

Assessing the Causal Effect of Curbside Collection on Recycling Behavior in a Non-randomized Experiment with Self-reported Outcome

Henning Best¹ · Thorsten Kneip²

Accepted: 15 March 2018 / Published online: 28 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract This paper aims at identifying the causal effect of reducing behavioral costs of participation in household waste recycling through curbside collection. Using propensity score matching and differences-in-differences estimation with individual-level panel data we estimate the effect of curbside collection, its variation between types of recyclables and sociodemographic background variables, and its elasticity with regard to the distance to collection containers in the bring scheme condition. We argue that in a quasi-experimental setting DD may be systematically upward biased due to the outcome variable being self-reported while DDD may be systematically downward biased in the presence of spillover effects. Accordingly, both estimators can be combined to derive upper and lower bounds of the true effect. We find that a curbside scheme has no effect on paper recycling but increases recycling participation by between 10 and 25% points for plastic and packaging. Moreover, we find systematic treatment effect heterogeneity with regard to pre-treatment distance to collection sites and individual environmental attitudes, but not by socio-demography. The results of our analysis therefore have important implications for effective and cost-efficient implementation of environmental protection policies in urban areas.

Keywords Environmental behavior · Environmental policy · Recycling · Natural experiment · Propensity score matching · Differences in differences · Causal analysis

✉ Henning Best
best@sowi.uni-kl.de

Thorsten Kneip
kneip@mea.mpisoc.mpg.de

¹ Department of Social Sciences, University of Kaiserslautern, Erwin-Schrödinger-Str. 57, 67663 Kaiserslautern, Germany

² MEA, Max Planck Institute for Social Law and Social Policy, Amalienstraße 33, 80799 Munich, Germany

1 Introduction

Over the past several decades, municipalities have faced enormous growth in solid waste output (World Bank 2012). Beyond the well-known global threats, that is environmental pollution and inefficient use of scarce resources, this development also relates to more practical problems of urban waste management. A reduction of total disposable waste, either by decreasing total output of packaging materials and other disposable goods or by increasing the level of recycling, is the generally preferred manner of improving waste management (see e.g. van den Bergh 2008; Abrate et al. 2014). Two policy instruments have been discussed to redirect waste quantities from landfills or incineration to recycling: Pricing systems in which fees depend on the actual amount of waste generated (see e.g. Bartelings et al. 2004; Fullerton and Kinnaman 1995, 1996) and curbside recycling schemes which reduce the effort required for individual participation. Curbside programs reduce households' recycling costs relative to a drop-off system by making recycling more convenient and less time-consuming. In particular, reducing the costs of storing and transporting recyclables can increase recycling participation (Ando and Gosselin 2005). With curbside recycling being a widely-used policy instrument (EPA 2015, pp. 131ff.; Hopewell et al. 2009, p. 2119), and a complement rather than an alternative to taxes and pricing systems, it is important to study the effectiveness of curbside programs in different settings.

This paper studies the effect of curbside recycling programs. In addition to estimating the overall effect of curbside programs, we are particularly interested in effect heterogeneity related to different levels of cost reduction: Is curbside recycling also effective when the previously used bring-in scheme used a dense grid of collection containers? And is curbside recycling equally effective for paper, plastics, and packaging? We also investigate possible effect heterogeneity with respect to socio-economic background characteristics (e.g. household composition and income). Estimating treatment effects for these different conditions allows us to comment on the efficiency of curbside recycling and may help policy makers in choosing an effective yet cost-efficient solution for the collection of recyclable household waste.

In general, identification of the causal effect of curbside programs, however, is difficult and requires a number of methodological difficulties to be overcome: In particular, studies need to account for the following problems: (1) (unobserved) heterogeneity between treatment group and controls, (2) policy endogeneity and self-selection into treatment, and—especially when using self-reported data—(3) treatment induced measurement error.¹

First, with cross-sectional data, researchers are bound to assuming that participation in a curbside collection program is effectively random when conditioned on observable characteristics. This strategy is particularly prone to biased estimates due to unobserved heterogeneity. For example, Reschovsky and Stone (1994) use cross-sectional data and probit models to analyze recycling participation. They control for several socio-demographic variables and access to several waste management programs in order to reduce bias, finding that curbside recycling is more effective than unit-based pricing. One of the most extensive studies (Jenkins et al. 2003) uses cross-sectional U.S. nationwide household-level data and distin-

¹ Strictly speaking, self-selection or policy endogeneity are only an issue for the control group if one is interested in estimating an average treatment effect on the treated (ATT). In empirical applications, however, the analytical distinction between heterogeneity—and particularly selection on the pre-treatment trend—and endogeneity may be of little relevance. This is because available control variables (like, e.g., socioeconomic status) can be seen as (proxy-)measures for antecedent conditions as well as expected costs or benefits of treatment (Gangl 2010). In effect, by successfully incorporating all relevant variables the ATT will also converge to the average treatment effect (ATE).

guishes five different materials. The authors estimate community-fixed effects logit models and conclude that access to curbside recycling significantly raises the percentage recycled for all materials irrespective of whether the program is mandatory or voluntary. Kinnaman and Fullerton (2000) model local governments' decisions about curbside recycling as a function of observable exogenous variables to control for possible endogenous policy choices. Their 2SLS approach helps in reducing bias, but unfortunately is restricted to observables as well. Though these studies laid important foundations for following empirical research, a cross-sectional design cannot adequately account for unobserved heterogeneity and endogeneity. This is mainly due to the lack of quasi-experimental data where units are observed at least once before and once after treatment.

Second, with longitudinal data, researchers basically need to assume common (hypothetical) trends in recycling participation for treatment and control group—had there been no treatment. Estimating fixed-effects models or calculating differences-in-differences offers an easy way to control for any unobserved background noise, which may lead to different levels of recycling for different groups. However, they might also still lead to biased results if relevant time-varying variables are omitted, pre-treatment trends are inadequately controlled for, or (self-)selection into treatment based on the expected treatment effect (i.e., in our case, policy endogeneity). Beatty et al. (2007) exploit within-county variation over time in their data to estimate the curbside effect. By using fixed-effects models they are able to exclude bias due to unobserved heterogeneity. They find a small impact of curbside programs only, which, in part, is due to changes in curbside access reducing returns to co-existing recycling centers. Similarly, Tsai and Sheu (2009) promote a difference-in-difference approach to identify the effect of unit pricing on garbage reduction and recycling. According to their results, the investigated fee-per-bag program significantly reduced garbage output but had no effect on recycling. However, both strategies rely on the parallel trends assumption. This assumption can be relaxed when conditioning on further—observable—controls in the DD estimation or, alternatively, by applying propensity score matching. Hence, many problems have been successfully addressed in previous research (see e.g. Dur and Vollaard 2015; Siddique et al. 2010; Kuo and Perrings 2010).

