Empirical Essays on the Political Economy of Public Finance and Social Policy

Johannes Hemker

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

© 2016
Johannes Hemker

All rights reserved

ABSTRACT

Empirical Essays on the Political Economy of Public Finance and Social Policy

Johannes Hemker

This dissertation comprises three empirical essays in political economy. The first essay analyzes the implementation of a French social program by subnational governments following a decentralization reform. Using program implementation data, it shows that local political environments strongly influence implementation decisions after decentralization, and that decentralization results in an overall tightening of benefits. The second essay reports the results of a conjoint field experiment involving German welfare offices. Using random assignment of cues about ethnicity and other characteristics in requests to welfare offices, it is shown that putative non-German applicants receive replies at the same rate as putative Germans, but are disadvantaged in terms of the substantive quality of responses. This suggests that minority populations do experience discrimination when attempting to access social benefits. Finally, the third essay uses micro-level voter file data from Illinois to measure whether property tax limitations reduce participation in local elections. In contrast with prior research, results from panel regressions with matching adjustments suggest that tax limitations do not affect political participation negatively. Together, these essays contribute to our understanding of public finance and social policy in contexts characterized by multi-level governance.

TABLE OF CONTENTS

List of Figures

II. Social Policy Implementation in France

III. Bureaucratic Discrimination in Germany (with A. Rink)

IV. Tax Caps and Political Participation in the U.S.

List of Tables

Acknowledgements

I am indebted and grateful to a great number of individuals and institutions that have made this dissertation possible.

First and foremost, Justin Phillips, Yotam Margalit and Olle Folke have advised me on my research ideas and their implementation and were always helpful and encouraging. I am very thankful for their support and their time. I would also like to thank Jeff Lax and Melissa Schwartzberg very kindly for their guidance during the dissertation seminar and for serving on my defense committee.

My debts to the academic community at Columbia and elsewhere are manifold. For teaching me political science and public finance, I would like to especially thank Timothy Frye, Macartan Humphreys, Ira Katznelson, Wojciech Kopczuk, Robert Lieberman and Isabela Mares. For their excellent instruction in statistics and quantitative methods, I thank Olle Folke, Jens Hainmüller, Jennifer Hill and Ethan Kaplan. I could never have completed my dissertation research without my marvelous French teachers: Maria Comsa, Madeleine Dobie, Morgan Labar, Noémie Ndiaye and Samuel Skippon. Finally, I also owe thanks to Matthew Braham, Hartmut Egger, Rainer Hegselmann, Walter Olbricht, Rudolf Schüßler and many others for my undergraduate education at the University of Bayreuth, which laid the foundation for my desire to pursue a Doctorate.

I am also grateful to Columbia as an institution. Not only has it been a wonderful academic home over the past half decade; it has also allowed me to have many enriching experiences far away from New York City. Being a visiting scholar at Sciences Po, Paris and the *Laboratoire interdisciplinaire d'évaluation des politiques publiques* (LIEPP) in 2014 and 2015 allowed me to conduct my research on France in a very supportive and stimulating environment. I am particularly grateful to Etienne Wasmer for receiving me at LIEPP, to Alexandre Biotteau for support with data requests, and to Elodie Luquet for facilitating participation in the exchange program. For hosting me as an exchange scholar at Stanford University, where I was greeted by both a very productive atmosphere and fascinating new friends and colleagues, I am very grateful to Bruce Cain, Stanford's Political Science Department, and the Institute on Poverty and Inequality. I would also like to thank Peter Haan and Ronny Freier for receiving me in the Public Economics department of the *Deutsches Institut für Wirtschaftsforschung* (DIW) in Berlin, which I found to be a terrific place to work.

I thank my friend and co-author Anselm Rink who it was a real pleasure to collaborate with. Special thanks are also due to Martin Stimming and Holger Hecler for providing me with a fantastic soundtrack throughout graduate school. Support for my dissertation research from Columbia University, the Global Public Policy Network, the Alliance Program and the French *Agence Nationale de la Recherche* (ANR) is gratefully acknowledged.

An especially big thank you goes out to my friends, old and new, in New York City and beyond. For inspirational conversation, laughter, music and dancing, I will be forever grateful.

Finally, I would like to warmly thank my wonderful parents. It is to them that I dedicate this dissertation.

To my parents

I. INTRODUCTION

In this dissertation, I study various areas of political economy in Western democracies. The first essay studies a decentralization reform to understand the politics of local social policy implementation in France. In the second essay, a field experiment is used to assess discrimination against foreigners by the German bureaucracy. The effects of property tax limitations on turnout in U.S. municipalities are the subject of the third essay.

In each of these contexts, interactions across different levels of government, and the distribution of responsibilities among them, are a key concern. This is perhaps most apparent in the essay on France, which focuses explicitly on the transfer of implementation and financing responsibility for a social program to subnational governments. But multi-level governance also looms large in the other contexts under study. A politically crucial aspect of the tax limitations studied in the essay on turnout is that they are imposed on local governments by state legislators or ballot measures. In the German case, the territorial organization of social policy bureaucracies matters for their ability to treat all individuals impartially.

In political economy, the study of interactions across different levels of government is most closely associated with the prominent literature on federalism. While this literature has inspired much of the thinking behind this dissertation, one goal of the essays is to contribute to the development of a more fine-grained perspective on how the allocation of responsibilities across levels of government matters for policy. In this perspective, interactions between levels of government can powerfully affect policy outcomes not just in federal systems, but even in textbook examples of unitary systems such as France. And even within federal systems, the territorial organization of public finance and social policy at the sub-state level may be just as important for policy outcomes as the interactions between states and federal governments that the literature on federalism tends to focus on.

While the primary interest of the essays is scientific, it is hoped that the lessons learned can also help inform debates about the policies under study and their reform.

The essay on the French *revenu minimum d'insertion* (RMI) suggests that the program's decentralization to the departmental level has effectively weakened social protection and increased regional heterogeneity in services provision. In the context of broad debate about the future of the departments and France's territorial administration more generally – and notwithstanding the fusion of the RMI with other programs into the *revenu de solidarité active* (RSA) in the meantime – the findings suggest that re-centralization could have substantively important effects on social program implementation.

The essay on Germany provides experimental evidence that putative foreigners are treated unequally when attempting to apply for social benefits that they are eligible for as part of the *Hartz IV* program. Against the backdrop of debate about the "intercultural competence" of the *Jobcenters* that administer the benefit, this suggests that additional efforts are indeed required to attain impartiality in social policy administration. As the observational evidence suggests, the territorial organization of *Jobcenters* could play a role in explaining this pattern. However, additional research would be required to more thoroughly assess this possibility.

Finally, the study of local turnout in the U.S., in contrast to prior research, suggests that property tax limitations do not appear to decrease turnout. Hence, the sizable effect of these limitations on tax revenue may come without deleterious consequences for the local democratic process. Whether and in how far this finding is contingent on the intertemporal distribution of tax cap's fiscal effects is a topic that would merit further study.

The three essays on France, Germany and the United States, are reproduced as chapters below in this order. Each chapter is self-containing, with its own abstract, bibliography and appendix, to facilitate reading.

II. SOCIAL POLICY IMPLEMENTATION IN FRANCE[1]

Abstract

This paper studies the political economy of social policy implementation. Leveraging the decentralization of a major French social program, I show that the local political environment strongly shapes subnational implementation choices when implementation and financing are delegated to subnational governments. This is despite the fact that laws and rules governing the program remain nominally national. I find that local executive partisanship and budgetary constraints are correlated with the generosity of benefit administration and investments in active labor market policies. Right-wing executives tend to tighten cash benefit administration but invest more in active labor market policies relative to left-wing executives. This is consistent with partisan conflict over the relative importance of activation versus de-commodification of recipients. Variation in budgetary constraints also gives rise to divergences in implementation, with richer jurisdictions keeping benefit administration relatively generous and investing more in active labor market programs. In the aggregate, decentralization results in tighter cash benefit administration and increased variance in active labor market

1 I would like to thank Justin Phillips, Yotam Margalit and Olle Folke for their continued help and support, and Francesc Amat, Tim Dorlach, Nicolas Duvoux, Andrew Gelman, Anselm Rink and panel participants at the 2015 APSA Annual Meeting for feedback. I am also very grateful to Etienne Wasmer for his feedback and for receiving me at SciencesPo's Laboratory for Interdisciplinary Public Policy Evaluation (LIEPP). In France, Cyprien Avenel, Blandine Destremau, Bernard Dolez, Patrick Le Lidec, Yannick L'Horty, Jean-Luc Outin and Jean-Louis Pepin were immensely generous in sharing their expertise on the subject. I am very indebted to Alexandre Biotteau at LIEPP, Odile Gaultier-Voituriez at CEVIPOF, Michèle Lelievre, Mathieu Calvo and Julie Labarthe at DREES (French Social Ministry), Stephane Donné at the CNAF, Isa Aldeghi at CREDOC, Laurent de Boissieu of france-politique.fr and Patrick Milhe Poutingon for helping me manage multitudinous data request processes. This project has benefited from funding by the Global Public Policy Network (GPPN), the Alliance program including an Alliance Doctoral Mobility Grant, and a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-11-LABX-0091, ANR-11-IDEX-0005-02).

policy effort. I discuss implications for the study of social policy implementation, the politics of active labor market policies, and differences in benefit take-up.

1. Introduction

In social policy as in other areas of state activity, what bureaucracies do on the ground is more than just a mirror image of written policy. Between abstract laws adopted by legislatures and the frontline bureaucrats facing benefit recipients, highly elaborate institutions are required to implement welfare programs (Lipsky (1980)). Exactly how this process is organized clearly has the potential to affect substantive policy content and outcomes.[2] However, implementation is inherently more difficult to observe than policy in laws and programs; hence, it is only rarely the object of systematic empirical research.

There is an increasing appreciation that this gap in scholarly attention is highly problematic. For example, Jacob Hacker has argued that the literature on welfare state retrenchment does not find evidence of retrenchment simply because it looks in the wrong place. This is because it seeks to unearth changes in laws and programs, and ignores "subterranean means of policy adjustment" like increased rejections of benefits by local officials, or "decentralized cutbacks" (Hacker (2004), p. 245). As a consequence, Hacker argues, in modern welfare states "it may become increasingly difficult to judge policy effects simply by reading statute books or examining disputes over policy rules. We will need to look at what really happens on the ground." (ibid., p. 247)

In an effort in this direction, the present paper studies the "subterranean" politics of implementation by examining how the local political environment shapes implementation choices in a major French social program; the *revenu minimum d'insertion* (RMI) or minimum

2 To take an extreme example, few would doubt that social protection would be weakened if welfare offices reduced their opening hours to half an hour per week and ceased to be reachable by phone.

integration income. The program combines cash benefit and labor market integration services, and its financing and implementation were decentralized to subnational governments headed by elected officials in a 2004 reform. Although policy parameters governing the benefit remained national, decentralization did enable officials to affect policy content using “subterranean” administrative measures that affect how easily benefits are accessible for recipients.

I use detailed data on the implementation of the program before and after the reform across all subnational governments to examine the political determinants of implementation decisions. In particular, I hypothesize that under decentralization, two aspects of the local political environment should affect how subnational governments implement national social policy: the partisanship of subnational executives and the budgetary constraints they face.

With regard to the partisanship of executives, research strongly suggests that parties maintain distinct positions on social programs (Nygård (2006), Slapin and Proksch (2008), Bakker et al. (2015)). Correspondingly, a large body of literature has assessed the role of partisan conflict in the making of social policy (Korpi and Palme (2003), Allan and Scruggs (2004), Iversen and Stephens (2008)), finding that government partisanship is strongly related to policy change. I argue that conditional on policy, this logic should extend to implementation as well: local partisan executives, if given power over implementation, should be expected to use this leverage to shift policy content towards their preferred outcome.

Analogously, budgetary conditions have been widely used to explain social policy making: Research examining both social policy expansion and retrenchment has frequently emphasized the role of budgetary factors in policy change (Pierson (1998), Pierson (2001)). I hypothesize that budgetary considerations should also affect implementation decisions: subnational executives exposed to more intense budgetary pressure should implement programs with a view to reducing program generosity and curtailing expenditure.

Both hypotheses are confirmed in my analyses of RMI implementation data. Following the decentralization of the program, implementation decisions are strongly associated with local partisanship and budgetary conditions: conservative executives tend to take administrative measures to suppress enrollment in cash benefits and increase controls of recipients, but expand active labor market policy relative to leftist executives. This, I argue, is consistent with partisan conflict over the relative priority of “activation” (the targeted use of benefits to push recipients onto the labor market and reduce long-term benefit receipt) versus “decommodification” (the insulation of recipients from market pressures through generous benefits) (Esping-Andersen (1990), Huo Nelson and Stephens (2008)). Budgetary factors also powerfully shape implementation in the expected direction: Executives faced with tighter budget constraints decrease labor market integration spending and tighten cash benefits. Finally, I document the aggregate effects of decentralization, showing that cash benefit administration is tighter and active labor market policy effort much more heterogeneous under decentralized implementation and financing of the program.

The study contributes to the political economy literature on several different levels.

First, it documents that local political environments shape social policy implementation within a nominally national policy, with important repercussions for the substance of social protection. This suggests that theories of policymaking which emphasize partisan and budgetary factors may usefully be extended to the study of implementation decisions – even in contexts featuring relatively strong bureaucratic institutionalization and stringent monitoring.

At the same time, implementation decisions of the type studied here may be an important mechanism driving previous findings relating partisanship to policy outcomes at the subnational level. A growing literature examining the reduced-form relationship between partisanship and policy outcomes in different subnational and local settings using regression discontinuity designs has tended to find partisan effects on policy (Petterson-Lidblom (2008), Folke (2014), Ferreira and Gyourko (2009), Leigh (2008)). However, what the outcome measures used in these types of studies often capture is a mix of policy setting and the implementation of a given policy. Therefore, it is a priori unclear which mechanisms account for the findings obtained. By contrast, the research design of the present study allows me to isolate implementation from policy, and results suggest that implementation decisions themselves give rise to partisan divergences in realized policy outcomes.

Second, the study contributes to our understanding of the role of "subterranean" reforms in welfare state retrenchment in France and elsewhere. I show that, along the lines of arguments made by Hacker (2004) and others, relatively unassuming changes to program organization can nevertheless lead to a substantively important weakening of social protection – a finding that might also apply in other contexts characterized by increasing social policy decentralization (Kuhlmann, Bogumil, Ebinger, Grohs and Reiter (2011)).

Third, the study is the first to conduct a detailed micro-level analysis of the RMI's reform, providing an opportunity to quell discord about the reform's effects in the extant literature. Some previous studies relying on highly aggregated data and case studies of a few departments have argued that decentralization "did not translate into divergences or inequalities in access to social policy rights" (Thierry (2008)), and that the "envisaged risk of a politicization of the management principles of the RMI does not seem to have played a role" (Pepin and Blandin (2007), p. 74). Likewise, an analysis of CAF data by Avenel (2007) suggests that implementation did not become politicized, stating that "the technical and regulatory character of the program (the criteria of which are nationally defined) has "neutralized" the leanings of departmental politics" (ibid., p. 35). These findings contrast starkly with statements in a more theoretical literature arguing that decentralization has engendered a rupture with the principles of legal equality and national solidarity in favor of the autonomy of territorial units (e.g. Destremau and Messu (2008), Donier (2010), Dubois (2012)). However, the existing evidentiary basis for both views is rather thin, leaving the debate at an impasse. The comprehensive analysis of implementation data at the subnational level undertaken here provides strong evidence of politically motivated implementation of the program after decentralization: the leanings of departmental politics do not, in fact, appear to have been "neutralized" by the technical and regulatory character of the program.

The remainder of the article proceeds as follows. Section 2 describes the RMI program and its decentralization, while Section 3 portrays the partisan and budgetary logics of the program. Section 4 describes the different data used in the study; Section 5 presents the empirical results obtained, and Section 6 concludes.

2. The RMI and its decentralization

Within the French social policy regime, the RMI marked a structural change when it was written into law in 1988. French welfare efforts, grouped under the label of continental, Christian democratic welfare states by Esping-Andersen (1990), had been focused on contributory systems of social insurance, based on fragmented funds separated between occupations, and managed jointly by employers and unions (Palier (2000)). With high structural unemployment in the 1980s, these systems did little for those citizens that had little or no connection to the labor market, resulting in increasing concerns about social exclusion. In 1988, the Socialist government responded with the creation of the RMI, a means-tested cash benefit that guaranteed a certain minimum income (the benefit shrinking with increasing income up to the eligibility threshold) to anyone over the age of 25.

This cash benefit was financed wholly by the national government, and benefit payments were made by the CAF (*Caisse d'allocations familiales*), a national welfare payments agency with local branches. Besides the cash benefit, the RMI law also provided for so-called integration (*insertion*) programs intended to combat social exclusion. As part of this integration program, every RMI recipient was supposed to sign an individualized integration contract (*contrat d'insertion*) with a social worker that would specify a plan of re-integration through training measures, participation in social programs and job search (Palier and Thelen (2010), Lafore (2003), Duvoux (2011)).

These integration programs were organized jointly by the central government and executives at the subnational level of so-called departments (*départements*) (Bouchoux, Houzel and Outin (2004)). Departments are a French subnational administrative and political unit since Napoleonic times. These bodies, of which there are 96 in mainland France, have their own tax revenues, budgets and elected councils. They are governed by the Council President who is elected by the new council after every council re-election (half of the councillors are replaced every 3 to 4 years) and who becomes the executive of the departmental bureaucracy, which numbers about 3,500 staff per department on average. Departments were responsible for budgeting and financing the integration programs, but a fixed minimum amount of expenditure (17% of cash benefits paid out) was mandated by law in order to prevent departments from saving money on integrating recipients.

Decentralization

Departments and their elected executives became crucial to the RMI when the program was decentralized by a conservative government in 2003 (Le Lidec (2011)). The reform dramatically changed the program's organization by giving departments complete *administrative and budgetary* responsibility over both the cash benefit and the integration parts of the program. The funding of the program as well as all practical decisions about implementation were now in the hands of departments, and specifically in the hands of their elected executives, the departmental council Presidents (Tuchszirer and Join-Lambert (2003), Kuhlmann et al. (2011), Inspection générale des affaires sociales (2007), Avenel (2007)).

