CESifo Working Papers

www.cesifo.org/wp

A Gift is not Always a Gift:
Gift Exchange in a Voucher Experiment

Sascha O. Becker
Dolores Messer
Stefan C. Wolter

CESIFO WORKING PAPER NO. 3488
CATEGORY 13: BEHAVIOURAL ECONOMICS
JUNE 2011

CESifo *Working Paper No. 3488*

A Gift is not Always a Gift: Gift Exchange in a Voucher Experiment

Abstract

Different from traditional gift exchange experiments, we study a field experiment where a random subsample of participants in the Swiss Labor Force Survey was sent vouchers to be used in adult training courses. Importantly for our purposes, actual voucher redemption can be traced. This gives the unique opportunity to study whether gift exchange in the form of participation in future rounds of the survey depends on the perceived usefulness of the gift. We find that the group of voucher recipients as a whole has significantly higher response rates in the survey six months after the vouchers were sent out. There is considerable heterogeneity, though. Our results point to a long-lasting gift exchange relationship for the sub-group that had redeemed their vouchers. Contrary to this group, the individuals who did not redeem their vouchers, had a response pattern that was not significantly different from the voucher non-recipients.

JEL-Code: C420, C930, Z130.

Keywords: gift exchange, reciprocity, field experiment, long-run effects.

Sascha O. Becker
Department of Economics
University of Warwick
Coventry / United Kingdom
s.o.becker@warwick.ac.uk

Dolores Messer
Department of Economics
University of Bern
Bern / Switzerland
dolores.messer@vwi.unibe.ch

Stefan C. Wolter
Department of Economics
University of Bern
Bern / Switzerland
stefan.wolter@vwi.unibe.ch

May 30, 2011
We thank Florian Englmaier, Armin Falk and participants at various seminars and conferences for useful comments and discussions. The authors thank the Swiss Federal Office for Professional Education and Technology for generous financial assistance and the Swiss Federal Statistical Office for data provision and the permission to use the Swiss Labor Force Survey. The usual disclaimer holds.

1 Introduction

In January 2006, a random subsample of 2,400 individuals who had participated in the year 2005 round of the Swiss Labor Force Survey (Swiss LFS) was sent a letter from the Swiss Federal Statistical Office. The letter, signed by the Director General of the Swiss Federal Statistical Office, contained a voucher that could be redeemed in any training activities of the recipient's choice. The letter stated that the voucher was part of a project to promote lifelong learning by the Federal Office for Professional Education and Training and that participants of the Swiss LFS were particularly well suited to receive this gift. The voucher could be redeemed for any ongoing or future training activity under the condition that the training started before the end of May 2006. In order to redeem their vouchers, recipients had to send proof of payment for training activities together with their voucher to the Swiss Federal Office for Professional Education and Technology. The take-up behavior of vouchers for participation in training activities is analyzed elsewhere (see Messer and Wolter, 2009) and shows that voucher recipients had an almost 20% higher training incidence than non-recipients.

Our interest in this paper is the survey response behavior in the Swiss Labor Force Survey in the summer 2006 and 2007, i.e. after the end of the voucher experiment. The Swiss LFS is structured according to a rotating panel principle in which the respondents are interviewed for five consecutive years. As a result, about one-fifth of the respondent population is replaced every year. Participation in the Swiss Labor Force Survey is voluntary and past respondents are not obliged to answer in subsequent rounds of the survey. Since the voucher was explicitly linked to past survey participation, voucher recipients might have been tempted to reciprocate towards the Swiss Federal Statistical Office by further active survey participation. It is important to note that the voucher experiment was *not* aimed at increasing survey participation,[1] so increased survey participation is an incidental by-product of the voucher experiment analyzed in Messer and Wolter (2009).

The advantage of our setup is that it constitutes a natural experiment, not an artificial experiment explicitly designed for the purpose of studying gift exchange. Our field data rather flows naturally from an ongoing survey and participants are not aware of being part of an experiment. Furthermore, we are outside a classical employer-employee setting with a formal contract between two parties, where the agent might perceive a wage increase (or a non-monetary gift) as an

[1] We discuss the literature on survey participation in Section 4.2.4.

invitation to enter a multi-stage game in which a worker's productivity increase will elicit further wage increases. In contrast, participants in the Swiss Labor Force Survey face a rather anonymous counterpart in form of the Swiss Federal Statistical Office and have no contractual obligation to participate in the survey. Higher participation rates for voucher recipients in future survey rounds are thus likely to constitute cleaner evidence on gift exchange because strategic motives can largely be excluded. In fact, we exploit the rotating panel nature of the Swiss Labour Force Survey and show that participants who are rotating out of survey participation have higher participation rates in their last survey round although they cannot expect to be contacted again.

A growing literature studies the relevance of reciprocity and gift exchange to understand patterns of interaction (see Fehr and Gächter, 2000, for an early overview). Initially, these issues were studied in laboratory experiments (e.g. Fehr, Kirchsteiger, and Riedl, 1993, Fehr, Gächter, and Kirchsteiger, 1997). Recently, several papers have analyzed gift exchange in field experiments. Gneezy and List (2006) study gift exchange at the workplace where in the "gift condition" wages were higher than previously announced (20 Dollars instead of 10 Dollars). They report that initially output is higher in the gift condition than in the no gift condition, but that over the work period of six hours the gift exchange relation eventually breaks down. Their study thus points to relatively short-lived gift exchanges.[2] Our study allows us to follow survey participants 6 months and 18 months after the initial gift (the training voucher). We find gift exchange even after 18 months for the group that had redeemed their vouchers, pointing to rather long-run effects.

In experiments that involve monetary gifts, e.g. higher wages in labor market experiments, the implicit assumption is that recipients attach a value to the gift. In settings with non-pecuniary gifts,[3] it is usually not possible to see whether recipients appreciate the gift. Falk (2007) studied the response to the Christmas mailing of a charity soliciting money, where some potential donors were randomly sent a "small" or a "large" gift. The small gift was one postcard plus envelope,

[2] Similarly short-lived effects are found by Bellemare and Shearer (2009) in a field experiment in a tree-planting firm. In contrast, Hennig-Schmidt, Rockenbach and Sadrieh (2009) do not find any productivity effect of an unanticipated pay rise for students performing field work as typists.

[3] Differences between pecuniary and non pecuniary gifts have also been tested in field experiments. Al-Ubaydli et al. (2008) find in a workplace experiment that reactions to both types of incentives are rather similar. However, Kube, Maréchal and Puppe (2010) find that a non-pecuniary gift shows a substantial increase in workers' productivity, whereas the reaction to a cash gift is ineffective although workers favour to receive the non-pecuniary gift's cash equivalent.

while the large gift consisted of a set of four postcards with four envelopes. The postcards showed colored paintings drawn by children. He finds gift-exchange to matter: the donation probability is lowest in the no gift condition, higher in the small gift condition, and highest in the large gift condition.

This design might hide further important heterogeneity because some members in the treatment group might not have even opened the letter from the charity or might have disposed of the postcards right away, not *perceiving* them to be a gift. The donation probability is likely to vary by the perception of the *intended* gift as a gift. Falk (2007) notes in his conclusion: "Ultimately the successful initiation of a gift-exchange relation depends on *attribution*, that is, how kind, generous, or fair a particular action or gift is *perceived* by the recipient.[4] A unique feature of our experimental design is that our setup allows us to distinguish between recipients that actually redeemed their training voucher and those who did not. We can thus analyze the degree of gift exchange as a function of the *perceived utility* of the gift, and we will be able to show that this distinction significantly alters the way one has to interpret the results of a gift exchange experiment.

Our paper thus contributes to the literature in several respects: first, and most importantly, the voucher experiment allows us to study reciprocity in an experimental setting in the field. It thus adds to a small but growing literature studying gift exchange outside standard laboratory experiments. Second, different from the previous literature, our experimental design allows us to study the long-run effects of gift exchange from "memorable" gifts. Third, we address the issue of *attribution* by analyzing differences between voucher recipients who redeem their training vouchers and those who do not.

Empirically, a challenge arises from the fact that voucher recipients who redeem their voucher and those who do not might systematically differ. In other words, whereas our experimental setting ensures that voucher recipients and non-recipients do not systematically differ, there is likely to be self-selection into training participation. To address this issue, we pursue a careful matching procedure to identify suitable control observations from the group of voucher-non-recipients. Furthermore, to address the issue of unobserved heterogeneity between the treat-

[4]In a similar vein, potential survey respondents are routinely given financial incentives to increase participation (see Laurie and Lynn 2008). Financial incentives typically take the form of banknotes sent along in the envelope containing the questionnaire or may come as vouchers that can be redeemed in high street stores. For instance, the British Household Panel Survey (BHPS) sends respondents £10 gift vouchers as a token of thanks (see Laurie 2007). These vouchers can be redeemed in 19,000 stores across Britain, the implicit assumption being that the vouchers are equivalent to cash and cash enters people's utility function positively, but their actual use is not followed up.

ment and control groups, we perform a statistical bounding analysis to assess how influential unobserved factors would have to be to overturn our findings.