Third, from a decision-theoretic perspective individual-household-level data would be preferable as households are the actual decision-making units targeted by recycling policies (Jenkins et al. 2003). However, the vast majority of previous longitudinal studies have used aggregate community-level data. Using individual level data offers some unique opportunities for research, e.g. analyzing mediator effects (see e.g. Best and Kneip 2011) and treatment heterogeneity with regard to socio-demographic or other factors. Treatment heterogeneity is particularly important from a policy perspective as it carries information about the generalizability of findings and the effectiveness of policy measures in specific settings. A potential caveat to using micro-level survey data is that one usually has to rely on self-reports. In the context of curbside recycling, survey data may be prone to systematic measurement error, namely an over-reporting of recycling due to an increased awareness or perceived social desirability to recycle *induced* by the treatment. If this were the case, treatment effects estimated from these data would be biased upwards. While this serious issue is often neglected in survey research it certainly needs to—and can—be addressed.²

² While the notion of misreported recycling behavior perceived as "socially desired" is a well-known argument against survey data, particularly among proponents of objective aggregate level data (e.g. Kuo and Perrings 2010), we are not aware of any study in the context of recycling, which has explicitly addressed this issue. Note that a general social desirability bias in the answering behavior of respondents does not create a problem when using differences-in-differences. If some people tend to always over-rate recycling participation it would only affect ξ [c.f. Eq. (3c) below], which is allowed to correlate with treatment status.

The present paper strives to tackle this issue. In order to estimate respective treatment effects in an unbiased way we use a semi-parametric differencing approach. As outlined in more detail below, we exploit individual-level panel data from a natural experiment and complement DD with propensity score matching to account for selection into treatment and control group. Using this approach, we are able to account for unobserved heterogeneity due to (self-)selection into the treatment group or policy endogeneity. Compared to standard DD, a triple-differences (DDD) estimator allows accounting for additional possible bias due to time-variant heterogeneity over groups. Particularly, this method effectively takes care of potential induced over-reporting, where DD would produce upward biased results. On the other hand, DDD may be downward biased in the presence of spillover effects. Consequently, both estimators can be combined to derive upper and lower bounds of the true effect.

The paper is structured as follows. In Sect. 2 we present the conceptual framework for our analyses and derive testable hypotheses. Section 3 describes the data and methods used. This includes a description of the underlying research design and the resulting data and central variables used for analyses, the delineation of the pursued analytic strategy, as well as some notes on how propensity matching was performed. Section 4 starts with a discussion of pre-treatment recycling rates in treatment and control groups. Subsequently, we present our estimations of treatment effects by type of recyclable, combining results obtained from a DD and DDD approach. Finally, we investigate effect heterogeneity with regard to individual pre-treatment conditions and socio-economic factors as outlined above. We conclude with a summary and discussion of our findings.

2 Theory and Hypotheses

2.1 Conceptual Framework

Previous research has developed conceptual frameworks to analyze effects of features of curbside recycling on recycling participation. We draw upon a model proposed by Kinnaman and Fullerton (2000) commonly applied in recent research, sometimes with slight modifications (e.g. Beatty et al. 2007; Jenkins et al. 2003; Siddique et al. 2010). According to this class of models, households maximize a utility function over consumption and waste disposal, subject to a budget constraint incorporating prices for different disposal options. This maximization process then yields demand functions d_o for different disposal options o . Essentially, these take costs of recycling (p_r), garbage disposal (p_g), and illegal disposal like dumping or burning (p_b) as well as socio-demographic characteristics (σ), including income, as arguments:

$$d_o = f_o(p_r, p_g, p_b, \sigma), \quad (1)$$

where $o \in \{r, g, b\}$. Prices may include fees (or, in the case of illegal disposal, fines) but also time and effort associated with the respective disposal options. Time costs may themselves be a function of σ . Socio-demographic characteristics may also influence other cost aspects of recycling participation like the volume of recyclables, cost for individual storage, or for transportation. For example, single households are likely to have lower waste output than families with children. The system of equations in (1) can serve as the basis for our empirical analysis. Since the right-hand-side variables in each demand equation are identical, the system can be estimated employing separate equations without introducing bias.

Policy measures like the introduction of curbside recycling can be seen as constituting a quasi-experiment. However, since the assignment of households to treatment and control group is non-random, the underlying selection process has to be accounted for as well. Suppose that a community shifts from a drop-off system for recycling in period $t = 0$ to curbside recycling in period $t = 1$, which is assumed to reduce recycling costs p_r . Further suppose that the costs for other disposal options (i.e. p_g and p_b) as well as socio-demographic characteristics (σ) remain stable over time. The (marginal) curbside effect on the optimal level of recycling is then given by the difference in demand for recycling between $t = 0$ and $t = 1$ if the policy measure is exogenous and if there is no time trend in recycling:

$$\Delta d_r \equiv d_r^{t=1} - d_r^{t=0} = f_r(p_r^{t=1}, p_g, p_b, \sigma) - f_r(p_r^{t=0}, p_g, p_b, \sigma). \quad (2)$$

2.2 Hypotheses

Based on our conceptual framework and the demand function given in (1), we expect the reduction in the cost of recycling due to a curbside scheme to lead to an increase in recycling participation. The reduction in cost is due to lower effort required in terms of time, storage, and transport for constant monetary cost of recycling. We can therefore formulate

H1 The introduction of curbside recycling increases the level of recycling participation.

This reduction in cost, however, is not necessarily constant over all respondents. Rather, it can be assumed to vary depending on respondent and household characteristics σ as well as on characteristics of the prior bring scheme. Clearly, we can expect the cost reduction to be lower in areas where the grid of collection containers under the bring condition was dense and the average distance to the nearest collection site was lower. Therefore,

H2 The lower the distance to collection containers at time t_0 , the lower the effect of a curbside scheme.

Further, the effort of recycling participation varies between kinds of recyclables because of variations in storage and transport costs. Both should be higher for plastic and packaging as compared to paper, therefore

H3 The effect of curbside recycling is more pronounced for plastic and packaging as compared to paper.

In addition to testing these hypotheses we explore possible effect heterogeneity by several sociodemographic variables. Knowing, which kinds of households respond to the increased ease of recycling is of practical interest for policy-makers and will also shed some light on the generalizability of our findings.

3 Data and Methods

3.1 Research Design

In most regions of Germany, a two- or three-stream curbside system has been used for collecting recyclables since the 1990s: in addition to bins for residual waste, households have bins for paper, as well as other bins (or bags) for packaging materials (mainly plastic, Tetra Paks and metal cans), and sometimes yet others for glass. In other municipalities, the

collection of glass is organized as a drop-off scheme with containers at street corners. By law, the industry is responsible for the collection and recycling of paper, plastics, and glass. The cost of recycling is added to products' prices; the consumer, therefore, pays for the recycling of the packaging materials when buying packaged products—regardless of his/her decision to recycle or not. Due to the upstream waste tax on recyclables, actual *participation* in recycling activities is free of charge. Residual waste, however, is charged with a volume-based downstream tax.

The city of Cologne relied on a drop-off system with drop-off containers at street corners for all kinds of recyclables. In 2006, the waste management authorities commenced a stepwise implementation of curbside collection, accompanied by a simultaneous closing of the drop-off stations and hence the bring-in system. Between February 2006 and October 2007, the drop-off scheme for recyclable waste was replaced by a curbside recycling scheme for paper and packaging.³ In one city district after another, households received blue and yellow bins for the collection of paper and plastic/metal cans free of charge. In one neighborhood, *Lindenthal*, the curbside scheme had already been implemented during a pilot study a few years earlier. This stepwise implementation provided an opportunity to design a field-experimental study with one treatment group and two control groups. In this natural experiment, the change in collection systems can be considered a (quasi-)experimental treatment to modify the behavioral cost of recycling (p_r). As noted above, there was no change in the collection system for glass bottles.