However, a legislative amendment that would have allowed departments to actually change key parameters of the policy, like benefit levels and eligibility criteria, failed. Therefore, the general legal rules governing the RMI remained unchanged, and benefits and eligibility were still determined by national law even after decentralization.

As a consequence, departments were empowered to make decisions on the implementation of policy, but not about the policy itself. The latitude offered to departments can, with some simplification, be conceptualized as consisting of two dimensions: the tightness of cash benefit administration, and the emphasis accorded to recipient re-integration.

As concerns the cash benefit administration dimension, departmental executives had to decide on a whole host of implementation questions following decentralization: they could decide where applicants would be able to apply for the benefit and delineate how they assessed compliance or non-compliance with integration contracts. They also decided whether and how severely to sanction non-compliance with the contract or no-shows at appointments by suspending cash benefits, how to treat administrative errors leading to overpayments, cash advances for expected future benefits in cases of hardship, how frequently to control recipients to verify eligibility, and a host of other issues. These aspects can seem small and mundane when considered in isolation. However, taken together, they can affect the substance of the program, especially in terms of how easy it is for potential recipients to apply for and continue receiving cash benefits.

As concerns the integration dimension of the RMI, spending requirements were lifted, and departments could freely decide on the total budget allocated to integration programs following decentralization. This gave them a free hand to prioritize or de-prioritize integration relative to all other areas of department expenditure. Within the integration budget, they could also freely choose which programs to offer: this implied deciding between labor market programs and other social programs less explicitly targeted at returning recipients to work, like health or housing benefits. Moreover, department executives were authorized to independently organize and instruct the department bureaucracy responsible for integration services. Crucially, this includes the definition of integration efforts that would be demanded in integration contracts, the intensity and frequency of follow-ups and meetings, et cetera.

3. Partisan and budgetary logics in the RMI

I hypothesize that two aspects of the local political environment – the partisanship of local executives and the budgetary situations they face – affect what executives do on the two dimensions of local implementation latitude. In this section I briefly explain the logic behind these hypotheses.

The partisan cleavage about the RMI program dates back to the moment of its introduction. While the left and right in the French National Assembly agreed on the need for a new social inclusion program, they had very different ideas about the benefits and obligations it should entail. As Eydoux and Tuchszirer (2010) describe in their analyses of the parliamentary records, conservative parties argued that the obligation to integrate fell on recipients. In their view, cash benefits should be the reward given to recipients in exchange for making efforts to re-integrate into the labor market and work. As a deputy from the conservative RPR argued, this required the possibility of "suspending the payment of the benefit if the person does not respect her commitments" (ibid., p. 7). For another deputy of the conservative UDF, "the integration contract must be perceived by the recipient as a firm commitment upon which the payment of the benefit is conditioned, since it is essential for the RMI not to become a new form of assistance" (ibid.). This worry about a "new kind of assistance" highlights concerns about disincentives and increased unemployment due to social policy that were and are frequently voiced in the debates about social policy in France.

By contrast, the governing Socialist Party in the French National Assembly saw the integration part of the policy as the expression of an obligation of society to integrate recipients instead of only giving them cash benefits. It therefore objected to making benefits conditional on "good behavior" or imposing a work requirement for recipients: giving aid to recipients took priority over concerns about disincentive effects. A compromise between these positions was eventually found, and the bill that passed did contain a requirement for recipients to sign an integration contract outlining measures they should take to integrate, though only *after* they had first applied for cash benefits.

This partisan cleavage, which accompanied political discussions of the RMI from its inception onward, is best understood as emanating from political conflict about the RMI's potential to decommodify recipients.[3] As Esping-Andersen (1990) has argued in developing this concept, a

3 Chemin and Wasmer (2004) provide evidence that the introduction of the RMI did have disincentive effects on recipients, increasing unemployment.

core characteristic of modern welfare states is the extent to which the social rights they grant allow citizens to live without relying on markets. But the extent to which the RMI could de-commodify its recipients depended quite directly on the details of its implementation: As a relatively generous cash benefit without an explicit legal work requirement, the RMI could substantially decommodify a considerable fraction of the French unemployed when administered leniently. However, with demanding integration contracts, frequent controls and tight conditionality of the cash benefit on recipient effort, the very same policy on paper could be geared much more towards recipient *activation* (in the terminology of Bonoli (2010) and Gingrich and Häusermann (2015)), and have its *decommodifying* effect reduced.

I hypothesize that giving partisan departmental executives latitude over RMI implementation effectively allowed them to move policy content closer to their preferred outcome by choosing whether to prioritize decommodification or activation. Importantly for this argument, the departmental party system very closely resembles the national party system in France, and party positions are strongly homogeneous across levels of government.[4] Consequently, the party affiliation of executives should affect how the RMI was implemented in departments: right-wing executives should administer cash benefits more restrictively than left-wing ones, and they should insist more on labor market integration effort and conditionality of benefits.

The second aspect of local political environments hypothesized to affect implementation is each department's budgetary situation. Assessing the role of budgetary pressures in the making of implementation decisions requires an understanding of the public finance changes that RMI

4 In fact, departmental politicians are not infrequently also national politicians in the French Senate or National Assembly simultaneously, a practice called mandate cumulation (*cumul des mandats*) which is currently being phased out.

decentralization engendered. In exchange for being burdened with full budget responsibility for cash benefits and integration, departments were allocated revenue from the domestic gasoline tax. This compensation, granted in the form of a fixed share of gasoline tax revenues, was calibrated to equal the level of spending that had been required in the year before decentralization. Importantly, this revenue allocation was *not* made contingent on the future development of RMI recipients or cash benefits. Any deviation, positive or negative, between this allocation and the real costs of running the RMI program would have to be absorbed by the departmental budget. Departments thus actually became fully responsible for funding in the sense that they could use extra revenue from other sources to increase social spending, but could also use funds not used for social spending on other purposes or to lower taxes.

The fiscal compensation from the gasoline tax turned out to grow much less than the expenditures on the RMI, resulting in a 450 million Euro deficit in the first year (Senat francais (2005), Kuhlmann et al. (2011)). In response to this, an additional budget allocation was made to departments for RMI spending from 2006, but its total amount was limited and it never fully compensated departments according to Audit Court reports (Cour des Comptes (2011a), Inspection générale des affaires sociales (2007)). Nationally, actual RMI costs to departments exceeded the transfers they received by about 13%, adding to their budgetary pressure. In summary, between 2004 and 2008, departments were only partially compensated for changes in their RMI expenditure and suddenly faced real trade-offs between social spending, other spending, and taxation decisions.

These budgetary trade-offs were faced by departmental executives whose budget situations were highly heterogeneous to begin with. In particular, the strength of tax bases and the importance of transfers and mandated expenditure from the national government vary widely across departments. I hypothesize that this should matter for RMI implementation decisions following decentralization: executives facing higher budgetary pressure should be more likely to use implementation levers — tighter cash benefit administration, decreased investment in integration services — to save on RMI expenditure.

It is worth pointing out that in the lead-up to decentralization, several actors voiced concerns along the lines of the hypotheses above, fearing that implementation decisions could become a function of partisanship and budgetary pressures. For example, in a parliamentary debate about the reform law, the Socialist Party (PS) deputy Gaëtan Gorce expressed worries about allowing executives to decide sanctioning policy:

"If we demand that suspension decisions be subordinated to a recommendation by the local bureaucracy [*instead of a decision by the elected President of the department council*], it's because we fear that the resources transferred by the national state will not cover the needs and that certain Presidents of departmental councils, having to choose between more firmness in the renewal of cash benefits and increasing their local taxes, will privilege the former option, which would contradict the requirements of solidarity" (Eydoux and Tuchszirer (2010), p. 11).

However, as discussed above, the extant empirical research literature on the subject has not found evidence of politically motivated implementation, concluding that the heavily bureaucratic character of the policy has “neutralized” departmental politics (Avenel (2007)). Dissenting research that claims decentralization to have engendered politicized program implementation is almost entirely legal-theoretical in nature and does not engage with the empirical data available (e.g. Destremau and Messu (2008), Donier (2010), Dubois (2012)). As a consequence, there is at present little in the way of evidence to adjudicate on the competing claims.

4. Data and empirical analysis

4.1. Data

In order to make an empirical contribution to this debate, I analyze four different types of data at the department level. They cover implementation and spending decisions made by departments as well as surveys of RMI program participants. Specifically, I analyze

- an internal administrative survey on the effects of decentralization conducted by the national benefit payments agency (CAF) in 2006,
- the results of an audit of a sub-sample of local social policy bureaucracies and their links with local governments conducted by the French Audit Court in 2011,
- an administrative panel dataset from the Ministry of Social Affairs covering spending on programs for social and labor market integration of recipients from 1996 to 2008 and
- a repeated cross-sectional survey of RMI recipients conducted by the research unit of the French Ministry of Social Affairs (DREES) in 2003 and 2006

The end of the observation period is 2009, when the RMI was amalgamated with other programs into the *revenu de solidarité active* (RSA).

Some of these sources also contain interesting qualitative data. I adduce this data and explain the quantitative measures in more detail in the respective sections. Throughout the study, I focus on the two dimensions of RMI implementation defined above, namely

(1) *tightness of cash benefit administration*: the degree to which implementation is favorable to applicants receiving cash benefits, including whether and how strongly receipt is conditioned on behavior

(2) *importance accorded to re-integration programs*: the intensity of efforts directed at integrating recipients in society and the labor market

In each of the datasets, I identify the measures that best match with these dimensions and use them as dependent variables.

To measure the partisan orientation of the departmental executive, I use data on the party of the President of the departmental council. Presidents are elected after every new election to the department (when half of all seats are renewed, every 3 or 4 years) and have wide-ranging powers in the local administration. When no single party attains a majority in the council, coalitions form to elect a President, and other offices are given to the party whose candidate is not the President. After RMI decentralization in 2004, the President's powers notably include final authority over all RMI decisions and the preparation of the RMI integration budget. On the basis of the President's party affiliation, I construct both a simple left/right dichotomous coding of departmental governments, and a more nuanced continuous measure of political orientation based on the Chapel Hill Expert Survey Survey (Bakker et al. (2015)). For this latter measure, I use the “economic policy” dimension of the expert coding, which maps parties on the left-right spectrum. The measure ranges from 0 (extreme left) to 10 (extreme right).[5]

To measure budget situations across departments, I use a measure of budgetary pressure that was fortuitously produced by a French Senate report in 2003, just before the RMI’s decentralization. Since the French fiscal system involves massive transfers between levels of government as well as heterogeneities in mandated expenditure, conditional and unconditional grants based on formulas and special programs, measuring a department’s budgetary situation is quite complex. The measure developed by the French Senate begins with a measure of the strength of the tax base of the departments, and then corrects for transfers and mandated expenditure to arrive at a continuous measure of the budgetary pressure on a department. This measure is dimensionless, and higher values indicate a more favorable budget situation.[6]

It is worth noting that the measures of partisanship and the budgetary situation are almost perfectly uncorrelated with each other (Pearson’s R of 0.01). In the regressions below, I include other covariates in order to rule out alternative explanations involving other characteristics of the departments. These include the level of unemployment, the relative size of the immigrant population, population density, and the number of RMI recipients per department, based on data from the statistical agency INSEE.[7] Descriptive statistics are available in Appendix section 7.1., Table 8.

5 An explanation of the partisan coding, as well as a list of the scores, can be found in Appendix section 7.2.
6 The measure is not available for Paris since its public finance situation is very particular.
7 For example, Duguet, Goujard and L’Horty (2009) document the important differences in labor market opportunities across municipalities. Administrations may take such factors into consideration when making implementation decisions; for example, punitive implementation might be easier to sustain in contexts with more labor market opportunities.

4.2. Internal CAF data

I first investigate data from an internal survey of the CAF. As explained above, the CAF is charged with actually paying the cash benefits to recipients, and dealing with benefit-related questions and problems. This means that it maintains databases of recipients, receives recipients in its offices, and administers changes to their status or payment modalities, including sanctions or deletions. Although the CAF has regional offices across France, it is part of the national bureaucracy and tasked primarily with implementing national programs.

Decentralization gave departmental authorities the power to demand administrative changes from the CAF. For example, departments could ask CAFs to increase controls on recipients, or allow the CAF to accept applications at its offices (instead of requiring applicants to visit the departmental social policy offices). Moreover, if departments wanted to sanction recipients or delete them from the rolls entirely, they needed to do so using a request to the CAF, since it maintains the recipient database. Therefore, regional CAFs are very well-informed about their department's position on RMI implementation, yet are not politically accountable to them, making them an ideal source of information for the purposes of this study.

The survey I analyze was administered by the national CAF to the management level of all regional CAF offices in 2006. The goal was to find out how decentralization affected the work of the departmental CAF offices, mostly for internal management purposes. The survey was filled out by very senior bureaucrats, typically the director or deputy director of the CAF in each department. These bureaucrats are highly familiar with policy and management issues and relationships with the departmental actors in social policy, though probably somewhat less familiar with the street-level work done by their colleagues given that an average CAF has about 315 staff. Owing to the nature of the survey, responses are available for all but 2 of the CAF offices, and there is almost no non-response (besides "no opinion" answers) on individual questions.

A broad overview of response averages was published (Avenel (2007)), but the full department-level results of the survey were not made publicly available. This means that responses are likely driven by a general concern to demonstrate the good functioning of the local branch to the central national office and desires to advance in the organization, rather than by worries about the scientific or political use of the data. Even more importantly, the results were not shared with departmental bureaucracies, and CAF bureaucrats freely answered questions about their cooperation with departmental bureaucracies.

Six questions in the survey relate to access, conditionality and integration efforts. Three questions asked about specific measures taken by departments and implemented by CAFs. They asked (1) whether, since decentralization, the regime of controls on recipients was tightened and by how much (2) whether, since decentralization, demands for temporary sanctions by the department went up or down (3) whether, since decentralization demands for permanent sanctions (deletion) by the department went up or down.

Three additional questions attempt to measure overall policy content from the perspective of recipients. They ask (4) whether, from the standpoint of recipients, access to benefits had improved or deteriorated since decentralization, (5) whether, from the standpoint of recipients, the management of benefits had improved or deteriorated since decentralization, and (6) whether, from the standpoint of recipients, integration services had improved or deteriorated since decentralization.

As the reader will have noted, these questions were all phrased to encourage comparison between the pre-centralization era and the time since 2004. Although this design arguably improves the measures over a pure cross-section taken after decentralization, it is not explicitly based on comparisons of two independent measurements before and after decentralization. Therefore, the degree to which these measures can succeed at eliminating the department fixed effect in comparisons cannot be ascertained.

In the 9 cases where there were two instead of one CAF districts per department in 2006, I employed the average response given by the two bureaucrats as the recorded outcome at the departmental level in order to use all available information.

In the national aggregate, the data give several insights into the effects of decentralization. First, the control regime applied to RMI recipients tightened considerably: all offices responded either no change or an increase of controls (Figure 1).

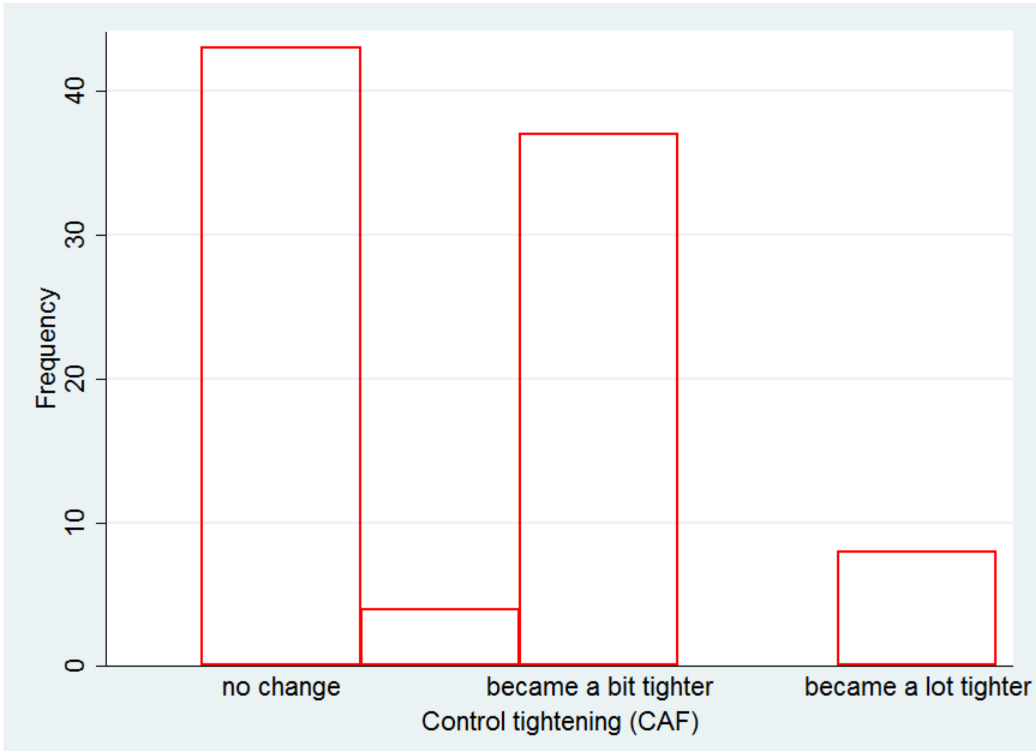


Figure 1: Frequency of CAF responses to question of whether the control regime tightened since the 2004 decentralization

Similarly, regarding sanctions, the aggregate data suggest that both temporary sanctions and deletions increased but that they did so unevenly across departments (Figure 2).[8] Overall, these responses are consistent with the hypothesis that decentralization led to a more punitive administration of benefits in the aggregate, though not homogeneously so across locales.

8 Sanctions consist of either a temporary suspension of payments, or a quasi-permanent deletion from the file and are typically pronounced based on non-compliance with an integration contract or a delay in supplying required documents. Again, local bureaucracies have ample latitude in interpreting the legal provisions regarding these sanctions, and can basically choose not to apply them at all, or to apply them as harshly as legally possible (though legal recourse is possible for beneficiaries). Importantly, the distinction between temporary and permanent sanctions is not extremely clear cut, partly since many temporary suspensions result in permanent sanctions over time, and because the precise handling of these sanctions differs across departments.