Our results show that voucher recipients have a significantly higher response rate than non-recipients by 5 percentage points in the first survey round after vouchers were sent out, but no effect is found in the second survey round 18 months after the initial gift. However, the results change remarkably once we distinguish between voucher recipients who do not redeem their vouchers, and voucher recipients who do redeem their vouchers. Voucher recipients who redeemed their vouchers before the expiry date have a significantly higher response rate than non-recipients by 25 percentage points in the first survey round and by 14 percentage points in the second survey round, pointing to long-lasting gift exchange effects six and eighteen months after the initial gift. In contrast, survey respondents who received the same training gift vouchers, but did not redeem them, show no difference in survey response rates compared to non-recipients. This suggests that the *perception* of the initial gift matters for the degree of gift exchange.

Our results might extend beyond our particular setting. Further training plays an important role in employer-employee relations and some firms consider employer-financed training as a gift substituting for a monetary bonus, or count on high degrees of reciprocity of those workers they invested in (see Leuven et al. 2005). If one accounts for the insight created by our experiment, workers' productivity might react quite differently depending on the perceived usefulness of the offer of continuous training.

The remainder of the paper is organized as follows. In the next Section, we present the details of the field experiment. Our empirical strategy is described in Section 3, results are contained in Section 4, and Section 5 concludes.

2 The experimental design

In 2005 the Swiss government mandated the Centre for Research in Economics of Education at the University of Bern to conduct a large scale, randomized field experiment with vouchers for adult education. In order to get a sufficient number of individuals for the experimental and for the control group, the research team had the chance to benefit from a planned reduction in the sample size in the Swiss Labor Force Survey. The Swiss LFS sample population had been raised significantly in the year 2002, but was scheduled to be reduced in subsequent years because of financial constraints. The Swiss LFS is structured according to a rotating panel principle in which the respondents are interviewed for five consecutive years. The financially induced reduction in the sample yielded the opportunity to select a random sample of 2,437 individuals who would otherwise have been dropped from future LFS rounds. For the purpose of the experiment, they were interviewed as if they continued to be part of the Swiss LFS. All individuals in the experimental group had been interviewed in the 2005 Swiss LFS, and most of them had also been interviewed in former years.

The experimental group was matched by a control group of 17,234 individuals who were interviewed by the Swiss LFS in 2005 and were scheduled to be interviewed again in 2006. The experimental design enables the use of longitudinal and cross-sectional information.

The only limitation that was imposed on the experimental sample refers to their age. Only subjects aged 20 to 60 were entitled to receive vouchers, since under-20s would be likely to be still undergoing education or training, and over-60s would be likely to be retired pensioners. There were no restrictions concerning employment status, as increasing the skills of non-employed and employed individuals was potentially a goal of the voucher policy.

The 2,437 randomly selected individuals received a letter from the Swiss Federal Statistical Office during the first days of January 2006 containing an adult education voucher (see Appendix). The letter stated that the voucher was part of a project to promote lifelong learning by the Federal Office for Professional Education and Training and that participants of the Swiss LFS were particularly well suited to receive this gift. In order to eliminate any doubts as to the legitimacy of the voucher, the letter was signed by the Director General of the Swiss Federal Statistical Office. No public-domain information was generated during the experimental period, to ensure that voucher recipients were unaware that the dispensing of the voucher was part of a field experiment.[5]

The voucher group was further split up according to the face value of the voucher. Vouchers had face values of 200 CHF, 750 CHF or 1,500 CHF (see Table 1), quite substantial amounts compared to other gifts used in the literature like postcards or 10 Pound banknotes. In our analysis, we look into the effect of voucher receipt as such, but can also analyze whether the face value of the voucher matters for the degree of gift exchange.

We exclude from our sample those survey participants that the interviewing agency was not able to locate anymore due to change of name, migration or death.

[5]The aspect that participants were unaware that they were participating in an experiment – usually difficult to achieve – is common practice in field experiments; see, for example, Gneezy and List (2006).

Table 1: Division of the experimental group (number of observations)

Face value: Advice	200 CHF[a]	750 CHF	1,500 CHF	Total
Yes	408	407	404	1,219
No	407	407	404	1,218
Total	815	814	808	2,437

Source: Messer and Wolter (2009).

[a] 1 CHF is equivalent to about 0.93 USD or 0.66 EUR.

Although sample attrition due to these reasons should affect the experimental and the control group randomly and therefore similarly, there might be non-random differences within the experimental group. The reason for voucher non-redemption, e.g. migration, might have been exactly the same as the one that lead to the sample attrition. Therefore, excluding those participants prevents a potential bias. Out of the 2,437 individuals receiving adult education vouchers, 354 could not be contacted anymore. In the control group 1,985 addresses out of the initial 17,234 could not be found anymore or were inactive.

Furthermore, for organizational reasons, the interview period was not identical for voucher-recipients and non-recipients. Voucher recipients were first contacted in June in order to guarantee a maximum voucher redemption period in the year 2006. Due to the large sample size of the Swiss LFS, the sample agency split the main LFS sample randomly into two groups for the interview every year. Half of the voucher non-recipients were first contacted in April and the remaining non-recipients were first contacted in May. They were all contacted at least once within the first two weeks of the respective month and there were further attempts to establish contact with voucher recipients during a maximum period of 70 days and 90 days with non-recipients. As the Swiss summer vacation lasts from July to mid-August, the Swiss LFS interview period overlapped with the summer vacation. Only the first month of the voucher recipients' interview period was prior to the vacation while in case of non-recipients either the whole interview period or two months did not overlap with the vacation.[6] To guarantee comparability of the

[6]Figure 1 (see Appendix) shows the cumulative response rates of voucher-recipients, non-recipients and of their matched controls in the years 2006 and 2007. The cumulative response rate of voucher-recipients persisted on a higher level than the rate of their matched controls, whereas for non-recipients the cumulative response rate falls below the rate of their matched controls after the first month since the beginning of the interviewing period. This is because the response rate in the remaining period is influenced by the fact that the interviewing of the voucher recipients is done during the period of the national summer holidays, whereas the interviewing of most of the control group had already been completed by then.

response rates, we limit the observed contacting period for our subsequent analysis to the first month of interviews. Consequently, responses are coded 1 if the interview was successfully administered within the first four weeks of interviews and zero otherwise.

The response rate of voucher recipients was 71.9% in the year 2006 and 64.7% in 2007, respectively. 66.6% of voucher non-recipients responded in the year 2006 and 63.3% in 2007, respectively.

3 Empirical Strategy

As vouchers were randomly assigned to survey participants, the comparison of the survey response rates between the treatment group (voucher receivers) and the control group (voucher non-receivers) establishes experimental evidence on gift-exchange. We are further interested in the difference in the degree of reciprocity between voucher recipients who appreciated the gift (voucher redeemers) and those who made not use of it (voucher non-redeemers) and therefore split the treatment group into these two distinct groups. However, voucher redeemers and non-redeemers might differ systematically. There is indeed clear evidence of heterogeneity between voucher recipients who redeemed their vouchers and those who did not, as Table 2 shows. For instance, females, holders of a non-academic tertiary degree, and training participants in the year before the voucher receipt are more likely to redeem their vouchers.

This heterogeneity in characteristics of those who redeemed their vouchers (*voucher & redeem* group) and those who did not (*voucher & don't redeem* group) implies that the comparison of the random control group with the combined treatment group does not give experimental evidence on gift exchange, as individuals are not randomly allocated to the two sub-groups of the treatment group. Raw comparisons of survey participation rates in 2006 and 2007 between the *voucher & redeem* group and the *full* control group will therefore not correctly measure the degree of reciprocity of those who appreciated the gift because the control group is made up of both types. The same is true for raw outcome comparisons between the *voucher & don't redeem* group and the *full* control group.