The inhabitants of the district of *Nippes* served as the experimental group. In this district, the curbside recycling scheme took effect in September/October 2006. Control groups came from two districts not subject to any change in recycling scheme over the relevant period. Inhabitants of Cologne-*Innenstadt* served as the first control group, as curbside pickup had not been introduced in that district until September/October 2007. Cologne-*Lindenthal* served as the second control group; in this district paper and plastic had been picked up at the curbside for some years. The use of two control groups offers the advantage to include *never-treated* and *ever-treated* groups in the comparison and hence capture possible heterogeneity. When estimating DD and DDD we pooled these groups in order to provide a single estimate for the treatment effect.

3.2 Data and Central Variables

All analyses in this paper are based on a two-wave panel postal survey. The participants were randomly selected from the population register of Cologne, distributed equally across the three selected districts: *Nippes*, *Innenstadt* and *Lindenthal*. The survey was designed following Dillman's tailored-design method (Dillman 2000), using incentives and two follow-up reminders. The first panel wave was conducted during July/August 2006 (that is well before the introduction of curbside pickup in the study group) and yielded a response rate of 64%. The second panel wave followed in May/June 2007 (that is after the introduction of curbside in the study group but well before the control group *Innenstadt*) with a retention rate of 83%. Overall, 1567 persons provided sufficient information in both waves of the panel (*Nippes*: 507, *Innenstadt*: 491, *Lindenthal*: 569).

The questionnaire of the first wave comprised questions on socio-demographic individual and household characteristics, a number of questions on environmental attitudes, the location of the collection containers for recyclables, and a detailed account of recycling behavior. For

³ The curbside scheme was not used for the collection of glass. Rather, recyclable glass continued to have to be brought to drop-off containers by participants.

each of the types of recyclable (paper, glass and packaging), the frequency of participation in recycling was to be indicated on a four-point ordinal scale. This is supposed to capture an increase in recycling *to the expense* of residual waste output.⁴ Note that, while we measure total recycling, the curbside scheme and the drop-off system were provided mutually exclusively. Hence, by design drop-off recycling is fully crowded out by curbside recycling and any observed changes in recycling have to be interpreted as changes in total recycling. In this sense, crowding-out or “cannibalization” of alternative systems, as reported by Beatty et al. (2007) is not a problem in our study. For the purposes of this paper, recycling participation was dichotomized, with persons declaring that they “always” participated in recycling being coded as 1, the rest as 0. We chose to do so due to reported participation being highly skewed (see Table 6 in the “Appendix”). However, we also replicated our analyses for different specifications of the outcome variable and results turned out to be robust (see Table 7 in the “Appendix”). In the second wave, the measurement of recycling behavior, the location of collection containers, and environmental attitudes was replicated, employing the same questions as in the first wave.

3.3 Analytic Strategy

Recall Eq. (2) showing the difference in demand for recycling between the two collection schemes. In order to test our hypotheses, it is necessary to isolate the effect of p_r on d_r or, more precisely, of the availability of curbside recycling [D in Eq. (3a) below]. An unbiased identification of the treatment effect requires a variation in D , holding p_g , p_b , and σ (as well as other factors affecting p_r apart from D) constant. Formally, the general identification problem can be described by the equation system

$$Y_t = f(D_t, X_t) + \varepsilon_t, \quad \varepsilon_t = \alpha + u_t \quad (3a)$$

$$D_t = g(X_t) + \eta_t, \quad \eta_t = \lambda + v_t \quad (3b)$$

$$Y_t^R = Y_t + \mu_t, \quad \mu_t = \xi + \omega_t, \quad (3c)$$

where (3a) is the outcome equation for recycling participation Y and (3b) gives the treatment assignment equation. Let X denote a set of observable factors which may affect Y as well as D , e.g. socio-economic characteristics like age, education, or income. ε_t and η_t reflect unaccountable variation in Y and D due to unobserved factors. The usual identifying assumption is conditional independence of the error terms, i.e. $E(\varepsilon_t \eta_t) = 0$, where ε_t can be conceptually decomposed into a time-invariant component α and a time-varying component u_t . As we have variation in D over time—induced by the implementation of a curbside scheme—we can use difference-in-difference estimation to considerably relax this assumption. At the same time, we can eliminate spurious effects due to aggregate changes in environmental awareness, large-scale policy changes, etc., as these would equally affect all three Cologne districts. The remaining identifying assumption is exogeneity of time-varying idiosyncratic errors, $E(u_t v_t) = 0$. Within the DD framework, this corresponds to the well-known parallel

⁴ Official aggregate-level administrative data back the validity of our self-reported measure: Between 2005 (i.e. the pre-treatment year) and 2007 (by which curbside collection was implemented city-wide) recycling output in Cologne increased substantially while at the same time residual output decreased, resulting in a 3% drop in total waste output (for details, see Table 5 in the “Appendix”).

trends assumption that in the absence of treatment the outcomes in treatment and control groups would have changed identically. In practice, (3a) is often estimated by means of regression analysis, which involves not only a selection of observed time invariant Xs but also assumptions about the functional form of $f(\cdot)$.

Not accounting for heterogeneity in X would not lead to bias as it would be captured by α , which is allowed to be correlated with the treatment. However, in the presence of non-parallel time trends for treated and untreated groups, DD would give a biased estimate. In our setting, this will be the case if changes in recycling participation unrelated to changes in the recycling scheme are a function of initial conditions that also influenced the likelihood of treatment assignment. Estimates will also be biased if certain groups react differently to the introduction of curbside collection than others and group membership is systematically related to the district of residence—and thus to treatment assignment. This concern can be easily addressed within the DD framework, e.g. by adding X as additional controls. Alternatively, a combination of DD with propensity score matching methods can be employed (Heckman et al. 1997; Ravaillon and Chen 2005). An advantage of this approach, e.g. vis-à-vis alternative regression methods, is that it is a largely non-parametric method of controlling for initial heterogeneity, thus avoiding potential bias due to a misspecification of the functional form of $f(\cdot)$.

Following this approach, we estimated DD and adjusted for pre-treatment differences in study and control groups using propensity score matching (Caliendo and Kopeinig 2008; Rosenbaum and Rubin 1983, 1985). The resulting DD matching algorithm (Gangl 2006; Heckman et al. 1997) provides a nonparametric estimate of the average effect of offering curbside recycling on the propensity to recycle material M in the treatment group

$$DD_M = \frac{1}{N_{E_1 \cap S}} \sum_{i \in E_1 \cap S} \left[\Delta Y_{i,M,T+1}^1 - \sum_{j \in E_0 \cap S} W_{ij} \Delta Y_{j,M,T+1}^0 \right], \tag{4}$$

where $\Delta Y_{i,M,T+1}^d = Y_{i,M,T+1} - Y_{i,M,T} | D = d$; E_d , treatment ($d = 1$) or control ($d = 0$) sample; S , area of common support; $N_{E_1 \cap S}$, number of observations in treatment group in the area of common support; D , causal factor of interest (curbside introduction); W_{ij} , kernel weight (Epanechnikov Kernel); and M , type of recyclable (paper, packaging, glass).

