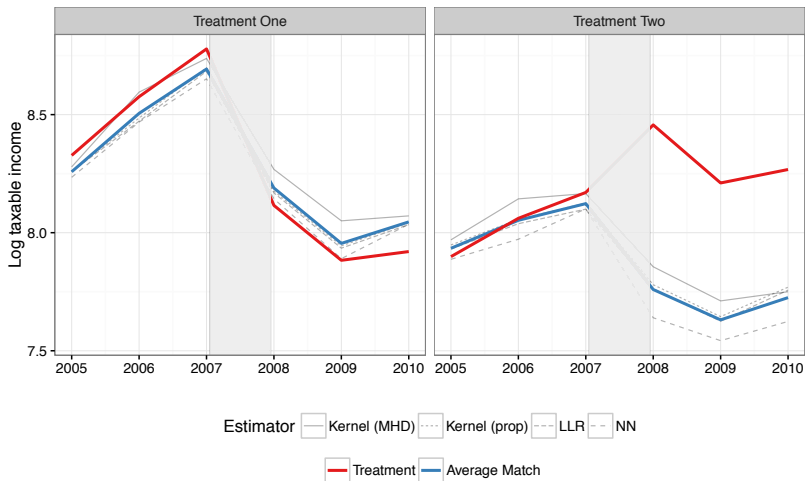
Similarity measures (between individual i and j):

- ▶ propensity score: $s_{ij} = |\rho_i - \rho_j|$
- ▶ Mahalanobis distance: $s_{ij} = [(X_i - X_j)S^{-1}(X_i - X_j)]^{0.5}$

weighting based on

- ▶ Nearest neighbor matching
- ▶ Kernel matching
- ▶ Local linear ridge regression
- ▶ Kernel with Mahalanobis distance

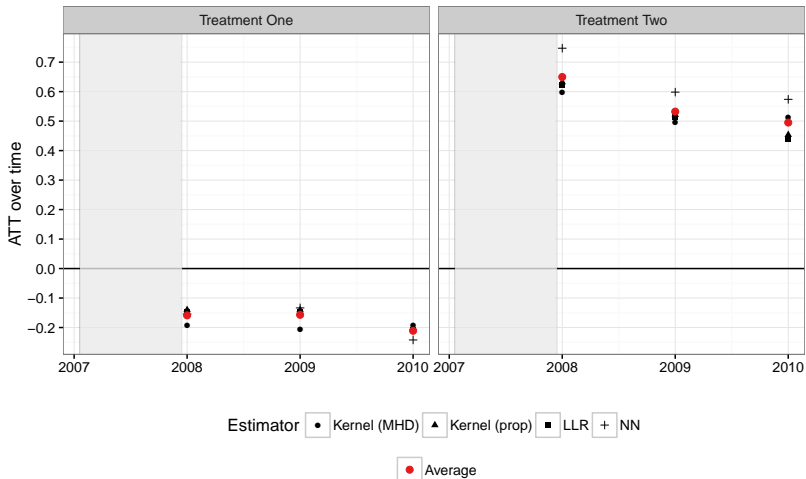
Audit response for compliant (Treatment One) and non-compliant (Treatment Two) taxpayers