In order to ensure comparability between individuals in the two disjunct treatment groups and those in the control group, we use propensity score matching (see Rosenbaum and Rubin 1983; for a survey, see Caliendo and Kopeinig 2008). Matching serves to select, in two separate analyses, suitable control groups for the two treatment groups. While the Appendix gives a more detailed overview of

Table 2: Descriptive Statistics

	Voucher redeemed		Voucher not redeemed	
	mean	s.d.	mean	s.d.
Age	40.91	9.89	41.94	10.35
Indic.: Female	.64	.48	.54	.50
Indic.: Swiss nationality	.91	.29	.88	.32
Indic.: Course participation in 2005[a]	.62	.47	.41	.49
Indic.: Non-employed	.14	.35	.17	.38
Indic.: French/Italian speaking area	.22	.41	.30	.46
Indic.: Education: Compulsory school	.07	.26	.15	.36
Indic.: Education: Vocational training	.44	.50	.52	.50
Indic.: Education: Matura[b]	.11	.31	.09	.29
Indic.: Education: Non-academic tertiary degree[c]	.25	.43	.15	.35
Indic.: Education: University	.13	.34	.10	.30
Indic.: Place of residence: city	.24	.43	.24	.42
Indic.: Place of residence: suburban area	.51	.50	.44	.50
Indic.: Place of residence: rural area	.25	.43	.32	.47
Number of Observations	427		1,656	

Data: Swiss Labor Force Survey, 2005 and 2006, and experimental data. Indicator variables (Indic.) take a value of one if the described condition is satisfied.

[a] Course participation in the year before voucher assignment.

[b] University-entry certificate.

[c] Degree at university of applied sciences or professional education and training at tertiary level.

propensity score matching, here we concentrate on the essential features.

Intuitively, for individuals in the *voucher & redeem* group, suitable controls will be those who would have the same (or very close) propensity of redeeming their voucher should they receive one.[7] Similarly, for individuals in the *voucher & don't redeem* group, suitable controls will be those who would have the same (or very close) propensity of *not* redeeming their voucher should they receive one.

We then estimate the average treatment effect on the treated (ATT), i.e. we compare the average outcome in the *voucher & redeem* group to the average outcome of those controls that have been *matched* to the *voucher & redeem* group. Similarly, we compare the average outcome in the *voucher & don't redeem* group to the average outcome of those controls that have been *matched* to the *voucher & don't redeem* group. These outcome differences are average treatment effects on the treated (see the Appendix for details).

Propensity-score matching removes the bias of non-random selection into treatment by comparing outcomes between treated and control units that are initially

[7]We employ single nearest-neighbor matching, i.e. we match treated individuals to that control unit with the closest propensity score.

identical and undergo treatment (receive a voucher). A crucial assumption is that observable covariates exhaustively determine selection into treatment. Since receipt of the voucher is randomized, this assumption is equivalent to assuming that redemption of the voucher is exhaustively determined by observed covariates. The wealth of information in our data—individual characteristics such as demographic, education and work variables, as well as controls for region of residence—comprehensively covers the pre-treatment conditions so that the assumption of selection on observables is not unreasonable. Still, to address remaining worries of selection on unobservables, we will calculate so-called Rosenbaum (2002) bounds to investigate how strong unobservable factors would have to be to overturn our results. That bounding analysis will show that unobservables would have to be unreasonably influential in order to change our findings. We discuss these extensions in Appendix A.4.

4 Results

Let us first consider the situation in which – similar to the existing literature – we would not have been able to observe the redemption of the vouchers. In this experimental situation, we would have only been able to compare the combined treatment group (all voucher recipients) with the control group (all voucher non-recipients). The results (see Table 3) show a significant average treatment effect on the treated (ATT) for all voucher values combined (first line) in the year 2006 survey, i.e. 6 months after the vouchers were sent out, and no significant effects in the year 2007 survey, i.e. after 18 months. We might interpret these results as evidence of gift exchange in the medium run (6 months after the original gift), but no gift exchange in the long-run (after 18 months).

Looking at different subgroups of voucher recipients by voucher value, these results are driven by a strong response by those receiving vouchers with high face value (750 or 1500 CHF) with no effect on those receiving a 200 CHF voucher.

Lacking knowledge about voucher redemption, we would not have been able to get further insight into differences in gift exchange behavior depending on how the voucher recipient values the gift.

We will now show that, based on the evidence for the combined treatment group, the interpretation of the effect size, of the duration of reciprocity as well as of strategic behavior would have been inadequate. The reason is that there is a significant difference in the pattern of reaction to the vouchers between the group of individuals who redeemed their vouchers and those who did not.

Table 3: AVERAGE TREATMENT EFFECT ON THE TREATED

	2006		2007	
	ATT	Std. Err.	ATT	Std. Err.
Treatment group: voucher received				
Voucher	.053	.015**	-.003	.018
Voucher 200	.010	.025	.034	.031
Voucher 750	.072	.025**	.026	.029
Voucher 1500	.064	.025*	-.004	.029

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data.
Note: ATT denotes the average treatment effect on the treated. See main text for details.

4.1 The importance of accounting for heterogeneity

We present results separately for voucher recipients that redeem their voucher (*voucher & redeem* group) and voucher recipients that do not redeem their vouchers (*voucher & don't redeem* group). Table 4 displays odds ratios and corresponding standard errors of logit propensity score estimates where the dependent variable is equal to one for voucher recipients that do redeem their vouchers. The control group is composed of all voucher non-recipients. The results indicate that older workers, women, those with education beyond compulsory education, those having participated in a course the year before the voucher assignment and from the German-speaking areas of Switzerland are more likely to redeem their vouchers.

We also run separate specifications according to voucher value. In all specifications, the control group *before* matching is composed of all voucher non-recipients.[8]

Before proceeding to the estimation of average treatment effects, we assess matching quality, i.e. we assess whether propensity score matching was successful in selecting individuals from the control group that are good matches to individuals in the *voucher & redeem* and *voucher & dont't redeem* group, respectively. Our two indicators of matching quality, the pseudo- R^2 of the propensity score logit regression and the median absolute standardized bias, indicate that observable characteristics of treated and *matched* control observations are well balanced after propensity-score matching (see Appendix for details).

A further interesting finding supporting our matching strategy is that the cumulative survey response rates of the two *matched control* groups (see Appendix, Figure 1) are nearly identical. This implies that – after controlling for observable

[8]To save on space, we do not display the further propensity score specifications.

Table 4: Propensity Score Specification

Dependent variable: voucher received and redeemed	Odds Ratio	Std. Err.
Age	1.023	.042
Age-squared	1.000	.0005
Indic.: Female	1.705	.183**
Indic.: Swiss nationality	1.364	.240†
Indic.: Course participation in 2005[a]	1.687	.178**
Indic.: Non-employed	.992	.149
Indic.: French/Italian speaking area	.483	.058**
Education variables: reference category "compulsory schooling"		
Indic.: Education: Vocational training	1.590	.324*
Indic.: Education: Matura[b]	2.094	.508**
Indic.: Education: Non-academic tertiary degree[c]	2.755	.609**
Indic.: Education: University	1.964	.470**
Residence variables: reference category "rural area"		
Indic.: Place of residence: city	.871	.127
Indic.: Place of residence: suburban area	1.145	.139
Marital status, child controls	yes	
Rotating panel controls[d]	yes	
Obs.	15,666	
Pseudo R^2	.050	

Significance levels : † : 10% $*$: 5% $**$: 1%

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data. Indicator variables (Indic.) take a value of one if the described condition is satisfied.

[a] Course participation in the year before voucher assignment.

[b] University-entry certificate.

[c] Degree at university of applied sciences or professional education and training.

[d] Controls for year in which respondent was first interviewed in the Swiss Labor Force Survey.

factors that influence voucher redemption – the control group for the *voucher & redeem* group and the control group for the *voucher & dont't redeem* show very similar response patterns in subsequent survey rounds. This result, which could not have been known prior to the matching exercise, is interesting for at least two reasons. First, it means that although the voucher-redeemers and non-redeemers differ substantially based on their observable characteristics, these differences do not seem to be relevant for the explanation of differences in the survey response patterns. Second, it also suggests that, if such observables like gender, education, age, etc. have no discernible influence on the response rates, it is hard to imagine which unobservables might then still differ between the voucher redeemers and non-redeemers that could have a substantial influence on the response behavior, except the fact that one group valued the gift and the other group did not.

However, after presenting the results from the ATT estimation, to test the sensitivity of our results with respect to hypothetical unobserved factors, we will use Rosenbaum (2002) bounds.