We used kernel matching (see Heckman et al. 1998) because of its relatively high efficiency and the possibility of bootstrapping standard errors of the treatment effect (see Abadie and Imbens 2008).⁵ We estimated the propensity score in a multinomial probit model. The selection model used cohabitation, presence of children in household, number of persons in household, education, labor-force participation, age, gender, income, nationality, migration background, environmental attitudes, and type of dwelling as covariates (see Table 4 in the “Appendix”). Common support of all values of the propensity score and kernel matching led to a very good adjustment of the propensity score distributions in treatment and control groups (see Fig. 2 in the “Appendix”). After matching, no statistically significant differences in the covariates of the selection model between the study and control groups remained. The standardized bias was below 2.5% for all matching variables except part-time employment (4.1%).

Assume that the model described in (4) yields the unbiased effect of curbside recycling on recycling behavior. Substituting *reported* for *actual* behavior [as in (3c)] would still yield an

⁵ Presented results are based on the Epanechnikov Kernel and a bandwidth of 0.06. We tested different varieties of propensity score matching (different bandwidth with the Epanechnikov Kernel (0.06 ± 0.04), different kernel types (gaussian, uniform, biweight), nearest neighbor matching, logit selection model instead of probit), and the results are robust against changes of these specifications.

unbiased estimate of the curbside effect if reported behavior is a function of actual behavior and the measurement error (or, more precisely, its time-varying component) is conditionally independent of the treatment, $E(\omega_t \nu_t) = 0$.⁶ With survey data, this assumption will, however, be violated in the presence of social desirability bias as the introduction of curbside recycling may act as a signal for recycling being socially desired. Consequently, DD estimates will be upward biased.

To tackle this problem, a DDD approach can be used, exploiting the fact that curbside collection was not offered for all materials. We constructed a triple-difference estimator by subtracting changes in participation in the collection of a recyclable *unaffected* by the introduction of a curbside scheme, namely glass. Equation (4) can easily be extended to obtain a DDD based estimation of the treatment effect from

$$DDDA = \frac{1}{N_{E_1 \cap S}} \sum_{i \in E_1 \cap S} \left[(\Delta Y_{i,A,T+1}^1 - \Delta Y_{i,U,T+1}^1) - \sum_{j \in E_0 \cap S} w_{ij} (\Delta Y_{j,A,T+1}^0 - \Delta Y_{j,U,T+1}^0) \right], \quad (5)$$

where A, type of recyclable affected by curbside introduction (paper, packaging); U, recyclable unaffected by curbside introduction (glass).

In the presence of a treatment-induced change in *over-reporting* model (5) will yield an unbiased estimate if this is equal for reports on all recyclables where model (4) required the complete absence of such effects. However, this comes at the cost of ruling out possible spillover effects by assumption: Curbside collection of paper and packaging might lead to an increase in actual glass recycling (e.g. due to increased recycling awareness or increased capacity to recycle glass at the container). So, while DD may be systematically upward biased due to the outcome variable being self-reported, DDD may be systematically downward biased due to over-control in the presence of possible spillover effects. As both estimators are immune to the other's respective weakness, they can be combined to derive upper and lower bounds of the true effect.

4 Results

In the following sections, we present the results of our empirical study. We start with a brief discussion of the pre-treatment setting, then identify the treatment effect and finally evaluate the elasticity of the treatment effect with regard to characteristics of the bring scheme.

4.1 Pre-treatment-Setting

In 2006, paper and packaging were collected at the curbside in *Lindenthal*, but had to be brought to containers at street corners in *Innenstadt* and *Nippes*. Glass was collected in a bring scheme in all three districts. As can be seen from Table 1, recycling rates differ substantially among materials and neighborhoods. Overall, recycling participation was relatively low for packaging (ca. 38–62%), as compared to glass (70–78%) and paper (71–86%). For all materials, rates were lowest in *Innenstadt* and highest in *Lindenthal*.

We also found that cross-district variation in rates is lowest for glass and substantially higher for packaging. As at T_0 curbside recycling had already been in place in *Lindenthal*

⁶ Note that standard errors will usually be larger when relying on reported instead of actual behavior due to random measurement error, leading to reduced statistical power. However, this qualification applies to survey research in general and we see no reason why there should be particular noise in the case of an everyday behavior like waste disposal.

Table 1 Recycling participation T_0 ("always recycling")

	Paper	Packaging	Glass	N
C1 (Innenstadt)	.705	.379	.701	491
C2 (Lindenthal)	.859	.624	.784	569
T (Nippes)	.787	.539	.771	507
Total	.788	.520	.756	1567

Table 2 Changes in recycling participation (D and DD, matched data)

	Paper	Packaging	Glass	N
C1 (Innenstadt)	.036	-.015	.022	491
C2 (Lindenthal)	.027	-.022	.002	569
T (Nippes)	.081	.178	.043	507
DD	.053 (.022)**	.196 (.028)***	.035 (.022)	1567

Bootstrapped standard errors in parentheses

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

for paper and plastic; these results could be regarded as a first indicator for the effectiveness of a curbside scheme: Not only were participation rates highest in the district with a curbside scheme, but the variation in rates by material was also larger when there was a variation in the recycling scheme (as opposed to glass recycling). Clearly, such an inference would be premature, as the results could well be due to endogenous policy or unobserved heterogeneity. Valid identification of the treatment effect of curbside recycling therefore requires that we turn to a discussion of changes over time in relation to a policy change.

4.2 Estimation of the Treatment Effect

Table 2 presents changes in recycling participation from 2006 to 2007 in the treatment and control groups using matched data. As the collection scheme for paper and packaging changed in *Nippes* from a bring to a curbside collection, one would expect an increase particularly in this group. Data shows that this indeed is the case: The increase in recycling participation is clearly the highest in *Nippes* for all materials. We estimate a DD of 5% points for paper recycling and 20% points for packaging, both statistically significant (using bootstrapped standard errors). For glass, there was no change in the collection system between observations; there were thus no immediate cost reductions for its recycling. However, as outlined above, a positive effect could result from possible spillover effects or over-reporting. The DD estimator for glass is estimated at 3.5% points, but statistically not significant. So, neither spillover nor over-reporting effects seem to be very large.

