Factors influencing the rate of sorted solid waste collection: An empirical analysis in Catalonian municipalities

L. IGNACIO SALDIVIA GONZATTI
LSALDIVI@UCM.ES

GRADO EN ECONOMÍA. FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
UNIVERSIDAD COMPLUTENSE DE MADRID

Trabajo de Fin de Grado/ Bachelor's Degree Dissertation

2020-2021
Convocatoria: Junio

Supervisor:
Jesús Barreal

Consent of dissemination

L. Ignacio Saldivia Gonzatti

June 17, 2021

I, the undersigned, enrolled in the Bachelor's degree of Economics at the faculty of Economics and Business, hereby authorize the Complutense University of Madrid (UCM) to distribute and make academic, non-commercial use of the Bachelor's degree dissertation: "Factors influencing the rate of sorted solid waste collection: An empirical analysis in Catalonian municipalities", written during the academic year 2020-2021 under the supervision of Jesús Barreal, from the Financial, Actuarial and Statistics department. Additionally, I authorize the UCM library to store it in the E-Prints Complutense Institutional Archive with the objective of increase its distribution, use, and impact on the internet, and ensuring its conservation and long-term access.

Abstract

A study of the factors influencing the recycling behaviour is crucial to assess the current waste management schemes and to determine additional, appropriate policies to implement. By estimating several econometric models on longitudinal data of Catalonian municipal solid waste covering the period 2000-2019, we demonstrate and quantify the geographic, demographic, and socioeconomic determinants of sorted waste collection rates at a municipal level. With respect to women population, absence of children in the household, population density, education, income, presence of door-to-door collection services, and political preferences, the results are consistent with prior findings. However, new insights are obtained on the effects of geographic conditions, unemployment, and foreign population on recycling behaviours.

The empirical evidence suggests that further recycling policies are needed to increase the rate of sorted waste collection rapidly enough to achieve the recycling goal set by the European Union legislation and the Spanish Waste Law. Additionally, the study highlights the importance of considering demographic trends and differences in recycling behaviour between social groups when tackling waste management issues.

Key words
Municipal solid waste
Recycling behaviour
Waste management
Longitudinal data

Table of Contents

Table of Contents I

Acknowledgements II

Dedication III

1 Introduction 1

2 Theoretical expectations and data 5

3 Methodology 13
- 3.1 Model 13
- 3.2 Imputation methods assessment 16

4 Results 20

5 Discussion 24

6 Conclusion 26

Bibliography 30

A Regressions 31

Acknowledgements

I would like to show my gratitude to D.F.S.G., who helped me keeping on track when disorientation hovered. Also, thanks to all of those who listened to, and discussed with me along the way. Lastly, thanks to all of those cited in this dissertation and to the Stack Overflow and the Stack Exchange communities.

Dedication

Dedicated to R.M.G.S., who used to wake up early on her days off to sort and recycle waste when no one else did; and to all of those who, like her, do their bit because they care. Thus, dedicated to those who care.

"Das größte Geschenk, das Sie anderen geben können, ist das Beispiel Ihres eigenen Lebens"

— Bertolt Brecht

1 Introduction

Recycling, as we know it today, has been around for almost a century. Since World War II, when the U.S. government encouraged citizens to salvage materials for the war (Cooper, 2008), there have been initiatives to reuse waste. However, since single-use items hit the market in the 50's, these initiatives have consistently fallen short. The rate of recycling in the European Union increased 154% from 1995 to 2019. And yet, only 47.7% of the municipal solid waste (MSW) was recycled in 2019 (European Statistical Office, 2021). Spain, in the same year, recycled 34.7% of the MSW; only two autonomous communities (AC), Catalonia and the Valencian Community, reached the recycling goal of 50% set by the Waste Law (BOE, 2011).

The environmental and economic benefits of recycling are generally recognized. Craighill and Powell (1996) combined a lifecycle assessment with an economic valuation technique in a case study of Milton Keynes (Central England) and concluded that the recycling system (recovery of materials and their subsequent use in new products) generally contributes less to global warming, acidification effects, and nutrification of surface water than the waste disposal system (landfill disposal of waste and use of primary materials). More recently, Ferreira et al. (2014) carried out a similar study in Portugal and concluded that the total economic and environmental benefits of the packaging waste management systems established in the country exceed the costs.

Additionally, recycling creates jobs and contributes to government tax revenue (EPA, 2020). According to an environmental study conducted in Spain, the waste handling and processing sector employed 140,000 people in 2009, 279% more than in 1998 (Fundación Forum Ambiental, 2011).

In this context, it is clear that the recycling sector must keep growing and that the increase of the recycling rate must accelerate to attain the environmental protection goals. If the administrations understand what factors affect the MSW collection rates at a local level, then the ongoing MSW programs can be improved, and new initiatives can be implemented appropriately.

Recycling policy goals in Spain are devised at the national level, but the AC are responsible for monitoring, inspecting, and sanctioning the MSW production and processing activities, and the municipalities are in charge of implementing the policies by collecting, transporting, and processing the MSW (BOE, 2011). This suggests that recycling actions and results vary across the country and within the AC due to policy differences. Furthermore, socioeconomic, geographic, and demographic factors might also produce collection rates disparities between local entities. Therefore, the objective of this study is to analyze the main factors of MSW collection rates in Catalonian municipalities; this AC was chosen for our case study because *(1)* although it is one of the two AC with the highest rate of recycling, there is large heterogeneity between its municipalities, as can be observed in fig. 1.1; and *(2)* because the available data for the region is extensive and easily accessible. Our analysis is carried out conducting econometric models of longitudinal data (also known as panel data), including geographic, demographic, socioeconomic, and policy related variables at the municipal level, and spanning twenty years (2000-2019).

Analysing MSW collection rates not only implies understanding the waste sorting behaviour of the households, but also the behaviour of the local governments and the companies responsible of waste collection. Several studies examine, in addition to household behaviour, the spatial cost-effectiveness of the policy design (e.g., Hage and Söderholm, 2008). However, our study does not focus on this aspect because of the inability to conduct unbiased, valid surveys. Albeit demographic, geographic, and socioeconomic variables are collected and published by the Catalonian administration and several statistical agencies, information about collection, transportation, and processing costs are not publicly available. Therefore, our study makes an important but partial contribution to the understanding of the sources explaining the MSW collection rate in Catalonia.

As this paper focuses on MSW collection, it is worth noting that urban waste accounts for a big part of the waste generated in Spain. In 2016, households generated 16.8% of the waste in the country, while the Services and Construction sectors generated 5.1% and 27.8%, respectively. The remaining waste was generated, in descending order, by the following sectors: industry; water provision, sanitation and waste management; and agriculture, livestock, forestry and fishing (INE, 2016).

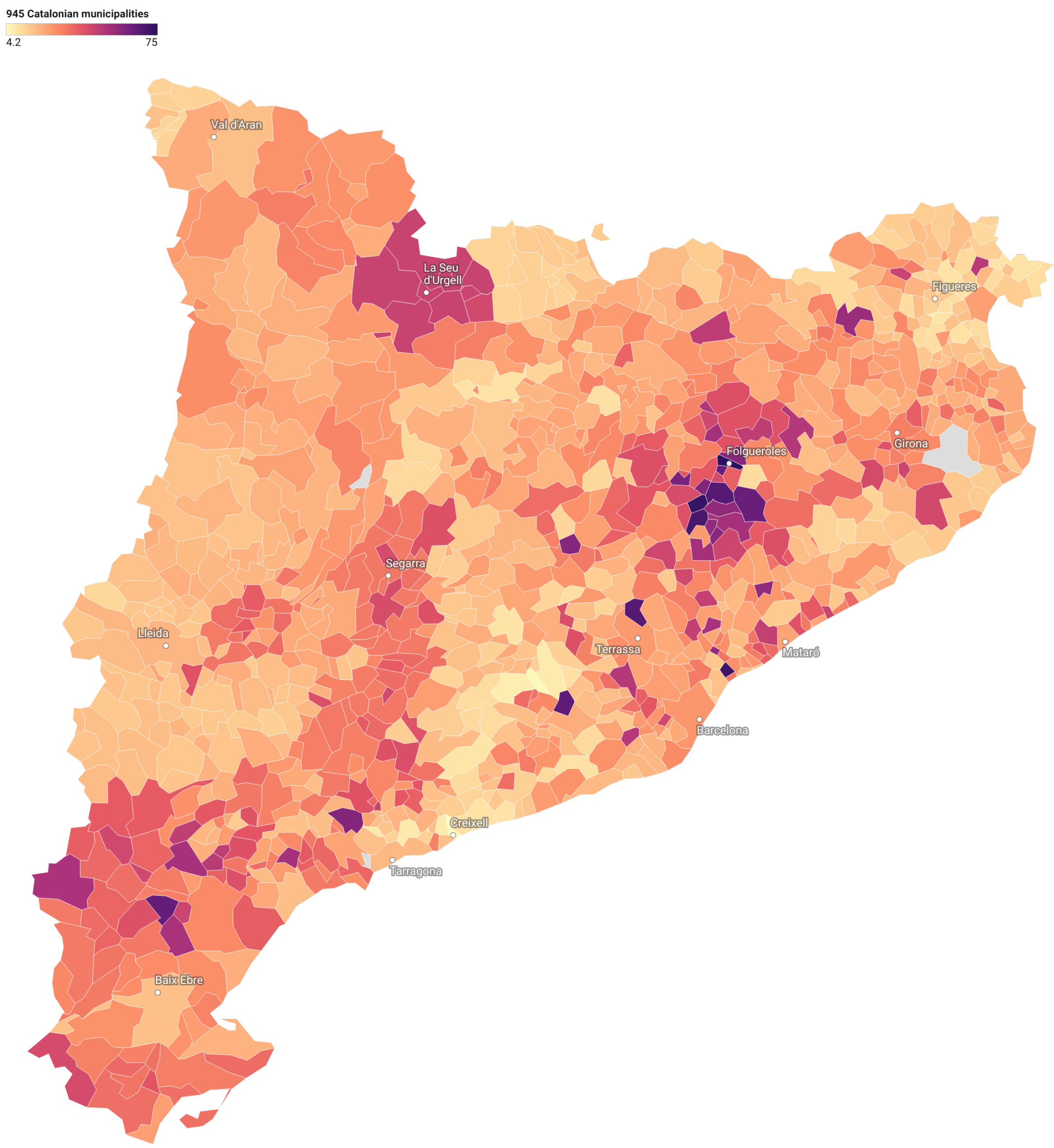


Figure 1.1: Average selective MSW collection rate in 945 municipalities, 2000-2019 (*source:* Dades Obertes Catalunya, 2021b)