Average treatment effects Having formed a matched sample of treated voucher redeemers and control individuals, we estimate the ATT. Table 5 shows that, in the summer 2006 wave of the Swiss Labor Force Survey, i.e. half a year after the vouchers were randomly assigned, the voucher recipients that redeemed their voucher have a survey response probability in the first four weeks of the interviewing period that is 25.5 percentage points higher than for comparable voucher non-recipients. After one-and-a-half years, in the summer 2007 round of the Swiss LFS, it is still 14.1 percentage points higher than for the control group of voucher non-recipients. These are very sizeable numbers that point to large and long-lasting effects of receiving the voucher gift and making use of it. The effect varies surprisingly little by voucher value.[9] In fact, for those who redeem their voucher, the response rates are uniformly higher across all voucher values. Survey respondents seem to attach value to the usefulness of a voucher, but reciprocity does not vary by voucher size.[10]

Remember that the vouchers with different values had been distributed randomly within the experimental group. This sheds new light on the results in in Table 3 for the treatment group as a whole (i.e. not distinguishing between those who redeem and those who don't redeem their vouchers). There, we found higher response rates for those with higher voucher values. Since, as we just discussed in the context of Table 5, reciprocity conditional on voucher redemption does not vary by voucher value, the higher survey response rates for those with higher voucher values can solely be attributed to the fact that the vouchers with higher values had higher redemption rates.[11] Given the fact that the voucher value as such has

[9] This particular finding is different from Falk (2007) who finds a significantly higher degree of reciprocity for more generous gifts (four Christmas cards instead of one). There might be decreasing "returns" to gifts in the sense that for small amounts (cash equivalent of a Christmas card), getting four times as much makes a large difference whereas for large amounts (value of 200 CHF for lowest face value of voucher are nearly 200 US Dollars) getting even more does not further increase the intensity of gift exchange.

[10] Note that recipients of a 200 CHF voucher do not know that other LFS respondents received vouchers of higher value and may be equally "thankful" of having received such a sizeable voucher at all.

[11] In fact, as shown by Messer and Wolter (2009), voucher redemption rates *are* increasing with voucher value. The redemption rate for vouchers with a face value of CHF 200 was 12.6%, whereas the redemption rate for CHF 750 and CHF 1500 vouchers was 21.0% and 21.7%, respectively.

no impact on the survey response rates but only on the redemption rates, the randomization of voucher values within the treatment group provides clean evidence that it is the *redemption* of the voucher that triggers the reciprocity.

The heterogeneity between those who redeem their voucher and those who don't, comes out clearly by looking at our second treatment group: voucher recipients who do not redeem their voucher. This group still exhibits slightly higher response rates in the 2006 round of the Swiss LFS than do comparable voucher non-recipients, but the difference is statistically insignificant. In the 2007 round, their response rates are lower than that of comparable non-recipients, but again the difference is statistically insignificant.

Whereas their response rates are indistinguishable from voucher non-recipients, they are considerably and significantly lower than those of voucher recipients that redeemed their vouchers. The *voucher & don't redeem* group thus exhibits a significantly lower degree of reciprocity (in fact zero) than the *voucher & redeem* group, a key finding of our study. The previous literature was not able to identify the value attached to a gift by the recipient. Our results point to considerable heterogeneity between those that value a gift and those that do not.

Table 5: Average Treatment Effect on the Treated

	2006		2007	
	ATT	Std. Err.	ATT	Std. Err.
Voucher & redeem group				
Voucher	.255	.028**	.141	.034**
Voucher 200	.255	.061**	.163	.072*
Voucher 750	.303	.044**	.144	.053**
Voucher 1500	.244	.045**	.118	.054*
Voucher & don't redeem group				
Voucher	.021	.017	-.012	.021
Voucher 200	.010	.028	-.018	.033
Voucher 750	.026	.029	.017	.034
Voucher 1500	-.058	.029*	-.057	.035

Significance levels : † : 10% $*$: 5% $**$: 1%

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data.

Note: ATT denotes the average treatment effect on the treated. See main text for details.

4.2 Interpretation of the results

4.2.1 Gift exchange versus individual-specific motivation

A potential threat to our interpretation could be that the response rate patterns are the result of explanations other than a gift exchange. One such alternative explanation could be that those who redeem a voucher are more motivated individuals who are more likely to participate in training activities (and use their voucher for that purpose) and, at the same time, are also more likely to participate in future survey rounds of the Swiss LFS. Note that several of our observable variables already try to control for these motivational differences: the dummy for course participation in 2005, the year before vouchers were assigned, measures previous training activities, for example, and to some extent captures such individual-specific traits.

But we can go one step further and test this alternative explanation by estimating the ATT *only* for the subsample of interviewees who had already been actively participating in continuous education in the year 2005. As Table 6 shows, the ATT is almost identical as in the first specification, indicating that the activity in continuous education is not likely to be the factor that explains the differences in the survey response rates.

4.2.2 Gift exchange versus strategic motives

The second alternative explanation we test for is the possibility that people respond to the survey not as an exchange for the initial gift, but out of strategic motives to secure the receipt of future vouchers. It is important to note that, as a matter of fact, our voucher experiment was the first and last time any kind of voucher was sent out to survey participants of the Swiss LFS. So, different from many household surveys where respondents are used to receiving small gifts to increase survey participation, the training vouchers came as a surprise to participants of the Swiss Labor Force Survey. Still, survey participants did not know for sure whether the sending of vouchers was a one-time gift or whether vouchers might now be a new regular feature, hence strategic motives are potentially an issue.

We can address the issue of strategic motives by exploiting the rotating panel structure of the Swiss LFS. Remember that participation in the Swiss LFS is for five consecutive years and that vouchers were sent out to a randomized group of individuals who had responded to the LFS in the year 2005. So, in 2006 there are no first-time respondents, but respondents are in their second, third, fourth or last

Table 6: Average Treatment Effect on the Treated

	2006		2007	
	ATT	Std. Err.	ATT	Std. Err.
Voucher & redeem group				
Voucher	.255	.028**	.141	.034**
Participation in Swiss LFS 2006				
second time	.234	.048**	.125	.057*
third time	.298	.054**	.119	.060*
fourth time	.245	.065**	.000	.089
last time	.135	.055*	–	–
Voucher & don't redeem group				
Voucher	.021	.017	-.012	.021
Participation in Swiss LFS 2006				
second time	-.031	.029	-.108	.035**
third time	-.017	.031	.005	.036
fourth time	-.014	.044	.017	.051
last time	-.003	.037	–	–

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data.
Note: ATT denotes the average treatment effect on the treated. See main text for details.

year of LFS participation. We can look at how response rates differ for respondents who only entered the Swiss LFS in 2005 and compare this to "experienced" LFS respondents. Furthermore, the group rotating out of the Swiss LFS is of particular interest as well. This is the group for whom the 2006 interview (i.e. after voucher receipt) is the *last* interview. Table 6 shows the ATT separately for response rates of survey respondents for whom the 2006 is their second, third, fourth or last LFS interview.

Only those respondents who had participated in the Swiss LFS for the *first* time in 2005 (and for whom the 2006 survey was their second round) might have been tempted to think that gifts in between survey waves could be a regular feature. To respondents who had been surveyed for the second, third or fourth time in 2005 the voucher gift must have come as a surprise. Interestingly, the ATT of voucher recipients who are "newcomers" (.234 with a standard error of .048) does not significantly differ from that of voucher recipients for whom the 2006 round was their third interview (.298 with a standard error of .054) or fourth interview (.245 with a standard error of .065). While this evidence is consistent with a gift exchange interpretation, it does not exclude strategic motives in so far as both newcomers and experienced LFS respondents might have acted in the hope of future rewards

Further evidence draws on respondents for whom 2006 is definitely their last round of LFS participation. This group of survey participants knew that in 2006 they would be interviewed for the last time and could not expect to receive new vouchers in the future. Although the ATT for people that participated for the last time in the Swiss LFS in 2006 is lower than for people that more recently joined the LFS population, Table 6 shows that the response rates of this subgroup in the *voucher & redeem* group is still significantly higher than the one of the control group. Furthermore the coefficient (.135) is not statistically significantly different from the one measuring participation for the fourth time (.245).

Although this result does not completely exclude strategic behavior, it contradicts the hypothesis that the differences in the response patterns between control and treatment group are the result of *exclusively* explained by strategic behavior to secure future vouchers.[12]

4.2.3 Gift exchange versus residual heterogeneity

Still, a remaining potential threat to our empirical strategy is that those who redeem their voucher and those who do not might differ in unobservable ways that matter for response rates. To address remaining worries of any sort of unobserved heterogeneity between our treatment groups and the control group, we perform a sensitivity analysis. We use Rosenbaum (2002) bounds to estimate how large the effect of a hypothetically unobserved confounding factor would have to be to overturn our ATT estimate (see Appendix A.4 for a detailed description).

