'Ghost citizens': Using notches to identify manipulation of population-based grants

Dirk Foremny* Jordi Jofre-Monseny† Albert Solé-Ollé‡

August 2017

Universitat de Barcelona
Institut d'Economia de Barcelona (IEB)

Abstract

This paper analyzes how local governments misreport population figures to obtain higher per capita grant allocations. In 1998, the allocation of a formula based grant in Spain switched from using the centrally administered census to local population registers administered by municipalities. The value of this per capita grant changes at fixed population thresholds for the entire local population. We exploit these notches to analyze the size distribution of municipalities to detect deliberate manipulation of the grant-assignment variable. This allows us to causally identify the effect of grant generosity on population over-reporting. We document an excess mass of municipalities to the right of the notch threshold and a density hole to the left of it: local registers included a proportion of 'ghost citizens', that is, people who presented no trace of actually residing in the municipalities which benefit the most from inflating population figures to pass the relevant threshold. We document that manipulation (rather than real population responses) is the mechanism at work. The main channel behind manipulation is the incorrect treatment of foreign residents to inflate total local population.

Keywords: Intergovernmental transfers, notches, bunching, enforcement.

JEL Codes: H7; H26; D7

* Corresponding author, Universitat de Barcelona, Facultat d'Economia i Empresa, C/ John M. Keynes, 1-11, 08034 Barcelona, foremny@ub.edu

† email: jordi.jofre@ub.edu

‡ email: asole@ub.edu.

We gratefully acknowledge helpful comments received by the Editor, two anonymous referees, Roland Hodler, Ami Glazer, and participants at the UEA Meeting (Washington DC), the European Public Choice Meeting (Groningen), the 2015 CESifo Public Economics Conference, the 2015 ZEW Public Finance Conference, the 2015 meetings of the European Economic Association (Mannheim), the 2015 IIPF (Dublin), and at seminars at UCI-Irvine and DIW Berlin. This research has received funding from projects ECO2012-37131 and ECO2013-41310 (Ministerio de Economía y Competitividad) and 2014SGR-420 (Generalitat de Catalunya).

1 Introduction

Local governments across the globe receive a substantial share of their revenues in the form of grants from higher-level governments. In 2010, for example, this share was as high as 70% in the United Kingdom and the Netherlands. In other countries the share is smaller and local governments have more autonomy over their own resources (Foremny, 2014). In Spain, nearly 40% of all local revenues are allocated as grants, high enough to warrant special attention. Not only the quantity, but also the way in which these grants are allocated to cities and municipalities differs across countries. The optimal allocation of grants has been discussed extensively (see Oates, 1972, 1999; Wildasin, 1986; Bird and Smart, 2002, for surveys) and it is typically concluded that they should be allocated as unconditional block grants, so that local spending and taxation decisions are not distorted. At the same time, it is argued that grants should be formula-based to avoid any political bias in their allocation (Dixit and Londregan, 1995; Persson and Tabellini, 2002). Most formula grants, however, are vulnerable to manipulation by recipients because information needed to apply the allocation criteria is distributed asymmetrically between levels of government (Bordignon, Manasse, and Tabellini, 2001; Huber and Runkel, 2006). Although the academic literature has not paid much attention to this issue, the difficulties in gathering reliable data to implement these formulas are a concern for both policy makers and advisers (see, e.g., and Bahl, 2000 and Boex and Martínez-Vázquez, 2007). For instance, grantors, whose equalization formulas rely on measures of tax capacity, need to be able to secure this information, since, if tax collection is in the hands of local governments, there might be incentives to under-report the tax base to the higher tiers of authority. Countries that use complex calculations of spending needs are also particularly vulnerable to incentives to withhold information (e.g., Australia, the UK, and the Nordic countries, see Kim and Lotz, 2008). Ultimately, the failure to address these problems can undermine the workings of such grants. According to Bahl (2000):

> "A major constraint to designing a formula grants system is finding the data to implement and update the system. An important underlying issue is this: formula grants are appealing because their transparency and objectivity. These advantages can be taken away if the data to allocate the funds are suspect." (p.14)

It has often been argued that the least distortionary and manipulable way of allocating grants are schemes based on the number of inhabitants (see again Bahl, 2000, and Boex and Martínez-Vázquez, 2007). After all, counting its citizens through periodical censuses is one of the most basic tasks performed by any country (see Brambor et al., 2016) and having information of comparable quality for other indicators is a task predominantly undertaken by more modern administrations. However, if the collaboration of local governments is required for the collection of population information, even this variable might potentially be manipulated. Such a problem may be especially acute in developing countries, where the central government lacks

the capacity to monitor population in the field effectively (as also pointed out by Boex and Martínez-Vázquez, 2007). Therefore, governments seeking to implement a system in which the allocation of grants is based on population face a trade-off between accuracy and manipulability. Increasing accuracy by relying on local government reports (as opposed to using expensive and outdated decennial census data) might create incentives to manipulate population figures, unless the central government can dedicate sufficient efforts to ensure compliance.

In Spain, municipalities are responsible for the administration of their population registers (*Padrón Municipal de Habitantes*). Since 1998, the population registered on January 1st determines the allocation of grants from the central government to the municipalities, while previously this allocation was based on decennial census data, which is administered by the Spanish Statistical Office (INE in what follows), a central government agency. In theory, this agency also coordinates and monitors the municipal register, with the aim of ensuring that population figures reflect the reality. However, for several years, central government monitoring was far from perfect and the registered population was systematically inflated. The population censuses recently carried out uncovered huge numbers of 'ghost citizens', i.e., individuals for whom there was no actual evidence of their residing in the respective municipality. Moreover, several scandals have been reported in which city councils have been accused of systematically manipulating population numbers. The most notorious of these took place in Santa Cruz de Tenerife (a major city in the Canary Islands). In an audit released in 2009, the INE found nearly 15,000 'ghost citizens' (around 7% of the population), with evidence of fraudulent use of ID documents, among other irregularities.[1] Immediately following the publication of the audit, the mayor of Las Palmas (the largest city in the Canary Islands) complained about the loss of grants suffered by his city due to the over-reporting of population in Santa Cruz and, more generally, about the harm done to the objectivity of the allocation of grants to municipalities in the region.[2]

Theoretically, for a given level of enforcement, over-reporting will be larger the more generous these grants are (i.e., the higher the amount of money linked to an additional resident). To the best of our knowledge, no previous attempt has been made to estimate the effect of grant generosity on population over-reporting. The reason for this is that it is very difficult to estimate this magnitude for a linear grant scheme, since the marginal effect of an additional resident on the amount of grants is constant across municipalities. Fortunately for our purposes, in Spain, per capita grants to municipalities jump at population thresholds. Specifically, the grant per inhabitant changes discontinuously at thresholds of 5,000, 20,000 and 50,000 inhabitants for the entire local population. Per capita grants that increase discontinuously at

[1] See, for example, "A city with 15,000 'ghosts' [Una ciudad con 15.000 fantasmas]", in *Canarias7*, 05/04/2013.

[2] The undue amounts received by Santa Cruz, during a ten-year period, were estimated at around 40 million euros, while the mayor estimated the amount of grants lost by Las Palmas at around 6.5 million euros (see "The mayor of Las Palmas demands compensation for municipalities because of the fraud in the Santa Cruz register [(El Alcalde de las Palmas) pide al Gobierno una compensación para los municipios por el fraude del padrón en Santa Cruz]", in www.eldiario.es, 07/02/2014).

population thresholds create incentives for municipalities to sort to the right of the threshold, as one additional inhabitant brings additional grants for all existing inhabitants. Although all municipalities might be tempted to misreport population figures, the incentives to misreport are stronger for municipalities close to the threshold. There is considerable anecdotal evidence from the municipalities citing the increase in grants to justify an aggressive policy to boost their populations above one of these thresholds.[3] Moreover, the jump in grants at the threshold provides exogenous variation in the marginal effect of an additional resident. This allows to identify the effect of grant generosity on population misreporting. This effect is estimated for municipalities close to the threshold but can be extrapolated to municipalities far from the threshold in order to quantify the potential effect of linear grant schemes on the incentives to misreport.

Policies that create jumps in governments' choice sets are commonly referred to as notches (Slemrod, 2010). Notches have been extensively used to study the effects of taxes on the behavior of individuals and firms (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013). This paper is, to the best of our knowledge, the first to exploit notches to study the response of local governments to the incentives introduced by intergovernmental grants. We make use of these notches to estimate the effect of grant generosity on population over-reporting. Specifically, we use the methods developed in the taxation literature to quantify bunching that, in our case, corresponds to the excess density found above the notch points.[4] The method allows us to estimate the implied responses of population over-reporting to grant generosity. We perform several heterogeneity analyses (by period and type of municipality) that, linked to a stylized theoretical model of population reporting, help us to interpret our results.

We find significant responses and heterogeneity in the responsiveness of municipalities to grants. To the left of the 5,000 threshold, there is a loss of mass of around 10% of the municipalities. In the most extreme cases, over-reporting can be as high as 1,000 additional residents, although on average it is around 60 residents. This implies that, for the most responsive municipalities, the elasticity of population over-reporting to grants is about 0.40, while the average response is around 0.013. The extrapolation of this last number to the whole distribution of municipalities indicates that the leakage of grant money due to population over-reporting could be in the order of 40 million euros per year, suggesting potential gains from improvements in audit policies. In fact, we find that the extent of over-reporting is higher during the period 1998-2005, and almost disappears after 2005, coinciding with an improvement in the enforcement of population numbers by the INE. We also show how the census is able to iden-

[3] This is the case, for example, of Manlleu, a city nearby Barcelona; in 2007 the city jumped over the 20,000 threshold and the effect on the amount of grants was discussed in the city council (see "The 20,000 effect [L'efecte 20,000 habitants]", in *El 9 nou*, 14/10/2013). In Cardona, another town nearby Barcelona, the mayor called the population to mobilize against the possibility of falling below the 5,000 threshold and losing grants from the central government (see "Cardona mobilized in search of more neighbors [La localidad de Cardona se moviliza para buscar más vecinos]", in *La Vanguardia*, 21/12/2014).

[4] For an updated review of papers studying bunching see Kleven (2016).

tify the non-compliers, that is, the municipalities that were shifted below the 5,000 threshold due to discrepancies between the register and the 2001 census. Non-compliers tend to have high shares of foreign population (i.e. immigrants) and second homes, which suggest that over-reporting was more difficult to detect in their case. They also have a low level of civic capital (as measured by electoral turnout), suggesting that the (political) costs of being caught to over-report were also lower. These results are consistent with our model's predictions and help to explain the heterogeneity we found in the responses. Several pieces of evidence suggest that manipulation (as opposed to real population responses) is the mechanism generating the observed bunching patterns. A number of robustness checks indicate that these bunching patterns are not likely to be explained by other institutional arrangements that also change discontinuously at population thresholds.

A few papers have already studied the incentives created by discontinuities in formula transfers. Litschig (2012) shows that grants allocated through a population-based formula in Brazil are not effectively shielded from special-interest politics. The central government, responsible for the administration of relevant population information, inflated the population numbers of its politically aligned municipalities. As a consequence, aligned municipalities received on average more grants than non-aligned municipalities. This indicates that higher-level governments may implement formulas that they can subsequently manipulate in their favor. In contrast, we do not find clear evidence of such political manipulations. Instead, we find that the behavior of Spanish municipalities is driven by the opportunities and incentives to manipulate population numbers. The findings in Litschig (2012) and our own results are similar in the sense that both point to the difficulty of avoiding manipulation of simple population-based formula grants, but would imply a different solution to the problem: in the case of Brazil the solution seems to be more transparency while in the Spanish case, it is the improvement of the administrative capacity of the central agency. Another paper similar to ours is De Witte and Geys (2015), which studies the incentives introduced by a similar grant in Belgium.[5] They do not find any evidence of over-reporting. However, they find that population and housing construction growth becomes larger as municipalities approach a threshold, which they interpret as evidence that municipalities use real policies to try to reach the threshold. We find no evidence of such an effect: the bunching pattern we find clearly indicates manipulation as it disappears when we look at census data and after an increase in enforcement. Furthermore, housing growth rates do not vary in any meaningful way as one approaches the threshold. The differences between the two cases might be indicative of the responses of municipalities to

[5] Grants which give higher weights per capita when population is larger are not uncommon. Other countries in Europe rely on similar grant schemes. For example, most German states use population based-grants with changing slopes, which could potentially raise the incentives to manipulate population numbers around kink points. These grants are allocated using municipal registers and, in fact, the 2011 census found a big discrepancy between register and census population. See "Declining population causes dispute over money [Bevölkerungsrückgang löst Streit ums Geld aus]", *Die Zeit*, 31/05/2013). As an example, Baskaran (2016) analyzes such discontinuities in the allocation formula for general-purpose transfers that the state government of Hesse pays to its municipalities.

changes in enforcement: when central government monitoring is strictly implemented, municipalities cannot manipulate population figures and must rely on other channels to try to influence population numbers.

The paper also contributes to the literature of regression discontinuity designs that study the effects of local policies that depend discontinuously on municipality size.[6] In a detailed study of the French, German and Italian cases, Eggers et al. (2017) identify the threats to identification that arise in this setting and provide guidance for addressing them. As in our study, they document discontinuous densities around the relevant population thresholds in all three countries. Hence, the manipulation of population figures by local authorities seems to be a widespread phenomenon in Europe. Our study differs from Eggers et al. (2017) in the following respect: while their study is concerned with the consequences of sorting when estimating the effects of policies that change discontinuously at population thresholds, we use the sorting induced by the policy discontinuity to quantify the effect of grants on population over-reporting.