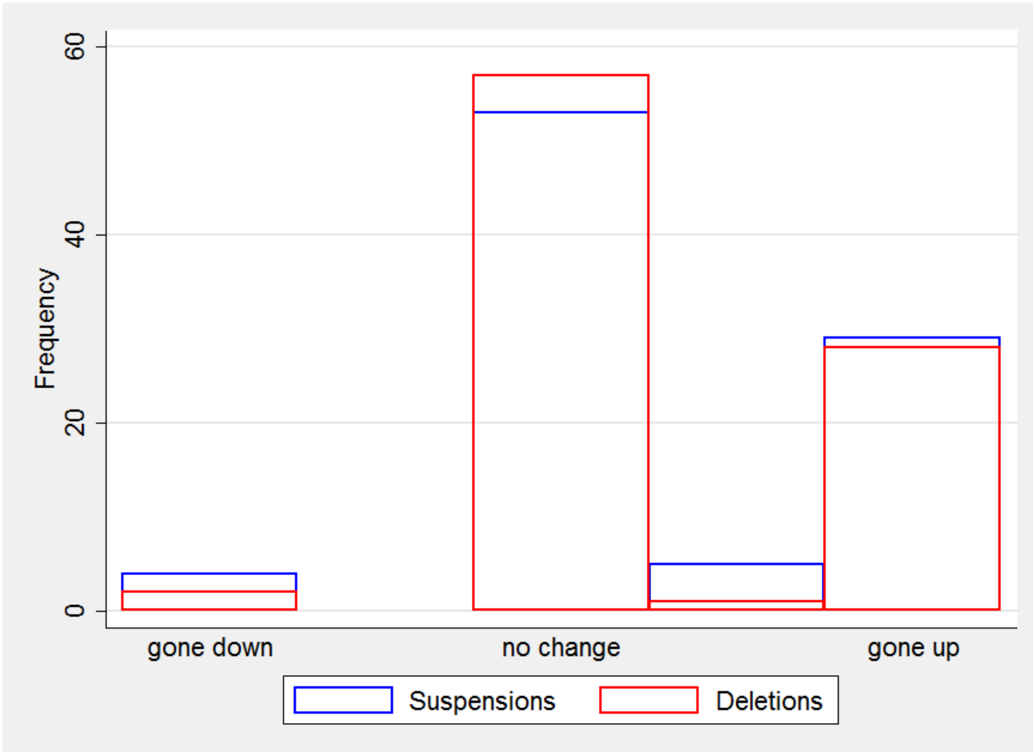


Figure 2: Frequency of CAF responses to questions of whether temporary (suspensions) or permanent sanctions (deletions) had increased or decreased since decentralization.

Figure 3 displays results of the questions asking how "from the standpoint of recipients", the program had evolved since decentralization. These questions are particularly interesting for the purposes of this study since they encourage respondents to produce a comprehensive assessment of policy content. The result shows that on balance, access to benefits and management are judged to have stayed roughly similar. However, they deteriorated in about twenty departments and improved in another twenty. Integration is the exception to this picture; it is judged to have improved significantly in most departments.[9]

9 However, it should be noted that the CAF is not involved in integration work itself (with some very minor exceptions), and this judgment is therefore to be interpreted as an "outside" judgment of the work of the departmental bureaucracy. Perhaps for this reason, the question on integration is the only one with significant "no opinion" response (by 35 CAF offices).

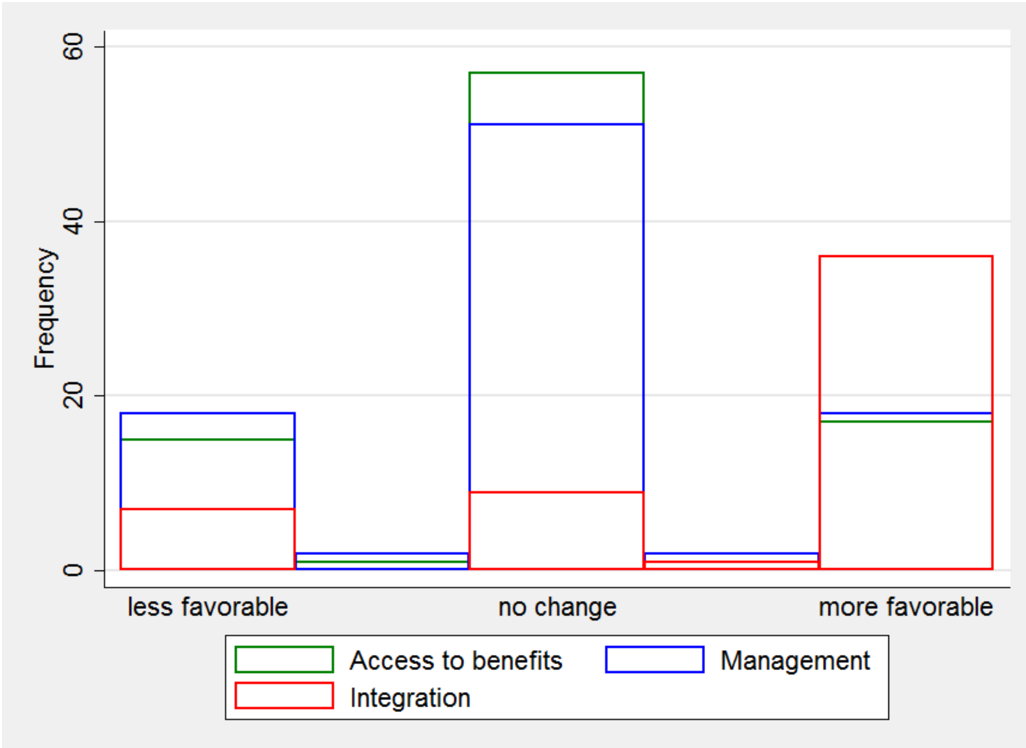


Figure 3: Frequency of CAF responses to questions of whether from the standpoint of recipients, access to benefits, benefit management, and integration, respectively, had improved or deteriorated since decentralization.

In order to test the hypotheses explained above and assess the role of local political environments in RMI implementation, I regress these measures on department executive partisanship and budgetary situation. Tables 1 and 2 contain the regression results for all outcomes. Below, graphs are displayed for the relationships that are statistically significant in those regressions.

Partisanship is strongly correlated with changes in the control regime and recipient's access to benefits. As Figure 4 indicates, departments governed by right-wing executives appear to have tightened control regimes more drastically than those governed by left-wing executives after decentralization.

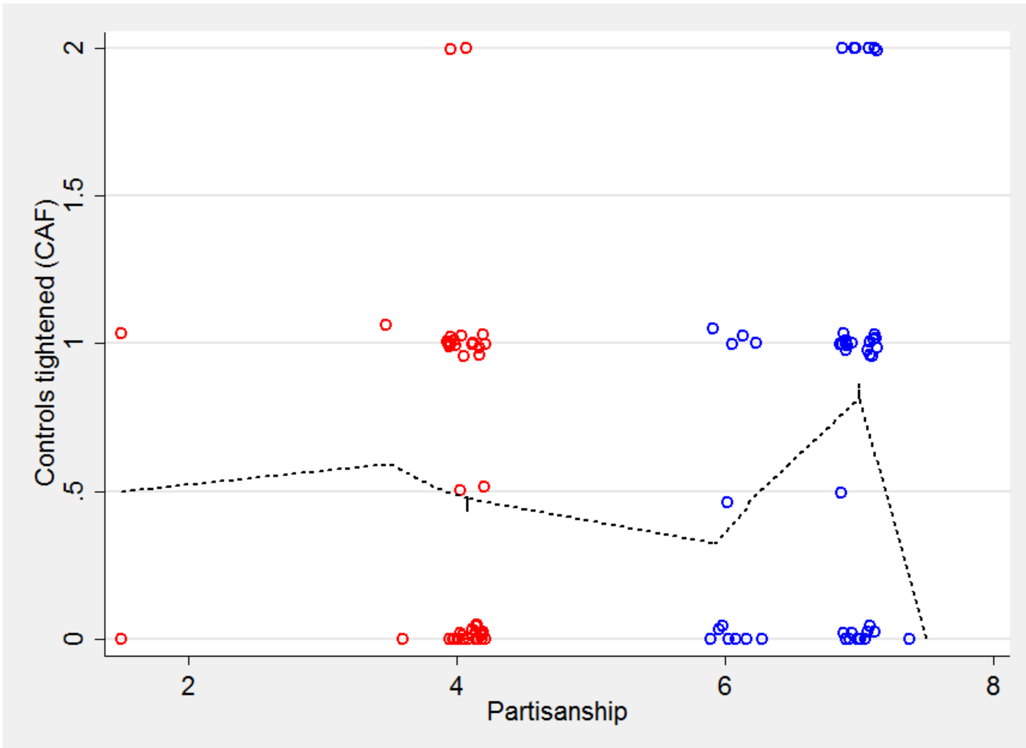


Figure 4: Partisanship of the department council President and change in the control regime since decentralization as assessed by CAF bureaucrats. Higher values on the x-axis indicate more conservative parties. Lowess regression, observations jittered to make distributions visible.

Moreover, access to rights is judged to have improved in jurisdictions governed by the left, whereas recipients in the average department governed by conservatives experienced a slight deterioration in access to benefits according to the survey (Figure 5). Both results are consistent with the hypothesis that partisan executives used their latitude over implementation to move policy content towards their preferred position.

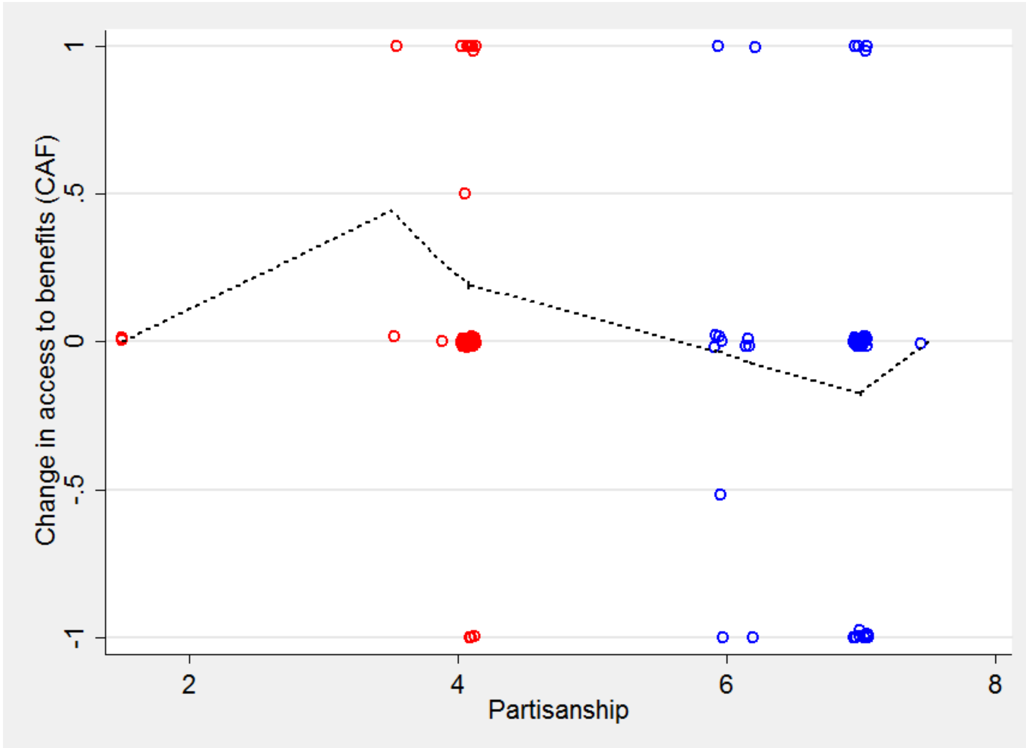


Figure 5: Partisanship of the department council President and change in access to benefits since decentralization as assessed by CAF bureaucrat heads. Higher values on the x-axis indicate more conservative parties. Higher (lower) values on the y-axis indicate improved (diminished) access to benefits. Lowess regression, observations jittered to make distributions visible.

The budget situation in departments is strongly associated with changes in the sanctioning policy and the quality of integration services, as assessed by the CAF. Suspensions of benefits generally increased following decentralization. However, they are judged to have increased relatively more in departments with higher budgetary pressure, as Figure 6 shows.

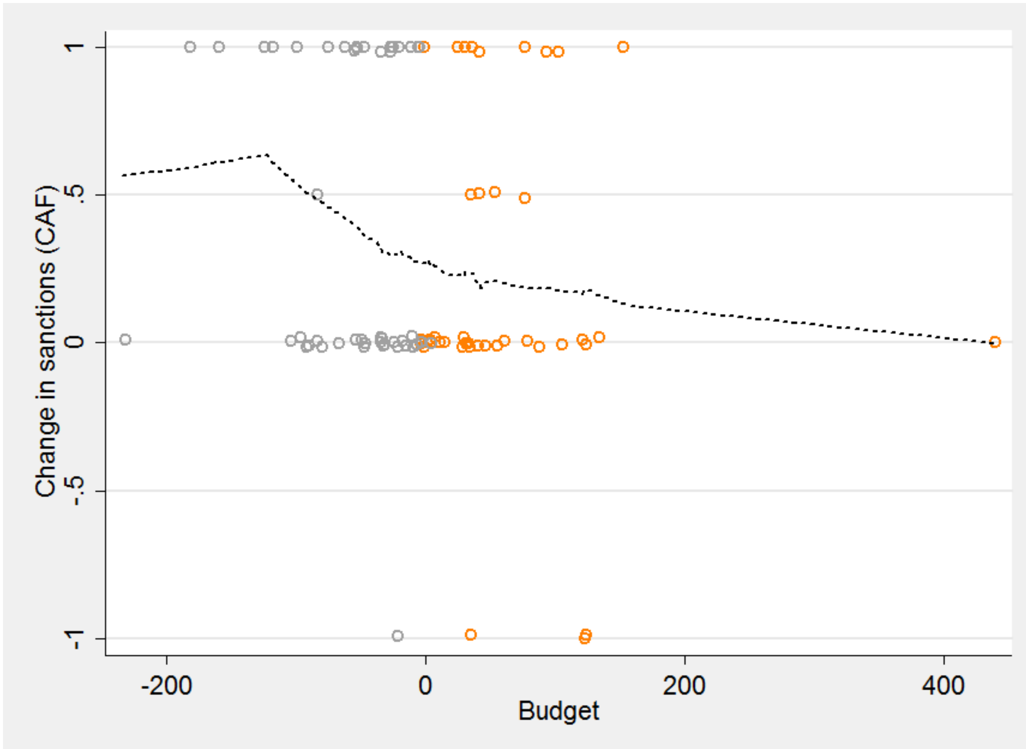


Figure 6: Budgetary situations in departments and change in sanctions demanded by department according to CAF, in departments with strong (gold) and weak (grey) budgets (cut at the median). Lowess estimate, observations jittered to make distribution visible.

Budgetary pressure also appears to have affected the quality of integration services offered to recipients: As Figure 7 shows, all of the departments that reportedly experienced deteriorations in integration policy score below the median (-2) in the budgetary pressure variable. Recall that mandatory spending thresholds for integration services were lifted in the reform, allowing departments to freely reduce spending if they so desired. This result from the CAF survey is consistent with the notion that departments with weaker budgets were more likely to make use of this possibility. I investigate spending on integration programs in more detail in section 4.4. below.

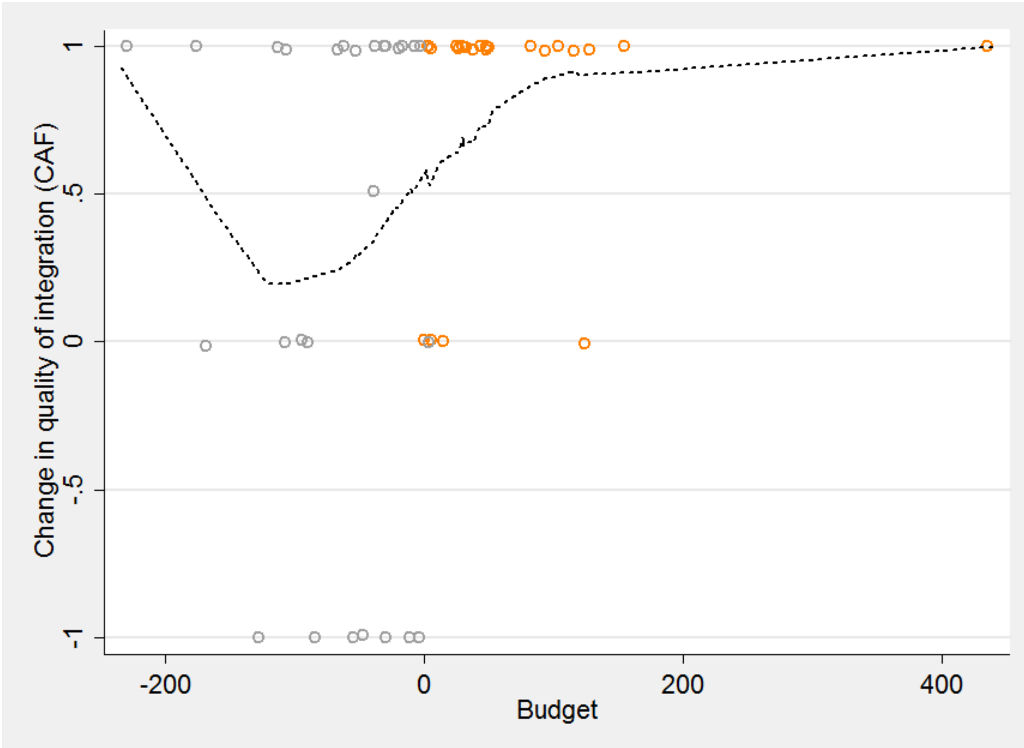


Figure 7: Budgetary situations in departments and change in quality of integration as assessed by CAF, in departments with strong (gold) and weak (grey) budgets (cut at the median). Lowess estimate, observations jittered to make distribution visible.