Note that for an unobserved variable to be a source of selection bias, it must affect the probability that an individual redeems the voucher *and* must affect the outcome. In particular, an unobserved variable that differentially affects subgroups of voucher recipients in the treatment group, but that does not have an effect on the outcome beyond the variables already controlled for, does *not* challenge the robustness of our results. Examples of such variables are motivational differences as just mentioned. Only if groups of individuals differ on unobserved variables that simultaneously affect the assignment to treatment and the outcome, a hidden bias may arise on unobserved heterogeneity. We want to determine how strongly a hypothetical unobserved variable would have to be to influence the selection process so that it could undermine the results of our matching analysis.

[12]Note that, in 2007, the ATT by interview wave has to be interpreted with caution as the sample size for subsamples of the *voucher & redeem* group becomes quite small due to attrition between 2006 and 2007.

We perform a sensitivity analysis for all statistically significant ATT effects. For this purpose, we gradually increase the level of the critical value of the odds ratio where inference about the treatment effect starts to be overturned. Table 7 displays the critical values for all ATT effects. For the *voucher & redeem* group, we find that the critical value, for which the statistically significant ATT effects in Table 5 would become statistically indistinguishable from zero, is well above 3 for most of our ATT estimates. Consider the effect for the *voucher & redeem* group when we do not distinguish by voucher amount. We find the critical odds-ratio value to be 3.75 in the summer 2006 LFS survey round. This means that all individuals with the same observed x-vector can differ in their odds of treatment by a factor of up to 3.75, or 375 percent, before the confidence band around the ATT estimate starts to include zero. This is a worst-case scenario. A critical value of 3.75 does not imply that there *is* indeed unobserved heterogeneity or that there is no effect of treatment on the outcome variable. This result only means that the confidence interval for the effect would include zero *if* an unobserved variable caused the odds ratio of treatment assignment to differ between treatment and control groups by 3.75 *and* if this variable's effect on the outcome is so strong that it almost perfectly determines the outcome in each pair of matched cases in the data. Table 4 gives an idea of what an odds ratio of 3.75 on a hypothetical binary variable compares to.

The largest numbers in that table are the odds ratios on the indicators of non-academic tertiary degree (2.755) and matura certificate (2.094), respectively. The *unobserved* motivation indicator would thus have to be far more influential (e^{γ}=3.75) than the *observed* difference between an individual with compulsory schooling (the reference group) and one with a completed tertiary degree (e^{γ}=2.755). While we cannot exclude with certainty that such an influential unobserved factor exists, we consider it implausible that motivational differences (for training participation), or any other unobserved factor outside our list of regressors, would exert such a strong impact on selection in *all* pairs of treated and matched controls. We therefore view the statistically significant ATT effects in the *voucher & redeem* group as pretty robust to hidden bias.

On the other hand, when we look at the *voucher & don't redeem* group, we find that the critical value is 1.13 for the only treatment effect that is statistically significant to begin with (voucher size 1500 in survey round 2006). The results for this particular group can therefore not be considered as very robust.

Table 7: Sensitivity Analysis with Rosenbaum Bounds

	γ_c	
	2006	2007
Voucher & redeem group		
Voucher	3.75	1.51
Voucher 200	2.47	1.30
Voucher 750	4.26	1.35
Voucher 1500	2.91	1.20
Voucher & don't redeem group		
Voucher	ns	ns
Voucher 200	ns	ns
Voucher 750	ns	ns
Voucher 1500	1.13	ns

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data.
Note: Table displays critical values of the statistically significant odds ratios e^{γ} based on the Mantel and Haenszel (1959) test statistic, as suggested by Rosenbaum (2002); "ns" for non-significant odds ratios. See Appendix A.4 for details.

4.2.4 Survey-response methodology

Our paper also relates to the methodological literature on how to increase response rates in surveys. Before discussing this literature, it is important to remind the reader that our voucher experiment was not designed with the aim of increasing survey response, but in order to study the role of financial support (vouchers) in increasing training participation. As such, our analysis of survey participation after receipt of training vouchers is an accidental by-product of that voucher experiment.

The survey methodology literature[13] has dealt with various ways to increase survey response. Underlying this is the assumption that respondents react to incentives due to e.g. "social exchange" (see Dillman (1978)). Interestingly, and related to the discussion about strategic motives above, there seems to be little concern that respondents will always expect incentives when they have once received them (Singer, Van Hoewyk, and Maher (1998); Singer et al. (1999)). As for the type of incentive, the literature finds larger pecuniary incentives to work better. Importantly, the range of pecuniary incentives used in household surveys is several orders of magnitude smaller than in our setup. As mentioned earlier, the British Household Panel Survey (BHPS) sends respondents £10 gift vouchers as a token of thanks (see Laurie 2007). Our training vouchers are worth between

[13] See e.g. Groves, Dillman, Eltinge and Little (2002) and Stoop, Billet, Koch and Fitzgerald (2010) for detailed overviews and further references.

200 Swiss Franks (ca. £140) and 1500 Swiss Franks (£1050).

The fact that in our experiment voucher recipients who do not redeem the training vouchers have no higher response rate compared to voucher non-recipients shows that even incentives with extremely high cash-equivalent values will be ineffective if the recipient does not value them.[14] To the best of our knowledge, this is a novel finding in both the gift exchange and survey methodology literatures.

5 Conclusions

This paper provides evidence on gift exchange in a unique field experiment in which a random subsample of the Swiss Labor Force Survey was sent training vouchers. Gift exchange comes in the form of voluntary participation in future rounds of the Swiss LFS. The results show a significantly higher response rate for the randomly selected treatment group (voucher recipients) compared to the control group (voucher non-recipients) in the first survey round 6 months after the vouchers were sent out, but no significant effect in the second survey round 18 months after the initial gift.

Different from the existing literature, we can also distinguish between participants that redeem their training voucher and those who did not. The difference in response rates in future survey rounds between those who redeem the voucher and those who don't, is substantial and points to considerable heterogeneity: individuals only reciprocate when they perceive a gift as a (useful) gift.[15]

A second unique feature of our experimental setting is that we can study gift exchange in the long run. Typical gift exchange experiments have a horizon of only several hours or days when studying "long-run" effects. We follow participants 6 and 18 months after the original gift (the training voucher).

Empirically, a challenge arises from the fact that voucher recipients that redeem their voucher and those that do not might systematically differ. In other words, whereas our experimental setting ensures that voucher recipients and non-recipients do not systematically differ, there is likely to be self-selection into training participation. To address this issue, we pursue a careful matching procedure to identify suitable control observations from the group of voucher-non-recipients. Furthermore, to address the issue of unobserved heterogeneity between the treat-

[14]Remember from footnote 4 that the actual use of gift vouchers is not followed up by survey agencies.

[15]Englmaier and Leider (2010) point to a different form of heterogeneity in their experiments. They show that an agent's effort depends not only on the pay received by the agent (the standard monetary "gift"), but also on the principal's payoff.

ment and control groups, we perform a statistical bounding analysis showing that unobserved factors would have to be unreasonably large to overturn our findings.

Our results show that survey response rates of training voucher recipients that redeem their vouchers exceed response rates of voucher non-recipients by 25 and 14 percentage points in the two survey rounds, half a year and one-and-a-half years after the voucher experiment, respectively. Voucher recipients that don't redeem their training vouchers have response rates that are not statistically different from those of non-recipients. The results do not vary much by the value of the voucher.

We show that if we had not been able to separately analyze the gift-exchange effects for those who value the gift (voucher) and those who did not, the interpretation of the magnitude and the duration of the effect would have been misleading. In conclusion, it is therefore equally essential for the interpretation of a gift exchange experiment to know how big the fraction of the treated group is that perceives the gift as a gift, as it is to know by how much treated people react to the gift by reciprocating. In all situations of a gift exchange that do not involve only cash money, this differentiation is therefore likely to be crucial for the interpretation of the results.

References

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, John A. List (2008) For love or money? Comparing the effects of non-pecuniary and pecuniary incentive schemes in the workplace. University of Chicago unpublished manuscript.

Bellemare, Charles, Bruce Shearer (2009) Gift giving and worker productivity: Evidence from a firm-level experiment. Games and Economic Behavior 67(1): 233–244.

Caliendo, Marco, Susanne Kopeinig (2008) Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys 22(1): 31–72.

Dillman, Don A. (1978) Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Englmaier, Florian, Steven Leider (2010) Gift Exchange in the Lab It is not (only) how much you give .... CESifo Working Paper No. 2944.

Falk, Armin (2007) Gift Exchange in the Field. Econometrica 75(5): 1501–1511.