Eventually, the paper also relates to the tax enforcement literature, see e.g. Slemrod (2007) and Besley and Persson (2013). Our results on the responses to grant generosity mirror those of papers providing evidence on the effect of tax rates on tax evasion (Fisman and Wei, 2004). Our findings on the effect of improved monitoring add to the evidence provided by many studies on the effects of third-party reporting, crosschecking and other tax enforcement technologies (Bird and Zolt, 2008; Kleven et al., 2011). The evidence that the non-compliers are municipalities with low civic capital is in line with other papers that show that tax compliance is influenced by social capital and political incentives (Casaburi and Troiano, 2015; Besley, Jensen, and Persson, 2015). Note also that, as in our case, these last two papers deal with non-compliance that materializes in missing houses and population.[7]

The rest of the paper is organized as follows. In Section 2 we describe the institutional background in detail. In Section 3, we develop a stylized model of population over-reporting to guide the empirical findings. The empirical methods and data are described in Section 4 while the results are presented and discussed in Section 5. In Section 6, we analyze the determinants of the discrepancies between census and register populations to shed further light on the mechanisms at work. Section 7 concludes.

[6] The papers by Casas-Arce and Saiz (2015) and Bagues and Campa (2016) which study the effects of political gender quotas that are binding for large municipalities are good examples of such studies in the Spanish case.

[7] The paper by Casaburi and Troiano (2015) analyzes the effect of an Italian enforcement program that used aerial photographs to uncover a massive number of 'ghost buildings' (i.e., property not included in the land registry and thus hidden to the tax authorities) that boosted local property tax revenues. In the second paper, Besley, Jensen, and Persson (2015) focus on the evasion of 'poll tax' liabilities in UK. After the introduction of the poll tax, more than four million of voters disappeared from the electoral register and some local authorities and academics believed that these missing voters were trying to evade the poll tax (see "'Missing' 3.4 million voters put MPs at risk: poll tax evasion could distort review", *The Independent*, 17/06/1993). Scottish local authorities are still searching their half million missing voters (see "Councils plan to track down poll tax dodgers among 'missing' million voters", *The Express*, 30/9/2014).

2 Institutional background

2.1 Local public finance in Spain

The size of the local government sector in Spain is moderate, with municipal budgets representing around 15 per cent of total public spending in our period of study. Spanish municipalities have spending responsibilities similar to those in other countries (e.g., water supply, refuse collection and treatment, street cleaning, lighting and paving, parks and recreation, traffic control and public transportation, social services, etc.), with the sole exception of education, which is the responsibility of regional governments. In terms of revenues, in 2012, inter-governmental grants represented 36 per cent of local operating budgets, and 68 per cent of these transfers were in the form of a formula-based block grant allocated by the central government (the *Participación Municipal en los Ingresos del Estado* - PIE). This grant - as all other financial concerns of Spanish local governments - is regulated by a law passed in 1988 by the central government (Law 39/1988 *Reguladora de Haciendas Locales*) and applies to all municipalities in the country. Although the variables and their exact weights have experienced (minor) changes over time, (weighted) population determined the allocation of 75 per cent of its funds during the years of our study.[8]

[Figure 1 about here]

To account for supposedly differential spending needs, a weight of 1 is applied to residents in municipalities with fewer than 5,000 inhabitants. This weight then jumps to 1.17, 1.3 and 1.4 as the population of the municipalities surpasses the thresholds of 5,000, 20,000 and 50,000 inhabitants, respectively.[9] As such, the weight or multiplier is not applied exclusively to the population in excess of the threshold, but rather to the entire population of the municipality. The effect of the weighting scheme on per capita resources depends on the size of the pot of funds to be distributed in each year, which is determined by the evolution of national tax revenues. Panel a) of Figure 1 illustrates this with data from 2005 budgets. In that year, the amount of per capita resources was 96, 113, 125 and 135 euros for municipalities below 5,000, between 5,001 and 20,000, between 20,001 and 50,000, and above 50,000 inhabitants, respectively. The

[8] The other variables used in the formula are fiscal effort, the inverse of tax capacity, and the number of school units, see Solé-Ollé and Bosch (2005).

[9] The law establishes a different funding scheme for municipalities with more than 75,000 inhabitants, which obtain a share of tax revenue in specific taxes (income, VAT and excises). Although the comparison between the two systems is not straightforward, the second one seems to be more generous, creating also a notch. There is anecdotal evidence of municipalities aiming at jumping over this threshold; examples are the cases of Benidorm, in the Mediterranean coast ("Benidorm estimates in 1 million euros the loss of transfer revenues due to the downward correction of 4,700 inhabitants [Benidorm cifra en 1 millón de euros lo que no ingresará del Estado al perder de un plumazo 4.700 habitantes]", www.LaMarinaPlaza.com 13/01/2015) and Alcalá de Guadaíra, a city close to Sevilla ("Alcalá de Guadaíra stays without the 'City' status in not passing the 75,000 threshold [Alcalá de Guadaíra se queda sin la calificación de Gran Ciudad al no llegar a los 75.000 habitantes]", *ABC (Sevilla)*, 28/01/2016). Unfortunately, the very small number of municipalities close to the 50,000 and 75,000 thresholds renders any analysis unfeasible.

absolute amount of resources involved are 85,000, 240,000 and 500,000 euros per year, providing strong incentives to municipalities to ensure they find themselves on the right side of the threshold. This is shown in Panel b) of Figure 1 for the 5,000 threshold.

2.2 Local population figures

Up to 1998, the population figures used to compute the formula grant were those of the census. The census quantifies the permanent population of each Spanish municipality by visiting all households in the country and is the responsibility of the Spanish Statistical Office (INE), an agency of the central government. The population census is carried out every tenth year, the last three corresponding to 1991, 2001 and 2011, although a similar operation took place in 1996 (known as the *Renovación Padronal*) so as to reduce the excessive time lapse between censuses. The population coming from the 1991 census was used to allocate grants for the period 1991-1997 whereas grants in 1998 were distributed in accordance with the *Renovación Padronal* of 1996.

Since 1998, the population registered on January 1st determines the allocation of grants from the central government to the municipalities.[10] In contrast to the census, the population register is administered by municipalities. Residents have to register in the municipality where they spend most of the year. Being registered is a condition for access to public services (such as education and health care, for example), and for both Spanish nationals and EU citizens determines where they are allowed to vote at local elections. Municipalities are responsible of registering people in the first place, of updating the register in the event of death or migration, of ensuring all people included in the register are actual residents, and of periodically reporting the number of registered residents to the INE. We must be clear about what the INE can and cannot do in this process. Although the INE can adjust downward the population of the municipalities, it is not the hierarchical superior of the municipalities and it cannot impose penalties on them if there is evidence of cheating. The INE or any other agent (e.g., the parties in the local opposition, a civic association) need to go to courts and a judge will decide on the case. Municipalities can also go to courts in case the downward adjustment imposed by the INE is considered arbitrary. The reputation of professionalism of the INE (and also of the court system) when dealing with these matters is substantial, and there are no cases of complaints by municipalities related to arbitrary decisions or to partisan favoritism.

The reasons to start using the register to allocate grants were the complaints of local governments regarding the gap between census information and reality, and the confidence in the capacity of the INE's computerized monitoring system to detect irregularities in the reporting of population. On a monthly basis the INE sends information to each municipality about

[10] Grants in 1999 and 2000 were distributed with the population measured in January 1st 1998 and 1999, respectively. From 2001 and onward, this lag disappears and population on January 1st determines the grants received in that same year.

births, deaths and internal migrations. The municipality uses this information and its own operations to update the register and sends a new count to the INE. If the municipality does not adjust the register to the information send by the INE and it if does not provide evidences that support this behavior, the INE will adjust the population numbers downwards. These numbers will also be downward adjusted in the event of new information revealed during a special audit procedure.

In practice, the automatic monitoring by the INE has been far from perfect for a number of years, which has led to an over-reporting of population numbers in some municipalities. These are the two main mechanisms at work:

- Unreported migrations abroad and over-registration of visitors. Internal movements of either Spanish nationals or foreigners are automatically detected by the INE when the person registers in another municipality. However, immigrants that return home (or move to another country) are more difficult to detect, since they seldom communicate that they have abandoned the country (and the municipality) of residence. In theory, the municipality is responsible for periodically checking that immigrants still live in town, but they often lack the incentives to do this job properly. This means that the number of immigrants is inflated in the register. A similar case is that of the EU citizens and also of Spanish nationals that spend only part of the year in the municipality (e.g., they own a second home). Municipalities tend to include them in the register without paying too much attention to the length of time they spend there.[11]

- Incorrect inclusion in the register. This might be due to administrative mistakes. For example, a resident might be included in the register more than once (i.e., in different cities) under slightly different names and/or with different documents giving proof of identity (which happens mostly in the case of immigrants). In some cases the discrepancy is due to overt attempts to manipulate the register.[12]

The central government had known about the existence of some of these problems since the implementation of the system in 1998, but their magnitude was probably unexpected. Note,

[11] In fact, there is ample anecdotal evidence of municipalities dedicating considerable efforts and resources to campaigns designed to convince these people to register. For instance, in Cardona (a small town near Barcelona), the mayor sent a letter to all residents telling them: "We encourage all the inhabitants of Cardona to lend a hand in enlarging our municipal register (...) with the aim of avoiding the loss of resources"; the letter included detailed instructions on how to register (see "Cardona mobilized in search of more neighbors [La localidad de Cardona se moviliza para buscar más vecinos]", in *La Vanguardia*, 21/12/2014).

[12] For example, the INE's internal investigation of the case of Santa Cruz de Tenerife revealed that "there were sheets and sheets of names in the register with thousands of false ID documents, while other names were duplicated one after another - entire blocks of neighbors that simply did not exist". The INE found "15,000 'ghost' residents after crosschecking the register with an ID database from the Ministry of the Interior. About 5,000 were Spanish nationals without a proper ID document, around 7,500 were Spanish nationals without any proper proof that they registered, and about 2,500 were foreigners that had left the city". See "Suspicion of fraud in the register of Santa Cruz de Tenerife. Commission set up to investigate more than 15,000 false IDs [Sospecha de fraude en el padrón municipal de Santa Cruz de Tenerife. Comienza una comisión para investigar la existencia de más de 15.000 DNIs falsos]", *El Confidencial Autonomico*, 17/04/2013.

for instance, that the huge immigration boom experienced in Spain during those years came as a surprise in 1998. During this period there was also a massive boom in housing construction along the coast which exacerbated the problems related to the monitoring of population in tourist areas. However, the INE has learned from these developments and during the period it has progressively improved enforcement through the implementation of several automatic checks and auditing operations:

- The most relevant automatic check is related to deal with the problems involved in monitoring immigrants. In order to deal with the under-reporting of out-migration and double counts, in 2005 the INE made it compulsory for non-EU immigrants to renew their registration every second year. In theory, all immigrants registered before December 2003 that failed to re-register were to be deleted from the register in December 2005. A similar mechanism was also implemented for EU residents who, from March 2009, must also re-register every fifth year.

- Another relevant audit operation was introduced after the 2001 census of population. The census operations revealed large discrepancies with the register and municipalities were required to justify these discrepancies or to make the necessary downsize adjustments. These operations were extended until the end of 2004. Also, to deal with problems of fraud involving Spanish nationals, in 2009 the INE started to cross-validate the register with the aforementioned ID database built by the Ministry of the Interior.

Three conclusions emerge from this discussion. First, it has been easier for some local governments to over-report population than it has been for others. This is the case for municipalities with high shares of immigrants (both non-EU and EU citizens) and those that are home to many second residences. Second, while some of the over-reporting can be attributed to statistical errors, there is ample anecdotal evidence of overt manipulation of population numbers. In some cases the manipulation is due to the municipalities shirking on their responsibilities to manage the register, while in other cases it involves committing illegal acts, and should be qualified as cheating or fraud. Under the Spanish law, the manipulation of the register is a crime of 'documentary misrepresentation' and might entail considerable penalties.[13] Although only fraud can be legally punished, both types of behavior can lead to some sort of punishment by voters, specially when the strategy to upgrade the status of the municipality fails.[14] This sug-

[13] 'Documentary misrepresentation' can entail fines, removal from public office, and prison of up to ten years. In the case of Santa Cruz, the attorney general initiated an investigation that proved the existence of such a crime; in another case in Gandía - a city on the Mediterranean coast that had jumped fraudulently over the 75,000 threshold, the public official responsible of the statistics department of the city council was convicted of this and other crimes; see "The arrested civil servant in Gandía falsified more than 200 registrations of immigrants [La funcionaria detenida en Ganía falseó más de 200 empadronamientos a inmigrantes]", www.lasprovincias.es, 29/08/2007.

[14] For example, in the case of Alcalá de Guadaíra cited above, the press reflected that the mayor announced the arrival of a grant 'mana' after reaching the 75,000 inhabitants just before having to admit that the resources would never come.

gests that we should find that, controlling for opportunities for over-reporting, places with less civic capital (i.e., were voters care less about complying with rules) and with politicians that are less subject to political control should over-report more. Third, the ability of the INE to monitor the municipal register seems to have improved over time (i.e., after 2005). If the excess mass observed to the right of the thresholds is explained by manipulation, this phenomenon should be less intense in the last years of our sample. Analyzing bunching in the 2001 and 2011 censuses provides further evidence to disentangle manipulation from actual population growth as the underlying mechanism at work.

3 A model of population over-reporting

In this section, we develop a stylized model that captures the story described in the previous section. The model is similar to a tax evasion model with endogenous probability of detection (Yitzhaki, 1987; Kleven et al., 2011), although adapted to the specific characteristics of the case in question. In the model, local governments decide on the population count reported, taking into account the possibility of a downward adjustment in these figures made by the central government, which arise as a result of either automatic checks or audits. When deciding on the amount of reported population, the local government has to weight the benefits derived from over-stating the count and the potential costs to be borne in case a correction to the population count is applied. This generates an inverted U-shape objective function which implies bunching in the presence of a notch.