While the DD estimates can be considered upper bounds for the true effects, the remaining DD effect for glass recycling can be used to calculate a more conservative estimate or lower bound, the DDD. It further relaxes the identifying assumptions with regards to the self-reported outcome measure but also eliminates possible spillover effects, leading to an underestimation of the true effect. After further differencing we obtained a DDD of 0.161 (0.031) for packaging and of 0.018 (0.025) for paper recycling (see line total in Table 3 below). The 95% confidence bounds of the combined effect are given in brackets. These results indicate that the implementation was far more effective for packaging than for paper. While the average treatment effect on plastic recycling is substantial and statistically signif-

Table 3 DD and DDD by distance to container (t_0), matched data

	Paper		Packaging		Glass
	DD	DDD	DD	DDD	DD
Total	.053** (.022) [-.031 .096]	.018 (.025)	.196*** (.028) [.098 .252]	.161*** (.032)	.035 (.022) [-.000 .070]
Low distance (0–100 m)	.022 (.031) [-.097 .083]	-.026 (.036)	.139*** (.038) [.011 .213]	.098** (.044)	.041 (.028) [-.005 .086]
High distance (above 100 m)	.075*** (.028) [-.010 .129]	.049 (.031)	.235*** (.036) [.131 .305]	.204*** (.037)	.031 (.027) [-.013 .076]
Difference (low-high)	-.053 (.039)	-.075* (.044)	-.096** (.047)	-.106** (.050)	.009 (.035)

Bootstrapped standard errors in parentheses. 95% confidence bounds in brackets. N=1567. *Note:* for some respondents the distance to collection containers differs between materials
 * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

icant, we cannot reject the null of no effect on paper recycling if we cannot credibly rule out that treated respondents over-reported their recycling behavior.

4.3 Heterogeneity in the Treatment Effect

Results in Table 3 further indicate heterogeneity in the effect of curbside recycling depending on the type of recyclable considered, which, as we have argued, exhibit different storage and transportation costs. Our data allowed us to investigate further possible heterogeneity across strata within the treatment group. The conditions prior to implementation of a curbside scheme were not identical for all households. Rather, some had to bring their recyclables to a distant container while others found a collection container next to their house. Therefore, the reduction in behavioral cost, $p_r^{t=1} - p_r^{t=0}$, was lower for the latter group than for the former.⁷ This gave us the opportunity to test the relative effectiveness of curbside recycling with different configurations of the bring scheme for different types of materials empirically.

Table 3 disaggregates the treatment effect of curbside recycling for prior distances to a collection container of 0–100 and more than 100 m. For paper, our estimates are too imprecise to draw a clear conclusion, but distance seems to play a role. We found the effect to be statistically significant only in the high-distance condition and only when estimating DDs. We thus cannot rule out that this is a true effect and not merely driven by response bias. The effect reduced to 4.9% points when estimating DDDs which was still significantly higher than in the low-distance condition. Regarding plastic recycling, we found a substantially strong and statistically significant treatment effect under both conditions. However, the effect was substantially (+ 10% points) and significantly stronger when the prior distance—and hence behavioral cost—had been high. Hence, the effectiveness of the introduction of curbside recycling varies greatly with materials and the *status quo ante*. For paper, a curbside scheme

⁷ For glass, distance to container does not capture a reduction but an increase in behavioral cost. In the presence of substantial spillover effects, we would thus expect the treatment effect to be greater in the low distance condition. For over-reporting, distance to container should be irrelevant.

does not seem to offer advantages over a dense grid of collection containers. For the collection of packaging, on the contrary, collection containers at street corners seem unsuitable. Here, a curbside scheme outperforms the bring scheme even under the condition of a dense grid of collection containers.

Our theoretical framework also allows for heterogeneity in the curbside effect with regard to socioeconomic characteristics σ . In the presence of effect heterogeneity, our findings might have little external validity. We therefore ran our matching analyses separately by strata of available socioeconomic background variables. Figure 1 shows how different kinds of households respond to the increased ease of recycling. Each panel displays three graphs, one for each type of recyclable, and each graph depicts the estimated ATTs from the stratified DD- (upper part) and DDD-matching models (lower part), respectively. Unfortunately, cell sizes did not allow for more fine-grained analyses.

By trend, curbside effects are larger when there are children living in the household. For paper, the effect is also only significant for households with children. This would be in line with benefitting more from curbside collection due to larger waste output of families with children. For housing situation, income, and educational level the patterns are less clear. Note that all differences between these sociodemographic strata are statistically insignificant (for details, see Table 8 in the “Appendix”). However, we found that pronounced pro-environmental attitudes lead to a weaker response to curbside recycling, whereas particularly those with intermediate attitudes react strongest. The difference is statistically significant for packaging ($p < 0.05$). This is in line with previous findings (Guagnano et al. 1995; Best and Kneip 2011) and reflects the relative unimportance of behavioral costs in the presence of strong attitudes.

5 Summary and Discussion

This study examined the effect of implementing a curbside scheme on participation in household waste recycling. Using a quasi-experimental design and a unique individual-level dataset, we estimated the treatment effect of a reduction in required effort, and thus a reduction in behavioral cost, on the probability of recycling activities. Our data also allowed us to additionally investigate treatment-effect heterogeneity.

We used a difference-in-differences matching strategy to identify the causal effect of implementing a curbside recycling scheme on recycling behavior. While methods which allow for causal identification under even weaker assumptions are generally available, they usually require multiple pre- and post-treatment measurements of the outcome (Winship and Morgan 1999; Lee 2016). The DD matching estimator is analogous to the standard DD regression estimator; however, it also accounts for selection into treatment based on observable characteristics without imposing functional restrictions in estimating the conditional expectation of the outcome variable. Our approach thus combines the advantages of differencing and matching methods. However, when using survey data and self-reports, the DD estimator may be upward biased due to over-reporting. By using triple differences, we can tackle this issue and additionally account for further unobserved heterogeneity. If treatment leads to over-reporting recycling participation but does so for *all* recyclables, DDD effectively controls for this by ruling out any effect on recycling of materials where recycling costs have not been affected by the treatment under study. The DDD estimator, however, may be downward biased in the presence of spillover-effects. The combination