Fehr, Ernst, Simon Gächter (2000) Fairness and Retaliation: The Economics of Reciprocity. Journal of Economic Perspectives, 14(3): 159–181

Fehr, Ernst, Simon Gächter, Georg Kirchsteiger (1997) Reciprocity as a Contract Enforcement Device. Econometrica 65(4): 833–860.

Fehr, Ernst, Georg Kirchsteiger, Arno Riedl (1993) Does Fairness Prevent Market Clearing? An Experimental Investigation. Quarterly Journal of Economics 108(2): 437–460.

Gneezy, Uri, John A. List (2006) Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets using Field Experiments. Econometrica 74(5): 1365–1384.

Groves, Robert M., Don A. Dillman, John L. Eltinge and Roderick J.A. Little (2002) Survey Nonresponse, Wiley Series in Probability and Statistics.

Hennig-Schmidt, Heike, Bettina Rockenbach, Abdolkarim Sadrieh (2009) In Search of Workers' Real Effort Reciprocity - A Field and a Laboratory Experiment. Journal of the European Economic Association 8(4): 817–837.

Kube, Sebastian, Michel André Maréchal, Clemens Puppe (2010) The Currency of Reciprocity - Gift Exchange in the Workplace. University of Zurich Institute for Empirical Research in Economics Working Paper No. 377.

Laurie, Heather (2007) The effect of increasing financial incentives in a panel survey: an experiment on the British Household Panel Survey, Wave 14. ISER Working Paper 2007-5. Colchester: University of Essex.

Laurie, Heather, Peter Lynn (2008) The Use of Respondent Incentives on Longitudinal Surveys. ISER Working Paper 2008-42. Colchester: University of Essex.

Lechner, Michael (2002) Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies. Review of Economics and Statistics 84(2): 205–220.

Leuven, Edwin, Hessel Oosterbeek, Randolph Sloof and Chris van Klaveren (2005) Worker Reciprocity and Employer Investment in Training. Economica 72: 137–149.

Mantel, Nathan, William Haenszel (1959) Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. Journal of the National Cancer Institute 22(4): 719–748.

Messer, Dolores, Stefan C. Wolter (2009) Money matters - Evidence from a large-scale randomized field experiment with vouchers for adult training. CESifo Working Paper 2548.

Rosenbaum, Paul R. (2002) Observational studies Second edition. Series in Statistics, New York and Heidelberg: Springer.

Rosenbaum, Paul R., Donald B. Rubin (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70(1): 41–55.

Sianesi, Barbara (2004) An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. Review of Economics and Statistics 86(1): 133–155.

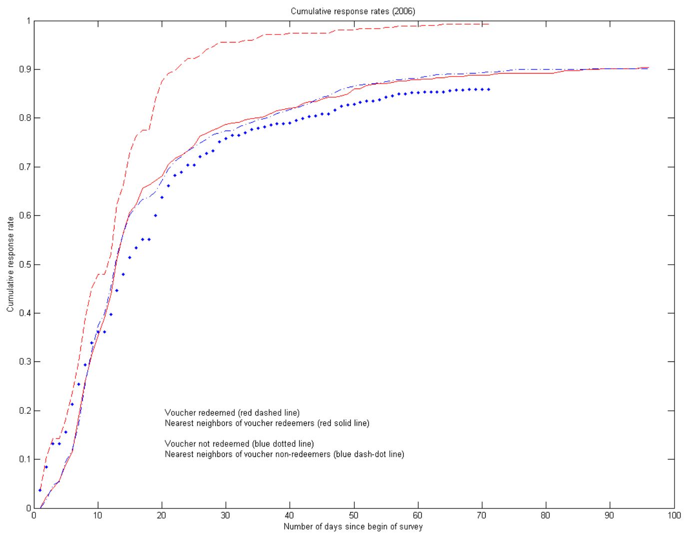
Singer, Eleanor, John Van Hoewyk, Nancy Gebler, Trivellore Raghunathan, and Katherine McGonagle (1999) The Effect of Incentives in Interviewer-Mediated Surveys. Journal of Official Statistics 15:217–230.

Singer, Eleanor, John Van Hoewyk, and Mary P. Maher (1998) Does the Payment of Incentives Create Expectation Effects? Public Opinion Quarterly 62:152–164.

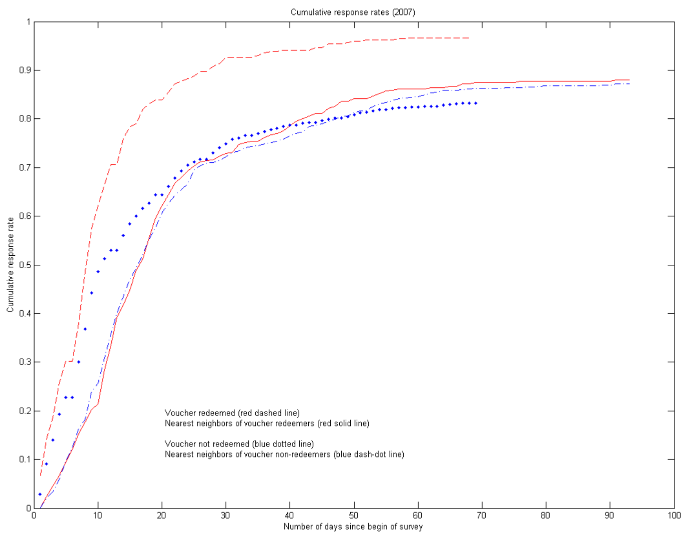
Stoop, Ineke, Jaak Billet, Achim Koch and Rory Fitzgerald (2010) Improving Survey Nonresponse: Lessons learned from the European Social Survey, Wiley Series in Survey Methodology.

A Appendix

A.1 Response rates



(a)



(b)

Figure 1: Response rates

A.2 Propensity score matching

We provide a brief methodological discussion in our context. Our estimator measures the average treatment effect on the treated (ATT) for two interesting treatment groups: those that receive a voucher and find it beneficial (*voucher & redeem* group) and those that receive a voucher which they do not redeem (*voucher & don't redeem* group). Whereas the *voucher & redeem* and the *voucher & don't redeem* groups *combined* are randomly drawn from the population, they are not individually random sub-populations because of the heterogeneity between those that redeem their vouchers and those that do not. Raw differences between the average outcomes in the *voucher & redeem* group and in the control group therefore give biased estimates of the effect of voucher recipience for those who value a voucher. Propensity-score matching reduces (and ideally removes) this bias. A crucial assumption is that observable covariates exhaustively determine selection into treatment. Since receipt of the voucher is randomized, this assumption is equivalent to assuming that redemption of the voucher is exhaustively determined by observed covariates. The wealth of information in our data—individual characteristics such as demographic, education and work variables as well as controls for region of residence—comprehensively covers the pretreatment conditions so that the assumption of selection on observables is not unreasonable. We address remaining worries of selection on unobservables by calculating Rosenbaum (2002) bounds as described in Appendix A.4.

Matching treated units on a vector of characteristics suffers dimensionality problems for large sets of characteristics. Propensity-score matching therefore summarizes pretreatment characteristics into a scalar, the propensity score. Exposing individuals with the same propensity score value (same ex ante probability to take training) to random treatment (voucher) eliminates the bias in estimated treatment effects. Define the propensity score as the conditional probability of receiving treatment given pretreatment characteristics,

$$p\left(\mathbf{x}_i\right) \equiv Pr(d_i\!=\!1 \mid \mathbf{x}_i) = \mathbb{E}\left[d_i|\mathbf{x}_i\right], \tag{1}$$

where d_i is equal to one for voucher recipients and $\mathbf{x}_i$ is the vector of pretreatment characteristics. (We omit time subscripts to save on notation.)