This preamble shows the importance of our study. There are big differences in the MSW collection rate across Catalonian municipalities. Moreover, the national recycling strategy has not achieved the goals of the Waste Law. To increase recycling rates, the administrations must increase collection from households, and to attain this goal, we must understand what affects the MSW collection rates.

Section 2 of the paper provides a description of the data, the manner in which a number of setbacks were handled, and the theoretical expectations. The econometric model and the methodology are presented in Section 3. The empirical results are depicted in Section 4 and discussed in Section 5. Conclusions and implications of the paper are presented in Section 6.

2 Theoretical expectations and data

Catalonia, as of 2020, has 947 municipalities (Idescat, 2020c). Because of geopolitical reasons, the data of three municipalities were manipulated in the following way: Medinyà's data was added to Sant Julià de Ramis', as the former was an independent municipality from 2015 to 2018, when it was rejoined to the latter (Rodríguez, 2018); La Canonja's data was added to Tarragona's because the former was segregated from the latter in 2010 (La Información, 2010); and Tiurana's data was deleted from the data set, as the municipality was founded in 2007 (CCMA, 2007). Thus, for our analysis, we have constructed a longitudinal data set of 18.880 observations: 945 Catalonian municipalities from 2000 to 2019. The explanatory variables are classified into five categories: geographic, demographic, socioeconomic, policy related, and political preferences. A list of the variables, their description, their source, and the available years of their data can be found in table 2.1. Additionally, a summary of the variables' statistics is depicted in table 2.2.

As regards the dependent variable, sorted MSW collected (SMSWC), we consider the following materials as sorted waste: organic, paper, glass, light packages (plastic, metal, non-metallic and compound packaging), electrical and electronic devices, cooking oil, textile, batteries, medicines, and other selective waste (small quantity of varied waste collected at recycling drop-off centers). The SMSWC rate is computed as the kilograms of the sorted waste over the total MSW collected annually. There is other selective waste, such as gardening and debris, that were not included in the computation because the management methods of the municipalities and climatic factors can produce large fluctuations in the data and distort the SMSWC rate. These data are collected by each municipality individually and provided to the Agència de Residus de Catalunya for its publication on the website of Dades Obertes Catalunya (2021b). A box plot grouped by years, which also illustrates the increase from 9% to 41% of the average SMSWC rate in Catalonia from 2000 to 2019, is illustrated in fig. 2.1.

Table 2.1 Variables definition, source and available years

Variable	Definition	Source
Dependent variable		
SMSWC	Kilograms of sorted waste over total MSW collected annually	DOC (2000-2019)
Geographic variables		
TOURESTA	Number of tourist establishments per square kilometer	Idescat (2000,2002-2019)
COAST	Dummy for coastal municipalities. 1 if coastal and 0 if not	DOC (2000-2019)
Demographic variables		
WOMEN	Women as a share of total population	Idescat (2000-2019)
NOCHILD	Percentage of nuclear families without children	Idescat (2001, 2011)
PDEN	Population density	Idescat, DOC (2000-2019)
ELDER	People above 64 years old as a share of total population	Idescat (2000-2019)
Socioeconomic variables		
EDUC	Percentage of the population with at least a bachelor's degree	Idescat (2001, 2011)
INCOME	Average general tax base as a measure of net disposable income	Idescat (2000-2018)
FOREIGN	Foreigners as a share of total population	Idescat (2000-2019)
UNEMP	Unemployment rate	SEPE, Idescat (2006-2019)
Policy variable		
DDC	Dummy of municipalities door-to-door waste collection service.	AMCRPP (2000-2019)
Political preferences		
POLT	Votes to left-wing parties as share of total votes in the general elections	Idescat (2000, 2004, 2008 2011-2016)

Note: DOC refers to Dades Obertes Catalunya.

Data of three independent variables are missing for substantial periods of time, either because of the nature of the variable or because their collection was sporadic. To take care of the missing values, we decided to apply three types of solutions and compare them through graphical diagnosis and the statistical quality of our models results: interpolation of the values; propagation of the closest valid observation, forward or backward depending on the available years; and iterative imputation, a tool designed in Python and inspired by the R MICE (Multivariate Imputation by Chained Equations) package, which uses available data to model the feature with missing values and uses the estimate to impute them (Pedregosa et al., 2011). Below we discuss the implications of having missing values, considering the characteristics of each variable.

The independent variables were selected either by influence of the literature review or by theoretical reasoning. The first geographic independent variable, Tourist

Accommodation Establishments (TOURESTA), is the sum of hotels, camping sites, and rural tourism establishments per square kilometer. As the values for 2001 were not available in the data, they have been propagated with the data of the following year; this filling mechanism should not alter the data significantly, as it is fairly stable over time. The data are provided by the Ministry of Business and Knowledge and published on the website of Idescat (2019). The impact of tourism on the rate of SMSWC has been previously analysed by Mateu-Sbert et al. (2013), who concluded that a resident in Menorca Island selectively collects on average 47.3% more than a tourist. Although the TOURESTA variable is not measured by actual tourist population, we expect to find a similar results, and we believe it is a relevant determinant of SMSWC rate.

Table 2.2 Summary statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
SMSWC	18,900	25.53	16.79	0.00	13.30	33.48	100.00
TOURESTA	18,900	0.20	0.48	0.00	0.01	0.20	9.12
WOMEN	18,900	48.39	2.65	23.19	47.28	50.03	71.94
NOCHILD	968	32.79	6.69	13.33	28.67	36.36	69.23
NOCHILD.Prop	18,900	32.82	6.75	13.33	28.65	36.40	69.23
PDEN	18,900	421.75	1,552	0.69	12.33	168.20	21,364
ELDER	18,900	21.15	6.73	5.26	16.03	25.74	58.54
EDUC	1,399	11.62	5.52	0.90	7.99	14.13	46.32
EDUC.Inter	18,900	11.63	5.39	0.90	8.14	14.17	46.32
EDUC.Itera	18,900	11.93	5.78	0.90	8.19	14.55	59.65
EDUC.Prop	18,900	11.22	5.25	0.90	7.69	13.79	46.32
INCOME	18,900	16,099	5,468	0.00	12,422	19,442	95,206
FOREIGN	18,900	8.32	6.59	0.00	3.54	11.20	51.85
UNEMP	13,077	7.19	3.49	0.00	4.62	9.38	28.81
UNEMP.Inter	18,900	6.19	3.43	0.00	3.64	8.18	28.81
UNEMP.Itera	18,900	6.92	3.33	0.00	4.54	8.94	32.84
POLT	18,900	47.14	13.61	6.76	36.86	58.23	92.86
Dummies (yes:1, no:0)							
DDC	18,900	0.088	0.284				
COAST	18,900	0.074	0.262				

COAST, the second geographic variable in our model, is a dummy variable indicating the coastal municipalities with 1 and the non-coastal municipalities with 0. The data are provided by the Department of Territory and Sustainability, and available on the website of Dades Obertes Catalunya (2021a). We included this binary variable because, as Milfont et al. (2014) suggest in their study, people who live closer to the coast expressed greater belief that climate change is real. Thus, we consider that, similarly, or consequently, people living in the coast have a higher propensity to recycle, i.e., we test if municipalities in the coast have greater SMSWC rates.