The section is organized as follows. First, we show how the local government determines the amount of population it will over-report in the case of a linear transfer schedule (i.e., a transfer allocated proportionally to population). Second, we analyze how the amount of over-reported population is affected by the presence of a notch in the transfer schedule. Finally, we introduce heterogeneity in the probability of detection and take into account potential downward adjustments made by the central government.

Linear transfer schedule

For the moment, we assume that the amount of transfers τ received by a local government is the product of the level of reported population r and per capita transfers α (i.e., $\tau = \alpha r$). Let's define $e = r - n$ as the level of over-reported population, where n is the number of residents in the municipality (which we assume to be exogenous) and r is reported population. The probability that population over-reporting is detected is denoted by $p(e)$. We parametrize p using the following Pareto distribution $p = 1 - e^{\gamma}$ where $-1 < \gamma < 0$. The local government

maximizes the following expected utility function:

$$u = (1-p)\alpha(n+e) + p(\alpha n - \theta e) \tag{1}$$

which can be re-written as:

$$u = \alpha n + \alpha e - p(\alpha + \theta)e \tag{2}$$

where θ measures the cost per over-reported citizen if caught. This specification means that costs are higher for greater as opposed to smaller deviations, given that massive over-reporting can hardly be attributed to error. These costs incurred in the event of detection, accounted for by θ, may include: (i) the possibility of losing resources already spent or compromised (the loss of transfers after the correction of the population numbers), (ii) the political costs related to the evidence of incompetence or unethical behavior, and (iii) formal penalties in case a judge decides there is evidence of unlawful behavior. The expected costs increase with the perceived probability of detection.

The first order condition that maximizes (2) is:

$$e^* = \left(\frac{\theta}{\alpha+\theta}\frac{1}{\gamma+1}\right)^{1/\gamma} \tag{3}$$

Given that $-1 < \gamma < 0$, it can readily be seen that over-reporting increases with grant generosity (α), decreases with the costs faced after over-reporting is detected (θ) and increases if the probability of detection is low (γ is small in absolute value).

Notched transfer schedule

Let's now consider the outcomes when the transfer schedule has a notch. The notched schedule can be written as:

$$\tau = [\alpha + \beta d(r > \hat{r})]r \tag{4}$$

Where $d(r > \hat{r})$ is a dummy variable equal to one if the reported population is larger than the threshold $\hat{r}$. According to this expression, per capita transfers are equal to α if $r \leq \hat{r}$ and equal to $\alpha + \beta$ if $r > \hat{r}$. At the notch threshold, increasing the reported population by one inhabitant causes the transfers to jump by $\beta\hat{r}$, thus creating powerful incentives to over-report. The utility of a municipality with a notched transfer is:

$$u = (1-p)[\alpha + \beta d(r > \hat{r})](n+e) + p[(\alpha + \beta d(r > \hat{r}))n - \theta e] \tag{5}$$

Figure 2 illustrates the expected utility level with a notched transfer scheme with a discontinuity at $\hat{r}$. The increase in utility at the notch is $(1-\hat{p})\beta\hat{r}$. That is, the increase in grants, $\beta\hat{r}$, times the probability of not being detected, $1-\hat{p}$, where $\hat{e} \equiv \hat{r} - n$ and $\hat{p} \equiv p(\hat{e})$. In the example shown in Figure 2, the municipality is close enough to $\hat{r}$ and the expected utility at the threshold, $u(\hat{r})$, is higher than at its interior equilibrium, $u(r^*)$. As it delivers higher utility, this municipality will report $\hat{r}$.

[Figures 2 and 3 about here]

Panel a) in Figure 3 shows all three possible cases. The dotted line depicts the case of a municipality that is too far from the threshold. Here, $u(r^*) > u(\hat{r})$ and so the municipality will report its interior equilibrium level, r^*. The dashed line corresponds to the case illustrated in Figure 2 in which the municipality is better off reporting $\hat{r}$ than its interior equilibrium r^*. Finally, the solid line defines the municipality that is indifferent between reporting r^* and $\hat{r}$. For this marginal municipality, which reports $\tilde{r} \equiv \tilde{n} + e^*$, the utility achieved at its interior optimum is equal to the utility at the notch, i.e., $u(r^*) = u(\hat{r})$, which implies:

$$\alpha\tilde{n} + \alpha e^* - p^*(\alpha+\theta)e^* = \alpha\tilde{n} + \alpha\hat{e} - \hat{p}(\alpha+\theta)\hat{e} + (1-\hat{p})\beta\hat{r} \tag{6}$$

Among the municipalities with $r^* < \hat{r}$, those with $n \leq \tilde{n}$ will report its interior equilibrium r^*, while those with $n > \tilde{n}$ will report $\hat{r}$. In turn, municipalities with $r^* > \hat{r}$ will just report r^*[15]. Hence, the presence of the notch modifies the reported population density function as shown in Panel b) of Figure 3. An excess mass appears at $\hat{r}$, which corresponds to the density hole between $\tilde{r}$ and $\hat{r}$.

The location of the 'marginal buncher' identifies the reported population change induced by the notch, dr, which amounts to $\hat{r} - \tilde{r}$. A measure of the change in grants faced by the marginal buncher enables us to identify the responsiveness of reported population to grants. When moving from $\tilde{r}$ to $\hat{r}$, grants increase by $\alpha(\hat{r} - \tilde{r}) + \beta\hat{r}$. Relative to the increase in reported population, this grant increase is $\alpha + \beta\hat{r}/(\hat{r} - \tilde{r})$. Given that just below $\tilde{r}$ grants increase by α, grants per reported inhabitant increase by $\beta\hat{r}/(\hat{r} - \tilde{r})$ between $\tilde{r}$ and $\hat{r}$. Hence, the elasticity of reported population to grants can be estimated as follows:

$$\epsilon_{r,\alpha} = \frac{dr/r}{d\alpha/\alpha} \simeq \frac{\frac{\hat{r}-\tilde{r}}{\tilde{r}}}{\frac{\hat{r}}{\hat{r}-\tilde{r}}\frac{\beta}{\alpha}} = \frac{(\hat{r}-\tilde{r})^2}{\hat{r}\tilde{r}}\frac{\alpha}{\beta} \tag{7}$$

[15] Inspecting equation 6 reveals that the expression can only hold for a unique level of population. Note that the $\alpha\tilde{n}$ terms drop out, yielding $\left(\alpha e^* - p^*(\alpha+\theta)e^*\right) = \left(\alpha\hat{e} - \hat{p}(\alpha+\theta)\hat{e}\right) + \left((1-\hat{p})\beta\hat{r}\right)$. The left-hand side of this equation is a constant term, since e^* and p^* do not vary across municipalities. It turns out that the right-hand side is an increasing function of n. For $e > e^*$, the term $\alpha e - p(\alpha+\theta)e$ diminishes with e. Hence, $\alpha\hat{e} - \hat{p}(\alpha+\theta)\hat{e}$ increases with n as $\hat{e} = \hat{r} - n$. The second term, $(1-\hat{p})\beta\hat{r}$, also increases with n because of the lower probability of being caught when reporting $\hat{r}$. Hence, there is a unique $\tilde{n} = \tilde{r} - e^*$ for which equation 6 holds.

From theory to empirics

The stylized model outlined above can be made more realistic by allowing γ_i to be heterogeneous across municipalities to reflect the fact that some municipalities might have more possibilities of over-reporting than others. To analyze what 'bunching' might look like in this scenario, we resort to simulations. The results are presented in Panel b) of Figure 4. Compared with the baseline case with no heterogeneity - shown in Panel a) - we see that now the range that goes from $\tilde{r}$ to $\hat{r}$ is not completely empty. Indeed, approaching $\tilde{r}$ there are many municipalities for which deviating from the optimum is now too costly; the proportion of these municipalities decreases steadily as one approaches the notch threshold. Note that in this more realistic setting, the responses to grants will be municipality-specific. This heterogeneity in responses implies that $\hat{r} - \tilde{r}$ and expression (7) would now correspond to over-reported population and the elasticity of reported population to grants of the most responsive municipalities. When discussing the results, we will also consider average responses in over-reported population, as well as in the elasticity of reported population to grants.

[Figure 4 about here]

Empirically, we expect that the heterogeneity in the possibilities of over-reporting is related to the presence of immigrants and second homes. Heterogeneity can also come from differences in the costs of over-reporting (i.e. heterogeneous θ_i's). The cost of being caught manipulating population figures can be related to the amount of civic capital (which might be higher in places with high turnout) and to the degree of control to which the local incumbent is subjected (which will be higher the lower her vote share is). Finally, over-reporting can also depend on whether the local government is politically-connected to the central government; a mayor aligned with the central incumbent might expect a lower monitoring effort or that some senior co-partisan will intercede if regularities are detected. Later on in the paper, we will investigate the determinants of over-reporting at the municipal level, by examining differences between register and census population figures. In that analysis, we will be able to assess which sources of heterogeneity matter the most.

The theoretical framework developed yields predictions in terms of reported population, i.e. $r = n + e$. In terms of data, this would correspond more closely to the population figures initially submitted by municipalities to INE. Unfortunately for our purposes, we can only observe official population counts, which amount to reported population by municipalities less any adjustments applied by INE. In terms of our model, we can think of the population figures that we observe as $r - pe$ or $n + (1 - p)e$. Panel c) in Figure 4 shows bunching for $r - pe$ in the case where the γ_i's are heterogeneous. For each simulated municipality, p is randomly drawn from a Bernoulli distribution with $p = 1 - e^{\gamma_i}$. For those municipalities that are caught, the adjusted population will be smaller by e inhabitants with respect to the initially reported population. Although the excess density (missing mass) to the right (left) of $\hat{r}$ does not disappear unless all

over-reporting is detected, there is more bunching in the population figures initially reported by municipalities - panel b)- than there is for the adjusted population figures -panel c)- which correspond more closely to the data examined in the empirical exercise.[16]

4 Bunching identification and data

4.1 Identifying bunching in practice

To quantify bunching behavior, we follow Kleven and Waseem (2013) whose procedure is summarized in Figure 5. Bunching at a notch point can be quantified by calculating the difference between the empirical density and an estimated counterfactual density.

[Figure 5 about here]

The model outline above predicts that all bunching municipalities will report exactly $\hat{r}$. In order to bring this sharp prediction to the data, we specify the population interval comprised between the notch ($\hat{r}$) and r_U to be the area suspected to show excess density. Likewise, r_L and $\hat{r}$ specify the population interval suspected to show missing density. Note that r_L is the empirical counterpart of $\tilde{r}$. Choosing r_L and r_U defines the excluded range (the area suspected to be affected by bunching behavior, where L stands for the lower bound and U for the upper bound). We group municipalities into population bins indexed by j. The baseline bin-size corresponds to 1 percent of the respective threshold, which results in 50 inhabitants-groups per bin at the 5,000 threshold, and 200 inhabitants-groups per bin at the 20,000 threshold. We provide other specifications with different bin-sizes as robustness checks. The counterfactual distribution is then estimated by fitting a polynomial of order q to the actual distribution of population bin counts (c_j), where dummies are included for the bins in the excluded range:

$$c_j = \sum_{h=0}^{q} \beta_h (r_j)^h + \sum_{i=r_L}^{r_U} \gamma_i \mathbf{1}[r_j = i] + \varepsilon_j \tag{8}$$

The (estimated) counterfactual distribution is built with the polynomial coefficient estimates obtained in (8), namely, $\widetilde{c_j} = \sum_{h=0}^{q} \widetilde{\beta_h} (r_j)^h$. The difference between the empirical and estimated densities provides us with a direct measure of B (the excess mass between $\hat{r}$ and r_U) and of M (the missing mass between r_L and $\hat{r}$). In practice, it might not always be obvious how to choose the excluded range, that is, r_U and r_L. The excess mass should be relatively concentrated to the right of the notch point and, following Kleven and Waseem (2013), we choose the upper limit r_U by visual inspection. Then, we choose r_L by equating the excess mass above the notch (B) with

[16] Since empirically we will not be able to observe neither e nor p, we do not know the empirical extent of the difference between the initially reported population ($n+e$) and the population net of the adjustments applied by INE ($n+(1-p)e$) as e and p could be both large or small.

the missing mass below the notch (M). The method also requires the researcher to specify q, the polynomial order of the counterfactual density. The city size distribution has been widely studied in the urban economics literature, and while there is some controversy as to which parametric distributions best fit the data (see, for example, Gabaix, 1999 and Eeckhout, 2004), the city size distribution is smooth and can be well approximated by a low-order polynomial. In our baseline specification, we choose a second-order polynomial (i.e. q=2), although we test the robustness of the results to higher-order polynomial specifications. To assess the statistical significance of the results, we follow Chetty et al. (2011) and compute standard errors using the residual bootstrap.

Besides the empirical methodology explained here, we also perform McCrary (2008)-tests to formally assess if population densities are continuous at the respective thresholds. We do so because a discontinuity in the population density is a prediction of our theoretical model of population manipulation. Besides, the empirical methodology developed by Kleven and Waseem (2013) becomes problematic to implement in situation where bunching is not obvious as the method still requires the researcher to specify r_U by visual inspection. In these cases, the McCrary (2008) test becomes a valuable statistic to confirm that there is no bunching in the data.

4.2 Data

As explained above, we can only observe official population counts, which amount to reported population by municipalities minus any adjustments applied by INE, i.e. $n + (1-p)e$. Hence, B, the excess density to the right of $\hat{r}$, is actually referred to municipalities that over-reported population to surpass $\hat{r}$ that went undetected.[17] However, note that the amount of grants received by municipalities is determined by 'undetected' over-reporting rather than by the population figures initially submitted by municipalities. Hence, from a policy perspective, the over-reporting responses to grants that we estimate are the relevant ones to assess the consequences in terms of misallocation of resources.