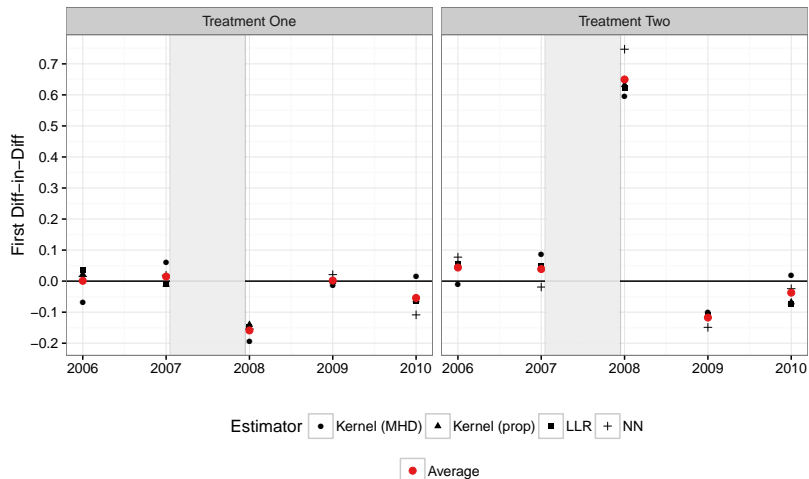
Estimated ATT one year after the audit						
Experimental Group	Treatment Group 1			Treatment Group 2		
Set of control variables	I	II	III	I	II	III
Matching estimator	(1)	(2)	(3)	(4)	(5)	(6)
Nearest neighbor	-0.154* (0.091)	-0.214** (0.094)	-0.153* (0.091)	0.609*** (0.142)	0.666*** (0.130)	0.604*** (0.142)
Kernel (Propensity score)	-0.140* (0.074)	-0.135* (0.074)	-0.149** (0.074)	0.626*** (0.111)	0.656*** (0.106)	0.650*** (0.111)
Local linear ridge	-0.148** (0.075)	-0.140* (0.075)	-0.161** (0.075)	0.622*** (0.112)	0.631*** (0.106)	0.631*** (0.112)
Kernel (MHD)	-0.173** (0.085)	-0.125* (0.075)	-0.136 (0.085)	0.591*** (0.109)	0.664*** (0.098)	0.673*** (0.109)
Average	-0.154** (0.073)	-0.154** (0.074)	-0.150** (0.073)	0.612*** (0.104)	0.654*** (0.100)	0.639*** (0.104)
Model uncertainty	1.24	1.71	1.18	1.06	1.06	1.11

*, **, and *** indicate significance at the 10%, 5% and 1% level. Bootstrapped standard errors in parentheses.

ATT over time



Placebo test



Policy implications, limitations, and scope for future work

- ▶ Risk-based audits preferable to random audits

Limitations

- ▶ What triggers audits, apart from DIF score?
- ▶ How do taxpayers respond during "normal" years?

Future research

- ▶ Drivers of reporting behaviour are in a black box
- ▶ combine administrative and survey data to understand psychological determinants

Thank You!

Sample selection

Number of taxpayers							
Step	Description	Control		Treatment		Total	
		#	% Δ (Step-1)	#	% Δ (Step-1)	#	% Δ (Step-1)
1	Baseline	10,964	-	5,472	-	16,436	-
2	No Audit in 2009	9,560	0.87	4,924	0.90	14,484	0.88
3	TY 2007 audited	9,560	1.00	3,768	0.77	13,328	0.92
4	Right Order	7,278	0.76	2619	0.70	9,897	0.74
5	Outlier	6,922	0.95	2,453	0.94	9,375	0.95

Sample selection - within treatments

Number of taxpayers					
Step	Description	Treatment One		Treatment Two	
		#	Δ (Step-1)	#	Δ (Step-1)
1	Baseline	2939	0.86	2533	0.88
2	No Audit in 2009	2659	0.90	2265	0.89
3	TY 2007 audited	2431	0.91	1337	0.59
4	Right Order	1771	0.73	848	0.63
5	Outlier	1652	0.93	801	0.94

Sample selection

Number of taxpayers					
Step	Description	Treatment One		Treatment Two	
		#	Δ (Step-1)	#	Δ (Step-1)
1	Baseline	2939	-	2533	-
2	No Audit in 2009	2659	0.90	2265	0.89
3	TY 2007 audited	2431	0.91	1337	0.59
4	Right Order	1771	0.73	848	0.63
5	Outlier	1652	0.93	801	0.94