Table 1: Regressions of CAF indicators on budgetary and partisan variables without controls.

	Suspensions	Controls	Access to benefits	Management	Integration
Partisan	0.035	0.090	-0.098	-0.022	-0.066
	(0.042)	(0.045)**	(0.039)**	(0.041)	(0.067)
Budget	-0.001	0.000	0.001	0.001	0.002
	(0.001)**	(0.001)	(0.001)	(0.001)	(0.001)**
_cons	0.106	0.121	0.538	0.118	0.924
	(0.234)	(0.238)	(0.212)**	(0.227)	(0.376)**
R^2	0.05	0.05	0.08	0.02	0.08
N	90	91	90	90	52

* p<0.1; ** p<0.05; *** p<0.01

Table 2: Regressions of CAF indicators on budgetary and partisan variables with controls.

	Suspensions	Controls	Access to benefits	Management	Integration
Partisan	0.044	0.090	-0.084	-0.007	-0.053
	(0.042)	(0.053)*	(0.040)**	(0.046)	(0.079)
Budget	-0.002	-0.000	0.001	0.001	0.002
	(0.001)*	(0.001)	(0.001)	(0.001)	(0.002)
Unemployment	-0.008	-0.056	0.028	0.054	-0.030
	(0.037)	(0.042)	(0.038)	(0.054)	(0.070)
No. of recipients	0.000	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Migrant share	-0.639	0.506	-0.498	-1.429	-0.005
	(1.807)	(2.011)	(1.492)	(2.244)	(3.690)
Population density	0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
_cons	0.103	0.499	0.215	-0.331	1.026
	(0.468)	(0.551)	(0.423)	(0.562)	(0.830)
R^2	0.06	0.07	0.12	0.04	0.10
N	90	91	90	90	52

* p<0.1; ** p<0.05; *** p<0.01

The point estimates in Table 1 are robust to the inclusion other covariates as controls in the regression (Table 2). However, standard errors increase in some cases, with the relationship between budgetary pressure and integration services becoming insignificant.

The CAF dataset also contains some interesting qualitative information in administrator's responses to open-ended questions. While the representativeness of these qualitative data is difficult to ascertain, they do allow one to get a perception of the bureaucracy's view on the RMI and its decentralization.

Based on their responses, quite a few administrators perceive some conflict between the goals of the social policy bureaucracy and those of the local executives. In response to a question about their relationship, an administrator in a department with a leftist executive writes that there are "two logics: one of financial control for the departmental council, and one of social and equitable administration for recipients no matter which benefits are concerned (for the CAF)". In the same vein, an administrator from a conservatively governed department states that “one notes a particular pre-occupation with RMI expenditure on the side of the departmental council generally”.

The imperative to reduce expenditure is at times directly related to implementation decisions taken, notably with regards to the policy of controls. As a manager from one department governed by the left writes, “the reinforcement of the control policy has its origins primarily in the desire to contain RMI expenditures”. Regarding controls, administrators from two conservative departments state that their council has demanded “severely increased strictness of controls”. A third administrator in a department with a conservative executive complains about the council’s demands for additional controls, stating that “the general council solicits us for presumed cases of fraud based on the views of elected officials that are founded not on administrative diagnoses, but on “sentiments”. The CAF should not participate in this approach.”. Conversely, in a department governed by the far left, an administrator notes that “aside from outright frauds, the demands for controls are, until now, much less important than before [decentralization]”.

Finally, in response to an open-ended question asking bureaucrats whether they have any other remarks they would like to make, an administrator from a department with a conservative executive states rather unequivocally: “the decentralization of the RMI has engendered a rupture with equal treatment for all beneficiaries nationally, and for complex cases”.

4.3. Audit Court data

In 2011, after the RMI had already been reformed again and amalgamated with other programs, the French Audit Court audited 17 departments to assess their management of the program and the effectiveness of their integration programs. To this end, it dispatched auditors from regional Audit Courts to the local bureaucracies and requested data and answers to a systematic questionnaire from them. The detailed reports and analyses for the whole of France and individual departments are available publicly, and contain both quantitative and qualitative data on the subset of departments audited. Importantly, the manner in which this subset was selected for audit is not known, and it is likely not random. In Appendix section 7.3., I document that there is a high degree of observational similarity between the audited and the non-audited departments.

The most interesting quantitative data gathered by the Audit Court concerns the number of sanctions pronounced against beneficiaries. This provides an alternative measure of sanction use, which was also tapped in the CAF data analyzed above. While the distinction between different types of sanctions is not consistently maintained in the data, and observations for many years are not available, one can get an overall estimate of the sanctioning intensity of departmental bureaucracies by averaging sanction rates for the years available and dividing them by the average number of beneficiaries. This results in the average frequency of sanctions per recipient per year. I use the average sanctioning intensity between the years 2004 and 2007, since they correspond to a term of the President of the department council.

Putting these figures in relation to the partisanship of departmental executives reveals a strong positive correlation between the conservatism of departmental executives and the sanctioning intensity of local bureaucracies, depicted in Figure 8. While there are practically no sanctions in departments governed by the French Communist Party, the average sanctioning intensity in departments governed by the most conservative parties is about 7%. Regression results, displayed in Table 3, show that the difference is statistically significant even in this small sample. Adding control variables increases the point estimate and standard errors. Conversely, the budgetary situation does not appear to be correlated with sanctioning intensity at all in the Audit Court data.

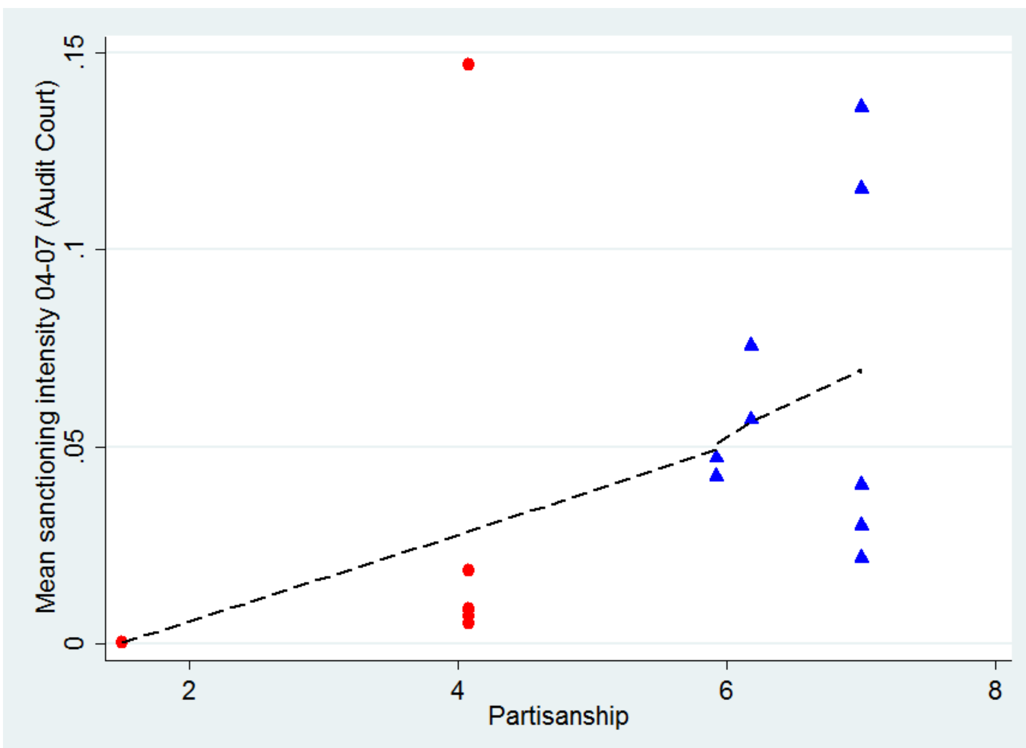


Figure 8: Average yearly sanctioning rate between 2004 and 2007, by partisanship of the departmental executive (higher values on the X axis indicate more conservative parties).

Recall that in the CAF data analyzed above, the finding was the inverse: changes in sanctions, as reported by CAF managers, appeared to be unrelated to partisanship, but strongly related to the budgetary situation. What might explain this divergence?

On the one hand, the CAF measure taps pre-/post-reform changes in sanctions while the Audit Court measured post-reform sanctioning levels. At some level it is therefore not surprising that the findings are different. Correlations in the cross-section might be driven by fixed attributes of departments; analyses of changes net out these department fixed effects and arguably lead to more credible inferences. To the extent that the CAF survey measure succeeds in absorbing the department fixed effect in sanction use, analyses based on CAF data are therefore more credible.

On the other hand, the Audit Court sanctioning measure is much less coarse than the CAF measure, providing administrative data on actual sanction density where it is available as opposed to a simple distinction between stable, increasing and decreasing sanctions. This would make the Audit Court measure seem preferable, especially if one is less confident about the ability of the survey instrument to net out the department fixed effect. I conclude that the evidence in favor of budgetary considerations in sanction use is stronger than that in favor of partisan considerations, but both data sources come with their advantages and disadvantages.

Table 3: Regression of average sanctioning rate per recipient (from Audit Court) on budgetary situation and partisanship indicator, with and without controls.

	Sanctioning intensity	Sanctioning intensity
Partisan	0.013	0.016
	(0.006)*	(0.011)
Budget	0.000	0.000
	(0.000)	(0.000)
Unemployment		0.009
		(0.012)
No. of recipients		-0.000
		(0.000)
Migrant share		-0.039
		(0.495)
Population density		0.000
		(0.000)
_cons	-0.025	-0.091
	(0.034)	(0.093)
R^2	0.21	0.28
N	17	17

* p<0.1; ** p<0.05; *** p<0.01

Besides sanctioning data, the Audit Court reports are also noteworthy for the qualitative information they contain on decision-making within the local bureaucracy. The contrast between RMI implementation in two departments in the Paris region offered by the audit reports is especially illustrating. Both Seine Saint Denis, an impoverished department in the northeast of Paris with a long legacy of communist local government, and Yvelines, one of the richest departments in France and governed by a succession of conservative parties, were subjects to audits by the Audit Court. In both cases, local administrations used, and in the opinion of the Court, overused their latitude in implementation, but in opposite directions.

In Seine Saint-Denis, the Court observed that the department was not implementing the law requiring departments to monitor insertion contracts and suspending individuals that were to be found in violation with them. As evidence, the report cites two internal papers circulated in the local bureaucracy. One from 2003 notes the “drift which could ensue after the decentralization reform in the context of ever-toughening budgetary requirements which could lead to systematic campaigns to suspend the benefits to beneficiaries without insertion contract or which do not follow every letter of such contract. *For the department of Seine Saint-Denis, adopting these positions is out of the question*. [emphasis added]” (Cour des Comptes (2011b), p.11).

The second internal position paper cited in the Audit Court report dates from 2006 and elaborates further on the position of the department with regards to sanctions. According to the Court report, “after evoking the obligation of the department to offer every beneficiary an integration contract, the [department bureaucracy's] paper explains that the establishment of a “general obligation to sign an integration contract” is excluded, since these contracts are “senseless and disconnected from the needs of the beneficiaries of public services”. In conclusion, the paper excludes deleting benefit recipients from the rolls in order to reduce their flow” (ibid., p.12).

Until 2007, not a single sanction was pronounced in Seine Saint-Denis. When the department slowly began pronouncing sanctions afterwards, it underlined its desire for a preventative approach using warning letters instead of direct sanctions and, vis-à-vis the Audit Court, remarked that information technology problems made it difficult to know which beneficiaries had signed an integration contract, impeding sanctioning.

Such problems did not seem to hinder the Yvelines department, on the other side of Paris, from pursuing a severely punitive sanctioning policy. As the Audit Court report notes, the department completely changed its local decision-making structure in response to decentralization (Cour des Comptes (2011c), p. 8). All local offices that had been previously tasked with pronouncing sanctions were deprived of this right, and a new office was created that maintained the sole prerogative over all sanctions in the department. As the report notes, this new office "pronounced almost 1000 permanent deletions from the welfare rolls in 2005, 989 in 2006 and 878 in 2005" (ibid.). The Audit Court held in its assessment that these steps were incompatible with the law, and the sanctioning powers were given back to the pre-existing local offices in 2008. As the court notes, still in 2008, the internal yearly integration plan for the department contained an "action sheet" entitled "Simplify the suspension procedure for RMI recipients related to the integration contract" (ibid.).

It is difficult to causally trace these two decisions back to the political partisanship of the decision-makers in departments in the years following decentralization. However, what the reports show quite strikingly is that, despite the same law governing the RMI in all of France, the reality "on the ground" was different for recipients in locales that are less than an hour away from each other by car, and that they were different for reasons that local actors could and actively sought to influence.

4.4. Integration spending data

Next, I consider the effects of decentralization on the RMI's integration programs using spending data compiled from reports issued by the Ministry of Social Affairs. Spending figures are yearly aggregates in constant Euros at the department level. Since the bulk of this spending is directed at labor market and professional integration, (i.e. on training and retraining programs and subsidized employment), I interpret spending as a measure of labor market integration effort. As I show in Appendix section 7.4., this interpretation is substantiated by more disaggregated data distinguishing labor market and social integration spending, which is available for a subset of departments from Audit Court reports.

In looking at the spending patterns, it is worth keeping in mind that the reform was in no small part designed to increase departmental integration efforts by giving departments stronger incentives to make recipients successful in the labor market. However, the reform also entailed the elimination of a minimum spending requirement on integration based on the argument advanced by departments that this requirement would be unnecessary once incentives were correctly aligned. Therefore, the effect of decentralization on spending is a priori unclear.

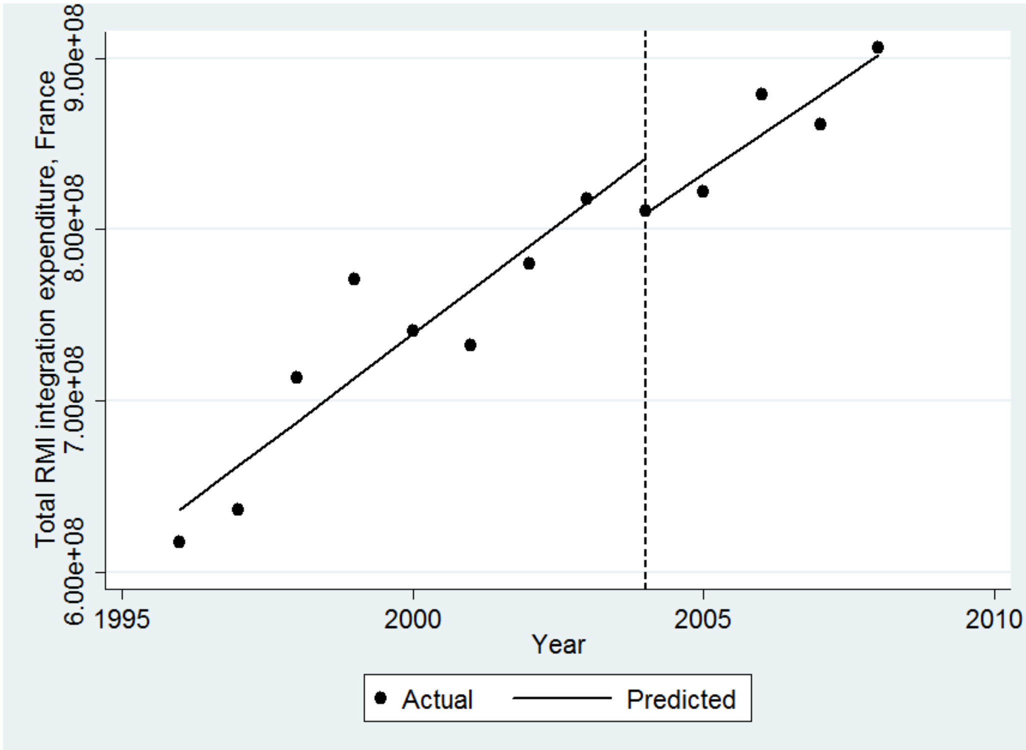


Figure 9: Total spending on integration of RMI recipients in constant Euros (Interrupted Time Series Analysis Graph).

As Figure 9 shows, the total level of spending stayed on its positive trend from the pre-decentralization era after 2004, without any notable acceleration or slowdown. However, as Figure 10 shows, the stability in overall spending levels masks a significant increase in regional heterogeneity of spending efforts that ensued in 2004 and after (Poutingon (2012)). After variation in integration effort between departments was slowly declining since 1996, the coefficient of variation in per-capita spending went from about 0.2 to about 0.45 within only four years following decentralization. Thus, although overall integration efforts did not change from a national perspective, the integration efforts directed at recipients depended much more on their place of residence than they previously had.

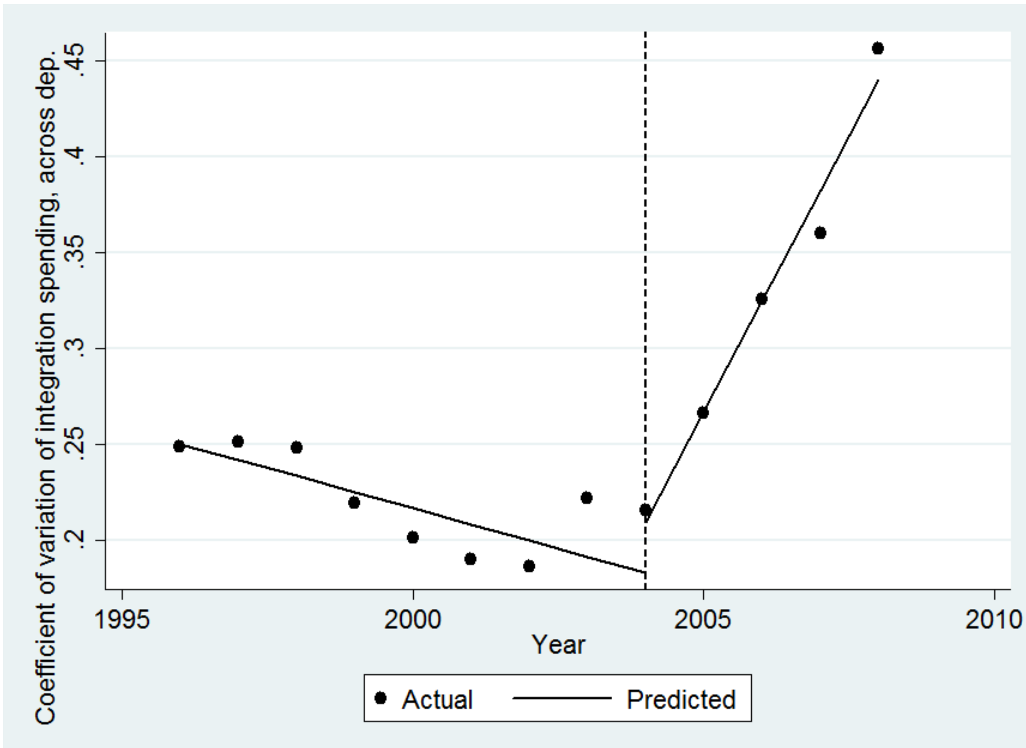


Figure 10: The coefficient of variation of per-capita integration spending, across departments (Interrupted Time Series Analysis Graph).