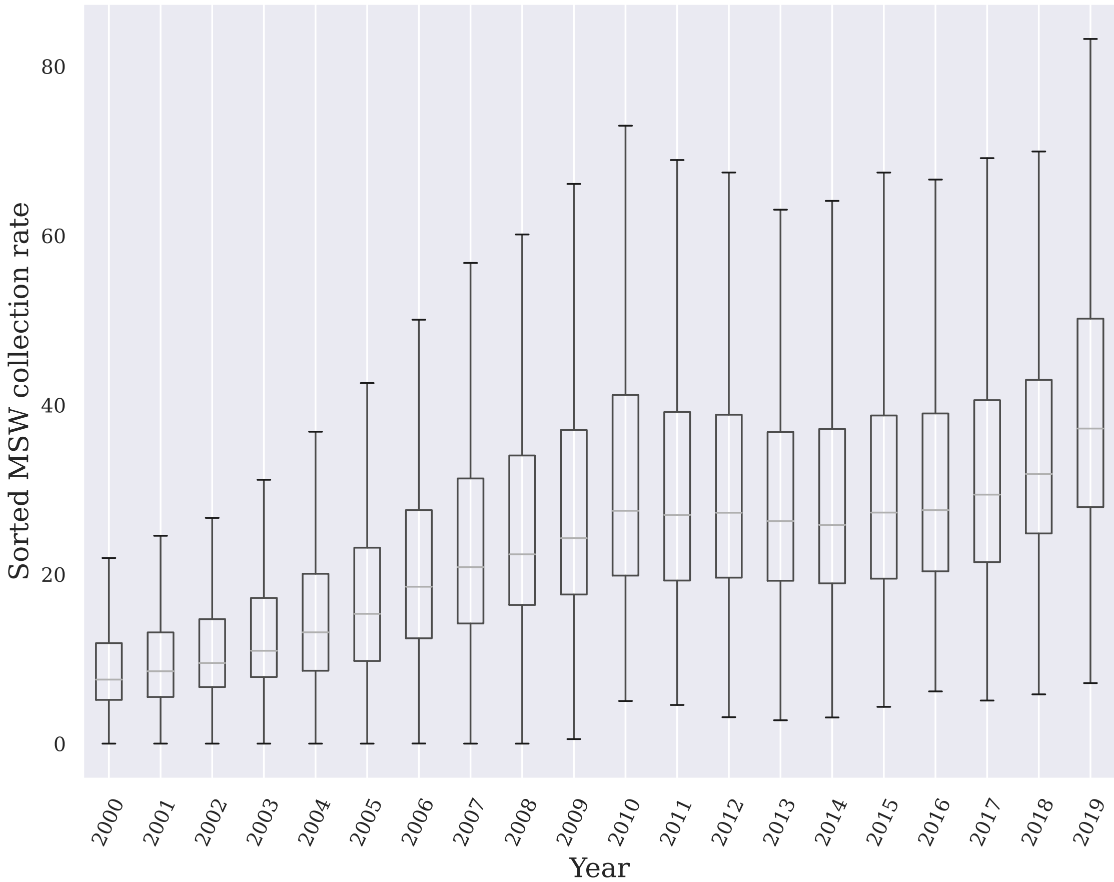


Figure 2.1: Average selective MSW collection rate in Catalonia (*source:* Dades Obertes Catalunya,2021b)

We have considered four relevant demographic variables. First, WOMEN, calculated as the number of women with respect to the entire population. These data are available on the website of the Spanish National Statistics Institute (2019). It has been concluded that women are more willing than men to pay for environmentally sound waste management (Bartelings and Sterner, 1999). Nevertheless, Hage et al. (2009) determined that gender do not affect recycling behaviour much. Thus, our intent is to confirm if there is evidence of women playing a role in the SMSWC.

NOCHILD, the second demographic variable, is defined as the number of nuclear families without children over the total amount of nuclear families. According to

Idescat, a nuclear family corresponds to a restricted conception of the family, limited to closer bonds. There are four types of nuclear families: couple without children, couple with one child or more, single mother with one child or more, and single father with one child or more. These data are collected in the population census that is carried out every ten years, and they are available on the website of Idescat (2014a). Due to unexplained reasons, Idescat published data for only 23 municipalities for the year 2011; thus, we have decided to propagate the data of 2001 across the entire period. This approach limits the quality of our variable considerably, as the population structure can vary greatly in twenty years. Nevertheless, we expect to see if the differences between municipalities are significant in our model. Johnstone and Labonne (2004) explain that the number of children in the household is important because consumption expenditure patterns for households with children are different, which can result in different patterns of MSW generation; and concludes that children have a negative influence on waste generation. Inversely, Knickmeyer (2020) notes that children can have a positive influence on the family's behaviour, as they are usually educated on environmental protection at school. Our analysis will corroborate one of these hypotheses or conclude that there is no significant effect in our case.

To calculate the third geographic variable, population density (PDEN), we used the population data available in the MSW data set (Dades Obertes Catalunya, 2021b) and the data of the area in square kilometres that is provided by Idescat (2020d). Population density is found to have a positive effect on the quantity of waste collected (Bartelings and Sterner, 1999). It has also been noted that investments in recycling infrastructure and education are cost-effective in densely populated areas (Grazhdani, 2016). Finally, Berglund and Söderholm (2003) show, for paper waste, that a high population density generally implies higher recovery rates. Therefore, we anticipate that more densely populated municipalities will have a higher SMSWC rate.

Lastly, in the demographic category, we selected an age effect variable, ELDER; from the register of inhabitants' data available on the website of Idescat (2021) we computed the share of the population that is over 65 years old, i.e., elderly people. It seems that the young population tends to be more environmentally aware. Nonetheless, Sidique et al. (2010) and Hage et al. (2009) concluded that age has a positive effect on the SMSWC rate because people who are older usually have more time to spend on recycling activities. Thus, we expect ELDER to be positively correlated with SMSWC.

EDUC, one of the four socioeconomic variables selected, is calculated as the population with a higher education over the total population in the census. Higher education is made up of *diplomatura*, which is an extinct degree of higher education; bachelor's degree; licentiate; and doctorate's degree. These data were collected by Idescat (2014b) in 2001 and 2011 (in the latter only for 465 municipalities). Because of its relevance, it was preferred to conduct an imputation of the data rather than leaving the variable out of the model. We applied interpolation, propagation, and iterative imputation to fill the missing values and conducted an analysis of our model with each of them to compare the results. A complication we encountered when filling the missing values was that, for the municipalities for which only the data of 2001 is available, all three methods behave in the same way; thus, with the iterative imputation and the interpolation methods, half of the municipalities maintain the EDUC variable constant throughout the analysed period, and the other half evolves as estimated by the methods. The rationale to include education as an independent variable is evident and well documented (Grazhdani, 2016; Sidique et al., 2010): we expect education to be positively correlated with the rate of SMSWC as more educated people are expected to be more aware of environmental issues, which would encourage them to selectively dispose of waste.

For our second socioeconomic variable, INCOME, Idescat only offers data on disposable household income from 2010 to 2018. Thus, we decided to use the average general tax base, which is available for the entire period, as a measure of disposable income. The general tax base consists of the total income obtained by taxpayers during the tax period; it includes income earned from work, rents from personal property, investment and business income, and capital variations. These data are collected by the Spanish Tax Agency and Idescat (2020b) publishes it in its website. As for the motives to include income in our model, it seems logical to expect that income has a negative relation with SMSWC rate, as people with higher income generally consume more and tend to generate higher amounts of waste (Grazhdani, 2016). Also, the opportunity cost for the high-income households is greater because recycling is a time-consuming activity. Nevertheless, Hage and Söderholm (2008) note that many empirical studies found that recycling efforts have a tendency to be positively correlated with income. Our analysis will shed light on the effect of income in the SMSWC rate.

The third socioeconomic variable, FOREIGN, indicates the number of foreigners out of the total population. Foreign refers to the nationality of the person: foreigners are considered to be those without Spanish nationality, according to the Spanish Immigration Law 8/2000 (BOE, 2000). The data are available on the website of Idescat (2020a) and its source is the register of inhabitants as of January 1, whose management corresponds to each municipality. Foreigners has been considered relevant for our study because immigrants are usually not familiar with local recycling and waste collection regulations (Hage et al., 2018). Additionally, there is previous evidence of a negative relationship between foreigners and recycling in Netherlands Dijkgraaf and Gradus (2004).

Unemployment rate (UNEMP), the fourth socioeconomic variable selected for our analysis, is defined as the number of people who are unemployed as a share of the labour force. The data of the unemployed people is available on the website of the Spanish Public Employment Service (2020) since 2006. The labour force, defined as the working age population (aged 15 to 64) was taken from the register of inhabitants' database provided by Idescat (2021). We have used backward propagation and iterative imputation to fill the missing values of the 2000-2005 period and conducted an analysis of the results, as shown in section 3.2. The rationale to include unemployment in our model is similar to that of ELDER: people who are unemployed usually have more time to spend on recycling activities (Hage et al., 2018).

The door-to-door collection (DDC) variable, considered the only policy variable and the second dummy variable of our model, takes the value 1 to indicate the presence of a door-to-door collection service and 0 to indicate its absence. As some municipalities offer DDC only in some of their neighbourhoods, only those municipalities in which at least 50% of the population is served with DDC were considered as having a DDC service. To the best of our knowledge, these data are not compiled by any governmental institution; these were taken from website of AMCRPP (2020), who collects and publishes this information. A door-to-door collection service generally accomplishes a higher level of selective waste collection than street containers collection (fixed drop-off points located along sidewalks) because the distance to be covered to deposit the waste is kept to a minimum (Johnstone and Labonne, 2004; Ventosa et al., 2013). Thus, we expect that municipalities with DDC service perform better and score higher on the SMSWC rate.