[Table 1 about here]

We use the official population counts (*Padrón municipal*) for the years 1998-2011. These population figures, referred to January 1st, are the ones used to determine the allocation of the transfer. As explained in Section 2, per capita grants change at thresholds of 5,000, 20,000, and 50,000 inhabitants. We include the range of +/40% of the threshold into our plots and estimations. We have chosen a 40% window because larger windows might include observations

[17] In some analyses, we will examine municipal population counts from (decennial) censuses, which are considered to be faithful reflections of the 'real' population (Jurado, 2004). In terms of the model, the census provide us with a measure of real population (n) but does not help us to back out r, the reported population initially submitted by municipalities.

which are affected by the previous or next threshold. Therefore, each graph is based on 40 bins to the left and another 40 bins to the right of each threshold. Table 1 reports the minimum and maximum number of yearly observations within +/- 40% of the relevant population over the period 1998 to 2011. We also report the figure for the 3,000 and 10,000 threshold, which we will use in a robustness check and for a placebo test, respectively. The table shows that the municipality size distribution is highly skewed, with the number of relevant observations decreasing rapidly as we move to the upper thresholds. Note that the empirical methods we employ here were developed in the context of millions of individual tax returns (see, for example, Chetty et al., 2011 and Kleven and Waseem, 2013), and, thus, they rely on thick densities. We study municipalities instead of individuals, which naturally provides us with fewer observations. Therefore, we primarily focus on the 5,000 threshold as it provides us with the highest number of observations possible, but provide similar evidence for the 20,000 threshold.

5 Results

5.1 Bunching at the 5,000-inhabitant threshold

This section identifies and quantifies bunching behavior at the 5,000-inhabitant threshold. Panel a) in Figure 6 serves as the empirical counterpart of Figure 5, used in illustrating the empirical methodology. The (appropriately binned) data show the empirical density and provide strong (suggestive) evidence that there are too many (few) municipalities to the right (left) of the threshold.

[Figure 6 about here]

Panel b) in Figure 6 shows the McCrary (2008) test and confirms that, indeed, the empirical density is not continuous at the 5,000-inhabitant threshold. Panel a) in Figure 6 also reports r_L and r_U, which determine the excluded range, while the solid line is the counterfactual density obtained if a second-order polynomial is specified (i.e.q=2). The corresponding bunching estimates are reported in Table 2.

[Table 2 about here]

r_U is chosen by visual inspection at 5,800 inhabitants, which is the first time that the empirical density crosses the threshold of the counterfactual density.[18] Summing up the difference between the empirical and counterfactual densities between $\hat{r}$ and r_U delivers B, the excess number of municipalities above the threshold. There are 362 municipalities that would be below 5,000 inhabitants in the absence of the notch. The standard error of B is 79.52, indicating

[18] Below, we provide estimates based on alternative choices of r_U.

that B is statistically significant at the 1% level. Remember that we choose r_L (3,850 in our application) to equate B and M; that is, the excess mass between $\hat{r}$ and r_U must equal the missing mass between r_L and $\hat{r}$. Hence, the number of missing municipalities below 5,000 inhabitants is also 362.

The fact that the density hole disappears at 3,850 inhabitants indicates that the most responsive municipalities over-report around 1,150 inhabitants. However, the population interval between r_L and $\hat{r}$ is far from being empty. Summing the counterfactual density between 3,850 and 5,000 inhabitants gives us the number of municipalities between r_L and $\hat{r}$ that would be observed in the absence of the notched transfer schedule. Dividing B by this number gives us the fraction of respondents. Only about 10% of all municipalities that would be in the 3,850-5,000 range crossed the 5,000 threshold to receive higher per capita grants.[19]

Within the density hole area, we know the counterfactual ($\widetilde{c_j}$) and observed (c_j) municipality counts in each bin j. Hence, we can compute the proportion of respondents in each bin as well as the reported population increase for those that responded; that is, $\hat{r} - r_j$. This allows us to compute an average response in terms of over-reported population:

$$dr = \sum_{j=r_L}^{\hat{r}} \tilde{w}_j \frac{(\widetilde{c_j} - c_j)}{\widetilde{c_j}} (\hat{r} - r_j) \tag{9}$$

where $\tilde{w}_j \equiv \frac{\widetilde{c_j}}{\sum_{j=r_L}^{\hat{r}} \widetilde{c_j}}$ is the share of counterfactual observations in each bin within the missing mass area.

Our bunching estimates imply that, on average, the notch increases over-reported population by 62 inhabitants. Note that this is more than ten times smaller than the effect found for the most responsive municipalities and is consistent with the fact that, between r_L and $\hat{r}$, only about 10% of municipalities respond to the incentives to over-report created by the notch.

As shown in Section 3, a measure of the density hole along with the change in grants induced by the notched schedule can be used to compute the elasticity of reported population to grants. Taking expression 7 and replacing $\tilde{r}$ for its empirical counterpart r_L, the elasticity of reported population to grants of the most responsive municipalities is:

$$\epsilon_{r_L,\alpha} = \frac{\frac{\hat{r}-r_L}{r_L}}{\frac{\hat{r}}{\hat{r}-r_L}\frac{\beta}{\alpha}} \tag{10}$$

Recall that the numerator is the percentage increase in reported population when moving from r_L to $\hat{r}$. In our application this percentage is 29%. As for the denominator, recall that

[19] Remember that the population figures that we examine are reported population by municipalities net of any adjustments made by INE, implying that the responses to grants that we compute are in terms of undetected population over-reporting. Hence, the set of municipalities found in the 3,850-5,000 interval includes two types of municipalities. Firstly, municipalities that did not report a population count above the 5,000 threshold to start with. Secondly, municipalities that having done so, their population count was downsized by INE and moved below the 5,000 threshold.

moving from 5,000 to 5,001 increases per capita grants by 85,000 euros (i.e., $\hat{r}\beta$). This increase becomes smaller for municipalities the further they are away from 5,001. For r_L, moving from 3,850 to 5,001 increases per capita grants by about 73%. As a result, the elasticity of over-reported population to grants is around 40% for the most responsive municipalities.

Our second measure reflects the average elasticity which amounts to:

$$\epsilon_{r,\alpha} = \sum_{j=r_L}^{\hat{r}} \tilde{w}_j \frac{(\widetilde{c}_j - c_j)}{\widetilde{c}_j} \frac{\frac{\hat{r}-r_j}{r_j}}{\frac{\hat{r}}{\hat{r}-r_j}\frac{\beta}{\alpha}} \tag{11}$$

The results indicate that the average elasticity is about 0.013. The two elasticities imply that doubling per capita grants would increase reported population by 40 and 1.3 per cent for the most responsive and the average municipality, respectively. Hence, one implication of our results is that there is a large degree of heterogeneity in the responsiveness of municipalities to grants, which compromises the fairness of grant allocation. A municipality with a real population of 3,850 inhabitants and which successfully over-reports 1,251 inhabitants (that is, a reported population more than 25% higher than its real population) obtains 50% more resources per 'real' resident than a municipality that has and reports 3,850 inhabitants. This extreme case illustrates just how severe the effects could be for a municipality facing extreme levels of opportunities and incentives to cheat. Thus, even if this value is not representative on average, it is a clear reminder that with imperfect auditing a municipality may massively over-report its number of residents. An average elasticity of 0.013 is the figure that can be considered most representative of the global effects of population over-reporting. For instance, let's assume for a moment that this elasticity is extrapolated to include the whole universe of Spanish municipalities. If we take the total amount of transfers allocated in 2001 (around 2,850 million euros), an elasticity of this dimension would mean that around 40 million euros were misallocated (i.e., 0.013×2,850); that is, that an extra 40 million euros would have had to be added to the pot of transfers in order to ensure a similar amount of transfers (per real resident) as that which would be available in a situation without manipulation. Although the impact on the level of transfers enjoyed by a typical (non over-reporting) municipality is not that huge, 40 million euros is by no means negligible.[20]

Columns 2 to 5 in Table 2 assess the robustness of the bunching estimates to the specification of bin size and polynomial order. In columns 2 and 3 we report the estimates obtained when

[20] Indeed, it would be interesting to compare this figure with the cost of improving the audit technology in order to put an end to this problem. In reality, this task is far from easy given the general lack of information within the INE regarding budgeting and the use of resources. However, in order to gauge the magnitude of the problem at hand, the 40 million euros can be compared to the cost of running a census (which can be considered a quasi-perfect audit, see next section), which was calculated at 162 million euros in 2001 (and given the ten-year interval between censuses, this amounts to 16 million per year, source: www.ine.es). Obviously, a census would be carried out even if there were no over-reporting of population, so the amount actually needed to keep this problem in check is probably much lower than the savings at stake.

the density is estimated grouping the observations in bin sizes of 40 and 60, respectively. This exercise shows that the results are not sensitive to the choice of the bin size. The baseline results are obtained by estimating the counterfactual density using a quadratic polynomial. Note that Figure 6 suggests that a second-order polynomial fits the empirical density well (which is corroborated by the Akaike information criterion). Nevertheless, to assess if the results are sensitive to this particular parametrization, we re-estimate the bunching using polynomials of order 3 and 4. The results, reported in columns 4 and 5 in Table 2 indicate that our results are robust to using higher-order polynomials, too.

Ultimately, we assess the robustness of our results to the choice of the upper limit r_U. Table 3 varies the limit in three different ways. In columns 1 and 2, we vary the limit which we have previously chosen by one bin down ($r_U = 5,751$) and one bin up ($r_U = 5,851$). In column three, we set the upper limit as half of the baseline specification ($r_U = 5,401$) to illustrate the implications of such a big change on our results. Note that the methodology described in Section 4 requires the researcher to choose r_U by visual inspection given the estimated counterfactual density. However, the estimation of this density also hinges on specifying the excluded range, i.e. r_U and r_L, which makes this problem a dynamic one. The first exercise (columns 1 and 2) is important as it shows that that the results are not greatly affected by first selecting r_U conditional on an estimated counterfactual. The second exercise has been carried out to illustrate the general sensitiveness of our results with respect to the choice of r_U.

[Table 3 about here]

In general, all results of Table 3 confirm our baseline findings. For the small permutation of $-/+$ one bin, changes are as expected small and the total number of municipalities found to manipulate does not change substantially. However, if we cut the upper limit in half to $r_U = 5,400$ as in column 3, the magnitude of manipulation (by construction) reduces, but significance remains almost unchanged as standard errors decrease as well. Other measures, such as the percentage of respondents and the average dr also remain relatively stable. However, the computed elasticities decrease such that the average municipality only over-reports 0.7% compared to 1.3% in the baseline estimation. At this stage we cannot justify the choice of $r_U = 5,800$ on other grounds than the criteria of crossing the counterfactual (see Figure 6), the next section provides indeed evidence that the higher limit is more likely to be correct than the lower one.

5.2 The census as an alternative counterfactual

If all irregularities in the distribution can be attributed to misreporting, data from the census provides information about the true distribution of municipalities. In such an instance, the census counts could be directly considered as the counterfactual. This eliminates the need to specify the bunching range, i.e. (r_L and r_U), to estimate the counterfactual density. The

ability to measure the counterfactual directly offers some substantial advantages over existing bunching literature, where counterfactuals cannot be directly observed.

Note, however, that this approach would be invalid if responses are real ones, i.e. municipalities attract real new residents. To this end, we first assess if the municipality size distribution is continuous around the 5,000 threshold in the 2001 and 2011 censuses. The graphical results are presented in Figure 7 while the results of the McCrary (2008) test are reported in Table 4. The analyses are performed for each census year separately as well as for the pooled data (2001 and 2011). No sign of discontinuity appears at the 5,000 threshold. These results indicate that the bunching patterns that we have documented are not explained by responses in real population, because in this case this should also be reflected in the census data. Eventually, the census data can be used to construct the counterfactual density, as it provides the true distribution of municipalities.

[Figure 7 and Table 4 about here]

However, using the census data to build the counterfactual density comes at the cost that any dynamics between the census waves would be neglected. Given that during our period of study only two censuses were performed, we keep the estimation of the counterfactual described in Section 4 as the baseline method and compare the results to those obtained by using the census. In order to capture some of the dynamics which might have occurred during this period, we pool the 2001 and 2011 census years, and give a weight of 7 to each observation (as we only have two years of data) in order to achieve a distribution which is comparable to the baseline specification which pools 14 years (1998-2011). Figure 8 shows graphically the results of this analysis.

[Figure 8 about here]

Note that the empirical distribution is exactly the same as before (see Figure 6) and the counterfactual has changed only slightly. The first important thing to note here is that the most obvious choice for the upper limit coincides with the one chosen under the previous approach, i.e. $r_U = 5,800$. Since the counterfactual here is estimated using all observations without excluding those in the area affected by bunching, the dynamic problem discussed above (i.e. choosing r_U conditional on an estimated counterfactual which in turn is affected by the choice of r_U) does not arise. We take this as evidence that an upper limit of 5,800 inhabitants seems to be the most appropriate upper limit here. Note next that the location of the counterfactual has changed only marginally while the shape seems to be flatter than it was before. Therefore, the overall count of cheating municipalities is reduced, but also precision has improved. The elasticities we obtain are very similar on average with $\epsilon_{r,\alpha} = 0.010$ but lower for the most responsive municipality ($\epsilon_{r_L,\alpha} = 0.177$) compared to the baseline case.

5.3 Bunching with improved enforcement over time

As outlined in Section 2, INE monitoring has improved over time. Municipalities were required to clean up their registers based on the results of the 2001 census. These operations were extended until the end of 2004. Subsequently, in 2005, to deal with problems related to the underreporting of out-migration and the double-counting of immigrants, the INE obliged non-EU immigrants to renew their registration every second year. In principle, all immigrants registered before December 2003 that failed to re-register were to be deleted from the register in December 2005. Thus, overall, we expect the opportunities to over-report population to be greater during the first years of our sample. To determine whether bunching is less evident with better monitoring, we split our sample into two different sub-periods (1998-2005 and 2006-2011). Figure 9 shows the bunching patterns for these two sub-periods in panels a) and c) together with the relevant McCrary (2008) test results in panels b) and d).