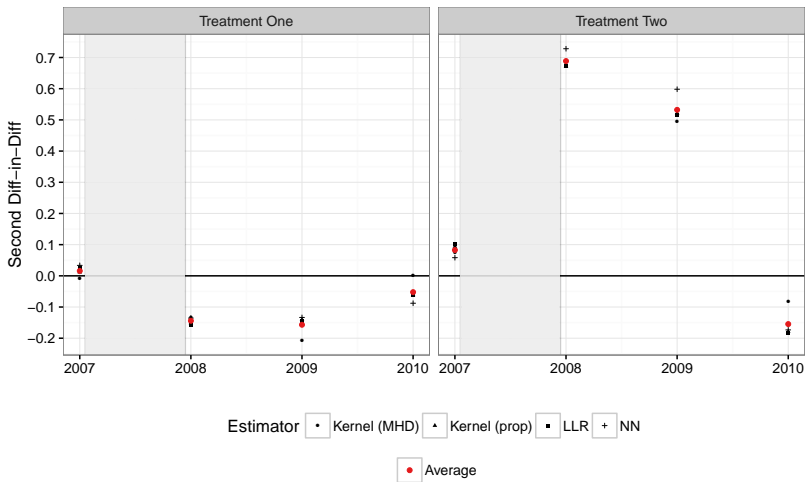
- ▶ 40% of positive audits related to tax returns before 2007
- ▶ results for *Treatment Two* less representative

Estimated ATT one year after the audit						
Experimental Group	Treatment Group 1			Treatment Group 2		
	I	II	III	I	II	III
Set of control variables						
Matching estimator	(1)	(2)	(3)	(4)	(5)	(6)
NN	-0.241** (0.105)	-0.242** (0.108)	-0.220** (0.105)	0.402** (0.166)	0.465*** (0.140)	0.506*** (0.166)
Kernel Prop	-0.201** (0.088)	-0.229** (0.088)	-0.204** (0.088)	0.421*** (0.123)	0.442*** (0.114)	0.450*** (0.123)
Local Ridge	-0.209** (0.088)	-0.240*** (0.089)	-0.213** (0.088)	0.418*** (0.123)	0.368*** (0.117)	0.420*** (0.123)
Kernel MHD	-0.191** (0.091)	-0.161* (0.086)	-0.175* (0.091)	0.515*** (0.119)	0.421*** (0.106)	0.480*** (0.119)
Average	-0.211** (0.085)	-0.218** (0.087)	-0.203** (0.085)	0.439*** (0.120)	0.424*** (0.111)	0.464*** (0.120)

*, **, and *** indicate significance at the 10%, 5% and 1% level. Bootstrapped standard errors in parentheses.

Estimated ATT one year after the audit						
Experimental Group	Treatment Group 1			Treatment Group 2		
	I	II	III	I	II	III
Set of control variables						
Matching estimator	(1)	(2)	(3)	(4)	(5)	(6)
Nearest neighbor	-0.154* (0.091)	-0.214** (0.094)	-0.153* (0.091)	0.609*** (0.142)	0.666*** (0.130)	0.604*** (0.142)
Kernel (Propensity score)	-0.140* (0.074)	-0.135* (0.074)	-0.149** (0.074)	0.626*** (0.111)	0.656*** (0.106)	0.650*** (0.111)
Local linear ridge	-0.148** (0.075)	-0.140* (0.075)	-0.161** (0.075)	0.622*** (0.112)	0.631*** (0.106)	0.631*** (0.112)
Kernel (MHD)	-0.173** (0.085)	-0.125* (0.075)	-0.136 (0.085)	0.591*** (0.109)	0.664*** (0.098)	0.673*** (0.109)
Average	-0.154** (0.073)	-0.154** (0.074)	-0.150** (0.073)	0.612*** (0.104)	0.654*** (0.100)	0.639*** (0.104)
Model uncertainty	1.24	1.71	1.18	1.06	1.06	1.11

*, **, and *** indicate significance at the 10%, 5% and 1% level. Bootstrapped standard errors in parentheses.



Aggregate audit impact