The last independent variable of our analysis, political preferences, is calculated as the number of votes to left-wing parties in the general elections as a share of total votes. The political parties are divided in the following way: Socialists' Party of Catalonia, Initiative for Catalonia Greens, Republican Left of Catalonia, In Common We Can, and Republican Left of Catalonia–Catalonia Yes are considered left-wing parties; and Convergence and Union, People's Party, Democratic Convergence of Catalonia, and Citizens are considered right-wing parties. The general elections take place every four years, thus, considering that the results of the elections are representative for the entire four-years term, we have propagated the values forward. These data are collected by the Ministry for Digital Policy and Public Administration and accessible on the website of Idescat (2017). Political preferences could be an important determinant of the SMSWC rate and might explain differences between municipalities, specifically, similarly to what Xiao and Buhrmann (2019) suggest, we expect to find that those municipalities in which the majority of votes are for left-leaning political parties tend to have a higher SMSWC rate.

To access, organize, and manipulate the data, we made use of the programming language Python. Specifically, we used the following modules, packages, and libraries: Datetime, Glob, NumPy (Harris et al., 2020), OS, Pandas (The pandas development team, 2020), PyPDF2, RegEx, Tabula, Webbrowser, and pywin32. Additionally, all figures depicted in this study were generated with Matplotlib (Hunter, 2007), Seaborn (Waskom, 2021), and Datawrapper (Datawrapper GmbH, 2021).

3 Methodology

3.1 Model

Fixed effects (FE) and Random effects (RE) models are used for the analysis. Also, pooled OLS (POLS) is computed as a tool to better compare our results. We specify a model of the following form:

$$y_{it} = \beta_0 + \beta X_{it} + a_i + u_{it}, \tag{3.1}$$

where y is the explained variable $SMSWC$, i denotes each municipality, and t denotes the time period. β_0 is the intercept, β is the $k \,\mathrm{x}\, 1$ matrix of coefficients, and X is the $1 \,\mathrm{x}\, k$ vector of explanatory variables ($TOURESTA_{it} + COAST_{it} + WOMEN_{it} + NOCHILD_{it} + log(PDEN_{it}) + ELDER_{it} + EDUC_{it} + log(INCOME_{it}) + FOREIGN_{it} + UNEMP_{it} + DDC_{it} + POLT_{it}$). The unobserved heterogeneity (or fixed effect), denoted as a_i, captures all unobserved, time-invariant factors that affect the explained variable. Finally, u_{it} is the idiosyncratic (or time-varying) error that represents unobserved factors that change over time and affect the explained variable.

As explained by Wooldridge (2015), if we assume that there is no endogeneity, i.e., correlation between the independent variables and u_{it} across all periods, the FE estimator is unbiased. Nonetheless, an inconvenience of the FE estimator is that because it allows for arbitrary correlation between the explanatory variables and a_i in any time period (heterogeneity is controlled), the explanatory variables that are constant over time for all i will be removed from the model by the FE transformation. In our case study, these variables are COAST and NOCHILD. The fixed (or within) transformation transforms our model into:

$$\ddot{y}_{it} = \beta \ddot{X}_{it} + \ddot{u}_{it}, \tag{3.2}$$

where the overdots indicate that the variable has been time-demeaned, i.e., the within-municipality time average for each variable is subtracted from the observed values of the variables. As we can see, by time-demeaning the data, the fixed effects estimator removes all time-constant variables. It is relevant to note that because the FE estimator uses the time variation in the independent and dependent variables within each cross-sectional observation, it will only explain differences within municipalities.

On the other hand, if we assume, additionally to all the assumptions of the FE model, that there is no correlation between a_i and the explanatory variables (whether time-variant or not), we find ourselves with a RE model. The RE estimator is computed in a way that allows for explanatory variables that are constant over time, which is an advantage over the FE model. In this sense, we will not have a problem including COAST and NOCHILD, but we must assume that these are uncorrelated with a_i. To see this better, we can transform eq. (3.1) to construct an adaptable RE model equation:

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta(X_{it} - \theta\bar{X}_i) + (v_{it} - \theta\bar{v}_i), \tag{3.3}$$

where $v_{it} = a_i + u_{it}$, the overbar denotes the time average of the variable, and θ is a parameter that measures the variance of a_i relative to the variance of u_{it}. When θ is close to zero, a_i is relatively unimportant and eq. (3.3) is similar to a POLS model, and when θ is close to one, a_i becomes important and the equation is close to the FE model described in eq. (3.2). Because the RE model jumps between OLS and FE depending on how much of the unobserved effect it attributes to the the error term, it can focus on differences between and within individuals.

Wooldridge (2015) suggests that, when applying FE and RE, it is also useful to compute the POLS estimates to compare the estimates and determine the nature of the biases caused by attributing the unobserved effect, a_i, entirely to the error term, as POLS does, or partially, as the RE transformation does. Nonetheless, we must remark that POLS does not consider time and individual characteristics; for it to generate a consistent estimator of the coefficients, we must assume that the unobserved effect, a_i, is uncorrelated with the explanatory variables, x_{it}. If it turns out that they are correlated, then POLS will be biased and inconsistent, which is possible considering that such an assumption is quite strict.

Considering that FE allows arbitrary correlation between a_i and the X_{it}, something that RE does not, the former seems to be a more reliable estimate. Also, as Wooldridge (2015) remarks, when observations are large geographical units, as in our case, we cannot treat the sample as a random sample from a large population, and it is logical to think of each a_i as a separate intercept to estimate for each cross-sectional unit, which is what the FE estimator does. Nonetheless, when the explanatory variable is time-invariant or has minimal within-unit variation, FE will not work properly, and only RE or POLS can be used (the former being more efficient than the latter). Thus, the use of propagated and interpolated data can be detrimental for the FE model, and we must keep this in mind when analysing the results.

In addition to these theoretical distinctions, we conduct several tests to assess the validity of each regression and make the appropriate adjustments and considerations. We use a Breusch-Pagan Lagrange multiplier test, which uses the residuals of the POLS model to test the null hypothesis that variances across entities are zero, i.e., that there are no significant differences across municipalities. Additionally, we employ a F test for individual and time effects based on the comparison of the FE and the POLS regressions. Following the literature, we formally test for statistically significant differences between FE and RE with the Hausman test. When such test rejects, it means that the main RE assumption —a_i is uncorrelated with each explanatory variable — is false. Nevertheless, we should not make the mistake of interpreting the rejection of the null as a sign to adopt the FE model, and the non-rejection as a sign to adopt the RE model (Baltagi, 2008). To test for serial correlation in the idiosyncratic error of the regressions, u_{it}, we make use of the Lagrange multiplier Breusch–Godfrey test, whose alternative hypothesis is that serial correlation is present in the model. As Baltagi (2008) suggests, long time-series longitudinal structures that do not account for cross-region dependence can result in misleading inference; thus, we implement a Pesaran's CD test to help us determine if there is cross sectional dependence in the specified models, i.e., if the municipalities are correlated due to the effect of some unobserved factor that is present in all municipalities and affects each of them in a similar or different way

(Tsionas, 2019). Finally, to determine if there is presence of heteroskedasticity in the model, we employ the Breusch-Pagan heteroskedasticity test, which tests if the variance of the errors is dependent on the values of the explanatory variables.

Because of the nature of our data—missing values in three explanatory variables—we compute five different models to be able to evaluate the validity of our results: *(1)* the complete model, including all independent variables except EDUC; *(2)* the complete model with all independent variables, using the propagated form of EDUC; *(3)* a model including only the 454 municipalities for which the iterative imputation of EDUC produces an increasing trend of the variable, and using the aforementioned imputation method to compare its results with the alternatives; *(4)* a model similar to the preceding one, with the difference that we use the interpolated form of the variable EDUC; and *(5)* a model including all municipalities and only taking values from 2006 to 2019, as to compare the results when using the imputed form of UNEMP with those produced when using only observed values of the variable. The first two are the main models and the ones we are more interested in, as they include all available observations; the last three are auxiliary models, useful to comprehend how well the filling mechanisms performed.

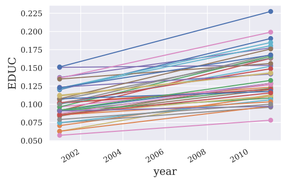
Time-dummies were included in all the regressions in order to grasp any unobserved, individual-invariant time effect not explained by the other explanatory variables (Baltagi, 2008); in our case these can be any specific, sudden event that affects the SMSWC rate. Furthermore, to simplify the interpretation of the coefficients, we use logarithmic transformation of the variables PDEN and INCOME; as it is easier to describe the effect of a percent increase in these variables than the effect of a unit increase. To compute the regressions, conduct the statistical tests, and produce the tables shown in this study, we made use of the programming language R. The libraries used were: plm (Croissant and Millo, 2018), lmtest (Zeileis and Hothorn, 2002), and stargazer (Hlavac, 2018).

3.2 Imputation methods assessment

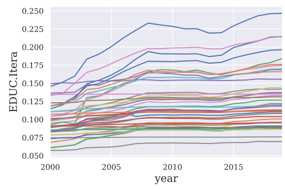
Imputation methods can be a useful tool when we require to include relevant data with missing values in our model. Nevertheless, if not done properly, imputation can be a source of undesired problems. Choosing the appropriate imputation method is paramount to assure the validity of our results. Thus, after computing several imputation methods of the independent variables EDUC, NOCHILD, and UNEMP, we carry out a visual assessment of the results to understand how our methodology can affect our results.