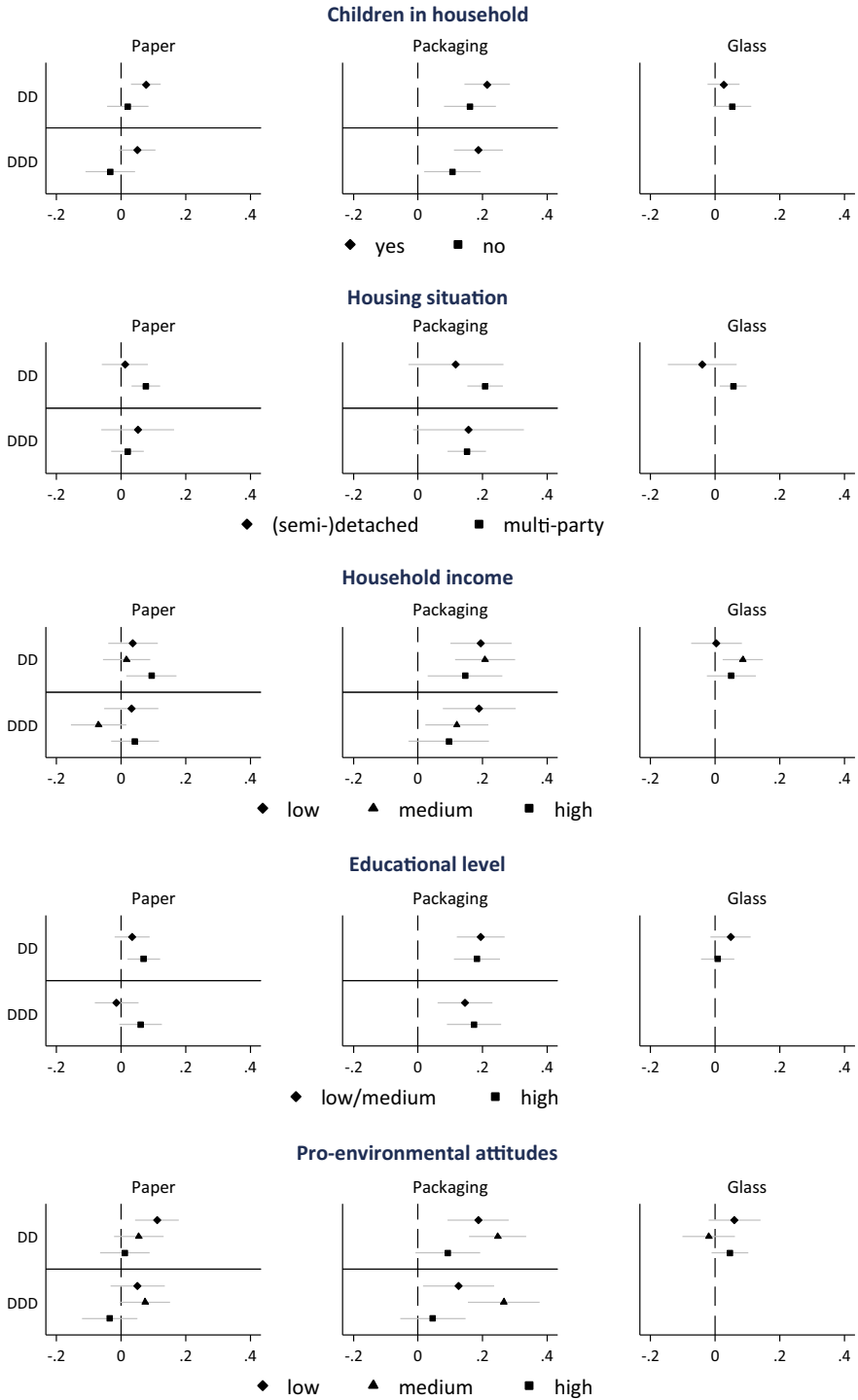


Fig. 1 Treatment effect heterogeneity by socioeconomic background

of DD and DDD estimators allows us to derive upper and lower bounds of the treatment effect.

After controlling for aggregate trends, unobserved heterogeneity, and self-selection into study and control groups and allowing for treatment induced over-reporting as well as possible spillover effects between differently affected recyclables, we found a substantial overall increase in recycling participation of between 10 and 25% points for packaging but no significant effect for paper. The effect of curbside recycling varied with distance to a collection container under the preceding bring scheme condition: For packaging, the estimated treatment effect is 10% points larger when the distance to the nearest collection container had been more than 100 m before curbside collection was introduced. The effect on paper recycling was generally smaller and never statistically significant. The DD estimator provides a significant effect also for paper under the high-distance conditions, though, amounting to about 7.5% points. However, we cannot rule out—other than by assumption—that this is a mere response effect and does not reflect actual behavior. We found no further systematic evidence for effect heterogeneity due to sociodemographic factors. Only environmental attitudes shaped the effect insofar that persons with strong pro-environmental attitudes reacted least to the treatment (this particular finding may well be due to a ceiling effect).

Despite our efforts to identify the causal effect of curbside recycling, potential shortcomings remain and must be addressed. To begin with, the estimation still relies on the assumption of exogeneity of time-varying idiosyncratic errors. Conditional independence of treatment and outcomes is required, ruling out endogenous selection into treatment based on agents' pre-treatment trends or predictions about treatment impact. This will be the case if treatment and control groups differ on unobserved characteristics not reflected in our propensity score model, which would have led to non-parallel time trends in recycling if there had been no implementation of curbside collection or which affected the likelihood of curbside collection being introduced in the treatment (rather than the control) group.

We have suggested combining DD and DDD to derive estimation bounds in order to jointly relax assumptions on treatment-related over-reporting and spillover effects otherwise necessary. Spillover effects could occur if, e.g., an increase in participation in glass recycling came along as a by-product of an increased participation in recycling other goods and should thus be considered as due to the introduction of curbside recycling. However, this comes at the cost of precision and statistical power. Another possible problem with our estimates relates to the distribution of our outcome variable. We consider the binary outcome of recycling participation versus non-participation so that the treatment effect can be interpreted as (additive) marginal effect on the participation rate. Given a pre-treatment participation rate of about 80% with regard to paper, some ceiling effect might occur. Thus, the differences in effect size for paper as compared to packaging could, to some extent, be a result of differential baseline probabilities of recycling participation.

A related concern pertains to the generalizability of our findings. We have argued that the effect of curbside recycling should vary over types of recyclables as well as over pre-treatment recycling options inasmuch as both affect relative cost savings for recycling. In our case, the implementation of the curbside scheme occurred when a rather well-planned bring scheme had been running for several years. In the presence of treatment-effect heterogeneity with regard to other factors, estimates may have limited external validity. Consequently, the treatment effect could only be interpreted locally, i.e. with regard to the specific characteristics of the population in the treatment group in our analysis. However, stratified analyses by several dimensions of socioeconomic background revealed little evidence for large effect heterogeneity, with the exception of a smaller curbside effect for the more environmentally concerned. This points to the importance of background characteristics of the

population under study, which go beyond socio-demographics and will usually require survey data.

Hence, our results point to some important implications for the implementation of environmental policies. We showed that curbside recycling can, in many circumstances, be an effective tool to enhance recycling rates—even when compared to an extended bring scheme. This is especially the case when the effort required for storage and transport of recyclables is substantial, as is the case with packaging (e.g. cans, yoghurt jars, Tetra Paks, plastic wrapping, and boxing). Such materials are quite bulky and therefore require substantial space to store and may additionally lead to nuisances either due to bad smells or, alternatively, to the necessity of frequent transport to the collection container. For paper, which is simple to store and relatively easy to transport, our results indicate that a bring scheme with a reasonably dense grid of collection containers may be as effective as a curbside collection scheme.

Acknowledgements This work was supported by the Fritz-Thyssen-Foundation, Cologne, Germany. We thank Tobias Rüttenauer, Susumu Shikano, Martin Spindler, several anonymous reviewers and the editor for helpful comments.

Appendix

See Fig. 2 and Tables 4, 5, 6, 7 and 8.