How is this variation related to the local political environment in departments? I first test the hypothesis that the budgetary situation affected spending decisions following decentralization. To this end, Figure 11 displays a lowess estimate of yearly changes in per-capita integration expenditure in relatively rich (gold) and poor (grey) departments. As the graph shows, pre-reform trends in integration spending were highly similar between rich and poor departments before 2004, but began to diverge notably after the reform. Departments with stronger budgets began to spend more per capita following the reform, leading to significant variation in policy effort by 2008. This divergence is of substantively important magnitude: while the richer half of departments actually spent slightly less on integration per capita than poor ones in 2003, they spent 34% more by 2008 (1049 vs. 779 Euros per recipient per year).

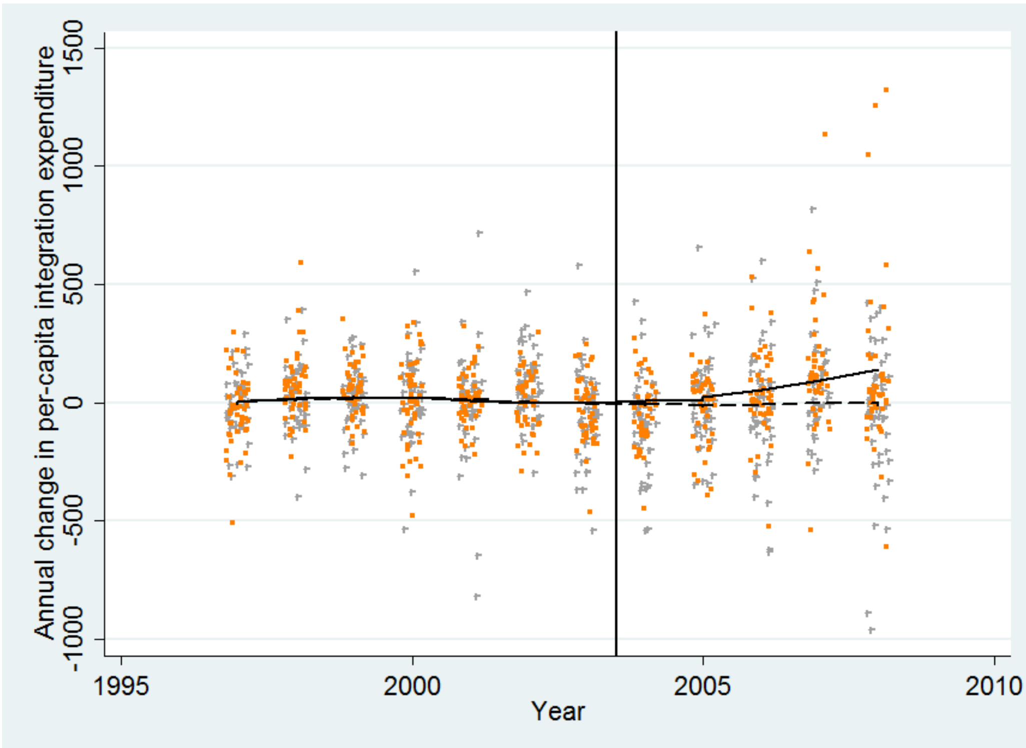


Figure 11: Year-on-year changes in per-recipient integration spending, in departments with strong (gold) and weak (grey) budgets (cut at the median). Lowess estimate, observations jittered to make distribution visible.

But the divergence in integration efforts is also related to the political make-up of the local governments that began making independent and unrestricted decisions about integration budgets in 2004. Figure 12 shows that while there were no partisan differences before decentralization, conservative executives appear to have increased per-capita spending on integration following the reform relative to left-wing ones.

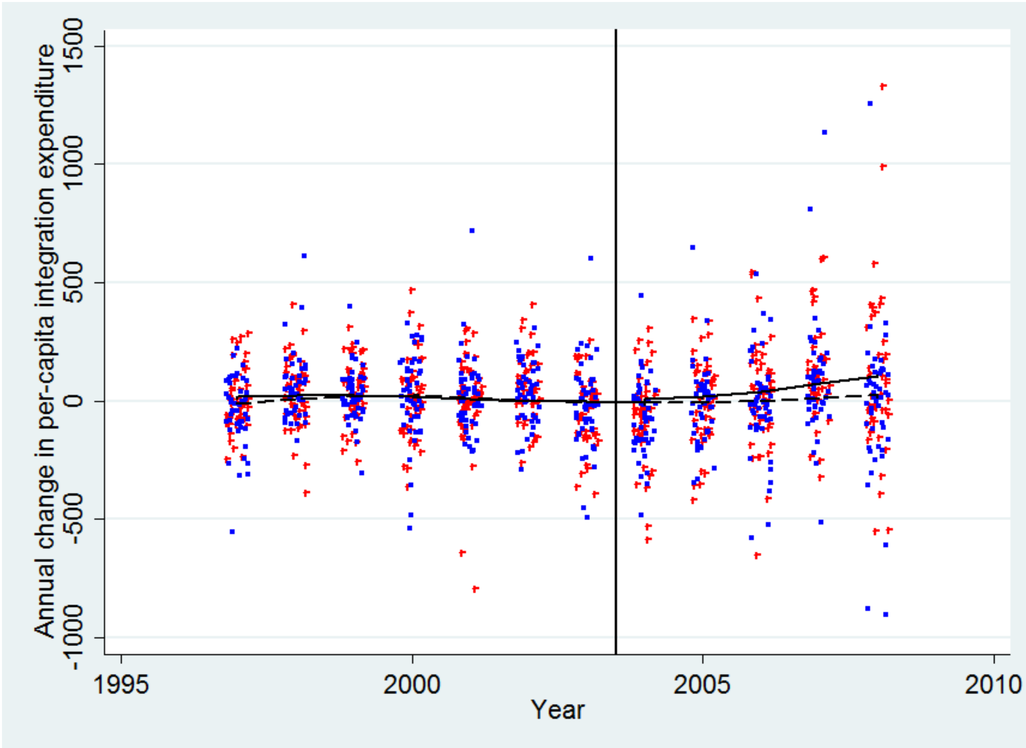


Figure 12: Year-on-year changes in per-recipient integration spending, in departments with left-wing (red) and right-wing (blue) executives. Lowess estimate, observations jittered to make distribution visible.

To demonstrate the role of decentralization in this partisan divergence, Figure 13 plots integration spending against partisanship before (2003) and after (2008) decentralization. While there was no relationship between partisanship and integration efforts in 2003, conservatively governed departments (higher values on partisan scale) spent significantly more on integration by the end of the observation period in 2008.[10]

10 It should be noted that budget situation and department partisanship are not at all correlated in the sample (Pearson's R of 0.01); the two results about budgetary situations and partisan orientations are therefore not mirror images of the same underlying correlation.

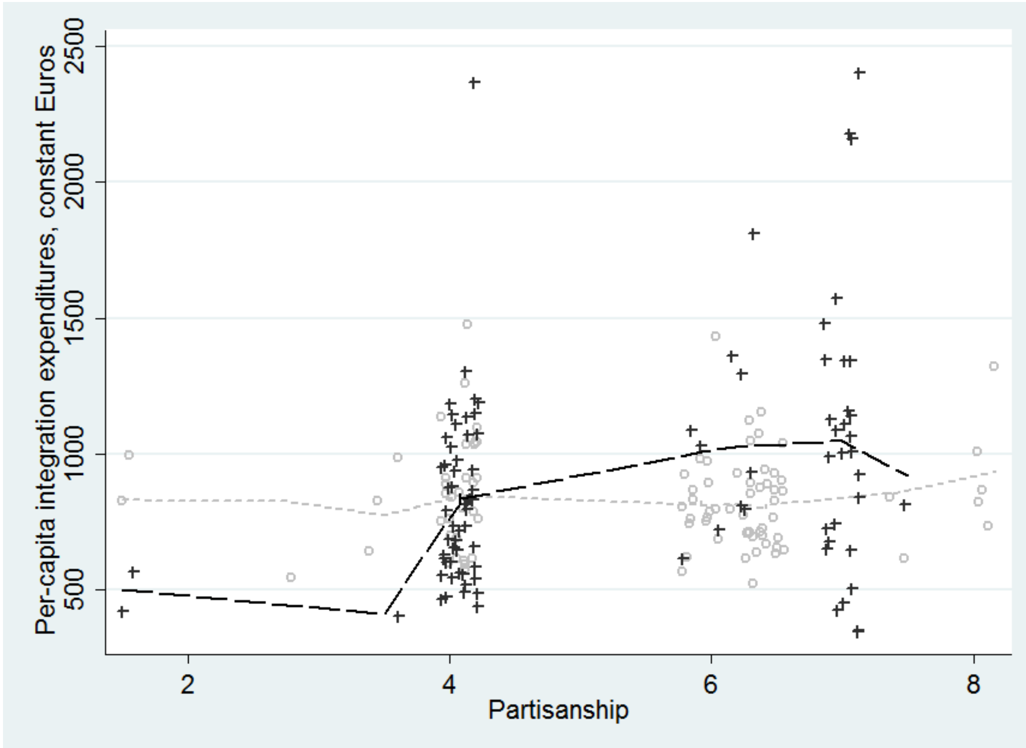


Figure 13: Integration spending per capita by department and partisan orientation in 2003 (light grey, circles and fine dashed line) and 2008 (black, crosses and long-dashed line). Higher values on the X axis indicate more conservative departmental executives. Lowess estimate, observations jittered to make distribution visible.

To more rigorously assess the role of these two factors in the integration spending divergence, I run regression models of yearly differences in integration spending on explanatory variables including year fixed effects, with and without additional control variables. The main independent variables of interest are the interaction of a post-reform dummy with the partisanship variable and the budget variable. Since partisanship varies over time, I can also control for partisanship separately (Column 2). As Table 4 displays, the budget variable is positive and significant in all specifications. However, the coefficient on partisan orientation, while significantly different from zero in the simplest specification, is slightly attenuated, and its standard error increases, when adding control variables.

Table 4: Regression of yearly differences in integration spending per capita on partisanship and budget indicators switched on after 2004 and control variables, per department, standard errors clustered at the department level.

	Integration spending	Integration spending	Integration spending
Partisan*post-reform	13.745	12.415	12.643
	(6.549)**	(7.892)	(7.911)
Budget*post-reform	0.227	0.227	0.203
	(0.119)*	(0.119)*	(0.121)*
Partisan		1.330	2.135
		(3.511)	(3.598)
Unemployment			-0.931
			(2.296)
No. of recipients			0.001
			(0.000)
Migrant share			26.155
			(84.761)
Population density			-0.001
			(0.001)
_cons	-17.846	-25.308	-27.396
	(14.256)	(26.821)	(40.313)
R^2	0.05	0.05	0.05
N	1,123	1,123	1,123

* p<0.1; ** p<0.05; *** p<0.01

4.5. Survey of RMI recipients from Ministry of Social Affairs

Finally, I use survey data to assess the implementation differences engendered by decentralization from the perspective of recipients.[11] The survey data was collected by the research directorate of the French Ministry of Social Affairs in 2003 and 2006, with some overlapping questions allowing to look at changes within locales. Interviews were conducted by drawing recipients from a national administrative dataset of individuals who had received the RMI about 14 months before. This has the advantage that one can study not only RMI recipients, but also former recipients who left the program in the months prior to the survey. Interviews were conducted face-to-face during home visits lasting about an hour. 2,000 and 3,600 RMI recipients were surveyed in 2003 and 2006, respectively. The clustered sampling strategy used is quite complex (see Appendix section 7.5.), and results in repeated survey observations from about 40 departments in 2003 and 2006.

I use responses to questions tapping the dimensions of access, conditionality and integration effort asked in both surveys, as well as responses to some additional questions that were only asked in 2006. All analyses use department-level means, not the underlying observations, meaning standard errors are conservative. Throughout, I weigh by the number of respondents in each department to reflect differences in variability of the estimated departmental means.

11 Although the survey data are noisy and imperfect, this is an important complement to the administrative data, partly since the categories used in the CAF survey are inherently contested and political. For example, when CAF administrators responded that integration benefits “improved from the standpoint of recipients”, it is far from clear whether the recipients concerned would agree with this characterization, a complication that the technical-bureaucratic language used within the bureaucracy tends to obscure. That being said, the social context of a survey of underprivileged benefit recipients by a government Ministry is also likely to generate artefacts of this kind (Bourdieu (1984)).

I first use the data from survey to display aggregate developments in four variables of interest: the fraction of RMI recipients (1) working, (2) having signed an integration contract, (3) perceiving the RMI as devaluing, (4) leaving the RMI.

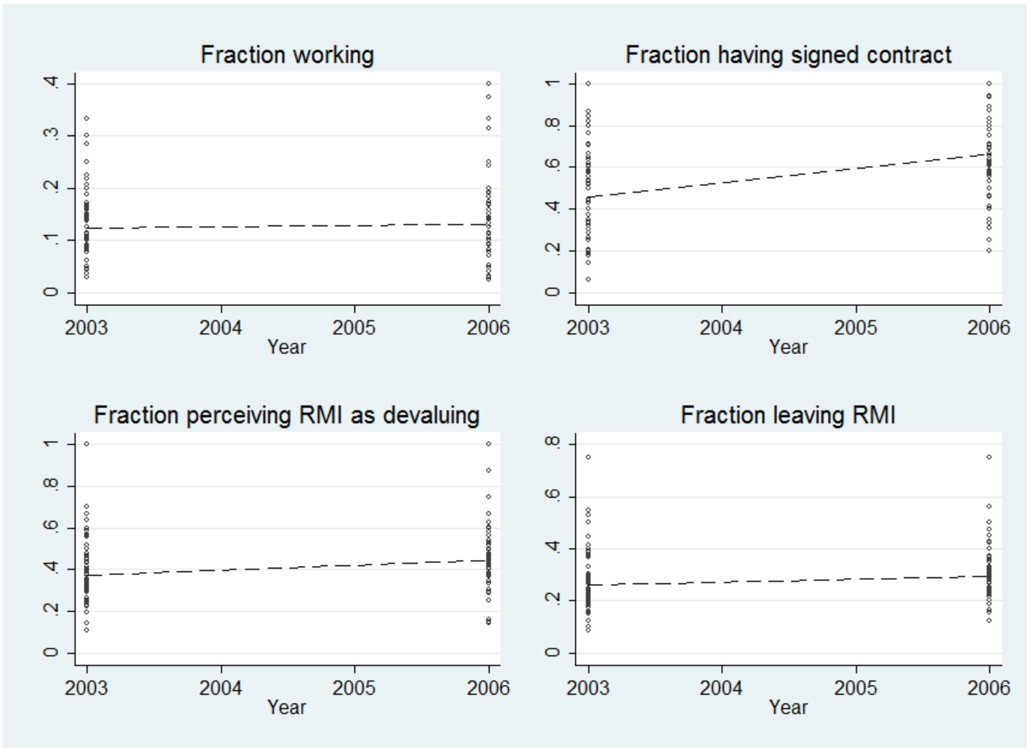


Figure 14: Fraction of RMI recipients working, having signed an integration contract, perceiving policy as devaluing, and leaving the program, in 2003 and 2006, at department level, weighted by number of respondents

The share of respondents having signed an integration contract increased quite strongly following decentralization, showing that at least on paper, departments indeed increased the focus on recipient's responsibility and integration goals. The share of respondents working, feeling that the RMI was devaluing, and leaving the RMI all increased very slightly in the 3 years between the surveys.

How did policy choices made in the departments affect substantive outcomes for recipients? Recall the argument above that departments that were able to increase their integration budget did so primarily with a view to advancing labor market integration, often through subsidized employment. If this is true, changes in integration budgets should correlate with changes in the employment of recipients. The left panel of Figure 15 shows that this appears to be the case in the survey data; integration budget expansion is mostly positively related to changes in the employment of recipients. However, as Tables 5 and 6 show, this relationship is not statistically significant.

As regards the success of integration budget increases in generating exits from the RMI into the labor market, the right-hand panel suggests that the departments that spent more on integration did not experience more frequent exits from the RMI than departments that cut their integration expenditure. This is consistent with evaluations showing that RMI labor market integration programs are not effective at increasing the chances of recipients in the labor market (Zoyem (2001)).