The following graphical analyses, based on previous studies on graphical diagnosis of imputation methods (Bondarenko and Raghunathan (2016); Nguyen et al. (2017), provide information about the distribution of the imputed values and will indicate whether these are reasonable or not. Clearly, the nature of the data to be imputed and the cause of the existence of missing values will also be determinant when deciding the most appropriate imputation method. Moreover, we understand that our interest lies in the relationship between the imputed explanatory variables and the explained variable, rather than in replacing the missing values.

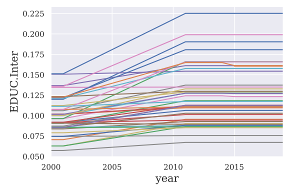
Figure 3.1 presents a graphical assessment of the imputation methods applied to the data for the variable EDUC. Figures 3.1a to 3.1d show the evolution of EDUC for the 42 counties (*comarcas*) that constitute Catalonia. We do not illustrate the evolution by municipalities, the unit of measure used in the regression, because the great number of observations would make the plots incomprehensible. It can be seen in fig. 3.1a that the average EDUC of all counties increased from 2001 to 2011, the years for which there is available data; thus, we can consider it a good sign if the imputed data present the same tendency, which they do, as we can see in fig. 3.1b and fig. 3.1c. In these two graphs we can also observe the consequences of the difficulty described in chapter 2, which can produce misleading, invalid results when including these data in our model. In fig. 3.1d we can see the jump effect produced by propagating the data. The density plots 3.1e, 3.1f, and 3.1g show that the three methods produce data with a distribution similar to that of the observed one, including the slight right skewness. Nevertheless, the iterative imputation shows the best result. Finally, scatter plots 3.1h, 3.1i, and 3.1j illustrate the correlation between the SMSWC rate and the observed data in red, and between the SMSWC rate and the imputed data in black. As we can see, all imputation methods change the best-fit line significantly. It is worth mentioning that because FE regressions do not work efficiently when presented with minimal variation variables, the iterative imputed data and the interpolated data are the best candidates for estimating the FE model. Further assessment of the imputation performance of EDUC will be provided in the next chapter.



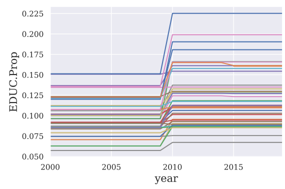
(a) Observed data



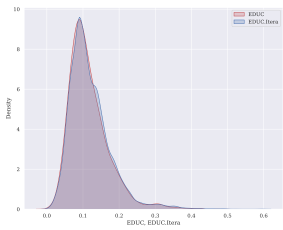
(b) Iterative Imputation



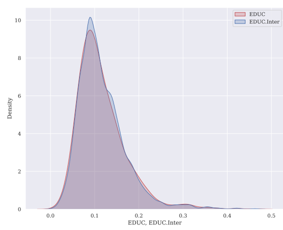
(c) Interpolation



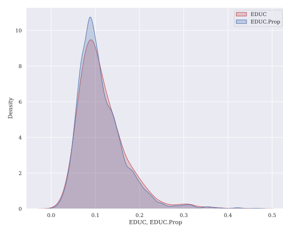
(d) Propagation



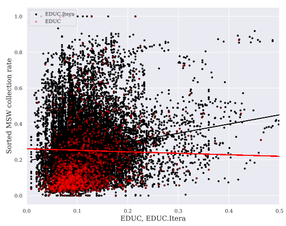
(e) Iterative Imputation



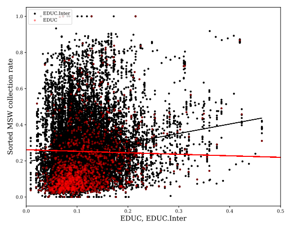
(f) Interpolation



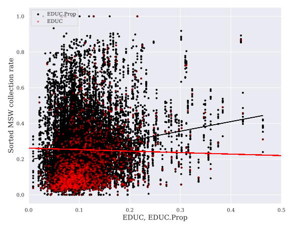
(g) Propagation



(h) Iterative Imputation



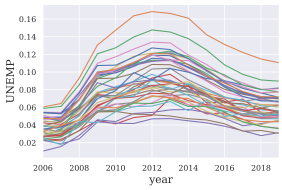
(i) Interpolation



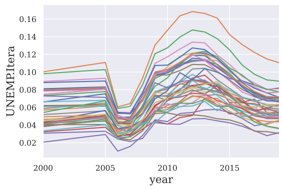
(j) Propagation

Figure 3.1: EDUC Imputation

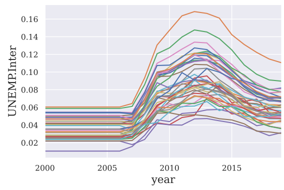
We follow the same assessment procedure with the variable UNEMP, as we can see in fig. 3.2. In this case, the observed data goes from 2006 to 2019, as shown in fig. 3.2a. The gentle rising behaviour created by the iterative imputed for the 2000-2005 period can be observed in fig. 3.2b, and the expected behaviour of the propagation method can be seen in fig. 3.2c. The density plots and the scatter plots illustrated in this figure make clear that the iterative imputation produces a distribution and a trendline closer to that of the observed data; hence, we discard the propagation method from our study.



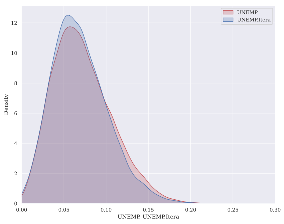
(a) Observed Data



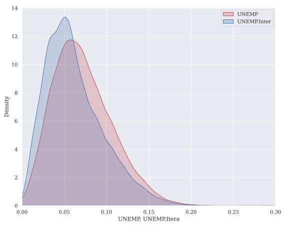
(b) Iterative Imputation



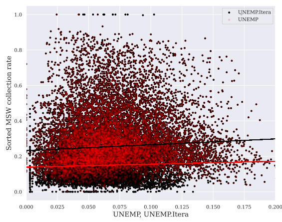
(c) Interpolation



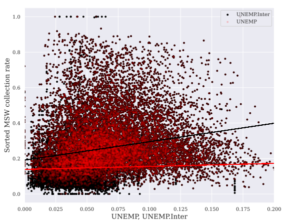
(d) Iterative Imputation



(e) Interpolation



(f) Iterative Imputation



(g) Interpolation

Figure 3.2: UNEMP Imputation

Lastly, in fig. 3.3 we compare the NOCHILD observed data with the propagated data and demonstrate that the imputed data have good properties, maintaining a distribution akin to the observed data and a trendline of the scatterplot with a similar slope. However, because the observed data is only available for 2001, the imputed data turns NOCHILD into a constant variable, with the disadvantages this entails.

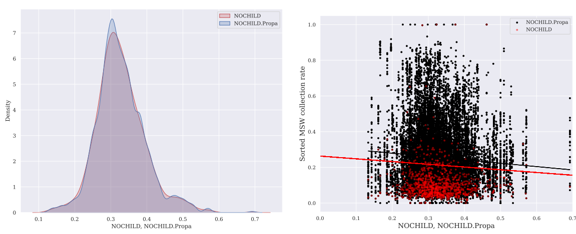


Figure 3.3: NOCHILD Propagation

4 Results

The results of the regressions are presented in tables A.1, A.2, A.3, A.4 and A.5. We find, from left to right, the results of the POLS, RE and FE regressions, and a Driscoll and Kraay estimator FE. The Lagrange Multiplier test for the POLS rejects the null hypothesis of not having significant time and individual effects in all five models. Thus, implementing a FE or a RE model should improve the efficiency of the results. Additionally, the F test comparing the POLS to the FE model verifies the existence of individual differences between municipalities.

The null hypothesis of no correlation between the idiosyncratic errors, u_i, and the regressors of the model is rejected by the Hausman test in all specified models, indicating that the RE model is inconsistent. Nevertheless, as explained in chapter 3, because FE cannot explain the effect of time-invariant variables on the explained variable, RE can come in handy to understand how COAST and NOCHILD possibly affect the explained variable. Furthermore, as Clark et al. (2015) concluded, the violation by the RE model of the supposedly fundamental assumption that the regressor and the individual effects are uncorrelated seems insufficient reason to rule out this model.

The Breusch–Godfrey test indicates the presence of serial correlation in the idiosyncratic error in all five models. Moreover, by means of the Pesaran's CD test, we reject the null hypothesis stating that the residuals do not present cross-sectional correlation. Finally, the Breusch-Pagan test indicates the presence of heteroskedasticity in all FE models. Hence, due to the presence of autocorrelation, crosss-sectional dependence, and heteroskedasticity in the models, we compute fixed effects SCC (spatial correlation consistent) estimates using the Driscoll and Kraay's robust covariance matrix estimators. This method was preferred to the Heteroscedasticity-Consistent Covariance Matrix Estimation (HAC) as the latter does not handle the problem of crosss-sectional dependence (Driscoll and Kraay, 1998).

All models show a F-statistic with $p < .01$, indicating that at least one coefficient is not equal to zero; in other words, there is at least one independent variable related to the explained variable SMSWC and, therefore, the selected variables improve our model. The R-squared, interpreted as the percentage of the variation in SMSWC across time that is explained by the model, is between .50 and .64 in all the specified models, except for model A.5, in which it falls to the range of .39-.40. It is important to note that adding year-dummies overestimate the R-squared, as each year coefficient explain some variation in the explained variable. Nevertheless, this measure provides an idea of how well our model fits the data.