[Figure 9 about here]

While qualitatively Figure 9 reveals that there is some evidence of manipulation in both periods, bunching patterns are stronger during the sub-period in which INE monitoring was less effective. In the period 1998-2005, the excess density to the right of the 5,000-inhabitant threshold totaled 155 observations while it reached just 35 in the period 2006-2011 period.[21] Our results confirm that weaker institutions give local governments more room to manipulate their population figures.

5.4 Real population responses?

In the theoretical model outlined in Section 3, we focused on one type of response to the incentives created by the transfer schedule, i.e., reporting population numbers that are larger than the number of real residents in the municipality. Besides the ample anecdotal evidence suggesting that this has been the main margin of adjustment in the Spanish case, three sets of results identify manipulation as the main mechanism underlying bunching behavior. First, Figure 7 and the results of Table 4 indicate that there is no bunching in census data, suggesting that real population responses are not explaining the bunching patterns observed in the population register. Second, as shown in Figure 9, as monitoring has improved over time (and detection probabilities have increased), the amount of bunching has decreased.

However, municipalities might also respond by enacting more expansionary zoning policies to attract newcomers.[22] To test this hypothesis directly, we use data from the municipal cadastres to examine the growth of the stock of dwellings in our period of consideration. The re-

[21] Note that this exercise is not one of decomposition, as r_U is chosen by visual inspection in each sub-period. It turns out that r_U is smaller in each of the two sub-periods than it is for the entire period.

[22] There is ample evidence that the preferences of voters and politicians determine the amount of developable land and other zoning instruments (see, e.g. Solé-Ollé and Viladecans-Marsal, 2012, 2013).

sults, which are deferred to the Online Appendix I, indicate no evidence of a differential growth around the 5,000 population threshold. All in all, the different pieces of evidence suggests that manipulation is the mechanism underlying the observed bunching patterns.

5.5 Bunching at the 20,000 threshold

As discussed above, the techniques used in this paper have been developed to analyze the thick densities of individual tax returns. For this reason, the paper has focused primarily on the first population threshold fixed at 5,000 inhabitants. However, in this subsection we analyze the bunching that occurs at the 20,000-inhabitant threshold. The results are presented in Figure 10. Panel a) shows the bunching metrics developed in Section 4.1 while panel b) reports the corresponding McCrary (2008) test results. Table 5 summarizes the information in a similar way as to that used for the 5,000-inhabitant threshold.

[Table 5 and Figure 10 about here]

The McCrary (2008) test indicates that the density is not continuous at the 20,000 threshold and that it is higher just above that threshold. The bunching estimates imply that the counterfactual density is distorted between 19,801 and 20,201 inhabitants. The estimate for B, that is, the number of municipalities above 20,000 inhabitants that should be below this threshold, is 18.36. The standard error of B is 10.001 which yields a p-value of 0.06. Recall that B is also the missing number of municipalities between 19,801 and 20,000 inhabitants (M). B amounts to 0.33 of the counterfactual density in the 19,801-20,000 range, and gives us the fraction of municipalities that respond to the incentives to bunch. In terms of responses, the most elastic municipalities over-report around 200 inhabitants, while the average response (dr) is 72. The corresponding elasticities are 0.002 and 0.0003 for the most responsive and the average municipality, respectively. Although this exercise might have limited power, the results obtained imply that the responses are relatively smaller than those obtained for the 5,000-inhabitant threshold. Note that while the density hole disappears at a point that is 200 inhabitants below the 20,000 threshold, the empirical density was below the counterfactual density at 3,850 inhabitants in the case of the analysis of the 5,000-inhabitant threshold. The results also suggest that there is less heterogeneity in the responses at the 20,000 threshold, with a third of municipalities responding, compared to the 10 per cent recorded at the 5,000 threshold. Interestingly, the average number of over-reported inhabitants is very similar in both cases: 62 and 72 in the 5,000- and 20,000-inhabitant thresholds, respectively. This is consistent with most municipalities being able to over-report, if needed, a few score of inhabitants irrespective of their size (rather than a fixed proportion of their population), as these individuals can be attributed to mistakes or randomness. A further, complementary, explanation might be that the INE monitors large municipalities more intensely, given that small municipalities are more expensive to monitor in per capita terms.

5.6 Other policy discontinuities

Eggers et al. (2017) and Ade and Freier (2011) have shown that in many instances several policies are simultaneously changing at a certain population threshold. Although anecdotal evidence from the municipalities engaged in manipulation indicates that in practice increased transfers are the only advantage considered, there are four other institutional arrangements that change at the 5,000-inhabitant threshold: spending responsibilities, maximum tax rates, council size and gender quota in local elections. We address these concerns in detail in the Online Appendix II.

First, we provide evidence that public spending on the relevant spending categories which become mandatory after reaching the 5,000 population threshold do not change discontinuously at this population level. Instead, the analysis shows that per capita spending on these items grows continuously with population. Second, in Section 6 we examine differences between census and register population figures. This analysis shows that municipalities that over-report more are not more indebted and do not set higher local tax rates, suggesting that maximum tax rates are unlikely to account for the observed bunching. Third, the number of representatives sitting on the council increases from 11 to 13 at the 5,000-inhabitant threshold. We show that this (moderate) change in council size at the 5,000-inhabitant threshold does not have a significant impact on council fragmentation. Furthermore, we conduct a placebo analysis at the 10,000-inhabitant threshold, where there is a larger increase in council size (from 13 to 17 seats) while per capita grants remain constant. No bunching at the 10,000 threshold is found (c.f. Table 5 and the Online Appendix II), providing further evidence that council size is unlikely to explain the bunching pattern observed at the 5,000 threshold. Finally, gender quotas were established for the first time in the 2007 local elections and require the two genders to be represented in at least 40% of the positions on every municipal electoral list in municipalities exceeding 5,000 inhabitants.[23] Although the gender quota cannot explain the more pronounced bunching patterns observed in the 1998-2005 period, this reform could explain why bunching is reduced towards the end of the period we study. To discard this possibility, we study bunching around the 3,000 threshold in 2011 as gender quotas were extended to this lower population threshold in that year. Results do not show any evidence of manipulation around that threshold, confirming that the introduction of gender quota did not reduce population manipulation.

6 The 2001 Census as an 'audit' operation

Our results clearly indicate that too many municipalities lie above the threshold, but are so far silent about what specific type of municipalities tend to over-report their population figures.

[23] Casas-Arce and Saiz (2015) and Bagues and Campa (2016) study the effects of the gender quota on female representation and on policy outcomes.

The 2001 census was the last field operation (including visits to all households) to produce municipal population figures (with base date November 2001) and can be considered to be a faithful reflection of the 'real' population (Jurado, 2004). This implies that comparing the census and register population figures is informative about the characteristics of the municipalities that manipulated their population data.

Focusing on the sample of municipalities that report a population level between $\hat{r}$ and r_U, we estimate two different models to analyze the channels through which manipulation might occur. First, we compute the percentage deviation between the January 2002 register population and the 2001 census population, where a positive value indicates the percentage (as a fraction of 100) in which the register population exceeds the corresponding census figure. Second, we estimate a logit model where the dependent variable is a dummy equal to one for a municipality reporting a population larger than 5,000 inhabitants in January 2002 while the census reported a population below 5,000 in November 2001. We control for the distance between reported population and the 5,000-inhabitant threshold in both regressions. We also control for population growth between 2001 and 2002 to proxy for the potential 'real' growth occurring between November 2001 and January 2002. Table 6 shows the results for both type of models, which we perform over a variety of potential indicators for the underlying channel. Each entry comes from a separate regression. In columns (1a) and (2a), we present the point estimates of the OLS and the average marginal effects for the logit model, respectively. Columns (1b) and (2b) show the increase of the R^2 (Pseudo-R^2 in the logit analysis) experienced by the model compared to a specification which only includes the above-mentioned control variables. This helps to interpret the relative contribution of each variable in explaining the variation of the two outcomes.

[Table 6 about here]

Among the municipal characteristics analyzed, we first consider factors that reflect differences in over-reporting opportunities. In terms of the model presented in Section 3, these differences would correspond to heterogeneous detection probabilities (the γ's). Specifically, we include the % of immigrants in 2001, row (i), and the % of second homes in 2001, row (ii).

Next, we consider political variables to proxy for differences in the incentives to over-report, related to the (political) cost of being caught over-reporting population (the θ's in Section 3). In row (iii), we introduce the vote turnout in the 2000 national elections. We do so because in places with a higher turnout, the level of civic capital of the population is higher (Putnam, 1993) and, thus, residents would not agree to report misleading population figures (even if this were to benefit the town in terms of extra transfers). There is a vast literature supporting the idea that tax compliance is higher in places with higher 'tax morale' (Slemrod, 2007), and some recent papers documented that tax enforcement is also higher in places with more social capital (see e.g., Casaburi and Troiano, 2015, and Besley, Jensen, and Persson, 2015). In row (iv),

we consider the incumbent vote share in the last municipal elections (see Besley, Persson, and Sturm, 2010). Here the prediction is ambiguous. On the one hand, we expect that an incumbent who won the election with a very small share of the vote would be subject to more stringent controls. Those incumbents may face a stronger opposition (which will be more active in the council asking for information and has more incentives to denounce irregularities to the press or the judiciary), and also find more difficult to sustain a majority in the council in order to retain the mayoralty.[24] Moreover, if the level of civic capital is high and voters dislike cheaters, the incumbent might also feel less safe electorally. Note, however, that if the level of civic capital is low, voters might actually reward population over-reporting, such that the incentives to cheat might actually increase the more competitive the elections are. In row (v), we introduce a partisan alignment indicator variable that takes a value of one if the mayor belongs to *Partido Popular*, the party in office at the national level during the period 1996-2004. We do so to test if, as Litschig (2012) has documented for Brazil, the over-reporting of population is higher in municipalities that are aligned with the central government. Note that in contrast to Brazil, in Spain, local population figures are reported by municipalities. Nevertheless, the central government might put less effort into monitoring aligned municipalities or might intercede in case that irregularities are detected.

Finally, in rows (vi) to (viii), we include fiscal incentives to determine whether over-reporting is related to situations of fiscal distress. Specifically, we test if municipalities with high debt service, row (vi), or high rates of property tax, row (vii), or business tax, row (viii), cheat more.

Our results underline mostly the importance of manipulation opportunities and, to a lower extent, political cost-related incentive channels. The percentage of immigrants seems to be particularly important. In both models the effect is statistically highly significant and indicates that municipalities with a high % of immigrants are more likely to manipulate the population register. This is consistent with the fact that the probability of detection is lower if the residents being over-reported are foreigners. The estimates imply that a one percentage point increase in this variable increases the difference between register and census by 0.577%. Similarly, the logit estimates imply that a one percentage point increase in the % of immigrants increases the probability of a municipality having over-reported its population so that it surpasses the 5,000-threshold by 0.027. The estimates imply that increasing the % of immigrants by one standard deviation (4.46) increases over-reported population by 2.58 percent, which amounts to 1.54 times the variable's mean (1.68). As for the probability of having over-reported population to surpass the 5,000-threshold, the corresponding effect is 0.12, which amounts to 0.71 times the variable's mean (0.17). The relevance of the % of immigrants is confirmed by looking at the variable's contribution to the R^2 and the Pseudo-R^2. Specifically, these two statistics increase by 0.42 and 0.11, respectively. The impact of the share of second residences

[24] See Solé-Ollé (2006) and Solé-Ollé and Viladecans-Marsal (2012) for evidence that local incumbents in Spain are sensitive to these incentives.

as another variable reflecting increased manipulation opportunities is much smaller and it is only statistically significant in the logit analysis. There, the results indicate that increasing the share of second homes by one standard deviation (12.09), increases the probability of having over-reported population to surpass the 5,000-threshold by 0.036, which is only 0.21 times the variable's mean. Accordingly, the results reported in columns (1b) and (2b) show that this variable adds little to the predictive power of the two models.

In the case of variables reflecting differences in the political costs of manipulation, municipalities with more civic capital (reflected in a higher voter turnout in national elections) are less likely to manipulate population numbers. Specifically, a one percentage point increase in turnout decreases the difference between register and census populations by 0.069%, which translates into 0.55% percent if one considers, instead, a standard deviation increase in the variable (7.99). This effect represents 0.33 times the variable's mean. The estimates of the logit model are qualitatively and quantitatively similar. A one percentage point increase in vote turnout decreases the probability that a municipality surpassed the 5,000-threshold by over-reporting population by 0.006, which increases to 0.048 if a one standard deviation is considered (7.99). This effect is 0.28 of the variables average value. Note that the R^2 and the Pseudo-R^2 statistics also convey the message that voter turnout plays a smaller role than the % of immigrants in predicting which municipalities over-report more. As for the rest of variables capturing political incentives, municipalities in which the position of the incumbent is relatively secure (measured by the incumbent vote share in the last local elections) are not found to over-report more nor less. As for the political alignment variable, although the estimated coefficients are positive, we do not find statistically significant effects. Finally, we find no evidence that more indebted municipalities or those that set higher local tax rates cheat more, indicating that manipulation is not driven by the need for fiscal resources.

