Pre-Analysis Plan

Using Intelligence from International Tax Cooperation to Improve Voluntary Tax Compliance: Evidence from a Swedish Field Study

Nikolay Angelov[*] Per Johansson[†]

May 4, 2020

---

[*] The Swedish Tax Agency, Uppsala Center for Labour Studies (UCLS) and Uppsala Center for Fiscal Studies (UCFS), nikolay.angelov@skatteverket.se

[†] Department of Statistics at Uppsala University, IFAU, and IZA, per.johansson@statistik.uu.se

1 Introduction

The Swedish Tax Agency receives information about parts of Swedish taxpayers' foreign income through OECD's Common Reporting Standard (CRS). The goal of the CRS is to combat tax evasion and it was approved by the OECD council in 2014. Within the CRS, tax authorities obtain information from financial institutions in their own jurisdiction and automatically exchange that information with other jurisdictions on an annual basis.

The purpose of the present study is to estimate the effect of a low-cost intervention (a digitally disseminated message) on the subsequent compliance of taxpayers who, according to CRS-data, had foreign dividends during the previous tax year.

2 Intervention

The intervention consists of sending the following text to the treated individuals and leaving the controls untreated:

> *Declaring foreign dividends/interest*
>
> Hi!
>
> We get many questions on how to declare foreign income and have therefore developed a new online app in order to make this easier.
>
> The Swedish Tax Agency has obtained information from a foreign tax authority that you have received dividends or interest from abroad during 2018.
>
> If you have received dividends or interest from abroad also during 2019, you can use the online app when you file your taxes. The app will help you with the correct amount to file and how much foreign tax offset you have the right to claim.
>
> You can find the app here:
>
> https://app.skatteverket.se/klient-sifu-segmentering/
>
> Sincerely,
>
> The Swedish Tax Agency

This message is sent in digital form to the treated taxpayers via a digital mailbox. In Sweden, about half of the population above 16 years of age have a digital mailbox.[1] This is a free service making it possible to receive mail in digital form from Swedish authorities and some large private firms. With regards to taxes, having a digital box implies that all communication from the Swedish Tax Agency that otherwise would have been sent as

---

[1] Source: https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2019/digitala-samhallstjanster/halften-av-svenskarna-har-en-digital-brevlada/

paper mail is sent digitally in a secure app. This includes pre-filled tax returns as well as various messages. Taxpayers can file their income tax declaration securely in the app.

The online app mentioned above does not require login and can be best described as a calculator. The user fills in the type of foreign income (dividends or interest), amount, currency, date of receiving the amount, country, and if applicable, the amount of foreign tax paid. Upon clicking *Calculate*, the app converts the amount to SEK, calculates the foreign tax offset, and indicates the specific tax declaration boxes where the amounts should be filed.

Data on the outcome variables (presented below) will be obtained at the earliest on May 5, 2020. The latest day for sending in the income tax declaration is May 4. Although unlikely, it is possible that some of the taxpayers declare some days after the official deadline.

3 Study population and sample

The original population consists of 10,344 individual taxpayers each of whom had foreign dividends summing to over 3,000 SEK (about 292 EUR) during tax year 2018, according to information from the CRS. For internal Tax Agency reasons, although the direct cost of the treatment is low, the number of treated was limited to 500 individuals at an early planning stage. For this and other reasons, the original sample was reduced according to the following steps:

1. Some of the individuals had not filed a tax declaration for tax year 2018 and were therefore removed from the sample (259 obs removed)

2. Individuals who did not have a digital mailbox on February 27, 2020, were removed (5,730 obs removed)

3. Observations with respect to several pre-experiment variables for the tax year of 2018 were defined as outliers and removed if the value of any single variable was larger than the third quartile plus 1.5 times the interquantile range (IQR) or smaller than the first quartile minus 1.5×IQR. The quartiles for all variables were calculated on the same data, i.e., after the previous step and before removing any outliers. The following variables were used for outlier detection (with *variable label* and number of removed observations in parentheses):

    - Foreign dividends (*fdiv*, 638)
    - Earnings including labor income, sick pay, pension, etc. (*earn*, 355)
    - Capital income (*capinc*, 725)
    - Foreign dividends as a share of capital income ($fdivrat = \frac{fdiv}{capinc+1}$, 513)

- Total tax paid (*tax*, 439)

The unique number of individuals removed due to outliers amounts to 1,546.

4. Individuals aged below 30 and above 75 were removed (112)

After these steps, there were 2,697 individuals left in the sample.

Our main outcome variable is capital income (*capinc*), where the amount of foreign dividends should be included. In addition, we plan to estimate the effect on total tax paid (*tax*). An increase in declared foreign dividends in *capinc* should lead to an increase in total tax paid unless no offsetting adjustments are made in the income tax declaration.