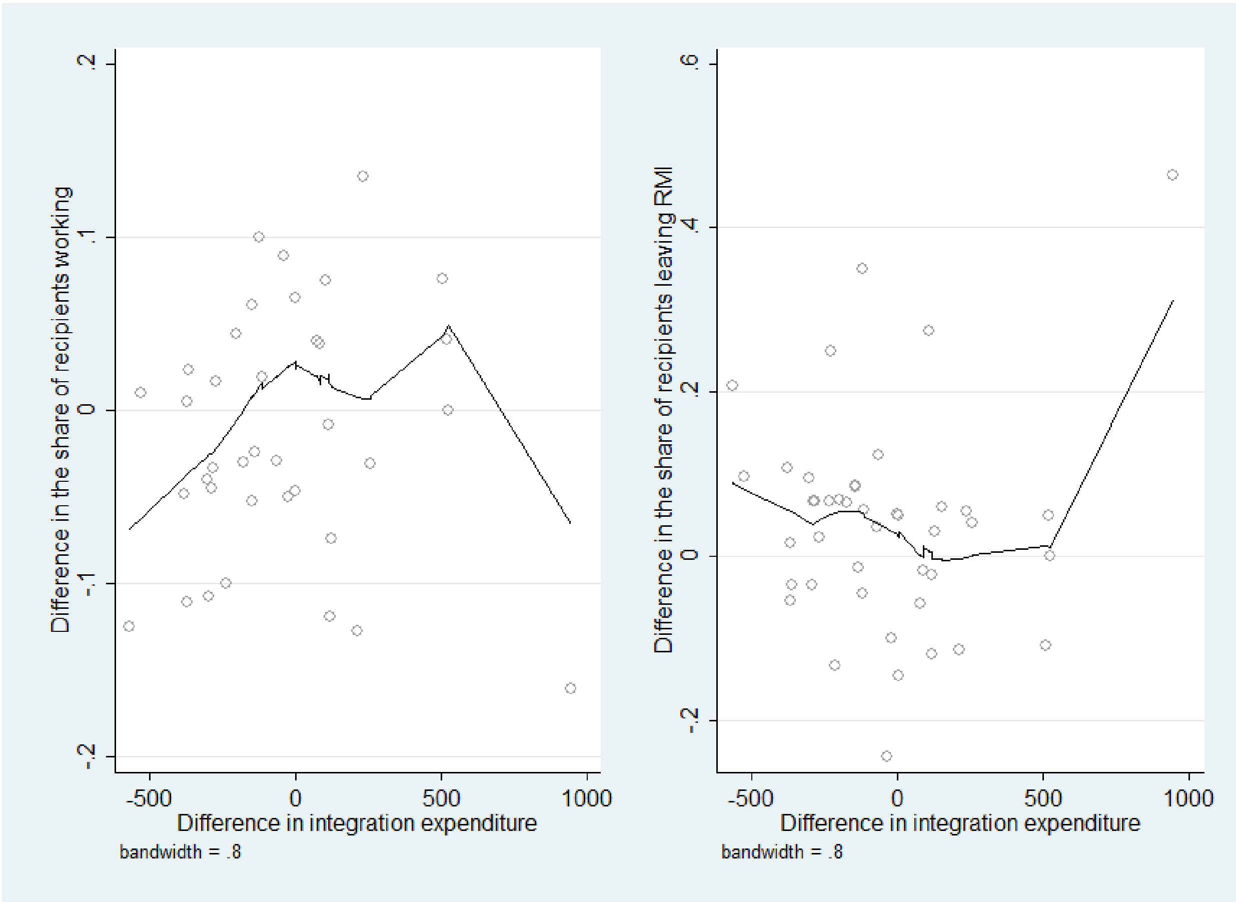


Figure 15: Change (2003-2006) in per-capita integration spending and change in fraction of RMI recipients working (left) and change in fraction of respondents having left the RMI (right). Differences at department level, weighted by respondents.

Table 5: Regressions of 2003-2006 differences in survey indicators on integration spending changes and budgetary and partisan variables without controls.

	Dif – Fraction working	Dif – Fraction having left RMI	Dif – Fraction working	Dif – Fraction having left RMI	Dif – Fraction considering RMI devaluing	Dif – Fraction signed integration contract
Dif – integration exp.	0.000	0.000				
	(0.000)	(0.000)				
Partisan			-0.007	0.008	-0.009	0.028
			(0.007)	(0.013)	(0.024)	(0.021)
Budget			-0.000	-0.000	-0.000	0.001
			(0.000)	(0.000)	(0.000)	(0.000)*
_cons	-0.014	0.041	0.020	-0.003	0.117	0.022
	(0.013)	(0.023)*	(0.035)	(0.063)	(0.111)	(0.102)
R^2	0.00	0.01	0.03	0.03	0.02	0.10
N	37	43	36	42	37	40

* p<0.1; ** p<0.05; *** p<0.01

Table 6: Regressions of 2003-2006 differences in survey indicators on integration spending changes and budgetary and partisan variables with controls.

	Dif – Fraction working	Dif – Fraction having left RMI	Dif – Fraction considering RMI devaluing	Dif – Fraction signed integration contract
Dif – integr. exp.	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Partisan	-0.006	0.005	-0.007	0.036
	(0.009)	(0.015)	(0.023)	(0.025)
Budget	-0.000	-0.000	-0.001	0.000
	(0.000)	(0.000)	(0.001)**	(0.000)
Dif – Unemp.	-0.071	0.025	-0.035	-0.020
	(0.042)	(0.069)	(0.102)	(0.084)
Dif – No. rec.	0.000	-0.000	-0.000	0.000
	(0.000)**	(0.000)	(0.000)	(0.000)
Migrant share	0.083	0.222	-0.501	1.393
	(0.299)	(0.440)	(0.744)	(0.962)
Population dens.	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)**	(0.000)
_cons	0.056	-0.016	0.166	-0.121
	(0.059)	(0.094)	(0.169)	(0.156)
R^2	0.19	0.04	0.24	0.22
N	36	42	37	40

* p<0.1; ** p<0.05; *** p<0.01

Some very interesting questions pertaining to the perception of the RMI by recipients and the implementation policy of the departments were only asked in 2006. I document results from these questions below, noting that results should be interpreted cautiously given the data's purely cross-sectional nature.

First, respondents were asked whether they were threatened with sanctions and payment suspensions, either due to their lack of a signed contract or due to their non-compliance with the contract. This is a very direct measure of one aspect of the conditionality of RMI administration. As Figure 16 shows, sanctioning threats as perceived by recipients appear strongly related both with partisanship (with conservative department executives threatening more) and with budget constraints (with more constrained department executives threatening more).

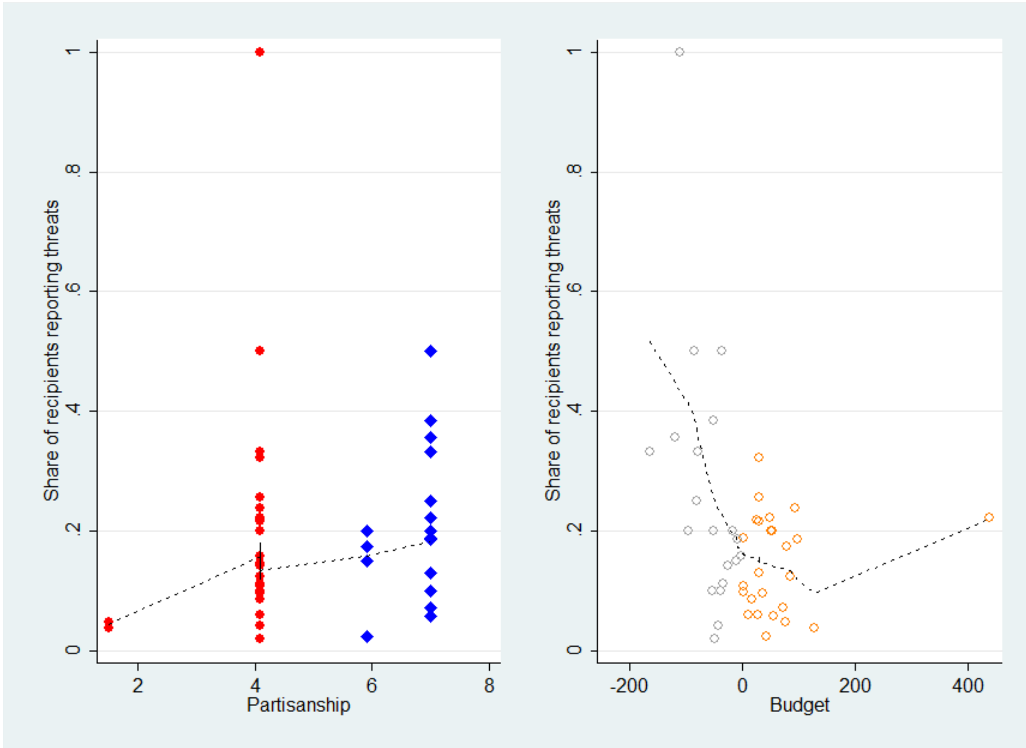


Figure 16: Partisanship of the department council President (left), budgetary situation of the department (right) and the share of recipients that report having been threatened with sanctions during their interaction with social policy officers. Higher values on the Y axis indicate more frequent threats.

A similar question posed in 2006 asks recipients whether they think that what is demanded of them in terms of integration efforts and the integration contract in return for their receipt of the RMI cash benefit is adequate or exaggerated. The pattern obtained is similar to the one for administration threats, with "exaggerated" demands more likely to be reported in conservatively governed and financially constrained departments (Figure 17). Finally, I also use the survey data to construct a measure of how frequently recipients visit the welfare office per year on average.

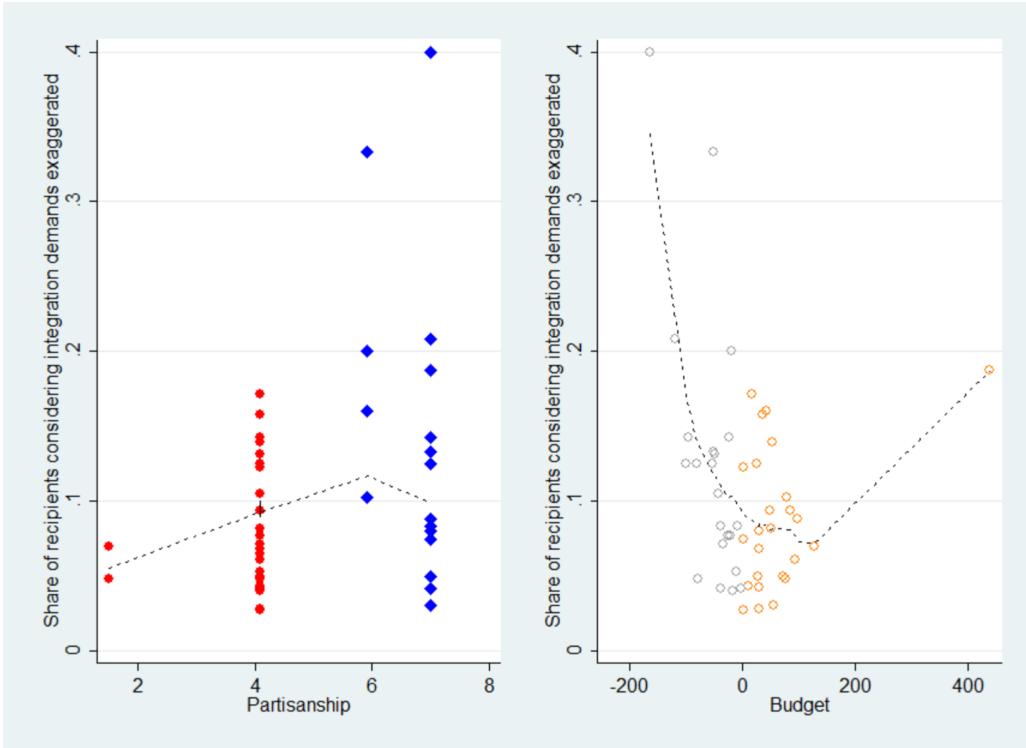


Figure 17: Partisanship of the department council President (left), budgetary situation of the department (right) and the share of recipients that report that demands made on their integration efforts by the administration are "exaggerated".

Table 7 presents the regression results for these cross-sectional data with and without control variables. Without control variables, only the effect of partisanship on the share of recipients reporting exaggerated demands from the administration is significant at conventional levels. This point estimate is robust to adding controls. In estimations with controls, there is also evidence that sanctions threats and excessive demands respond to budgetary conditions (in the expected direction), and that office visits are more frequent in conservatively governed departments.

Table 7: Cross-sectional regressions of questions asked only in 2006, with and without controls.

	Fraction threatened with sanctions	Fraction considering demands excessive	Frequency of office visits per year	Fraction threatened with sanctions	Fraction considering demands excessive	Frequency of office visits per year
Partisan	0.011	0.015	0.204	0.004	0.015	0.271
	(0.015)	(0.007)**	(0.149)	(0.017)	(0.007)**	(0.162)*
Budget	-0.001	-0.000	0.002	-0.001	-0.000	-0.002
	(0.000)	(0.000)	(0.002)	(0.000)*	(0.000)*	(0.003)
Unemployment				-0.023	-0.011	-0.071
				(0.018)	(0.008)	(0.123)
No. of recipients				-0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)
Migrant share				0.024	-0.491	11.536
				(0.423)	(0.180)***	(3.165)***
Population dens.				0.000	0.000	0.000
				(0.000)	(0.000)***	(0.000)
_cons	0.150	0.035	4.413	0.378	0.143	3.786
	(0.088)*	(0.030)	(0.740)***	(0.217)*	(0.080)*	(1.510)**
R^2	0.11	0.14	0.05	0.20	0.34	0.15
N	45	47	58	45	47	58

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

5. Conclusion

This study presents first systematic evidence that RMI decentralization led to the kind of "subterranean" policy changes described by Hacker (2004), with important repercussions for recipients and the degree of decommodification achieved by the policy, but also for the homogeneity of policy towards citizens across the French territory.

As the study shows, the implementation of the RMI became considerably more heterogeneous across French departments, and some of these divergences are related to two key aspects of the departmental political environment: variations in budgetary pressure and partisanship of the departmental executive in charge of implementation. Specifically, I find evidence that changes in control regimes, integration demands made by authorities and overall access to benefits are associated with partisanship, and that integration spending is associated with the departmental budget's condition. The findings regarding sanction use are mixed, with the more credible evidence suggesting that sanction use is related to budgetary conditions.

These findings have several implications for the study of social policy in advanced democracies.

First, by documenting political sources of variation in social policy implementation within a nominally national policy, the study suggests that theoretical explanations frequently hypothesized to account for policy formulation may usefully be extended to questions of implementation. To clarify this argument it may be helpful to contrast the findings obtained here with a literature on U.S. state politics which finds that state government partisanship predicts how states design welfare programs that they have jurisdiction over. For example, Hill, Leighley and Hinton-Andersson (1994) show that states with higher working-class mobilization and Democratic political control tended to have higher AFDC benefits between 1978 and 1990. In a similar analysis, Barilleaux and Miller (1988) found partisan differences in state medicaid policy.[12] The crucial difference between these studies and the present one is that U.S. state legislators are legally and de facto able to set policy by writing laws defining eligibility conditions and benefit amounts. By contrast, the variation in policy outcomes analyzed here is is due only to varying implementation of the very same program; nevertheless, the patterns obtained are broadly similar to those found in institutional contexts where subnational governments are formally enabled to define policy.

12 Another example concerns adoptions of Medicare expansion at the state level following passage of the Affordable Care Act, which are also known to follow a strongly partisan pattern.

Second, the study sheds new light on the sources of political support for active labor market policies. A growing body of research has begun investigating the political determinants of active labor market policy (ALMP) effort, using data on the share of GDP spent on ALMP measures at the country level. This literature has produced both evidence that left-wing governments spend more on ALMP policy (Huo Nelson and Stephens (2008)), and evidence that partisanship is not associated with ALMP (Rueda (2006)). While the former set of authors claim that the significant association shows that left-wing governments desire high labor market participation and low unemployment, Rueda (2006) argues that social democratic parties have incentives to focus on the constituency of employed and protected 'insiders' instead of spending public funds on outsiders, explaining his finding of no association. One of the major problems with interpreting the results from both sets of studies is that they compare aggregate national spending on programs which are extremely heterogeneous (Bonoli (2010)). Another perhaps even more fundamental problem is that spending on any budget category, like ALMP, is mechanically linked to spending on other budget categories through the government's budget constraint. Thus, the finding of no partisan effects on ALMP in Rueda (2006) may be explained by his insider outsider theory of social democracy ("social democrats do not want to help outsiders"), or alternatively it may be the case that social democrats choose to expand targeted cash benefits rather than active labor market policies when in power ("social democrats do help outsiders, but use other policies"). The decentralization of the RMI allows us to look in more detail at what exactly partisan executives do when faced with trade-offs between different components of social policy. The evidence described above suggests that in the French case, left-wing executives actually decrease spending on the ALMP component of the RMI following decentralization relative to right-wing executives, instead focusing on keeping cash benefits accessible. This is perfectly consistent with partisanship theory once trade-offs between multiple policy goals are acknowledged, but shows that attempts to statistically explain cross-national variation in "ALMP spending" without considering the alternative uses of public money can be highly misleading.

Third, politically induced variation in implementation could help explain the puzzle of low and regionally heterogeneous social policy take-up. Research examining welfare programs in different contexts has found somewhat surprisingly that very large fractions, often around 50%, of the population eligible for benefits do not actually receive any (Currie (2004), van Oorschot (1991)). Moreover, take-up rates are frequently very heterogeneous across regions and over time within the same policy for reasons that are not fully understood (Blank and Card (1991)). The present study suggests that by driving differentials in implementation, political and budgetary factors may explain regional variation in take-up within the same policy.

One important caveat accompanying the findings is that it is difficult to empirically disentangle the various aspects of local political environments. For example, while I focus on the partisanship of executives making implementation decisions, it is also plausible that executives respond to the ideological position of their local constituencies when making these decisions. Separating out these two mechanisms, which are not necessarily mutually exclusive, is empirically challenging in the context of this study given data constraints and the scarcity of partisanship switches after decentralization reform. Further research could use more advantageous institutional settings to isolate the precise mechanisms linking local political environments to implementation decisions.

6. References

Allan, J. P., and Scruggs, L. (2004). Political partisanship and welfare state reform in advanced industrial societies. *American Journal of Political Science*, 48(3), 496-512.

Avenel, C. (2007). Les conséquences de la décentralisation sur la gestion du RMI [The consequences of decentralization for RMI administration]. *Recherches et Prévisions*, 87(1), 25-37.

Bakker, R., de Vries, C., Edwards, E., Hooghe, L., Jolly, S. Marks, G. Polk, J. Rovny, J. Steenbergen, M., Vachudova, M. (2015). Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999-2010. *Party Politics*, 21(1), 143-152.