7 Conclusion

In this paper we analyzed the notches generated by a transfer system to local governments. We find significant evidence that the incentives generated by the grant schedule result in the manipulation of population figures. Municipalities, which are responsible for the registration of local residents, systematically inflate their population figures, thereby making them eligible for higher per capita transfers. Specifically, we found a missing mass of around 10% of the municipalities to the left of the 5,000-inhabitant threshold, where transfers per capita jump discretely. Municipalities in this range over-report by an average of 60 more residents than those actually living in the town. However, in some municipalities the number of 'ghost' residents reaches 1,000. The amount of misallocated transfers is far from negligible (we estimate a figure of around 40 million euros for 2001), which suggests more resources should be assigned to audit policies. Indeed, we find evidence that enhanced audit policies are effective in curbing

this problem. For instance, we found that bunching disappeared after the national statistical agency introduced several audit improvements (including, for example, using the census to cleanse the population register, and requiring immigrants to re-register every two years). Our analysis of the discrepancies between the census and the register also reveals that the municipalities more prone to over-reporting are those in which there are greater opportunities (a higher share of immigrants and second homes) and incentives (less civic capital) to do so. These results are consistent with the predictions of the model presented in this paper and also suggest that the reason municipalities bunch to the right of the transfer thresholds is population over-reporting. Moreover, the fact that we find no bunching in census figures is a further proof that municipalities do not respond to transfers by attracting more residents (or if they do so, they are largely ineffective).

The general conclusion to be drawn from our results is that even very simple formula transfers (such as those based on population) can be manipulated by recipients if their collaboration is needed in supplying information, and if the control policies operated by the central government are imperfect. While it is true that the size of the effects that we find are not especially large (the loss of resources is not dramatic, even without proper auditing systems) and that the problem was resolved when the national government intervened, they nevertheless serve as a warning to governments and consultants advising on similar transfer systems in developing countries - if the system can be manipulated in a developed country (with a modern, albeit inefficient, administration and a certain level of civic capital), what might happen in less favorable environments?

References

Ade, Florian and Ronny Freier. 2011. "When Can We Trust Population Thresholds in Regression Discontinuity Designs?" Discussion Papers of DIW Berlin 1136, DIW Berlin, German Institute for Economic Research.

Bagues, Manuel and Pamela Campa. 2016. "Can Gender Quotas Empower Women? Evidence from a Regression Discontinuity Design." Mimeo.

Bahl, Roy. 2000. "Intergovernmental Transfers in Developing and Transition Countries: Principles and Practice." Background Series 21097, The World Bank.

Baskaran, Thushyanthan. 2016. "Intergovernmental Transfers, Local Fiscal Policy, and the Flypaper Effect: Evidence from a German State." *FinanzArchiv: Public Finance Analysis* 72 (1):1–40.

Besley, Timothy and Torsten Persson. 2013. "Taxation and Development." In *Handbook of Public Economics, Handbook of Public Economics*, vol. 5, edited by Martin Feldstein Alan J. Auerbach, Raj Chetty and Emmanuel Saez. Elsevier, 51 – 110.

Besley, Timothy, Torsten Persson, and Daniel M. Sturm. 2010. "Political Competition, Policy and Growth: Theory and Evidence from the US." *The Review of Economic Studies* 77 (4):1329–1352.

Besley, Timothy J., Anders Jensen, and Torsten Persson. 2015. "Norms, Enforcement, and Tax Evasion." C.E.P.R. Discussion Papers 10372, Centre for Economic Policy Research.

Bird, Richard M. and Michael Smart. 2002. "Intergovernmental Fiscal Transfers: International Lessons for Developing Countries." *World Development* 30 (6):899–912.

Bird, Richard M. and Eric M. Zolt. 2008. "Technology and Taxation in Developing Countries: From Hand to Mouse." *National Tax Journal* 61 (4):791–821.

Boex, Jameson and Jorge Martínez-Vázquez. 2007. "Designing intergovernmental equalization transfers with imperfect data: concepts, practices, and lessons." In *Fiscal equalization: challenges in the design of intergovernmental transfers*, edited by J. Martínez-Vázquez and Bob Searle. Springer, 291–343.

Bordignon, Massimo, Paolo Manasse, and Guido Tabellini. 2001. "Optimal Regional Redistribution under Asymmetric Information." *American Economic Review* 91 (3):709–723.

Brambor, Thomas, Agustín Goenaga, Johannes Lindvall, and Jan Teorell. 2016. "The lay of the land. Information capacity and the modern state." STANCE Working Paper Series 2.

Casaburi, Lorenzo and Ugo Troiano. 2015. "Ghost-House Busters: The Electoral Response to a Large Anti Tax Evasion Program." *Quarterly Journal of Economi* forthcoming.

Casas-Arce, Pablo and Albert Saiz. 2015. "Women and Power: Unpopular, Unwilling, or Held Back?" *Journal of Political Economy* 123 (3):641–669.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics* 126 (2):749–804.

De Witte, Kristof and Benny Geys. 2015. "Strategic Housing Policy, Migration and Sorting around Population Thresholds." CESifo Working Paper Series 5639, CESifo Group Munich.

Dixit, Avinash and John Londregan. 1995. "Redistributive Politics and Economic Efficiency." *The American Political Science Review* 89 (4):pp. 856–866.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94 (5):1429–1451.

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini. 2017. "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions." *American Journal of Political Science* (forthcoming).

Fisman, Raymond and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China." *Journal of Political Economy* 112 (2):471–500.

Foremny, Dirk. 2014. "Sub-national deficits in European countries: The impact of fiscal rules and tax autonomy." *European Journal of Political Economy* 34 (0):86 – 110.

Gabaix, Xavier. 1999. "Zipf's Law For Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3):739–767.

Huber, Bernd and Marco Runkel. 2006. "Optimal Design of Intergovernmental Grants Under Asymmetric Information." *International Tax and Public Finance* 13 (1):25–41.

Jurado, Ana. 2004. "Diferencias entre Censo de Población y Padrón Municipal." *Índice - Revista de Estadística y Sociedad* 3 (3):12–13.

Kim, Junghun and Jorgen Lotz. 2008. *Measuring Local Government Expenditure Needs - The Copenhagen Workshop 2007*. The Korea Institute of Public Finance and the Danish Ministry of Social Welfare.

Kleven, Henrik J. 2016. "Bunching." *Annual Review of Economics* 8:435–464.

Kleven, Henrik J., Martin B. Knudsen, Claus Thustrup Kreiner, Sï¿½ren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79 (3):651–692.

Kleven, Henrik J. and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *The Quarterly Journal of Economics* 128 (2):669–723.

Litschig, Stephan. 2012. "Are rules-based government programs shielded from special-interest politics? Evidence from revenue-sharing transfers in Brazil." *Journal of Public Economics* 96 (1144):1047–1060.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142 (2):698–714.

Oates, Wallace E. 1972. *Fiscal federalism*. New York: Harcourt Brace Jovanovich.

———. 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37 (3):1120–1149.

Persson, Torsten and Guido Tabellini. 2002. *Political Economics: Explaining Economic Policy*. MIT Press Books. The MIT Press.

Putnam, Robert. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. New Jersey: Princeton University Press.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3):180–212.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives* 21 (1):25–48.

———. 2010. "Buenas Notches: Lines and Notches in Tax System Design." *University of Michigan Workin Paper* .

Solé-Ollé, Albert. 2006. "The effect of party competition on budget outcomes: empirical evidence from local governments in Spain." *Public Choice* 126 (1-2):145–176.

Solé-Ollé, Albert and Núria Bosch. 2005. "On the Relationship between Authority Size and the Costs of Providing Local Services: Lessons for the Design of Intergovernmental Transfers in Spain." *Public Finance Review* 33 (3):343–384.

Solé-Ollé, Albert and Elisabet Viladecans-Marsal. 2012. "Lobbying, political competition, and local land supply: Recent evidence from Spain." *Journal of Public Economics* 96 (1-2):10 –19.

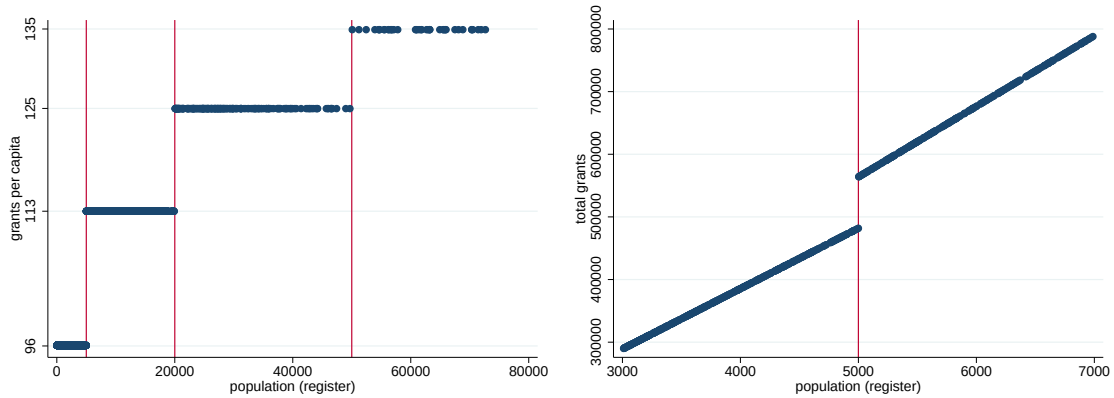
———. 2013. "Do political parties matter for local land use policies?" *Journal of Urban Economics* 78 (C):42–56.

Wildasin, David E. 1986. *Urban Public Finance*. Routledge.

Yitzhaki, Shlomo. 1987. "On the Excess Burden of Tax Evasion." *Public Finance Review* 15 (2):123–137.

Graphs and tables

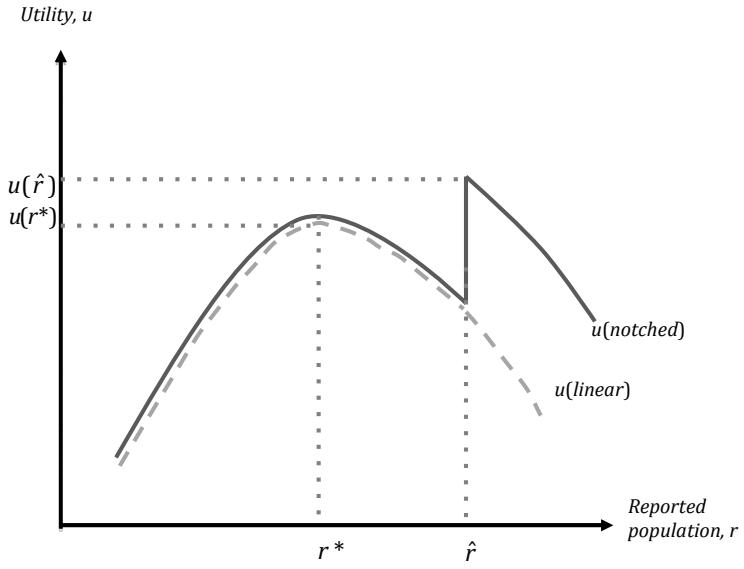
Figure 1: Distributed grants in 2005



(a) grants per capita (b) grants in absolute values

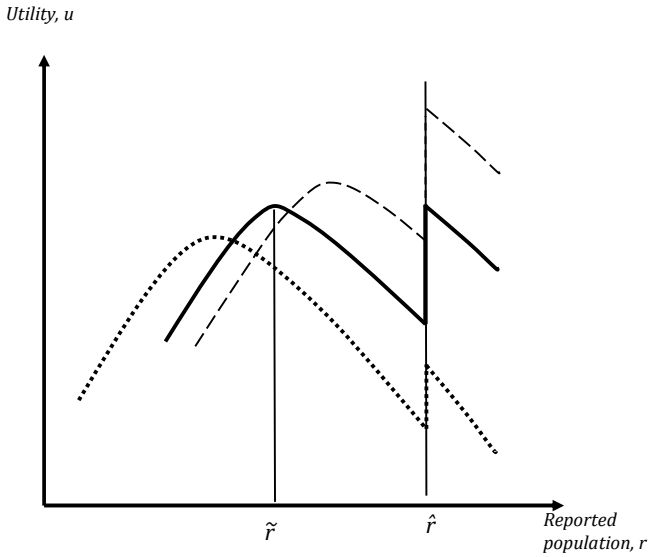
Notes: 75% of PIE allocated according to municipalities' population. Official data for 2005.

Figure 2: Utility with a notched transfer scheme

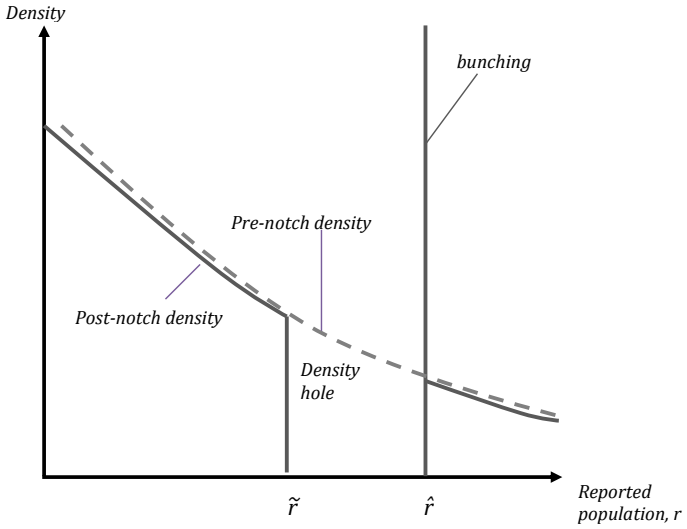


Notes: *u(linear)* is utility under a linear transfer scheme while *u(notched)* is utility under a notched transfer scheme with notch at $\hat{r}$. $u(r^*)$ and $u(\hat{r})$ are the utility levels when reporting the population levels that correspond to the interior equilibrium r^* and the notch $\hat{r}$. At the notch, utility increases by $(1-\hat{p})\beta\hat{r}$ where $1-\hat{p}$ is the probability of not being caught over-reporting $\hat{r}$ and β is the additional transfers received if reporting $\hat{r}$.