4 Experimental design

Balanced designs, that is where the number of treated and controls is equal, are preferable to unbalanced designs in both Fisher (Chung and Romano, 2013) and Neyman-Pearson (Freedman, 2008) inference. For this reason, we decided to equalize the number of treated and controls, meaning that the sampling frame is set to 1,000 individuals.

Even though the estimators from well-conducted experiments are unbiased in expectation, the estimates from any single experiment may still be far from being unbiased (in the colloquial sense) due to an unlucky, albeit random, allocation. For this reason the experimental design is stratified on gender and historical tax compliance; two variables that we believe can be important determinants for the two outcomes under investigation. A second reason for stratification is that we are interested in testing for group differences in behavior as a consequence of the intervention. As we also have continuous covariates (historical data on the outcomes, earnings, age, etc.) we also would like to balance them within each stratum.

To this end, we used the rerandomization stategy suggested by Morgan and Rubin (2012). The idea is to remove from consideration allocations with imbalance in observed covariates between treated and control units and then randomize within the set of allocations with balance on these covariates. Call the set of all allocations $\mathcal{A}$ and the set of acceptable allocations $\mathcal{A}_a$. As in Morgan and Rubin (2012) the Mahalanobis distance is used as the criterion for choosing $\mathcal{A}_a$. The four strata are formed by the interaction of the two binary variables $\mathbf{1}[woman]$ and $\mathbf{1}[compliant]$, where $\mathbf{1}[\cdot]$ is the indicator function. We do not have access to a perfect measure of previous compliance. As a proxy, we use information about whether the amount of foreign dividends obtained from the CRS was less than or equal to total capital income, i.e., $\mathbf{1}[compliant] \equiv \mathbf{1}[fdiv \leq capinc]$. The logic behind this is that for compliance with the tax code, the amount of $fdiv$ should be included along with other capital income sources in the declared *capinc*. Therefore, although $fdiv \leq capinc$ is not necessarily a sign of compliance, $fdiv > capinc$ is a clear measure of non-compliance.

Using the 2,697 individuals left in the sample after the procedure used in the previous section, we proceed as follows:

1. Divide the sample in a compliant and a non-compliant group (number of observations within parentheses):
    (a) $\mathbf{1}[compliant] = 1$ (1,759)
    (b) $\mathbf{1}[compliant] = 0$ (938)
2. Draw two simple random samples, each of size 500, from (a) and (b) respectively. These 1,000 individuals constitute the sampling frame of the trial.
3. Create four strata defined by the interaction $\mathbf{1}[woman] \times \mathbf{1}[compliant]$
4. Within each stratum, we randomly select an allocation with a Mahalanobis distance between treated and controls means of six covariates to be less than 0.17. (Note: $Pr\left(\chi^2(6) < 0.17\right) = 0.0001$.) In practice, this is achieved by randomly drawing new allocations until the criterion is fulfilled. The following covariates measured for the pre-study income year 2018 were used in the rerandomization:
    - age: the taxpayer's age
    - $fdiv$: see above
    - $capinc$: see above
    - tax: see above
    - $earn$: see above
    - $finc \equiv \mathbf{1}[has\ foreign\ income]$: categorical variable based on a check box in the tax declaration

Table 1 shows group means by stratum after performing the rerandomization as well as the resulting number of observations in each stratum.

Table 1: Group averages after rerandomization

Stratum		*age*		*fdiv*		*capinc*		*tax*		*earn*		*finc*		*#obs*	
$\mathbf{1}[compliant]$	$\mathbf{1}[woman]$	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*	*C*	*T*
0	0	50.93	50.57	7.46	7.39	1.81	1.82	248.59	253.11	615.49	625.63	0.08	0.08	167	167
0	1	51.86	51.58	6.51	6.48	1.58	1.52	207.97	206.92	548.71	544.53	0.07	0.06	83	83
1	0	52.75	52.65	6.35	6.35	11.21	11.23	339.85	341.63	829.46	829.56	0.06	0.07	200	200
1	1	51.44	51.90	7.63	7.76	13.26	13.85	306.35	310.08	769.90	774.46	0.18	0.18	50	50

Note: C denotes controls and T denotes treated individuals. The measure units of the variables are as follows: *age* is measured in years, $finc$ is valued 0 or 1, *#obs* denotes number of observations, and the rest of the variables are expressed in 1,000s SEK which is approximately equal to 97 EUR.