Barilleaux, C. and Miller, M. (1988). The political economy of state medicaid policy. *American Political Science Review*, 82(4), 1089-1107.

Blank, R. and Card, D. (1991). Recent trends in insured and uninsured unemployment: is there an explanation? *Quarterly Journal of Economics*, 106(4), 1157-1189.

Bonoli, G. (2010). The political economy of active labor-market policy. *Politics & Society*, 38(4), 435-457.

Bouchoux, J., Houzel, Y., and Outin, J. L. (2004). Revenu minimum d'insertion et transitions: une analyse des inégalités territoriales [Minimum integration revenue and transitions: An analysis of regional inequalities]. *Revue française des affaires sociales*, 4(4), 105-132.

Bourdieu, P. (1984). L'opinion publique n'existe pas [Public opinion does not exist]. In *Questions de sociologie* (pp. 222-235) [Sociology in question]. Paris: Les Éditions de Minuit.

Chemin, M., and Wasmer, E. (2009). Regional difference-in-differences in France using the German annexation of Alsace-Moselle in 1870-1918. In J. Frankel, and C. Pissarides (Eds.), *NBER International Seminar on Macroeconomics 2008* (pp. 285-305). Chicago, IL: University of Chicago Press.

Cour des Comptes (2011a). *Du RMI au RSA: La difficile organisation de l'insertion; Constats et bonnes pratiques* [From RMI to RSA: the difficult organization of integration; diagnoses and best practices]. Public Report.
URL: http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/114000403.pdf

Cour des Comptes (2011b). *Rapport d'observations definitives: RMI/RSA, Seine Saint Denis* [Final report: RMI/RSA, Seine Saint Denis]. Public Report. URL:
https://www.ccomptes.fr/content/download/24096/386750/version/file/IFR2011-32.pdf

Cour des Comptes (2011c). *Rapport d'observations definitives: RMI/RSA, Yvelines* [Final report: RMI/RSA, Yvelines]. Public Report. URL:
https://www.ccomptes.fr/content/download/23981/384955/version/file/IFR201010.pdf

Currie, J. (2004). *The Take Up of Social Benefits*. National Bureau of Economic Research Working Paper 10488. URL: http://www.nber.org/papers/w10488.pdf

Destremau, B. and Messu, M. (2008). Le droit à l'assistance sociale à l'épreuve du local [The right to social assistance tested by the local]. *Revue française de science politique*, 58, 713-742.

Donier, V. (2010). Garantir les droits sociaux dans le cadre de la décentralisation [Guaranteeing social rights in the framework of decentralization]. *Informations sociales*, 162, 108-116.

Dubois, V. (2012). *Le role des street-level bureaucrats dans la conduite de l'action publique en France* [The role of street-level bureaucrats in French public policy]. Unpublished manuscript. HAL Open Archives, URL: https://halshs.archives-ouvertes.fr/halshs-00660673

Duguet, E., Goujard, A., and L'Horty, Y. (2009). Geographic inequality of access to employment in France: an investigation based on comprehensive administrative sources. *Economie & Statistique*, 415, 17-44.

Duvoux, N. (2011). Le rmi et les dérives de la contractualisation [The RMI and the outgrowth of decentralization]. In S. Paugam (Ed.), *Repenser la solidarité* (pp. 451-472). Paris: Presses Universitaires de France.

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism.* Princeton, NJ: Princeton University Press.

Eydoux, A. and Tuchszirer, C. (2010). *Du RMI au RSA : les inflexions de la solidarité et de la gouvernance des politiques d'insertion* [From RMI to RSA: the bending of solidarity and of integration policy governance]. Working Paper, Centre d'etudes de l'emploi. URL: http://www.cee-recherche.fr/sites/default/files/webfm/publications/docdetravail/134-rmi-rsa-inflexions-solidarite-gouvernance-politiques-insertion.pdf

Ferreira, F., and Gyourko, J. (2009). Do Political Parties Matter? Evidence from US Cities. *The Quarterly Journal of Economics*, 124(1), 399-422.

Folke, O. (2014). Shades of brown and green: party effects in proportional election systems. *Journal of the European Economic Association*, 12(5), 1361-1395.

Gingrich, J., and Häusermann, S. (2015). The decline of the working-class vote, the reconfiguration of the welfare support coalition and consequences for the welfare state. *Journal of European Social Policy*, 25(1), 50-75.

Hacker, J. (2004). Privatizing Risk without Privatizing the Welfare State: The Hidden Politics of Social Policy Retrenchment in the United States. *American Political Science Review*, 98(2), 244-260.

Hill, K, Leighley, J. and Hinton-Anderson, A. (1995). Lower-Class Mobilization and Policy Linkages in the U.S. States. *American Journal of Political Science*, 39, 75-86.

Huo, J., Nelson, M. and Stephens, J. (2008). Decommodification and activation in social policy: resolving the paradox. *Journal of European Social Policy*, 18(1), 15-20.

Inspection générale des affaires sociales (2007). *Rapport de synthèse sur la gestion du revenu minimum d'insertion (RMI)* [Synthetic report about the administration of the minimum integration income (RMI)]. Public Report. URL: http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/074000760/0000.pdf

Iversen, T., and Stephens, J. D. (2008). Partisan politics, the welfare state, and three worlds of human capital formation. *Comparative Political Studies*, 41(4-5), 600-637.

Keiser, L. R., Mueser, P. R., and Choi, S. W. (2004). Race, bureaucratic discretion, and the implementation of welfare reform. *American Journal of Political Science*, 48(2), 314-327.

Korpi, W., and Palme, J. (1998). The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries. *American Sociological Review*, 63(5), 661-687.

Kuhlmann, S. Bogumil, J., Ebinger, F. Grohs, S. and Reiter, R. (2011). *Dezentralisierung des Staates in Europa. Auswirkungen auf die kommunale Aufgabenerfüllung in Deutschland, Frankreich und Großbritannien* [Decentralization of the state in Europa: Effects on local task fulfillment in Germany, France and the United Kingdom]. Wiesbaden, Germany: VS-Verlag.

Lafore, R. (2003). Le contrat dans la protection sociale [Contracts in social protection]. *Droit social*, (1), 105-114.

Le Lidec, P. (2011). Les relations financières entre l'État et les collectivités territoriales: un sauvetage des conseils généraux orchestré au prix fort [Financial relations between the national state and the subnational entities: a rescue of general councils, engineered at a high price]. *Informations sociales*, 162(6), 32-40.

Leigh, A. (2008). Estimating the impact of gubernatorial partisanship on policy settings and economic outcomes: A regression discontinuity approach. *European Journal of Political Economy*, 24(1), 256-268.

Lipsky, M. (1980). *Street-level Bureaucracy: Dilemmas of the Individual in Public Services*. New York, NY: Russell Sage Foundation, 1980.

Nygård, M. (2006). Welfare-Ideological Change in Scandinavia: A Comparative Analysis of Partisan Welfare State Positions in Four Nordic Countries, 1970–2003. *Scandinavian Political Studies*, 29(4), 356-385.

Palier, B. (2000). Defrosting the French Welfare State. *West European Politics*, 23(2), 113-136.

Palier, B. and Thelen, K. (2010). Institutionalizing Dualism: Complementarities and Change in France and Germany. *Politics & Society*, 38(1), 119-148.

Pepin, J.-L. and Blandin, O. (2007). La décentralisation du RMI : quels effets, quels enjeux pour les CAF? [The decentralization of the RMI: What effects, what stakes for the CAF?]. *Recherches et Prévisions*, 88, 71-81.

Pettersson-Lidbom, P. (2008). Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association*, 6(5), 1037-1056.

Pierson, P. (1998). Irresistible forces, immovable objects: post-industrial welfare states confront permanent austerity. *Journal of European public policy*, 5(4), 539-560.

Pierson, P. (2001). *The new politics of the welfare state*. Oxford: Oxford University Press.

Poutingon, P.M. (2012). *La décentralisation des politiques sociales est-elle conciliable avec l'exercice de la solidarité nationale ? L'exemple du revenu minimum d'insertion* [Can the decentralization of social programs be reconciled with the exercise of national solidarity? The example of the minimum integration income], Unpublished M.A Thesis, ETSUP (École Supérieure de Travail Social).
URL: http://cediasbibli.org/opac/doc_num.php?explnum_id=1597

Rueda, D. (2006). Social Democracy and Active Labour-Market Policies: Insiders, Outsiders and the Politics of Employment Promotion. *British Journal of Political Science*, 36, 385-406.

Sénat francais (2005). *Rapport d'information fait au nom de l'Observatoire de la décentralisation sur la décentralisation du revenu minimum d'insertion (RMI) par M. Michel Mercier* [Informational Report done in the name of the Decentralization Observatory about the decentralization of the minimum integration income (RMI) by Mr. Michel Mercier]. Public Report. URL: http://www.senat.fr/rap/r04-316/r04-316.html

Slapin, J. B., and Proksch, S. O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3), 705-722.

Soss, J., Fording, R. and Schram, S. (2011). *Disciplining the poor: Neoliberal paternalism and the persistent power of race*. Chicago, IL: University of Chicago Press.

Thierry, M. (2008). La décentralisation du RMI: responsabiliser les départements, mais pas seulement les départements [The decentralization of the RMI: holding departments accountable, but not just the departments]. In M. Lelièvre and E. Nauze-Fichet (Eds.), *RMI, l'état des lieux* (pp. 243-248) [RMI: the state of affairs]. Paris: La Découverte (Recherches).

Tuchszirer, C. and Join-Lambert, M. (2003). Décentralisation du RMI et du RMA: les risques d'une réforme [Decentralization of the RMI and the RMA: the risks of a reform]. *L'Économie politique*, 19, 25-32.

van Oorschot, W. (1991). Non-Take-Up of Social Security Benefits in Europe. *Journal of European Social Policy*, 1(1), 15-30.

Zoyem, J. P. (2001). Contrats d'insertion et sortie du RMI: Évaluation des effets d’une politique sociale [Integration contracts and exits from the RMI: Evaluation of the effects of a social program]. *Economie et statistique*, 346(1), 75-102.

7. Appendix

7.1. Descriptive statistics

Table 8: Descriptive statistics.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
CAF (2006)					
Change in suspensions	91	0.302	0.537	-1	1
Change in deletions	88	0.301	0.506	-1	1
Change in controls	92	0.598	0.639	0	2
Change in access to benefits	91	0.0220	0.601	-1	1
Change in benefit mgmt	91	0	0.641	-1	1
Change in integration prog.	53	0.557	0.718	-1	1
Audit Court (2004-2007)					
Sanction intensity	221	0.0450	0.0458	0.000216	0.147
Integration program (Ministry, 1996-2008)					
Per-capita spending on integration program	1,222	805.5	229.1	224.1	2,413
Ministry Survey (DREES, 2003 & 2006)					
Dif. Recipients working	37	-0.0142	0.0717	-0.161	0.135
Dif. Contract signed	41	0.173	0.228	-0.321	0.635
Dif. Left RMI	43	0.0394	0.131	-0.244	0.464
Dif. Thinks policy devaluing	38	0.0620	0.195	-0.417	0.732
Frequency of visits ('06)	767	5.454	1.530	1	10
Adequacy of demands ('06)	624	0.105	0.0720	0.0270	0.400
Threats of sanctions ('06)	598	0.197	0.165	0.0192	1
Independent variables (1996-2008)					
Partisanship	1,222	5.436	1.404	1.500	8.170
Budgetary situation	1,209	0.624	84.55	-234	440
No. of recipients	1,222	10,410	12,114	450	76,363
Unemployment	1,222	8.298	2.078	3.700	15.50
Migrant share	1,222	0.0744	0.0462	0.0196	0.299
Population Density	1,222	564.4	2,513	15	21,602

7.2. Partisanship scores

To score the partisanship of parties, I use 2002 the economic policy dimension ratings from the Chapel Hill Expert Survey. The few parties that are not rated by the dataset in 2002 receive, in this order:

1) their rating from the closest possible time period outside 2002

2) the rating of related parties, for parties that developed from or into other parties over time

3) the average rating of parties of their leaning (left/right), for the few parties where 1 and 2 are infeasible.

The resulting ratings are: PCF 1.05, PS 4.08 , PRG 3.5, VERTS 2.75, RPR 5.92, FN 6, UDF 5.92, MoDem 5.92, DL 8.17, UMP 7, MPF 7.5, PR 8.17, RPF 7.42, MDC 2.75, RPR 6.42, MRG 5.41, DVG 3.58, RDG 3.58, DVD 6.18, MLM 6.18, Mouvement libéral et modéré 6.18

7.3. Representativeness of the audited departments

Unfortunately, no information could be obtained about the procedure by which the 17 departments audited by the Public Audit Court were selected. In order to get a sense of the representativeness of the audited departments, I conducted simple t-tests on political and economic covariates across the two samples. As can be seen in Table 9, audited departments do at not appear to be very different from non-audited departments on observed covariates. Importantly, divergences on the budget and partisan variables are substantively small. However, standard errors are not small enough to rule out meaningful divergence in terms of the number of recipients per department: auditors may plausibly have focused their efforts on departments with a relatively large RMI population.

Table 9: Observable differences between audited and non-audited departments.

	Mean in audited departments (s.e.)	Mean in non-audited departments (s.e.)	Difference (s.e.)
Partisanship (04-07)	5.25	5.36	0.11
	(0.39)	(0.16)	(0.42)
Budget	14.17	-2.41	-16.58
	(17.43)	(10.07)	(20.13)
Number of recipients	14248	9864	-4384
	(4613)	(1190)	(4765)
Unemployment	7.58	7.93	0.35
	(0.47)	(0.17)	(0.51)
Migrant share	0.08	0.07	-0.01
	(0.01)	(0.00)	(0.02)
Population Density	574	562	-12
	(375)	(307)	(485)

* p<0.1; ** p<0.05; *** p<0.01

7.4. The composition of integration spending

Throughout the paper, I use aggregate integration spending (per recipient) as an outcome variable since more detailed data on spending subcomponents are not available. However, for the subset of departments investigated by the Public Audit Court, a breakdown of integration spending by function (labor market, social, housing, health) is available between 2004 and 2008. These data come with the caveat that the allocation of spending on programs to these four areas is somewhat arbitrary, a fact which is criticized in multiple Audit Court reports. Inspecting these data suggests that the share of labor market integration spending in total integration spending is high and rising over time, namely from an average of 58% in 2004 to an average of 71% in 2008.

More importantly since I interpret increases in integration spending as increased effort to reintegrate recipients into the labor market, the data allows me to show that the share spent on labor market integration does not move inversely to the overall amount spent on integration, a possibility which would invalidate my interpretation. As Figure 18 shows, the opposite is the case: changes in overall per-recipient integration spending are, if anything, weakly positively correlated with the change in the share of this expenditure directed at labor market integration according to the Audit Court Report. While this is based only on the non-random subsample of departments that were audited, it strengthens the interpretation of spending changes as changes in labor market integration effort.

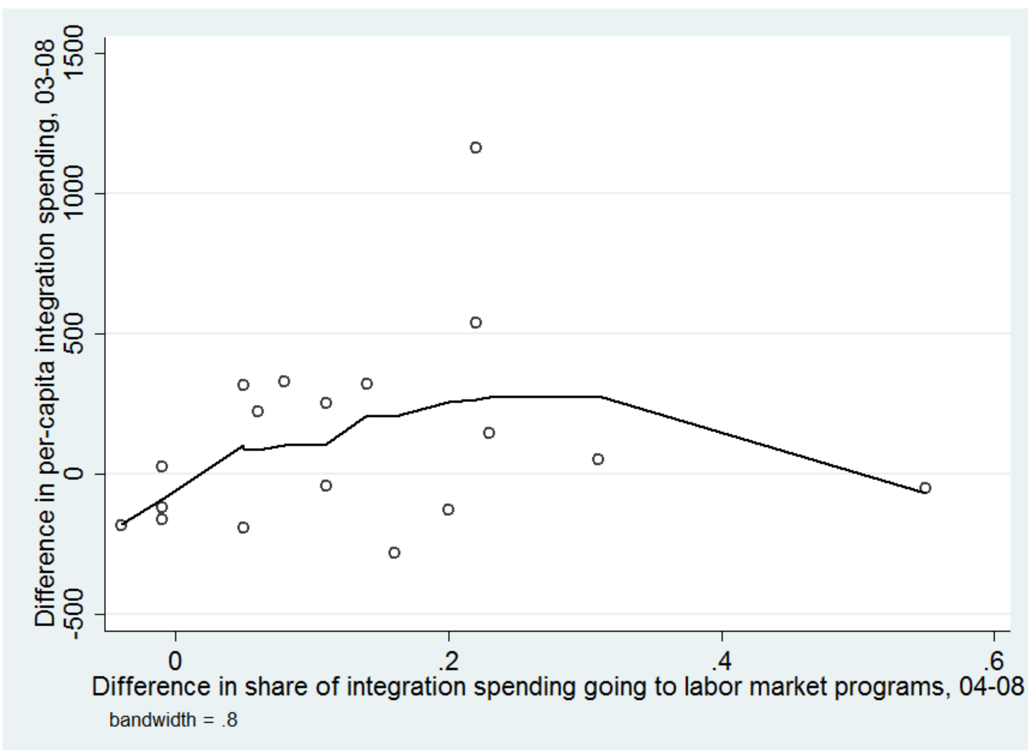


Figure 18: Change in share of integration spending on labor market integration (between 2004 and 2008, according to Audit Court data) and change in spending on integration (between 2003 and 2008, according to DREES data) at the department level for the subset of departments audited by the Audit Court.

7.5. Sampling strategy for DREES survey of RMI respondents

Summary translation of the survey sampling strategy employed by the DREES:

The sampling strategy for the 2003 and 2006 surveys involved three levels of stratification: departments, municipalities and individuals.

1. Departments were drawn from 5 strata: (1) the 8 departments of the Paris metro area, (2) the 6 departments with the highest total numbers of benefit recipients. Of these two strata, all departments were selected. Among (3) very urban departments with urbanization above 65% (4) medium-urbanized departments with urbanization between 50% and 65% and (5) rural departments with urbanization below 50%, 15 departments were picked within each stratum, proportionally to their size in terms of benefit recipients.