Rosenbaum and Rubin (1983) show that, if the exposure to treatment is random within cells defined by $\mathbf{x}_i$, it is also random within cells defined by the values of the scalar propensity score $p\left(\mathbf{x}_i\right)$. Rosenbaum and Rubin (1983) also show that, if the propensity score $p\,(\mathbf{x}_i)$ is known, the ATT can be defined as

$$\begin{aligned} ATT &\equiv \mathbb{E}\left[y_{1i} - y_{0i} | d_i = 1\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[y_{1i} - y_{0i} | d_i = 1, p\,(\mathbf{x}_i)\right]\right] \\ &= \mathbb{E}\big[\,\mathbb{E}\left[y_{1i} | d_i = 1, p\,(\mathbf{x}_i)\right] - \mathbb{E}\left[y_{0i} | d_i = 0, p\,(\mathbf{x}_i)\right] \;\big|\, d_i = 1\big], \end{aligned} \tag{2}$$

where outer expectations are over the distribution of $p\,(\mathbf{x}_i)|d_i = 1$, and y_i is the outcome taking a value of one iff the individual participates in future survey rounds. To denote the two counterfactual situations of, respectively, treatment and no treatment, we use shorthand notations $y_{1i} \equiv (y_i|d_i = 1)$ and $y_{0i} \equiv (y_i|d_i = 0)$. The derivation of the ATT estimator requires two intermediate results to hold. First, the pretreatment variables need to be *balanced* given a valid propensity score (Rosenbaum and Rubin (1983), lemma 1): If $p\,(\mathbf{x}_i)$ is the propensity score, then

$$d_i \quad \perp \quad \mathbf{x}_i \quad | \quad p\,(\mathbf{x}_i). \tag{3}$$

As a consequence, observations with the same propensity score have the same distribution of observable characteristics independent of treatment status. Put differently, exposure to treatment is random for a given propensity score so that treated and control individuals are, on average, observationally identical. The orthogonality of d_i and $\mathbf{x}_i$ conditional on the propensity score is empirically testable. We perform according balancing tests and compare changes in the goodness of fit for alternative sets of pretreatment variables $\mathbf{x}_i$.

Second, the assignment of the treatment needs to be unconfounded conditional on observable characteristics (Rosenbaum and Rubin 1983, lemma 2). If assignment to treatment is unconfounded, that is if

$$y_{1i}, y_{0i} \quad \perp \quad d_i \quad | \quad \mathbf{x}_i, \tag{4}$$

then assignment to treatment is unconfounded given the propensity score, that is

$$y_{1i}, y_{0i} \quad \perp \quad d_i \quad | \quad p\,(\mathbf{x}_i). \tag{5}$$

Equation (4) is a maintained assumption of our method.

We estimate the propensity score $Pr(d_i = 1\,|\,\mathbf{x}_i) = F(h(\mathbf{x}_i))$ under the assumption of a logistic cumulative distribution function $F(\cdot)$, where $h(\mathbf{x}_i)$ is, in principle, a function of linear and higher-order terms of the covariates. We find linear terms on our comprehensive set of covariates to suffice for balancing (3) to be satisfied and omit higher-order terms.

To implement an estimator for the ATT (2), we use the estimated propensity scores to pick pairs based on nearest-neighbor matching. Denote by $\mathbb{C}(i)$ the set of control units matched to the treated unit i with an estimated value of the propensity score of p_i. Nearest-neighbor matching assigns $\mathbb{C}(i) \equiv \min_j \parallel p_i - p_j \parallel$, a singleton unless there are ties (multiple nearest neighbors). In the non-experimental sample, we observe y_{1i} only for treated individuals and y_{0i} for untreated individuals. The estimator therefore uses y_i^T from the treated subsample as treated outcome and y_j^C from the control sample as counterfactual outcome y_{0i}. We denote the number of controls matched to observation $i \in T$ by N_i^C and define weights $w_{ij} \equiv 1/N_i^C$ if $j \in \mathbb{C}(i)$, and $w_{ij} = 0$ otherwise. Then, the nearest neighbor estimator of the ATT is:

$$ATT^{NN} = \frac{1}{N^T} \sum_{i \in T} \left[y_i^T - \textstyle\sum_{j \in \mathbb{C}(i)} w_{ij} y_j^C \right], \tag{6}$$

where N^T denotes the number of treated and N^C the number of control observations. Our propensity score estimator is the mean difference in outcomes over matched pairs. The specification of $h(\mathbf{x}_i)$ satisfies the balancing hypothesis and is more parsimonious than the full set of interactions needed to match cases and controls on the basis of observables. The propensity score therefore reduces the dimensionality problem of matching treated and control units on the basis of the multidimensional vector $\mathbf{x}_i$.[16]

A.3 Matching quality

Covariate balancing assesses matching quality. Table 8 shows matching quality indicators. Our first matching statistic, the pseudo R^2 from logit estimation of the conditional probability of voucher redemption, indicates the degree to which regressors $\mathbf{x}_i$ predict the treatment probability (columns 3 and 4). After matching, regressors $\mathbf{x}_i$ should have no explanatory power for selection into treatment if the treatment and matched control samples have balanced characteristics. Our results show that this is the case. The pseudo R^2 statistics drop from 5.0 to 1.0 percent in the *voucher & redeem* group when we do not distinguish by voucher value.

Rosenbaum and Rubin (1985) propose a comparison between (standardized)

[16] It is important to note that the outcome plays no role in the algorithm for the estimation of the propensity score. This is equivalent, in this context, to what happens in controlled experiments in which the design of the experiment has to be specified independently of the outcome.

Table 8: Covariance Balancing, Before and After Matching

	No. of treated	No. of controls	Logit ps. R^2 before	Logit ps. R^2 after	Median bias before	Median bias after
	(1)	(2)	(3)	(4)	(5)	(6)
2006						
Voucher & redeem group						
All vouchers	427	15,239	.050	.010	9.896	4.799
Voucher 200	94	15,239	.073	.033	19.237	5.452
Voucher 750	165	15,239	.046	.015	8.630	5.528
Voucher 1500	168	15,239	.045	.012	11.715	5.769
Voucher & don't redeem group						
All vouchers	1,656	15,239	.019	.004	2.542	3.018
Voucher 200	592	15,239	.014	.007	4.113	2.925
Voucher 750	532	15,239	.024	.009	3.938	3.444
Voucher 1500	532	15,239	.015	.006	5.110	3.438
2007						
Voucher & redeem group						
All vouchers	391	10,256	.066	.010	10.165	4.717
Voucher 200	86	10,256	.094	.026	18.906	4.829
Voucher 750	153	10,256	.065	.020	10.657	5.284
Voucher 1500	152	10,256	.054	.025	9.926	5.847
Voucher & don't redeem group						
All vouchers	1,276	10,256	.032	.005	3.024	2.344
Voucher 200	451	10,256	.026	.009	5.820	3.981
Voucher 750	418	10,256	.036	.015	5.051	4.503
Voucher 1500	407	10,256	.027	.009	5.911	3.733

Data: Swiss Labor Force Survey, 2005, 2006 and 2007, and experimental data.
Note: Column (5) displays $B_{before}(\mathbf{x}_i)$ and column (6) displays $B_{after}(\mathbf{x}_i)$.

treated unit means and (standardized) control unit means before and after matching as a second evaluation method for covariate balance.[17]

As is commonly done in the evaluation literature, we show the median absolute standardized bias before ($B_{before}(\mathbf{x}_i)$) and after matching ($B_{after}(\mathbf{x}_i)$), over all

[17] The standardized differences (standardized biases) between the means for a covariate $\mathbf{x}_i$ are defined as:

$$\begin{aligned} B_{before}(\mathbf{x}_i) &= 100 \cdot \frac{\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i0}}{\sqrt{V_1(\mathbf{x}_i) + V_2(\mathbf{x}_i)/2}} \\ B_{after}(\mathbf{x}_i) &= 100 \cdot \frac{\bar{\mathbf{x}}_{i1M} - \bar{\mathbf{x}}_{i0M}}{\sqrt{V_1(\mathbf{x}_i) + V_2(\mathbf{x}_i)/2}}, \end{aligned}$$

where $\bar{\mathbf{x}}_{i1}$ denotes the treated unit mean and $\bar{\mathbf{x}}_{i0}$ the control unit mean for covariate $\mathbf{x}_i$.

regressors $\mathbf{x}_i$ that enter the propensity score estimation (columns 5 and 6). In the main specification, matching reduces the median absolute standardized bias by three quarters (from 9.896 to 4.799). There seem to be no formal criteria in the literature to judge the size of standardized bias. Yet the remaining bias between 3 and 6 percent is in the same range as in microeconomic evaluation studies (e.g. Lechner (2002) and Sianesi (2004)).[18]

Overall, observable characteristics between treated and control observations are well balanced after propensity-score matching.