The POLS estimates in all the specified models suggest that most variables are statistically significant at a one percent level; TOURESTA is not significant in models A.1 and A.2, FOREIGN is not significant in models A.2 and A.5, and log(INCOME) is not significant in models A.3, A.4, and A.5. As explained in chapter chapter 3, POLS can suffer from heterogeneity bias if the unobserved effects, a_i, is correlated with some observed regressor. Additionally, the estimates of the POLS model do not account for differences between municipalities. Thus, we understand the limitations of these results and consider them for the comparison of all estimated models.

RE results are more distinctive. COAST is significant in models in which all municipalities are included (A.1, A.2, and A.5), and has a negative relation with the explained variable in all models. The effect of WOMEN is not decisive, although it has a significant, small, positive effect on SMSWC in two of the main models. NOCHILD is significant and negatively related to SMSWC in all models. As for log(PDEN), this is significant at a ten percent level in model A.2, but its effect is minor: a one percent increase in the population density decreases the rate of sorted MSW collected by .003 units, which translates to a decrease of .003 percentage points, as this is the unit of measure of the explained variable. EDUC is statistically significant and positively related to the explained variable in the main models and in the two auxiliary models used to assess the result of the imputation methods. The magnitude of its effect is remarkably different depending on the imputed method considered: a one percentage point increase in EDUC increases SMSWC by .062, .234, and .185 percentage points in models A.2, A.3, and A.4, respectively. log(INCOME) is positively related to SMSWC, and it is only significant in the models in which all municipalities are considered, which include the main models. FOREIGN is statistically significant at a five percent level and it is positively related to the dependent variable. UNEMP is negatively related to SMSWC in all models, and it is significant only in those models which consider half the municipalities (A.3 and A.4). Finally, DDC and POLT are positively related to the explained variable and are statistically significant in all RE regressions. DDC produces a relatively large effect: municipalities provisioned with door-to-door sorted waste collection have, in average, a sorted MSW collection rate .27 percentage points higher than those who are not provisioned with it.

As regards the FE regression results, we focus on the Driscoll and Kraay's method results because of their robustness. WOMEN is statistically significant at a ten percent level in the main models, and it has a positive effect of around .09 percentage points increase in SMSWC for every percentage point increase in the share of women population. Population density is significant in all models and it affects the rate of sorted MSW collection negatively, with an approximate effect of $-.07$ percentage points. The significance of EDUC in the FE regressions is similar to those of the RE estimates, with the difference that the Driscoll and Kraay's transformation wipes out the significance of the variable in its propagated form (model A.2); as for its magnitude, the coefficients are approximately fifty percent larger than those of RE. INCOME has a significant effect in the main model A.2 and in the auxiliary model A.5, and it is positively related to the dependent variable in the models which include all municipalities. As for FOREIGN, it is significantly related to the explained variables in models A.1 and A.2, and it is positively related to the explained variable in all models; its considerable high coefficient in the aforementioned models is also noteworthy: a percentage point increase in the rate of foreign population increases SMSWC by .11 percentage points. UNEMP is significant at a ten percent level in the main models, A.1 and A.2, but it is not significant when the variable is included in its non-imputed form (model A.5), and it is negatively related to the dependent variable. The policy variable DDC is significant and positively related to SMSWC in all models, with coefficients close to those of POLS and RE regressions. Lastly, the rate of left-wing votes in general elections is positively related to the explained variable, but it is not significant in the main models; notably, this is the only variable whose significance is removed from both models A.1 and A.2 by the Driscoll and Kraay's standard errors estimates.

5 Discussion

The results described above shed light on which of our predictions are corroborated empirically and on how the specified theoretical model considerations manifest in a case study. We can see the considerable differences in the results produced by the POLS, RE, and FE models. As we are not willing to assume that there is no unobserved heterogeneity, the POLS model is considered to be inconsistent. Moreover, even if POLS was consistent, it would still suffer from serial correlation, especially with a long period of time such as the one considered in this study. This serial correlation is controlled for by the adaptable RE eq. (3.3), which makes the RE estimates more efficient than the POLS estimates. Nevertheless, RE estimates still require that the fixed effect is not correlated to the independent variables. Thus, although the FE model comes with limitations, we consider it the safest model to draw conclusions.

Despite of what our literature review indicates, it seems that municipalities in the shore show a lower rate of sorted MSW collection than the inland municipalities. This could be explained by the fact that, as coastal areas are more touristic, the rate of recycling is smaller; as we can observe in the regressions results, the variable which measures tourism, although not significant, is negatively related to SMSWC. FE and RE estimates provide similar, small coefficients for the rate of women population, results which coincide with our findings mentioned in chapter 2. With respect to families having children at home, it seems that these recycle more; this can be thanks to the environmental education taught at schools, as Knickmeyer (2020) suggested, although more causes could play a part. Indisputably, and contrary to what other authors concluded, population density has a negative effect on the rate of SMSWC in Catalonian municipalities. We could consider a $-.07$ percentage points decrease (per one percent increase in population density) a relatively small effect, but population density can vary substantially across time. As an example, the population density of Barcelona increased eight percent from 2000 to 2019.

The average general tax base, used as a measurement of average income, is positively correlated with the rate of recycling, as Hage and Söderholm (2008) estimated. It is worth noting its coefficient in model A.5, as this is between two and three times larger than in the main models. A possible explanation is that the 2009 financial crisis affected the rate of recycling of low income municipalities, which at the time could not allocate as much money into recycling solutions as before, and this effect has a greater weight in the model A.5.

As expected, the percentage of the population with higher education has a positive effect on the sorted MSW collection rate. As for the imputation methods, these seem to perform adequately. The model with the iterative imputed data produced the largest coefficient and, as mentioned in section section 3.2, it seems to be the most adequate filling mechanism when computing FE estimates. Overall, we are inclined to consider the coefficient of education as large as model A.3 estimates it.

Perhaps the most surprising result is that of the effect of foreign population. It is contrary to what previous studies indicate and it has a relatively large, positive coefficient of .11. There are many possible reasons to find this results; for instance, the implementation of different immigrant integration programs in the municipalities could be one of them, although further research is required to uncover more possible causes of this. Another unexpected result is the negative effect of unemployment rate on the rate of SMSWC. Concerning the results of the imputation methods implemented, these are sufficiently good, as we can see by comparing models A.1 and A.2 with model A.5, and by observing the results in fig. 3.2.

Door-to-door collection is, along with education, the variable with the largest effect on the explained variable, and it is surely significant. This result not only corroborates previous studies but gives support to the implementation of door-to-door collection schemes in more municipalities. Lastly, although its coefficient is not significant in the FE SCC model, the percentage of votes to left-wing parties has a positive effect on recycling rates in Catalonian municipalities, as concluded by previous authors.

6 Conclusion

Using longitudinal data of solid waste in Catalonian municipalities, this study has demonstrated how geographic, demographic, socioeconomic, policy related, and political preferences variables determine the rate of sorted waste collection. A careful evaluation of pooled OLS, random effects, and fixed effects models were used to conclude that the presence of coast, the population density, the unemployment rate, and the absence of children in the household contribute negatively to the rate of sorted MSW collection. Inversely, the rate of women population, the average income, the percentage of foreigners, the door-to-door collection services, and the rate of votes to left-wing parties are positively related to the rate of sorted MSW collection. Additionally, we found that the effects of the rate of elderly population and of the number of tourist establishments per square kilometre do not affect the rate of sorted MSW collection significantly.

The results second the implementation of door-to-door collection services, as its effect on recycling rates is substantial. Moreover, our findings can help administrations and public agents evaluate the implementation of programs and the enactment of policies to encourage the appropriate groups to sort waste at a household level and increase the rate of recycling. Lastly, in order to maintain a high quality of solid waste management service, attention should be given to the changes of population density within municipalities.

Despite the relevant contributions of this study to the understanding of the factors influencing the rate of sorted MSW collection, there are limitations to our analysis that should be considered in future research. The use of imputed data, although useful in certain cases, rises considerable drawbacks and should be used as a last resource. Finally, if data on transportation and processing costs of MSW can be accessed, their incorporation in the estimations could produce new, relevant findings.

Bibliography

Association of Catalonian towns for door-to-door selective waste collection AMCRPP. Door-to-door selective waste collection by municipalities, 2020. URL www.portaaporta.cat/ca/municipis.php. Accessed: March 31, 2021.

Badi Baltagi. *Econometric analysis of panel data.* John Wiley & Sons, 2008.

Heleen Bartelings and Thomas Sterner. Household waste management in a swedish municipality: determinants of waste disposal, recycling and composting. *Environmental and resource economics*, 13(4):473–491, 1999.

Christer Berglund and Patrik Söderholm. An econometric analysis of global waste paper recovery and utilization. *Environmental and Resource Economics*, 26(3):429–456, 2003.

Boletín Oficial del Estado BOE. Ley 8/2000. https://boe.es/buscar/pdf/2000/BOE-A-2000-23660-consolidado.pdf, 2000. Accessed: May 22, 2021.

Boletín Oficial del Estado BOE. Ley 22/2011. www.boe.es/buscar/pdf/2011/BOE-A-2011-13046-consolidado.pdf, 2011. Accessed: March 31, 2021.

Irina Bondarenko and Trivellore Raghunathan. Graphical and numerical diagnostic tools to assess suitability of multiple imputations and imputation models. *Statistics in medicine*, 35(17):3007–3020, 2016.