5 Estimation and hypotheses

A drawback with the rerandomization strategy is that the mean difference estimator no longer is asymptotically normally distributed (Li, Ding, and Rubin, 2018). However, Li and Ding (2019) showed that standard asymptotic inference can be conducted using ordinary least squares (OLS) by regressing the outcome on the covariates and the treatment together with the Eicker-Huber-White (EHW) robust standard error estimator (Eicker, 1967; Huber,1967; White, 1980) for the inference. To be specific, let $\boldsymbol{x}_i$ be the covariate used in the Mahalanobis distance for individual i, W_i be the treatment indicator and Y_i the outcome. The treatment effect is the estimated coefficient on W_i in the ordinary least squares (OLS) regression of Y_i on W_i, $\boldsymbol{x}_i$ and $W_i(\boldsymbol{x}_i - \bar{x})$. To construct asymptotically valid confidence intervals one should use the EHW robust standard error estimator. However, the results from Zhang and Johansson (2019) suggest that for sample sizes below 400 the EHW estimator is downward biased. The HC2 covariance matrix estimator is shown to have good small-sample performance. In this estimator the OLS residuals $\widehat{u}_i$ from the EHW covariance matrix are replaced with $\widehat{\epsilon}_i = \widehat{u}_i/\sqrt{1-h_i}$, where h_i is the ith diagonal element of the projection matrix (for details see MacKinnon, 2013).

A balanced design simplifies the analysis and the tests for effect differences across strata. The reason is that the four effects estimated in the saturated linear regression model are the same as the four between-group differences in mean estimates, which are unbiased (Schultzberg and Johansson, 2019). This holds also in a rerandomization design.

The analysis will be conducted by estimating

$$\begin{aligned} Y_i =& \alpha_0 + \tau W_i + \alpha_w \mathbf{1}[woman] + \alpha_c \mathbf{1}[compliant] \\ &+ \alpha_{wc}\mathbf{1}[woman] \times \mathbf{1}[compliant] + \boldsymbol{\beta}_0' \widetilde{\boldsymbol{x}}_i + \boldsymbol{\beta}_1' W_i \widetilde{\boldsymbol{x}}_i + \varepsilon_i, \end{aligned} \tag{1}$$

where $\widetilde{\boldsymbol{x}}_i \equiv \boldsymbol{x}_i - \bar{\boldsymbol{x}}$. Asymptotic inference will be performed using the HC2 covariance matrix.

We have two outcomes (*capinc* and *tax*) and the test for overall effect for each outcome is $H_0 : \tau = 0$ against the alternative $H_1 : \tau > 0$. We let the overall risk level for judging whether we have an effect or not to be 5% which means that each single test will be conducted at the 2.5% risk level. If H_0 is rejected, we will test for differences in behavior across groups by estimating

$$\begin{aligned} Y_i =& \alpha_0 + \tau W_i + \alpha_w \mathbf{1}[woman] + \alpha_c \mathbf{1}[compliant] + \alpha_{wc}\mathbf{1}[woman] \times \mathbf{1}[compliant] \\ &+ \tau_w \mathbf{1}[woman] W_i + \tau_c \mathbf{1}[compliant] W_i + \tau_{wc}\mathbf{1}[woman] \times \mathbf{1}[compliant] W_i \\ &+ \boldsymbol{\beta}_0' \widetilde{\boldsymbol{x}}_i + \boldsymbol{\beta}_1' W_i \widetilde{\boldsymbol{x}}_i + \varepsilon_i. \end{aligned} \tag{2}$$

Testing for effect heterogeneity across groups will be performed as a standard F-test, i.e. $H_0: \ \tau_c = \tau_w = \tau_{wc} = 0$ at the 5% level. Rejecting H_0 means that at least one of the parameters is different from zero. We will provide point estimates along with confidence intervals for all parameters, and if H_0 is rejected we will discuss the sign and magnitude of the point estimates in the text.

References

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, volume I, pages 59–82. University California Press, Berkeley, CA.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, volume I, pages 221–233. University California Press, Berkeley, CA.

MacKinnon, J. G. (2013). Thirty years of heteroskedasticity-robust inference. In Chen, X. and Swanson, N. R., editors, Recent Advances and Future Directions in Causality, Prediction, and Specication Analysis: Essays in Honor of Halbert L. White Jr., pages 437–461. Springer, New York.

Zhang, J. L.; Johansson, P (2019). A Comparison of Methods of Inference in Randomized Experiments from a Restricted Set of Allocations Working Paper 2019:5 Department of Statistics

White, H. (1980). Using least squares to approximate unknown regression functions. International Economic Revi, 21(1):149–170.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. Annals of Statistics, 41(2):484–507.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. Annals of Statistics, 40(2):1263–1282.

Li, X. and Ding, P. (2019). Rerandomization and regression adjustment. To appear in Journal of the Royal Statistical Society, Series B.

Li, X., Ding, P., and Rubin, D. B. (2018). Asymptotic theory of rerandomization in treatment Control experiments. Proceedings of the National Academy of Sciences of the United States of America, 115(37):9157–9162.

Freedman, D. A. (2008a). On regression adjustments to experimental data. Adv. in Appl. Math. 40 180–193.

Schultzberg, M. and Johansson, P. (2019) “Rerandomization: a complement or substitute for stratification in randomized experiments?”, Working Paper 2019:4, Department of Statistics, Uppsala University.