A.4 Rosenbaum bounds for binary outcomes

We outline the idea behind Rosenbaum (2002) bounds. Rewrite the probability that individual i with observed characteristics $\mathbf{x}_i$ receives a voucher and redeems it (treatment 1) or receives a voucher but does not redeem it (treatment 2):

$$p(\mathbf{x}_i) = Pr(d_i = 1|\mathbf{x}_i) = F(\beta\mathbf{x}_i + \gamma u_i), \tag{7}$$

where u_i is the unobserved variable of concern (intrinsic motivation, for instance) and γ is the effect of u_i on the treatment probability. If the estimator is free of hidden bias, γ is zero and the participation probability is solely determined by $\mathbf{x}_i$. However, if there is hidden bias, two individuals with the same observed covariates x have differing chances of receiving treatment. Take a matched pair of observations i and j, and consider the logistic distribution F. The odds that the individuals receive treatment are $p(\mathbf{x}_i)/(1-p(\mathbf{x}_i))$ and $p(\mathbf{x}_j)/(1-p(\mathbf{x}_j))$ so that the odds ratio is given by

$$\frac{\frac{p(\mathbf{x}_i)}{1-p(\mathbf{x}_i)}}{\frac{p(\mathbf{x}_j)}{1-p(\mathbf{x}_j)}} = \frac{p(\mathbf{x}_i)(1-p(\mathbf{x}_j))}{p(\mathbf{x}_j)(1-p(\mathbf{x}_i))} = \frac{\exp(\beta\mathbf{x}_i + \gamma u_i)}{\exp(\beta\mathbf{x}_j + \gamma u_j)} = \exp[\gamma(u_i - u_j)]. \tag{8}$$

If both individuals share the same observed covariates after propensity-score matching, the x-vector cancels. The individuals nevertheless differ in their odds of receiving treatment by a factor that involves the parameter γ and the difference in the unobserved variable u. It is the objective of sensitivity analysis to evaluate how inference about the treatment effect is altered by changing the values of γ and $(u_i - u_j)$.

Assume for the sake of simplicity that the unobserved covariate is an indicator variable with $u_i \in \{0, 1\}$ (indicating the acquisition of an ownership advantage).

[18]Rosenbaum and Rubin (1985) suggest that a value of 20 is "large."

Rosenbaum (2002) shows that equation (8) then implies the following bounds on the ratio of the odds that either of the two matched individuals will receive treatment:

$$\frac{1}{e^{\gamma}} \leq \frac{p(\mathbf{x}_i)}{p(\mathbf{x}_j)} \frac{(1 - p(\mathbf{x}_j))}{(1 - p(\mathbf{x}_i))} \leq e^{\gamma}. \tag{9}$$

The two matched individuals have the same probability of being treated only if the odds ratio $e^{\gamma} = 1$. If the odds ratio $e^{\gamma} = 2$, then individuals who appear to be similar (in terms of x), could differ in their odds of receiving the treatment by as much as a factor of 2.

We compute critical values of the odds ratio e^{γ} based on the Mantel and Haenszel (1959) test statistic, as suggested by Rosenbaum (2002). The Mantel and Haenszel test statistic assesses the strength of hidden bias that would be necessary to overturn our ATT estimate.

The non-parametric Mantel and Haenszel (1959) test compares the successful number of individuals in the treatment group to the same expected number under the null hypothesis that the treatment effect is zero. Denote with N_{1s} and N_{0s} the numbers of treated and non-treated individuals in stratum s, where $N_s = N_{0s} + N_{1s}$. y_{1s} is the number of treated individuals with a positive outcome (survey participation), y_{0s} is the number of non-treated individuals with a positive outcome, and y_s is the total number of positive outcomes in stratum s. The MH test-statistic Q_{MH} asymptotes the standard normal distribution and is given by

$$Q_{MH} = \frac{|y_1 - \sum_{s=1}^{S} E(y_{1s})| - .5}{\sqrt{\sum_{s=1}^{S} Var(y_{1s})}} = \frac{|y_1 - \sum_{s=1}^{S} (\frac{N_{1s} y_s}{N_s})| - .5}{\sqrt{\sum_{s=1}^{S} \frac{N_{1s} N_{0s} y_s (N_s - y_s)}{N_s^2 (N_s - 1)}}}. \tag{10}$$

Our propensity-score matching procedure minimizes differences between treatment and control group observations so that the MH test (designed for random samples) is applicable. Take the possible influence of a binary hidden variable with an effect $e^{\gamma} > 1$ on the outcome. For fixed $e^{\gamma} > 1$, Rosenbaum (2002) shows that the MH test statistic Q_{MH} can be bounded by two known distributions. If $e^{\gamma} = 1$, the bounds are equal to the baseline scenario of no hidden bias. With increasing e^{γ}, the bounds move apart, reflecting uncertainty about the test statistic in the presence of unobserved selection bias.

Consider two scenarios. First, let Q_{MH}^{+} be the test statistic given that we overestimate the treatment effect and, second, let Q_{MH}^{-} the case where we under-estimate the treatment effect. The two bounds are then given by:

$$Q^+_{MH} = \frac{|y_1 - \sum_{s=1}^{S} \widetilde{E}^+_s| - .5}{\sqrt{\sum_{s=1}^{S} Var(\widetilde{E}^+_s)}} \tag{11}$$

and

$$Q^-_{MH} = \frac{|y_1 - \sum_{s=1}^{S} \widetilde{E}^-_s| - .5}{\sqrt{\sum_{s=1}^{S} Var(\widetilde{E}^-_s)}}, \tag{12}$$

where $\widetilde{E_s}$ and $Var(\widetilde{E_s})$ are the large sample approximations to the expectation and variance of the number of successful participants when the hidden variable is binary and γ given.[19]

A.5 Training voucher and letter

[19]The large sample approximation to $\widetilde{E}^+_s$ is the unique root of the quadratic equation $\widetilde{E}^2_s(e^\gamma - 1) - \widetilde{E}_s[(e^\gamma - 1)(N_{1s} + y_s) + N_s] + e^\gamma y_s N_{1s}$, after addition of $max(0, y_s + N_{1s} - N_s \leq \widetilde{E_s} \leq min(y_s, N_{1s}))$ to select the root. $\widetilde{E}^-_s$ follows by replacing e^γ with $1/e^\gamma$. The large sample approximation to the variance is $Var(\widetilde{E}_s) = [1/\widetilde{E}_s + 1/(y_s - \widetilde{E}_s) + 1/(N_{1s} - \widetilde{E}_s) + 1/(N_s - y_s - N_{1s} + \widetilde{E}_s)]^{-1}$.

Voucher Nr. 19789db
for
Hanny Sample

of the value of CHF 1500.-

for the participation in continuous education
valid until 31 May 2006

The voucher can be redeemed on presentation of the filled-in confirmation of course participation. Several courses can be attended.

Please send the voucher and the course participation by 31 July 2006 to:

LINK Marketing Services
Keyword: Voucher
Spannortstrasse 7/9
6000 Luzern 4

The redemption of the voucher is voluntary. I agree with the statistical use of the information of the voucher, respectively with the confirmation of the course participation (without name) and the linkage of the information coming from the telephone interview and the Swiss Labor Force Survey.

Place, Date: ……………………Signature:……………………..

Confirmation of course participation 1:

We confirm, that Hanny Sample has attended the following course:

Topic according to announcement…………………………………………………………..

From: …………….. to: ……………………………….

Total of course lessons: ………………….

Total amount invoiced: ……………

Organiser:
Name:………………………………………….
Street:………………………………………..
Zip-code Place:………………………………………

Date:……………

Stamp and signature of the organiser: ………………………………………

More course confirmations on the backside

Figure 2: 1500 CHF training voucher (English translation)

OFS BFS UST
Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

Statistique suisse Statistik Schweiz Statistica svizzera Statistica svizra Swiss Statistics

Mrs
Hanny Sample
Ystreet 88
9999 xlingen

Neuchâtel, January 2006
[15-04.20 BG/ AB]

Voucher for continuous education

Dear Mrs. Sample

On behalf of the Swiss Federal Office for Professional Education and Technology (OPET) we send you today a voucher for one or several continuous training courses. The voucher is part of a new research project of the Federal government. In the attachment you will find explanations how to redeem the voucher.

The Federal Statistical Office, on behalf of the OPET will be responsible for the statistical implementation of the project and data protection. You have been participating in the past in the Swiss Labor Force Survey (SLFS). This participation makes you particularly suited for this project. In the summer of 2006 you will be asked to participate in a telephone interview on employment and continuous education, carried out by the LINK institute on behalf of us. After the completion of the survey all connections between your name and your responses will be deleted. Our staff will only transmit the reports with the statistical results to the OPET, preserving absolute anonymity.

Participation in this project is voluntary. However, for a good success of the project your disposition to participate in the survey is important.

Espace de l'Europe 10
CH - 2010 Neuchâtel
www.statistik.admin.ch

OFS BFS UST

Seite 2/2

If you have questions related to this project or this letter, please contact the Federal Statistical Office, Tel. 032 713 61 91.

Yours sincerely,

FEDERAL STATISTCAL OFFICE
The General-Director

Dr. Adelheid Bürgi-Schmelz

Attachements:
- Voucher with confirmation of participation
- Explanation how to redeem the voucher

Figure 3: Letter that was sent along with the voucher (English translation)