CCMA. Medinyà perd la independència per ordre del Tribunal Constitucional. *CCMA*, 2007. URL www.ccma.cat/324/tiurana-estrena-el-nou-poble/noticia/213470/. Accessed: May 23, 2021.

Tom S Clark, Drew A Linzer, et al. Should i use fixed or random effects. *Political science research and methods*, 3(2):399–408, 2015.

Tim Cooper. Challenging the 'refuse revolution': war, waste and the rediscovery of recycling, 1900–50. *Historical research*, 81(214):710–731, 2008.

Amelia L Craighill and Jane C Powell. Lifecycle assessment and economic evaluation of recycling: a case study. *Resources, conservation and recycling*, 17(2):75–96, 1996.

Yves Croissant and Giovanni Millo. *Panel Data Econometrics with R: the plm package.* Wiley, 2018.

Dades Obertes Catalunya. Data of the urban map of catalonia, 2021a. URL https://analisi.transparenciacatalunya.cat/es/Urbanisme-infraestructures/Dades-del-mapa-urban-stic-de-Catalunya/epsm-zskb. Accessed: March 31, 2021.

Dades Obertes Catalunya. Data of municipal waste, 2021b. URL https://analisi.transparenciacatalunya.cat/es/Medi-Ambient/Estad-stiques-de-residus-municipals/69zu-w48s. Accessed: March 31, 2021.

Datawrapper GmbH. Datawrapper. https://www.datawrapper.de/, 2021.

Elbert Dijkgraaf and Raymond Henri Jean Marie Gradus. Cost savings in unit-based pricing of household waste: the case of the netherlands. *Resource and energy economics*, 26(4):353–371, 2004.

John C Driscoll and Aart C Kraay. Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, 80(4):549–560, 1998.

Environmental Protection Agency EPA. Recycling economic information (rei) report, 2020. URL https://www.epa.gov/sites/production/files/2020-11/documents/rei_report_508_compliant.pdf. Accessed: April 4, 2021.

European Statistical Office. Recycling rate of municipal waste, 2021. URL https://ec.europa.eu/eurostat/databrowser/view/t2020_rt120/default/line?lang=en. Accessed: April 4, 2021.

S Ferreira, M Cabral, Nuno F da Cruz, and Rui Cunha Marques. Economic and environmental impacts of the recycling system in portugal. *Journal of cleaner production*, 79: 219–230, 2014.

Fundación Forum Ambiental. Estudio del sector económico del medio ambiente en España 2011, 2011. URL http://wp.forumambiental.org/wp-content/uploads/2017/11/esemae2011.pdf. Accessed: April 4, 2021.

Dorina Grazhdani. Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in prespa park. *Waste Management*, 48:3–13, 2016.

Olle Hage and Patrik Söderholm. An econometric analysis of regional differences in household waste collection: the case of plastic packaging waste in sweden. *Waste management*, 28(10):1720–1731, 2008.

Olle Hage, Patrik Söderholm, and Christer Berglund. Norms and economic motivation in household recycling: Empirical evidence from sweden. *Resources, Conservation and Recycling*, 53(3):155–165, 2009.

Olle Hage, Krister Sandberg, Patrik Söderholm, and Christer Berglund. The regional heterogeneity of household recycling: A spatial-econometric analysis of swedish plastic packing waste. *Letters in spatial and resource sciences*, 11(3):245–267, 2018.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Marek Hlavac. stargazer: Well-formatted regression and summary statistics tables, 2018. URL https://CRAN.R-project.org/package=stargazer. R package version 5.2.2.

John D Hunter. Matplotlib: A 2d graphics environment. *IEEE Annals of the History of Computing*, 9(03):90–95, 2007.

Statistical Institute of Catalonia Idescat. Households by number of families, 2014a. URL http://www.idescat.cat/pub/?id=censph&n=302&by=mun&t&lang=en. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Population aged 16 and over according to level of education, 2014b. URL http://www.idescat.cat/pub/?id=censph&n=15&by=mun&t. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Results of the general elections by party votes, 2017. URL http://www.idescat.cat/pub/?id=elecd&n=386&by=mun&t. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Tourist accommodation establishments, 2019. URL https://www.idescat.cat/pub/?id=turall&n=6030&by=mun&lang=en&t=201900. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Foreign population by municipalities, 2020a. URL https://www.idescat.cat/poblacioestrangera/?b=6. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Personal income tax (irpf), 2020b. URL https://www.idescat.cat/pub/?id=irpf&n=4070&by=mun. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Number of municipalities and population, 2020c. URL https://www.idescat.cat/pub/?id=aec&n=924. Accessed: May 22, 2021.

Statistical Institute of Catalonia Idescat. Altitude, area and population by municipalities, 2020d. URL https://www.idescat.cat/pub/?id=aec&n=925. Accessed: March 31, 2021.

Statistical Institute of Catalonia Idescat. Population by sex and age in large groups, 2021. URL https://www.idescat.cat/pub/?id=pmh&n=1181&by=mun. Accessed: March 31, 2021.

Instituto Nacional de Estadística INE. Otras cuentas medioambientales: Cuenta de los residuos. año 2016, 2018. URL https://www.ine.es/prensa/cma_2016_res.pdf. Accessed: April 6, 2021.

Nick Johnstone and Julien Labonne. Generation of household solid waste in oecd countries: an empirical analysis using macroeconomic data. *Land Economics*, 80(4):529–538, 2004.

Doris Knickmeyer. Social factors influencing household waste separation: A literature review on good practices to improve the recycling performance of urban areas. *Journal of cleaner production*, 245:118605, 2020.

La Información. Medinyà perd la independència per ordre del Tribunal Constitucional. *La Información*, 2010. URL https://www.lainformacion.com/espana/el-municipio-de-la-canonja-se-independiza-de-tarragona_5odmxexqriw4nuwfdbxbw6/. Accessed: May 23, 2021.

Josep Mateu-Sbert, Ignacio Ricci-Cabello, Ester Villalonga-Olives, and Elena Cabeza-Irigoyen. The impact of tourism on municipal solid waste generation: The case of Menorca Island (Spain). *Waste management*, 33(12):2589–2593, 2013.

Taciano L Milfont, Laurel Evans, Chris G Sibley, Jan Ries, and Andrew Cunningham. Proximity to coast is linked to climate change belief. *PLoS One*, 9(7):e103180, 2014.

Cattram D Nguyen, John B Carlin, and Katherine J Lee. Model checking in multiple imputation: an overview and case study. *Emerging themes in epidemiology*, 14(1):1–12, 2017.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Marta Rodríguez. Medinyà perd la independència per ordre del tribunal constitucional. *El País*, 2018. URL https://cat.elpais.com/cat/2018/02/01/catalunya/1517497328_466858.html. Accessed: May 23, 2021.

Shaufique F Sidique, Satish V Joshi, and Frank Lupi. Factors influencing the rate of recycling: An analysis of Minnesota counties. *Resources, Conservation and Recycling*, 54(4):242–249, 2010.

Spanish National Statistics Institute. Population data of municipal register, 2019. URL https://www.ine.es/dynt3/inebase/es/index.htm?padre=517. Accessed: March 31, 2021.

SEPE Spanish Public Employment Service. Registered unemployment and contracts by municipalities, 2020. URL https://www.sepe.es/HomeSepe/en/que-es-el-sepe/estadisticas/datos-estadisticos/municipios.html. Accessed: March 31, 2021.

The pandas development team. pandas-dev/pandas: Pandas, February 2020. URL https://doi.org/10.5281/zenodo.3509134.

Mike Tsionas. *Panel Data Econometrics: Empirical Applications*. Academic Press, 2019.

Ignasi Puig Ventosa, Jaume Freire González, and Marta Jofra Sora. Economic balance of door-to-door collection and road containers waste collection for local authorities and proposals for its optimisation. *Associació de municipis catalans per a la recollida Porta a Porta*, 2013. URL https://www.portaaporta.cat/documents/arxiu_portaaporta_173.pdf.

Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL https://doi.org/10.21105/joss.03021.

Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Cengage learning, 2015.

Chenyang Xiao and Jan Buhrmann. Ideas to action: environmental beliefs, behaviors, and support for environmental policies. *Journal of Environmental Studies and Sciences*, 9 (2):196–205, 2019.

Achim Zeileis and Torsten Hothorn. Diagnostic checking in regression relationships. *R News*, 2(3):7–10, 2002. URL https://CRAN.R-project.org/doc/Rnews/.