Figure 3: Effect of a notched transfer scheme on reported population density



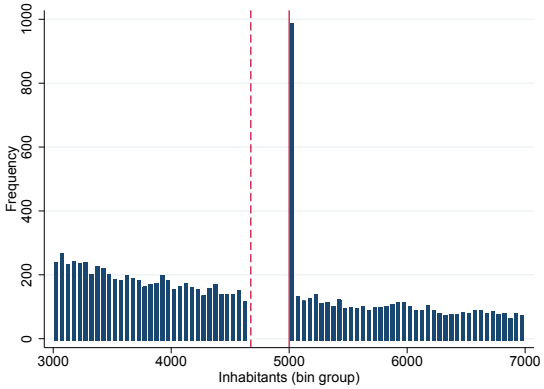
(a) Utility



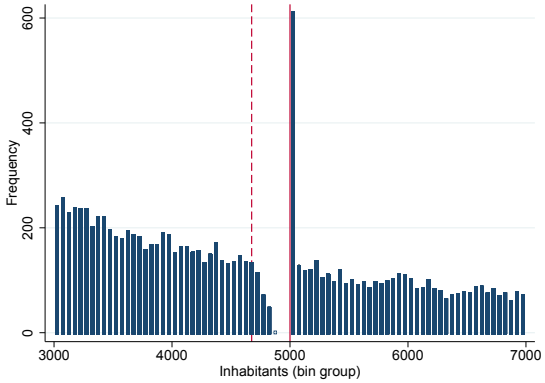
(b) Density function

Notes panel (a): utilities for different population levels. For a population level $\tilde{n} \equiv \tilde{r} - e^*$, the utility level of reporting $\tilde{r}$ equals the utility level of reporting $\hat{r}$. For a population level lower than $\tilde{n}$ (dotted line), the interior equilibrium yields higher utility than reporting $\hat{r}$. In contrast, for a population level higher than $\tilde{n}$ (dashed line), reporting $\hat{r}$ yields higher utility than reporting the interior equilibrium. Notes panel (b): pre-notch density is the density that would be observed in the absence of the notched transfer scheme. Post-notch density is the density that emerges once municipalities have responded to the incentives to bunch at $\hat{r}$.

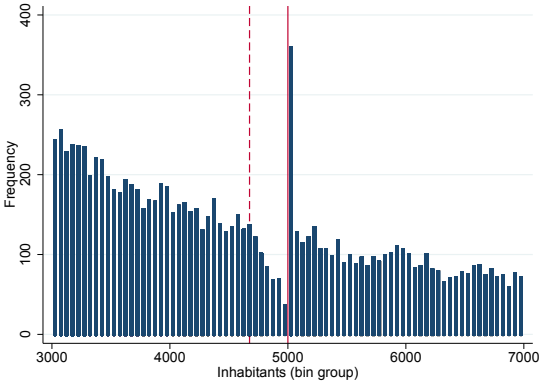
Figure 4: Simulated bunching in reported and observed population in the presence of heterogeneity



(a) Baseline case, reported population, r



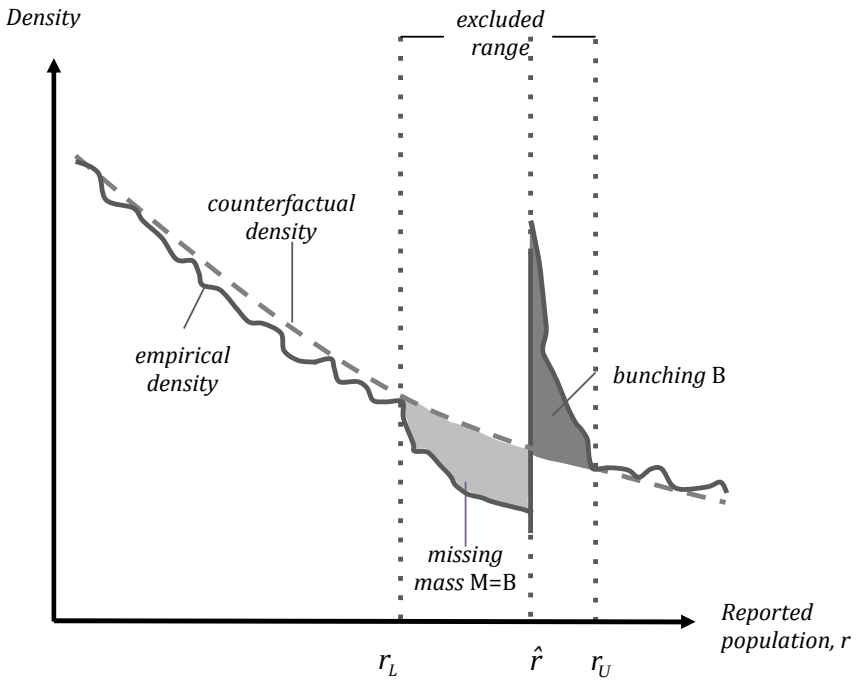
(b) Heterogeneity, reported population, r



(c) Heterogeneity, observed population, $n+(1-p)e$

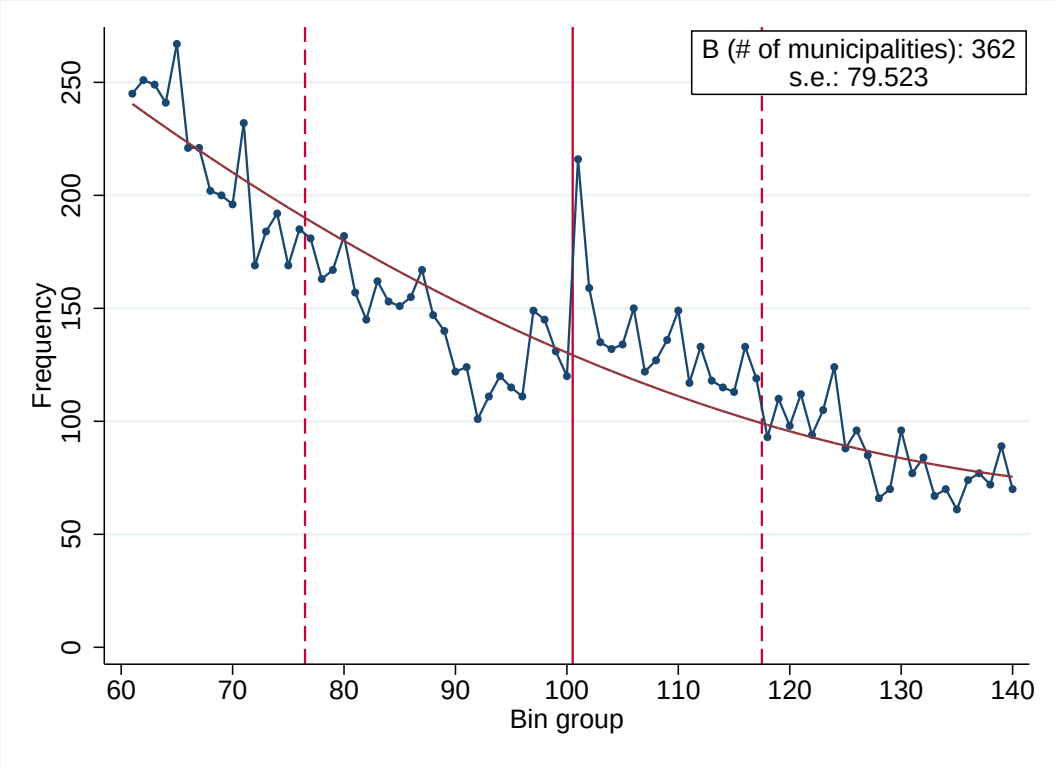
Notes: Densities obtained from simulation based on 100k draws. n drawn from a log-normal distribution with mean and standard deviation equal to 6.458 and 1.840. α, β, θ equal to 100, 16, 400. In panel (a) - no heterogeneity -, γ is -0.1 while $\gamma_i \sim U(-0.2,-0.1)$ in Panels (b) and (c). Panel (c) plots observed population, that is, reported population less the adjustment made by INE in case that the municipality is caught over-reporting, i.e. $n+(1-p)e$. For each simulated municipality, p is a random draw from a Bernoulli distribution with $p = 1-e^{\gamma}$.

Figure 5: Theoretical illustration of the procedure used to identify bunching

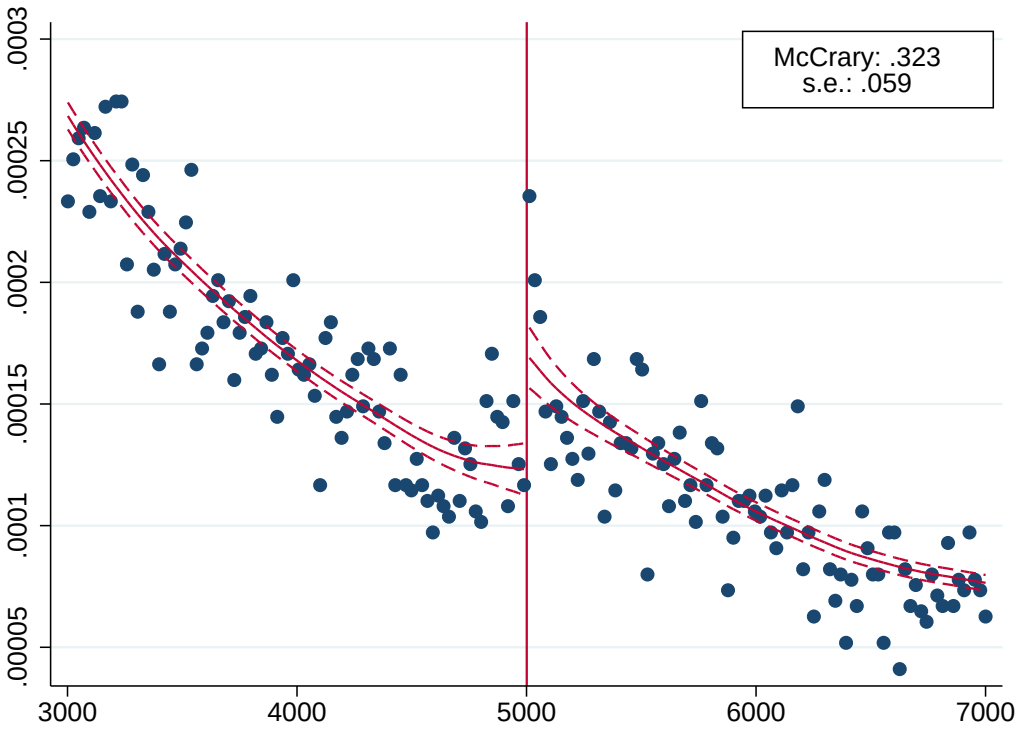


Notes: The lower and upper population bounds (r_L and r_U) define the excluded range, that is, the population interval suspected to be affected by bunching behavior. The counterfactual density is fitted excluding the data points in the excluded range. r_U is determined by visual inspection while r_L is set to ensure that the excess mass (B) in the $\hat{r}$-r_U range equals the missing mass (M) in the r_L -$\hat{r}$ interval.

Figure 6: Baseline graph: 5,000-threshold



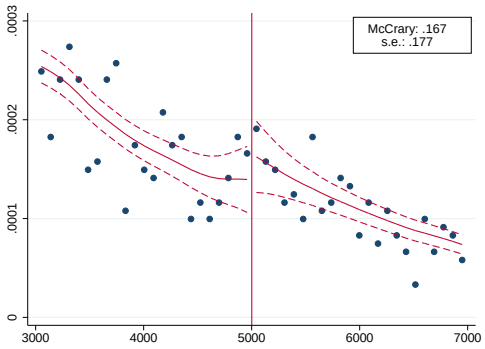
(a) bunching



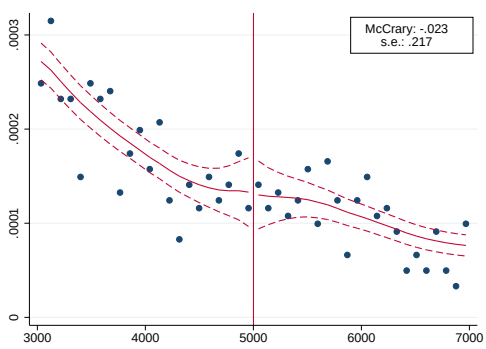
(b) McCrary

Notes: Notching around 5,000 inhabitants. Bins of 50 inhabitant size in panel a). Parameters chosen automatically in panel b). Pooled for 1998-2011. On the vertical axis, bunching is expressed in frequencies and the McCrary (2008) test in densities. The range of the horizontal axis is identical in both graphs, but expressed either in the bins included (bunching) or in population (McCrary, 2008). Dashed lines show 95% confidence intervals.