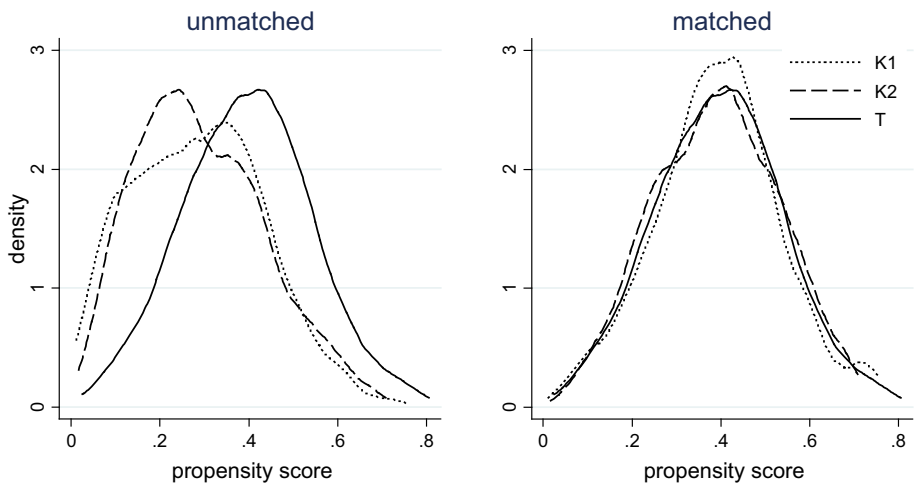


Fig. 2 Common support and distribution of propensity scores

Table 4 Selection models for estimating the propensity score

	Innenstadt		Lindenthal	
	b	se	b	se
Intercept	-0.179	(0.653)	1.568	(0.611)
Age (deciles) (ref 18–32)				
32–36	-0.914***	(0.264)	-0.974***	(0.268)
37–40	-1.281***	(0.269)	-1.104***	(0.272)
41–44	-1.275***	(0.269)	-1.177***	(0.269)
45–48	-1.247***	(0.281)	-1.214***	(0.281)
49–53	-1.487***	(0.292)	-1.273***	(0.284)
54–60	-1.133***	(0.286)	-0.755***	(0.280)
61–66	-1.101***	(0.309)	-0.721**	(0.304)
67–73	-1.423***	(0.323)	-0.685**	(0.315)
74–93	-1.340***	(0.338)	-0.356	(0.325)
Missing	-1.275***	(0.295)	-0.767***	(0.288)
Female	0.095	(0.115)	0.228**	(0.111)
Secondary education (ref low)				
Medium	0.128	(0.172)	0.477***	(0.168)
High	0.374**	(0.183)	0.755***	(0.179)
Missing	0.272	(0.447)	0.900**	(0.425)
University degree	-0.052	(0.148)	0.191	(0.143)
Missing	-0.021	(0.331)	-0.459	(0.355)
Employment (ref unemployed)				
Full time employed	-0.083	(0.157)	-0.068	(0.152)
Part time employed	-0.394**	(0.192)	-0.272	(0.185)
Missing	-0.308	(0.299)	-0.337	(0.289)
Income group (ref no income)				
> 400 €	-0.079	(0.550)	-0.545	(0.590)
400–749 €	-0.130	(0.408)	-0.140	(0.398)
750–999 €	-0.158	(0.345)	-0.734**	(0.351)
1.000–1.249 €	0.381	(0.334)	-0.427	(0.341)
1.250–1.499 €	0.170	(0.299)	-0.124	(0.284)
1.500–1.749 €	0.502	(0.308)	-0.203	(0.303)
1.750–2.999 €	0.163	(0.293)	-0.039	(0.275)
2.000–2.249 €	-0.069	(0.305)	-0.327	(0.284)
2.250–2.499 €	0.190	(0.300)	0.014	(0.281)
2.750–3.999 €	0.174	(0.315)	-0.118	(0.293)
3.000–3.249 €	0.424	(0.312)	-0.230	(0.302)
3.250–3.499 €	0.154	(0.329)	-0.138	(0.304)

Table 4 continued

	Innenstadt		Lindenthal	
	b	se	b	se
3.500–3.749 €	0.999***	(0.362)	0.670**	(0.340)
3.750–4.999 €	0.734**	(0.354)	0.311	(0.327)
4.000–4.249 €	0.622*	(0.366)	0.230	(0.334)
4.250–4.499 €	0.698*	(0.396)	0.535	(0.261)
4500 € and more	0.760***	(0.288)	0.517**	(0.261)
Missing	0.198	(0.369)	–0.223	(0.349)
Cohabitation				
Yes	–0.057	(0.155)	0.092	(0.151)
Missing	0.665	(0.570)	0.030	(0.589)
Children in HH				
No	–0.051	(0.139)	0.179	(0.135)
Missing	–0.624	(0.762)	–0.914	(0.712)
Persons in HH	–0.072	(0.076)	0.005	(0.073)
German citizen	–0.695***	(0.272)	–0.323	(0.273)
Migration background (parents)	–0.257	(0.187)	–0.239	(0.185)
Apartments (ref detached)				
2–4 apartments	1.347***	(0.235)	–0.237	(0.180)
5–10 apartments	1.700***	(0.208)	–0.020	(0.147)
11–15 apartments	1.947***	(0.248)	–0.025	(0.205)
16+ apartments	1.712***	(0.258)	–0.041	(0.216)
Specific environmental concern	0.040	(0.099)	–0.171*	(0.095)
General environmental concern	0.027	(0.108)	–0.139	(0.102)

N = 1567. Coefficients from multinomial probit; standard errors in parentheses

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

Table 5 Development of waste output in Cologne 2005–2007

Year	Recyclables				Residual waste	Total waste
	Paper	Packaging	Glass	Total		
2005	52,646	17,509	17,767	87,922	314,279	402,201
2006	55,481	18,640	18,134	92,255	306,794	399,049
2007	59,877	19,610	18,077	97,564	293,375	390,939
Absolute change	7231	2101	310	9642	−20,904	−11,262
Relative change	13.7%	12.0%	1.7%	11.0%	−6.7%	−2.8%

Source: Office for Environment and Consumer Protection Cologne; <http://www.stadt-koeln.de/mediaasset/content/pdf57/9.pdf> (Last accessed 22 March 2018)

Table 6 Proportions of materials recycled, t1

Material	Percentage and number of respondents recycling...				Total
	Never	Sometimes	Often	Always	
Paper	5.42%	5.49%	10.34%	78.75%	1567
	85	86	162	1234	
Packaging	19.27%	12.76%	16.02%	51.95%	1567
	302	200	251	814	
Glass	6.45%	7.4%	10.53%	75.62%	1567
	101	116	165	1185	

Note: The respective question asked was “some materials in your waste can be recycled, e.g. glass, paper, cans or plastic. We would like to know whether you collect these materials separately in your household. How often do you separate the following materials?” (translated from German original)

Table 7 Robustness by specification of outcome variable

	Paper		Packaging		Glass
	DD	DDD	DD	DDD	DD
Always versus rest	.053** (.022)	.018 (.025)	.196*** (.028)	.161*** (.032)	.035 (.022)
Always/often versus rest	.054*** (.016)	.009 (.017)	.151*** (.024)	.106*** (.025)	.044** (.019)
Linear	.117*** (.038)	.019 (.039)	.403*** (.057)	.305*** (.056)	.098** (.042)