A Regressions

Table A.1 Regression results for sorted MSW collection rate in 945 Catalonian municipalities

	Dependent variable:			
	SMSWC			
	POLS	RE	FE	FE SCC
TOURESTA	−0.125	−0.048	−0.198	−0.198
	(0.205)	(0.298)	(0.325)	(0.216)
COAST	−3.439***	−2.786**		
	(0.392)	(1.169)		
WOMEN	0.255***	0.081*	0.093*	0.093*
	(0.038)	(0.047)	(0.050)	(0.050)
NOCHILD.Prop	−0.152***	−0.246***		
	(0.015)	(0.043)		
log(PDEN)	0.468***	−0.239	−6.535***	−6.535***
	(0.073)	(0.170)	(0.576)	(0.984)
ELDER	0.100***	0.051*	0.009	0.009
	(0.018)	(0.028)	(0.031)	(0.058)
log(INCOME)	1.158***	0.487**	0.206	0.206
	(0.242)	(0.196)	(0.198)	(0.139)
FOREIGN	−0.033**	0.057**	0.113***	0.113***
	(0.016)	(0.023)	(0.025)	(0.028)
UNEMP.Itera	−0.337***	−0.063	−0.080*	−0.080*
	(0.036)	(0.039)	(0.041)	(0.045)
DDC	27.664***	26.961***	26.738***	26.738***
	(0.313)	(0.339)	(0.350)	(1.127)
POLT	0.046***	0.044***	0.040***	0.040
	(0.010)	(0.014)	(0.015)	(0.030)
Constant	−11.936***	7.454**		
	(2.929)	(3.332)		
Observations	18,900	18,900	18,900	
R^2	0.503	0.589	0.600	
Adjusted R^2	0.502	0.589	0.578	
F Statistic	636.311***	27,078.390***	959.298***	

Note: Time dummies were included in all regressions. *p<0.1; **p<0.05; ***p<0.01.

Table A.2 Regression results for sorted MSW collection rate in 945 Catalonian municipalities

	Dependent variable:			
	SMSWC			
	POLS	RE	FE	FE SCC
TOURESTA	−0.248	−0.015	−0.141	−0.141
	(0.204)	(0.298)	(0.326)	(0.255)
COAST	-3.906^{***}	-2.895^{**}		
	(0.392)	(1.162)		
WOMEN	0.232^{***}	0.075	0.086^{*}	0.086^{*}
	(0.038)	(0.047)	(0.050)	(0.052)
NOCHILD.Prop	-0.149^{***}	-0.240^{***}		
	(0.015)	(0.042)		
log(PDEN)	0.388^{***}	-0.292^{*}	-6.965^{***}	-6.965^{***}
	(0.072)	(0.172)	(0.603)	(0.916)
ELDER	0.150^{***}	0.048^{*}	−0.006	−0.006
	(0.019)	(0.028)	(0.032)	(0.053)
EDUC.Prop.	0.250^{***}	0.062^{**}	0.092^{**}	0.092
	(0.020)	(0.031)	(0.038)	(0.124)
log(INCOME)	0.419^{*}	0.492^{**}	0.227	0.227^{*}
	(0.248)	(0.196)	(0.198)	(0.134)
FOREIGN	−0.021	0.057^{**}	0.112^{***}	0.112^{***}
	(0.016)	(0.023)	(0.025)	(0.027)
UNEMP.Itera	-0.226^{***}	−0.060	-0.081^{**}	-0.081^{*}
	(0.037)	(0.039)	(0.041)	(0.045)
DDC	27.642^{***}	26.968^{***}	26.735^{***}	26.735^{***}
	(0.312)	(0.339)	(0.350)	(1.125)
POLT	0.052^{***}	0.047^{***}	0.043^{***}	0.043
	(0.010)	(0.014)	(0.015)	(0.029)
Constant	-8.556^{***}	7.046^{**}		
	(2.928)	(3.330)		
Observations	18,900	18,900	18,900	
R^2	0.507	0.589	0.600	
Adjusted R^2	0.506	0.589	0.578	
F Statistic	626.408^{***}	$27{,}073.070^{***}$	926.671^{***}	

Note: Time dummies were included in all regressions. *p<0.1; **p<0.05; ***p<0.01.

Table A.3 Regression results for sorted MSW collection rate in 454 Catalonian municipalities

	Dependent variable:			
	SMSWC			
	POLS	RE	FE	FE SCC
TOURESTA	−0.560**	0.240	−0.069	−0.069
	(0.220)	(0.315)	(0.342)	(0.242)
COAST	−3.348***	−1.831		
	(0.421)	(1.252)		
WOMEN	0.504***	−0.469***	−0.565***	−0.565***
	(0.116)	(0.142)	(0.149)	(0.160)
NOCHILD.Prop	−0.432***	−0.651***		
	(0.034)	(0.090)		
log(PDEN)	0.566***	−0.877***	−11.957***	−11.957***
	(0.125)	(0.297)	(0.911)	(0.976)
ELDER	0.345***	0.053	−0.252***	−0.252*
	(0.039)	(0.061)	(0.072)	(0.132)
EDUC.Itera	0.253***	0.234***	0.332***	0.332***
	(0.028)	(0.045)	(0.057)	(0.054)
log(INCOME)	0.220	0.389	−0.058	−0.058
	(0.486)	(0.380)	(0.382)	(0.369)
FOREIGN	0.040*	0.063*	0.034	0.034
	(0.021)	(0.035)	(0.040)	(0.036)
UNEMP.Itera	−0.459***	−0.486***	−0.576***	−0.576***
	(0.067)	(0.082)	(0.086)	(0.110)
DDC	29.268***	27.411***	26.900***	26.900***
	(0.402)	(0.453)	(0.467)	(1.394)
POLT	0.074***	0.209***	0.232***	0.232***
	(0.021)	(0.029)	(0.031)	(0.037)
Constant	−15.557**	45.195***		
	(6.907)	(8.400)		
Observations	9,080	9,080	9,080	
R^2	0.571	0.631	0.643	
Adjusted R^2	0.570	0.629	0.623	
F Statistic	389.096***	15,447.250***	534.647***	

Note: Time dummies were included in all regressions. *p<0.1; **p<0.05; ***p<0.01.

Table A.4 Regression results for sorted MSW collection rate in 454 Catalonian municipalities

	Dependent variable:			
	SMSWC			
	POLS	RE	FE	FE SCC
TOURESTA	−0.539**	0.279	−0.002	−0.002
	(0.220)	(0.315)	(0.342)	(0.258)
COAST	−3.324***	−1.731		
	(0.422)	(1.254)		
WOMEN	0.528***	−0.471***	−0.591***	−0.591***
	(0.116)	(0.142)	(0.149)	(0.169)
NOCHILD.Prop	−0.430***	−0.659***		
	(0.034)	(0.090)		
log(PDEN)	0.580***	−0.805***	−11.527***	−11.527***
	(0.125)	(0.299)	(0.914)	(0.867)
ELDER	0.332***	0.045	−0.247***	−0.247*
	(0.039)	(0.061)	(0.072)	(0.128)
EDUC.Inter	0.243***	0.185***	0.252***	0.252**
	(0.030)	(0.052)	(0.069)	(0.119)
log(INCOME)	0.371	0.426	−0.046	−0.046
	(0.487)	(0.380)	(0.383)	(0.367)
FOREIGN	0.037*	0.056	0.024	0.024
	(0.021)	(0.035)	(0.040)	(0.035)
UNEMP.Itera	−0.480***	−0.506***	−0.589***	−0.589***
	(0.068)	(0.082)	(0.086)	(0.106)
DDC	29.274***	27.400***	26.888***	26.888***
	(0.403)	(0.454)	(0.468)	(1.393)
POLT	0.070***	0.206***	0.234***	0.234***
	(0.022)	(0.029)	(0.032)	(0.038)
Constant	−17.774**	45.635***		
	(6.917)	(8.407)		
Observations	9,080	9,080	9,080	
R^2	0.570	0.630	0.642	
Adjusted R^2	0.569	0.629	0.622	
F Statistic	387.587***	15,409.380***	532.697***	

Note: Time dummies were included in all regressions. *p<0.1; **p<0.05; ***p<0.01.

Table A.5 Regression results for sorted MSW collection rate in 945 Catalonian municipalities (2006-2019)

	Dependent variable:			
	SMSWC			
	POLS	RE	FE	FE SCC
TOURESTA	−0.565**	−0.325	−0.452	−0.452**
	(0.263)	(0.331)	(0.356)	(0.223)
COAST	−4.246***	−3.106**		
	(0.493)	(1.387)		
WOMEN	0.331***	0.121**	0.090	0.090
	(0.047)	(0.059)	(0.063)	(0.068)
NOCHILD.Prop	−0.160***	−0.206***		
	(0.018)	(0.051)		
log(PDEN)	0.501***	0.039	−8.313***	−8.313***
	(0.092)	(0.210)	(1.260)	(1.408)
ELDER	0.259***	0.029	−0.100**	−0.100*
	(0.025)	(0.039)	(0.048)	(0.061)
EDUC.Prop	0.261***	−0.069*	−0.128***	−0.128***
	(0.024)	(0.037)	(0.045)	(0.041)
log(INCOME)	0.695	1.017***	0.709**	0.709***
	(0.469)	(0.346)	(0.352)	(0.259)
FOREIGN	−0.015	0.005	0.068	0.068
	(0.018)	(0.033)	(0.042)	(0.052)
UNEMP	−0.331***	−0.055	−0.023	−0.023
	(0.046)	(0.041)	(0.043)	(0.034)
DDC	27.434***	27.421***	27.250***	27.250***
	(0.343)	(0.413)	(0.440)	(2.243)
POLT	0.092***	0.067***	0.063***	0.063
	(0.013)	(0.017)	(0.018)	(0.039)
Constant	−9.331*	9.039*		
	(5.127)	(4.876)		
Observations	13,230	13,230	13,230	
R^2	0.403	0.392	0.395	
Adjusted R^2	0.402	0.390	0.348	
F Statistic	356.482***	8,500.377***	348.419***	

Note: Time dummies were included in all regressions. *p<0.1; **p<0.05; ***p<0.01.