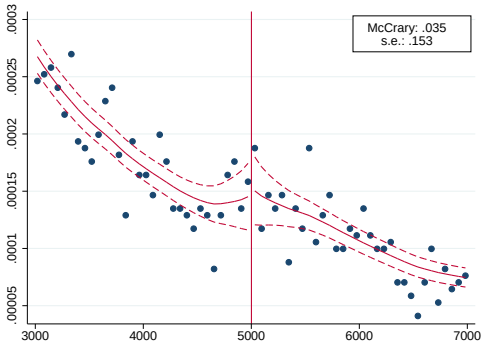
Figure 7: Bunching in the census population figures



(a) 2001



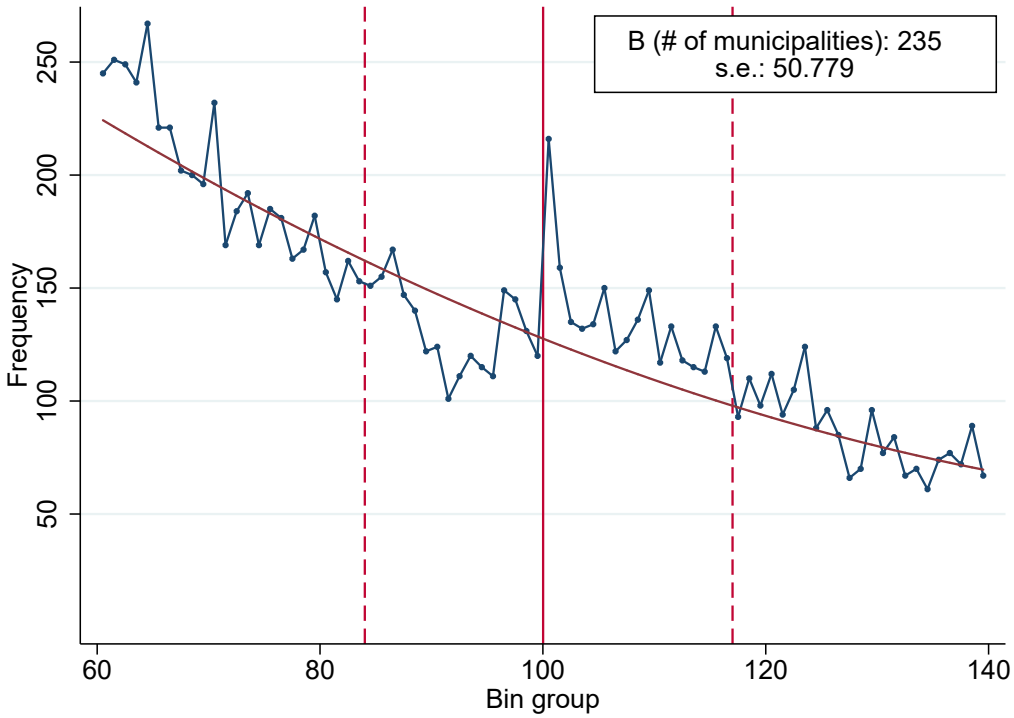
(b) 2011



(c) 2001 and 2011 pooled

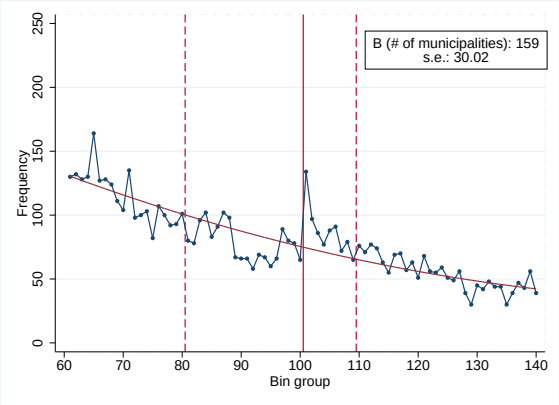
Notes: McCrary (2008) test on 2001, 2011, and pooled census information. Bandwidth and binwidth automatically determined by the routine.

Figure 8: 5,000-threshold with census counterfactual

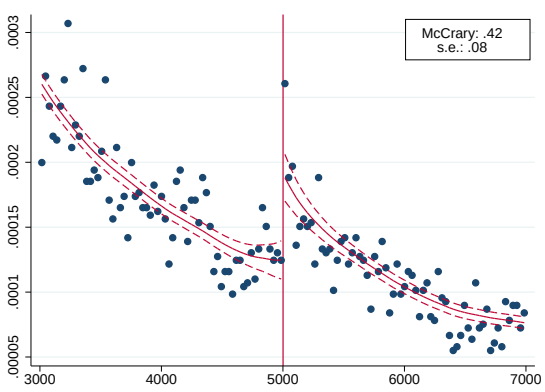


Notes: Notching around 5,000 inhabitants. Bins of 50 inhabitant size in panel. Pooled for 1998-2011. On the vertical axis, bunching is expressed in frequencies.

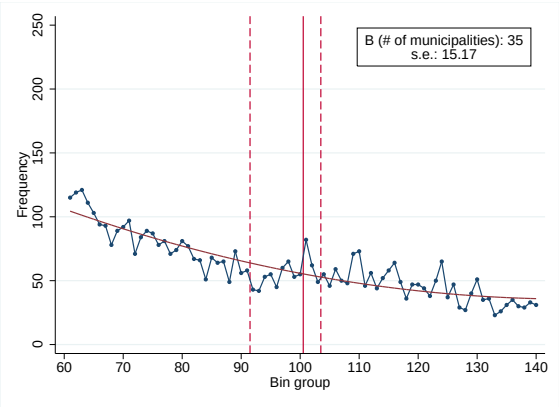
Figure 9: Bunching with varying levels of enforcement: 5,000-threshold



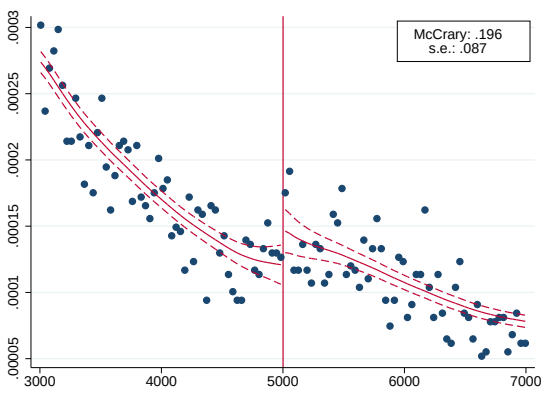
(a) 1998-2005 -bunching



(b) 1998-2005 - McCrary



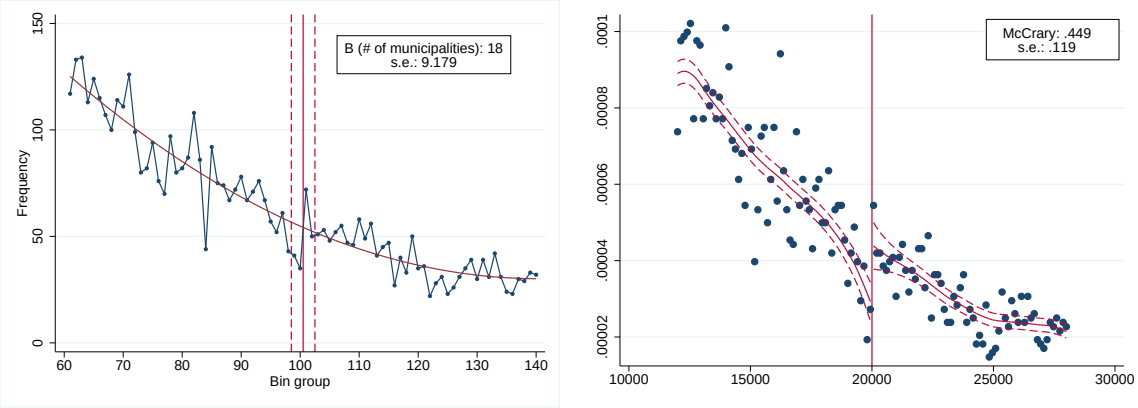
(c) 2006-2011 - bunching



(d) 2006-2011 - McCrary

Notes: notching around 5,000 (panel a and b) inhabitants. The size of each bin is 1% of the threshold (i.e. 50 inhabitants). Pooled for 1998-2005 (panel a) and 2006-2011 (panel b). On the vertical axis, bunching is expressed in frequencies and the McCrary (2008) test in densities. The range of the horizontal axis is identical in both graphs, but expressed either in the bins included (bunching) or in population (McCrary, 2008). Dashed lines show 95% confidence intervals.

Figure 10: Bunching at the 20,000 threshold



(a) 1998-2011 - 20,000 inhabitants (b) 1998-2011 - 20,000 inhabitants

Notes: Notching around 20,000 inhabitants. The size of each bin is 1% of the threshold (i.e. 100 or 200 inhabitants, respectively). On the vertical axis, bunching is expressed in frequencies and the McCrary (2008) test in densities. The range of the horizontal axis is identical in both graphs, but expressed either in the bins included (bunching) or in population (McCrary, 2008). Dashed lines show 95% confidence intervals.

Table 1: # of observations around thresholds (+/- 40%)

threshold	# yearly observations
transfer thresholds	
5,000	785-811
10,000	503-566
20,000	326-375
50,000	105-145
other thresholds	
3,000	1,097-1,119
10,000	503-566

Table 2: Baseline results - Bunching at the 5,000 threshold

Polynomial order (q)	2	2	2	3	4
Bin size	50	40	60	50	50
Range $[r_L, r_U]$	[3850, 5800]	[3840, 5800]	[3860,5840]	[3550, 5800]	[4400, 5800]
# municipalities (B)	362***	338***	366***	287***	191**
s.e. (B)	79.523	72.908	76.236	77.581	91.275
% of respondents	0.100	0.092	0.102	0.060	0.110
Average dr	63.319	71.270	66.270	71.934	53.957
$\epsilon_{r_L,\alpha}$	0.403	0.408	0.398	0.688	0.100
$\epsilon_{r,\alpha}$	0.013	0.015	0.014	0.021	0.006
McCrary			0.323***		
s.e. (McCrary)			0.059		

Notes: Pooled observations 1998-2011. The McCrary (2008) test is based on the sample until the next population threshold is reached and calculated with automatic bandwidth and bin size determination. The standard error for B is obtained by (residual) bootstrapping. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Variation of r_U - Bunching at the 5,000 threshold

Polynomial order (q)	2	2	2
Bin size	50	50	50
Range $[r_L, r_U]$	[4001, 5751]	[3701,5851]	[4451, 5401]
# municipalities (B)	320***	362***	188***
s.e. (B)	80.647	91.415	47.804
% of respondents	0.105	0.086	0.120
Average dr	60.766	74.344	56.272
$\epsilon_{r_L,\alpha}$	0.327	0.590	0.092
$\epsilon_{r,\alpha}$	0.012	0.020	0.007

Notes: Pooled observations 1998-2011. The standard error for B is obtained by (residual) bootstrapping. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: McCrary (2008)-test of census population

	2001	2011	2001/2011
McCrary	0.167	-0.023	0.035
s.e. (McCrary)	0.177	0.217	0.153

Notes: McCrary (2008) test on 2001, 2011, and pooled census information. 40% window around the threshold.

Table 5: Bunching at the 20,000 and 10,000 threshold

Threshold	a) 20,000	b) 10,000
Range $[r_L, r_U]$	[19800, 20200]	[9200, 10200]
# municipalities (B)	18**	21
s.e. (B)	9.179	20.080
% of respondents	0.335	0.031
Average dr	72.357	18.056
$\epsilon_{r_L,\alpha}$	0.002	.
$\epsilon_{r,\alpha}$	0.0003	.
McCrary	0.449***	0.104
s.e. (McCrary)	0.119	0.082

Notes: Pooled observations 1998-2011. The McCrary (2008) test is based on the sample until the next population threshold will be reached and calculated with automatic bandwidth and bin size determination. No elasticities reported at 10,000 as grants remain unchanged. The standard error for B is obtained by (residual) bootstrapping. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: The determinants of over-reporting

	(1a)	(1b)	(2a)	(2b)
	percentage deviation		Probability(cheating)	
	Coeff.	ΔR^2	MFX	$\Delta Pseudo-R^2$
cheating opportunities				
(i) % of immigrants	0.577***	0.422	0.027***	0.111
	(0.053)		(0.007)	
(ii) % of second homes	0.034	0.013	0.003**	0.027
	(0.025)		(0.001)	
political incentives				
(iii) vote turnout	-0.069*	0.023	-0.006**	0.032
	(0.038)		(0.003)	
(iv) incumbent vote share	0.001	0.002	0.002	0.007
	(0.024)		(0.002)	
(v) partisan alignment / right wing	0.479	0.005	0.083	0.018
	(0.693)		(0.052)	
fiscal incentives				
(vi) debt service	0.260	0.002	-0.134	0.001
	(4.211)		(0.320)	
(vii) business tax rate	0.019	0.003	0.035	0.001
	(1.448)		(0.116)	
(viii) property tax rate	0.001	0.003	0.098	0.003
	(1.962)		(0.144)	

Notes: Total observation count is 160. Column 1a presents OLS estimates where the dependent variable is $100 \times [\text{register}_{2002}/\text{census}_{2001}] - 1$; Mean (SD) of 1.677 (3.831). Column 2a presents average marginal effects (MFX) of logit estimates. The binary dependent variable is =1 if the January 2002 register figure is above the 5,000 threshold while the November 2001 figure is below; There are 27 cheaters. All regressions include as controls a constant, the distance to the 5,000 threshold and the population growth between 2000 and 2001. % of immigrants is the share of foreign inhabitants in total population in 2001; Mean (SD) of 2.521 (4.465). % of second homes is the share of homes characterized as second residences in 2001; Mean (SD) of 14.378 (12.093). Vote turnout is the turnout at the 2000 national elections at the municipal level; Mean (SD) of 72.658 (7.993). Incumbent vote share is the vote share of the mayor's party in the local council at the 1999 elections; Mean (SD) of 48.419 (12.960). Partisan alignment/ right wing indicates whether the mayor is right wing; Mean of 0.262. Debt service is the sum of interest and principal debt payments as a share of total current revenue in 2000; Mean (SD) of 0.077 (0.073). Property tax rate is the 2000 statutory rate in the "Impuesto sobre la Propiedad Inmueble" - Mean (SD) of 1.143 (0.356) - while the business tax rate is the "Coeficiente de Incremento" is the 2000 tax rate in the "Impuesto sobre Actividades Económicas"; Mean (SD) of 0.535 (0.199). Columns 1b and 2b report the increase in ΔR^2 ($Pseudo-R^2$ in the logit analysis) experienced by the model compared to specification which only includes the above-mentioned controls. Standard errors in parentheses. *** $p < 0.01$,** $p < 0.05$, * $p < 0.1$.