Bootstrapped standard errors in parentheses

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

Table 8 Treatment effect heterogeneity by socioeconomic background

		Paper		Packaging		Glass
		DD	DDD	DD	DDD	DD
<i>Children in household</i>						
(1)	Yes	.077*** (.028)	.050 (.034)	.214*** (.043)	.187*** (.046)	.027 (.031)
(2)	No	.020 (.039)	-.033 (.047)	.161*** (.049)	.107** (.053)	.054 (.036)
(1)-(2)		.057 (.049)	.083 (.058)	.053 (.064)	.080 (.069)	-.027 (.048)
<i>Housing situation</i>						
(1)	(Semi-)detached	.012 (.043)	.052 (.069)	.117 (.090)	.157 (.104)	-.040 (.065)
(2)	Multi-party	.076*** (.027)	.020 (.031)	.208*** (.034)	.151*** (.036)	.056** (.025)
(1)-(2)		-.064 (.052)	.031 (.075)	-.090 (.095)	.006 (.109)	-.096 (.070)
<i>Household income</i>						
(1)	Low	.036 (.046)	.032 (.051)	.194*** (.058)	.189*** (.068)	.005 (.047)
(2)	Medium	.017 (.045)	-.070 (.052)	.207*** (.056)	.120** (.059)	.087** (.038)
(3)	High	.094* (.047)	.043 (.045)	.146** (.070)	.095 (.076)	.051 (.046)
(1)-(2)		.019 (.066)	.101 (.071)	-.013 (.080)	.069 (.091)	-.082 (.062)
(1)-(3)		-.057 (.064)	-.011 (.068)	.049 (.091)	.095 (.100)	-.046 (.066)
(2)-(3)		-.077 (.065)	-.113 (.069)	.061 (.088)	.025 (.094)	.036 (.058)
<i>Educational level</i>						
(1)	Low/medium	.034 (.033)	-.014 (.041)	.194*** (.045)	.146*** (.051)	.049 (.038)
(2)	High	.070** (.031)	.060 (.040)	.182*** (.044)	.173*** (.051)	.009 (.031)
(1)-(2)		-.035 (.045)	-.075 (.058)	.012 (.062)	-.027 (.072)	.040 (.050)
<i>Pro-environmental attitudes</i>						
(1)	Low	.111*** (.041)	.051 (.051)	.186*** (.058)	.126* (.067)	.060 (.049)
(2)	Medium	.055 (.047)	.074 (.047)	.247*** (.054)	.266*** (.068)	-.019 (.049)

Table 8 continued

		Paper		Packaging		Glass
		DD	DDD	DD	DDD	DD
(3)	High	.011 (.046)	−.035 (.052)	.092 (.061)	.046 (.061)	.046 (.034)
(1)–(2)		.056 (.062)	−.023 (.071)	−.061 (.079)	−.140 (.096)	.080 (.072)
(1)–(3)		.100* (.061)	.086 (.074)	.094 (.085)	.080 (.093)	.014 (.061)
(2)–(3)		.044 (.067)	.109 (.071)	.154* (.080)	.220** (.092)	−.066 (.061)

Bootstrapped standard errors in parentheses

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

References

- Abadie A, Imbens GW (2008) On the failure of the bootstrap for matching estimators. *Econometrica* 76:1537–1557
- Abrate G, Erbetta F, Fraquelli G, Vannoni D (2014) The costs of disposal and recycling: an application to Italian municipal solid waste services. *Reg Stud* 48:896–909
- Ando AW, Gosselin AY (2005) Recycling in multifamily dwellings: does convenience matter? *Econ Inq* 43:426–438
- Bartelings H, Dellink RB, van Ierland EC (2004) Modeling market distortions in an applied general equilibrium framework: the case of flat fee pricing in the waste market. In: Van den Bergh JCJM, Janssen MA (eds) *Economics of industrial ecology*. MIT Press, Cambridge
- Beatty TKM, Berck P, Shimshack JP (2007) Curbside recycling in the presence of alternatives. *Econ Inq* 45:739–755
- Best H, Kneip T (2011) The impact of attitudes and behavioral costs on environmental behavior: a natural experiment on household waste recycling. *Soc Sci Res* 40:917–930
- Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *J Economic Surv* 22:31–72
- Dillman DA (2000) *Mail and internet surveys: the tailored design method*, 2nd edn. Wiley, New York
- Dur R, Vollaard B (2015) The power of a bad example: a field experiment in household garbage disposal. *Environ Behav* 47:970–1000
- EPA (2015) *Advancing sustainable materials management: facts and figures 2013*. United States Environmental Protection Agency, Washington, DC. https://www.epa.gov/sites/production/files/2015-09/documents/2013_advncng_smm_rpt.pdf. Last Accessed 22 March 2018
- Fullerton D, Kinnaman TC (1995) Garbage, recycling, and illicit burning or dumping. *J Environ Econ Manag* 29:78–91
- Fullerton D, Kinnaman TC (1996) Household responses to pricing garbage by the bag. *Am Econ Rev* 86:971–984
- Gangl M (2006) Scar effects of unemployment: an assessment of institutional complementarities. *Am Sociol Rev* 71:986–1013
- Gangl M (2010) Causal inference in sociological research. *Annu Rev Sociol* 36:21–47
- Guagnano GA, Stern PC, Dietz T (1995) Influences on attitude-behavior relationships. A natural experiment with curbside recycling. *Environ Behav* 27:699–718
- Heckman JJ, Ichimura H, Todd PE (1997) Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Rev Econ Stud* 64:605–654
- Heckman JJ, Ichimura H, Todd PE (1998) Matching as an econometric evaluation estimator. *Rev Econ Stud* 65:261–294
- Hopewell J, Dvorak R, Kosior E (2009) Plastics recycling: challenges and opportunities. *Philos Trans R Soc B* 364:2115–2126

- Jenkins RR, Martinez SA, Palmer K, Podolsky MJ (2003) The determinants of household recycling: a material specific analysis of recycling program features and unit pricing. *J Environ Econ Manag* 45:294–318
- Kinnaman TC, Fullerton D (2000) Garbage and recycling with endogenous local policy. *J Urban Econ* 48:419–442
- Kuo Y-L, Perrings C (2010) Wasting time? Recycling incentives in urban Taiwan and Japan. *Environ Resour Econ* 47:423–437
- Lee M (2016) Generalized difference in differences with panel data and least squares estimator. *Sociol Methods Res* 45:134–157
- Ravaillon M, Chen S (2005) Hidden impact? Household saving in response to a poor-area development project. *J Public Econ* 89:2183–2204
- Reschovsky JD, Stone SE (1994) Market incentives to encourage household waste recycling: paying for what you throw away. *J Policy Anal Manag* 13:120–139
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55
- Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 39:33–38
- Sidiqie SF, Joshi SV, Lupi F (2010) Factors influencing the rate of recycling: an analysis of Minnesota counties. *Resour Conserv Recycl* 54:242–249
- Tsai T-H, Sheu S-J (2009) Will unit-pricing enhance recycling? *Int J Sustain Dev World Ecol* 16:102–108
- van den Bergh JCJM (2008) Environmental regulation of households: an empirical review of economic and psychological factors. *Ecol Econ* 66:559–574
- Winship C, Morgan SL (1999) The estimation of causal effects from observational data. *Ann Rev Sociol* 25:659–706
- World Bank (2012) What a waste. A global review of solid waste management. Urban development series knowledge papers. World Bank, Washington, DC