2. Within selected departments, municipalities are grouped in three strata: (1) an exhaustive stratum of the four largest municipalities in terms of benefit recipients (all four largest are chosen). (2) Those cantons (sub-departmental districts) or fractions thereof classified as "urban" by the INSEE (national statistical agency) (3) Those cantons or fractions thereof classified as "rural" by the INSEE.

For each departments, 6 fractions of cantons are selected, proportionally to the number of benefit recipients in them: the distribution of fractions of cantons between strata (2) and (3) varies with their relative importance in each department.

Paris is treated separately: 10 of 20 neighborhood districts (*arrondissements*) are chosen randomly with probabilities proportional to their size in terms of the number of benefit recipients: there is no exhaustive stratum.

3. Individuals are systematically sampled on the basis of a dataset sorted by age and gender, within each commune or fraction of canton that is selected.

III. BUREAUCRATIC DISCRIMINATION IN GERMANY (WITH A. RINK)[13]

Abstract

A growing experimental literature uses response rates to fictional requests to measure discrimination against ethnic minorities. This paper argues that restricting attention to response rates can lead to faulty inferences about substantive discrimination depending on how response dummies are correlated with other response characteristics. We illustrate the relevance of this problem by means of a conjoint experiment among all German welfare offices, in which we randomly varied five traits and designed requests to allow for a substantive coding of response quality. We find that response rates are statistically indistinguishable across treatment conditions. However, putative non-Germans receive responses of significantly lower quality, potentially deterring them from applying for benefits. We also find observational evidence suggesting that discrimination is more pronounced in welfare offices run by local governments than in those embedded in the national bureaucracy. We discuss implications for the study of equality in the public sphere.

13 Research for this paper was done jointly with Anselm Rink. We are grateful to Alex Coppock, Don Green, Jens Hainmüller, and Macartan Humphreys for help and feedback during different stages of the project. Insightful comments from seminar participants at WZB Berlin are also gratefully acknowledged. This study was approved by Columbia University's Institutional Review Board (AAAO5013) and pre-registered at EGAP (20150122AA).

1. Introduction

The equal application of the law to all citizens is a cornerstone of democratic political systems. Persistent findings of in-group favoritism, however, highlight that equality before the state does not arise naturally from interactions of democratically-minded citizens and state officials; it requires institutions to produce and maintain (Hewstone, Rubin and Willis (2002), Jolls and Sunstein (2006)). A long line of scholarship has ascribed modern state bureaucracies, characterized by their routinized and impersonal decision-making procedures, a central role in this process (Rothstein and Teorell (2008), Weber (2009), Rosanvallon (2011)). Yet, the extent to which modern bureaucracies succeed in treating all citizens fairly and equally is a question of significant debate.

A large body of social science research has produced convincing evidence that state officials exhibit significant partiality in spheres including legal bureaucracy (Mustard (2001), Rachlinski, Johnson, Wistrich and Guthrie (2009)), election administration (Atkeson, Bryant, Hall, Saunders and Alvarez (2010), Cobb, Greiner and Quinn (2010)), and human services bureaucracy (Lipsky (1980), Keiser, Mueser and Choi (2004)).[14] In order to isolate the causal link between ethnicity and bureaucratic behavior, recent scholarship has increasingly employed field experiments. Using fictitious information requests, studies have shown that putative members of historically disadvantaged groups are less likely to receive a response from state officials in settings ranging from U.S. states to South African and Chinese bureaucracies (Butler and Broockman (2011), Distelhorst and Hou (2014), Giulietti, Tonin and Vlassopoulos (2015)). Related studies show that in-group favoritism is present in both directions: white officials in both the United States and

14 A related but arguably distinct research agenda investigates discrimination on labor and housing markets and in civil society (Fix and Struyk (1993), Bertrand and Mullainathan (2004), Pager and Quillian (2005), Pager and Shepherd (2008), Humphreys, Fang and Guess (2014)).

South Africa are more likely to respond to requests from putative whites, while black politicians favor putative blacks (Broockman (2013), McClendon (forthcoming)).

The present study questions how informative response rate differentials are about substantive discrimination. Our argument is straightforward. The quality of a bureaucrat's response is inherently multidimensional — whether a response is received is only one dimension. Depending on how the multiple dimensions of quality are correlated with one another, restricting attention to response rates can produce misleading or outright false conclusions about discrimination. The bias can work in different directions: ignoring the content of a response may lead researchers to find "false null" results of no discrimination, despite the fact that discrimination is in fact present. This occurs, for example, when a minority applicant has the same likelihood of receiving a response, but is sent emails of substantively lower quality. Even more problematically, one group may receive more frequent but less helpful responses than another, falsely leading researchers to conclude that this group is being treated favorably when attention is restricted to response rates.

We illustrate the relevance of these considerations using a conjoint experiment involving all 408 German welfare offices. After pre-registering the design, hypotheses and measurement strategy at EGAP, we sent fictional requests to all offices, in which applicants inquired about the application process for cash benefits. The requests were designed to elicit responses that would be interpretable in terms of their helpfulness to the applicants. We randomly varied five traits including the putative requester's ethnicity. 321 responses were received, implying a response rate of 78 percent. Importantly, the response rates were virtually identically distributed across the treatment conditions.

However, it would have been premature to conclude that no discrimination occurred. In assessing the quality of the answers, we find that putative non-Germans (Turks and Romanians) received responses that were substantively inferior. Responses to non-Germans score 0.4 points lower on our 5-point quality measure — a 21 percent decline. Specifically, non-Germans were more likely to receive responses stating that applying for cash benefits was more administratively burdensome than it is according to national policy. The results hence demonstrate that non-Germans are not treated equally by the German social policy bureaucracy, though in dimensions that are more subtle than the response/non-response margin typically analyzed in correspondence trials.

We also assess the degree to which discrimination relates to an institutional feature of the German social policy bureaucracy, whereby a minority of offices is run not by the national welfare bureaucracy but by local governments. Observational evidence indicates that discrimination is more pronounced in these types of welfare offices, suggesting that bureaucratic centralization may be helpful in reducing inequality.

2. Multiple dimensions of discrimination

When applying for jobs, housing, or benefits, individuals prefer receiving a response to being ignored. Based on this logic, a growing literature uses response rates in correspondence trials to assess whether requesters of different backgrounds are being treated equally. In many cases, this empirical strategy is reasonable. When applying for a job, a callback signals employer interest. Following an ostensibly similar logic, researchers interested in discrimination by state officials have sent requests for information or help and adopted response rates as their primary outcome measure.

There are at least three reasons for this decision. At the level of the individual request, it is plausible that any response is preferable to no response (though exceptions are conceivable). Moreover, whether or not a response was received can be coded transparently, and with little effort involved. Finally, analyses of response rates do not face thorny missing data problems since non-response is a well-defined outcome. Taken together, these factors make response rates a simple and easily implementable measure of how willing state officials are to help applicants of different backgrounds.

However, the helpfulness of replies received is likely to vary significantly. In the studies cited above, Butler and Broockman (2011) sent emails in which citizens asked legislators for help in signing up to vote, while Distelhorst and Hou (2014) sent emails in which citizens requested information about a basic welfare program. Even with such simple requests, many kinds of outcomes are possible, ranging from simple responses asking requesters to come to the office to a detailed explanation of the program or procedure. Bureaucrats may even go so far as to solve the problem raised by requesters entirely.

This variation in response quality matters for the substantive question of whether some requesters receive better treatment than others. Good treatment is inherently multidimensional.

However, this variation is not considered when restricting attention to response rates. To see this, consider three stylized scenarios.

In the first scenario, assume that the probability of receiving a response is positively correlated with its latent quality. Here, our concern reduces to a measurement error problem. Bureaucrats are characterized by their “helpfulness”, which may or may not be a function of the ethnicity of an applicant. Bureaucrats only respond to queries if their helpfulness exceeds a given threshold. If a response is given, its quality is increasing in the bureaucrat’s helpfulness. If a subset of bureaucrats is more (or less) helpful to one ethnic group than another, this materializes as a response rate differential across the groups. The fact that quality is not fully proxied by the response dummy, then, induces measurement error. Importantly, however, differences in response rates *are* indicative of a helpfulness differential between ethnic groups for the average bureaucrat.

However, suppose in a second scenario that bureaucrats are obligated to respond to every incoming query. This could be the case, for example, if there is an effective monitoring system in place and bureaucrats are held accountable by their superiors. In this scenario response rates would be 100 percent across the board and hence uncorrelated with the underlying helpfulness of the bureaucrat. Yet, this does not necessarily imply that no discrimination takes place. Drafting responses and doing the requisite research requires effort. Bureaucrats might well be less likely to expend this effort when responding to one group rather than another. In a similar fashion, bureaucrats might consider some groups undeserving of help, and may therefore be less likely to write helpful answers (Van Oorschot (2006), Applebaum (2001)). To be plausible, this scenario requires response quality to be less easily monitored by the hierarchy than whether any response is sent. This is likely to be the case given that the monitoring problem faced by a hierarchy is the same faced by social scientists: to monitor quality, one would be required to read all messages, while monitoring response rates only requires measuring whether an email was sent. Thus, although an experiment would find identical response rates across groups, this would *not* be indicative of equal treatment, possibly producing a misleading study result.

Finally, consider a third scenario in which there are two types of bureaucrats: “racists” and “white guilt” administrators. “Racists” only respond to requests from putative majority requesters. “White guilt” administrators respond to everyone, but put extra effort into their responses to requesters from the minority group. In this scenario, minority requesters as a group get fewer, but qualitatively superior responses than requesters from the majority group. Across groups, response rates and response quality are inversely correlated. In this situation, a correspondence trial restricting attention to response rates would yield measurable discrimination effects. However, considering the divergence in response rates and response quality across groups, it would be difficult to determine which group is being discriminated against.

In short, focusing exclusively on response rates in correspondence trials may be a mere measurement imperfection without problematic implications for findings (scenario 1). However, it may also lead to unwarranted or at least misleading findings of no discrimination (scenario 2) or of discrimination (scenario 3). Many more scenarios are, of course, imaginable.

These arguments do not, in our view, render the literature on response rates obsolete. Its findings are empirically credible, especially when interpreted narrowly: as results about response rates. However, at a minimum, additional research assessing the quality dimension of bureaucratic responsiveness in correspondence trials seems warranted.

3. Experimental design

3.1. Setting

To move the study of discrimination beyond response rates, the present paper focuses on the German bureaucracy. The German case exhibits two main features that render it of particular interest in this context.

First, after a history of authoritarianism and genocide in the 20th century, social norms surrounding tolerance are by many accounts very strong in contemporary Germany (Art (2005)). Discrimination based on race, ethnic origin or religion is outlawed according to the "General Equal Treatment Law" (*Allgemeines Gleichbehandlungsgesetz*). Moreover, the German bureaucracy is widely considered to be highly professional and effective, owing partly to Max Weber's influential scholarship on bureaucratic institutions in Prussia (Mayntz (1965)). Whether these background conditions lead to impartial bureaucratic behavior is therefore of particular interest.

Second, contemporary Germany is increasingly ethnically diverse. Almost 20 percent of the population are either of non-German nationality or have a migrant background (i.e. at least one migrant parent) according to the 2011 Census. The recent influx of immigrants from countries outside of the European Union will further increase the diversity of the resident population. There is a lively academic debate about the consequences of this diversity and the role of prejudice in the German public sphere. On the one hand, mainstream political discourse has been increasingly accommodating over the past decades. Notably, the political interpretation of immigration as a short-term phenomenon has given way to a narrative of Germany as a "country of immigrants". Moreover, antiprejudice norms are judged to be quite strong by many observers (Blinder, Ford and Ivarsflaten (2013)). On the other hand, survey research shows that a significant fraction of the German population exhibit xenophobic biases. Krumpal (2012) shows that 27 percent of survey respondents agree that "Germany is dangerously swamped by foreign influences". This rate increases to 35 percent when a randomized response technique is employed to elicit truthful reporting. Consistent with this finding, experimental research the German labor and housing markets has found that immigrants face significant hurdles to equality (Kaas and Manger (2012), Schmid (2015)).

To our knowledge, no study has systematically examined discrimination against non-Germans by the German welfare state. This is surprising given that the receipt of welfare benefits by non-Germans is a politically salient and contested issue. As in other European countries, parties have campaigned on a platform of reducing welfare use by non-Germans, insinuating that migration is driven primarily by a desire to receive cash benefits (Thränhardt (1995), Spiegel (2014)).

Our study focuses on a component of the German welfare state that is particularly contested: the system of means-tested cash benefits frequently dubbed *Hartz 4*. These benefits, which are not contingent on previous contributions to the welfare system, are collected by 6 million individuals (about 7 percent of the population) and constitute the backbone of the German social safety net. In 2014, 1 out of 6 million recipients were non-German nationals.[15] Participating households receive, on average, 735 Euros ($840) per month in benefits. In order to receive benefits, individuals have to present themselves at local welfare offices called *Jobcenters* to apply. Continued receipt of benefits is conditional on compliance with rules designed to encourage labor market integration.

News reports have repeatedly suggested that non-Germans face discrimination in welfare offices: offices have inter alia been accused of withholding application documents from non-German nationals (Deutsche Welle (2014)). A study of three Berlin-area welfare offices by a private consulting company suggests that migrant applicants are singled out by more than 45 percent of welfare office staff as the most "problematic" group of clients (Stern, Wecking and Reinecke (2008)). At the same time, admittedly coarse aggregate data on benefit take-up rates do not suggest that eligible non-Germans are less likely to receive benefits than eligible Germans, as one might expect if welfare offices systematically discriminated against non-Germans (Castronova, Kayser, Frick and Wagner (2001), Bruckmeier and Wiemers (2012)). This finding of no take-up differential between natives and immigrants contrast with studies from other countries including the U.S., which have found immigrants to be less likely to take up benefits conditional on eligibility (Currie (2004)).

15 The largest groups are Turkish (350,000), Polish (69,000), Italian (62,000) and Iraqi (51,000) nationals.

3.2. Quality measure

In order to systematically study response quality, responses elicited by the request must be interpretable with little ambiguity and high transparency.[16] After initial research, we decided to base our instrument on two questions that measure the applicant's trustworthiness in the eyes of bureaucrats, and the effort bureaucrats are willing to expend to respond to the question.[17]

The first question was designed to induce the bureaucrat to state a legal fact about the application process to the applicant. Specifically, the applicant — having stated that he/she lives in a shared apartment with co-tenants who are his/her friends — asked whether paperwork from his/her co-tenants was required when submitting an application for benefits. Importantly, national legislation clearly states that documents about non-relatives who are co-tenants do not have to be submitted. Yet, applicants have an incentive to hide family or partner relationships with other members of their households to receive additional benefits. When preparing an application, bureaucrats therefore have to trust that the information applicants provide is accurate. The given answer can thus be interpreted as a signal about whether the bureaucrat deems the applicant trustworthy.

The second question was designed to measure the bureaucrat's effort. Specifically, the applicant asked what kind of paperwork was required in order to apply for benefits. The documents

16 In the pre-registration document, we postulated the following guideline for our instrument: "1. The request must induce meaningful variation within responses. Simple response rates, which are a common outcome in the existing literature, are problematic as the response margin may be the quality of the help given. For example, if a bureaucratic organization is legally bound to respond to every inquiry received, but individual case workers are prejudiced against minorities, this could lead to them answering all requests but treating them differentially in terms of the helpfulness of the response, or in terms of encouraging the applicant to apply. 2. The request must be realistic in order to measure the way that everyday requests are handled by the offices. 3. The request must be relatively simple to answer in order to avoid wasting the time of bureaucrats."

17 To ensure that the putative applicants would actually be eligible for benefits, rather than first receiving contributory unemployment insurance, we included a statement suggesting that the applicant had worked independently before, but that he/she had had to stop their independent activity because of a lack of business.

required can vary to some degree according to individual circumstances. Nevertheless, all applicants are required to produce a rental contract, documents on the non-rental cost of housing (heating, water, electricity bills, etc.), bank statements, their social security certificate, their health insurance card, proof of income for the past three months, as well as statements relating to ownership of assets. The more items bureaucrats list in response to this question, the more they prepare the applicant for the application process.

3.3. Treatment

To study the effects of ethnicity and other request characteristics on response quality, the instrument randomly varied five aspects of the request: the applicant's ethnicity, gender, and skill level, as well as the formality of the email and whether the applicant mentioned a lawyer. The treatments and instruments are summarized in Table 1. Interacting these characteristics yields 48 unique types of requests, which were randomly assigned to all welfare offices.

The key treatment of interest, **A**, varies the applicant's putative ethnicity among three groups: Germans, Turks and Romanians. Turks represent the largest group of immigrants in Germany (Schönwälder (2013), p. 637) and have long been the focus of the discourse on immigration and discrimination. Romanian migrants have only become the subject of political debate more recently when complete labor force mobility between Romania (and Bulgaria) and Germany took effect in January 2014. Since then — despite total migration from Romania being modest — politicians and media pundits have discussed widespread "poverty migration". Romanians have therefore come under discriminatory pressure, including accusations of "welfare cheating".

The second treatment, **B**, varies the applicant's gender, resulting in six aliases in total. We used lists of common names in Germany and in Romania, as well as among Turks in Germany, to find aliases that are reasonably similar to each other while clearly signaling ethnicity.[18]

The third treatment, **C**, varies the applicant's putative skill level. Here, we varied the profession the applicant claims to have practiced before. We chose two professions which are frequently exercised independently and relatively gender-balanced empirically. Our choice of the skilled job was physiotherapist, which requires extensive skills training and a certificate, while our choice of the unskilled job was a cleaning person.

Table 1: Treatment instruments.

	Treatment 1	**Treatment 2**	**Treatment 3**
A - *Ethnicity*	German (Schaefer)	Turkish (Yilmaz)	Romanian (Ionescu)
B - *Gender*	Female (Anna / Ayse / Ana)	Male (Michael / Mustafa / Mihai)	
C - *Skill*	Unskilled (Cleaning person)	Skilled (Physiotherapist)	
D - *Formal*	Informal (Typos, less formal tone)	Formal (No typos, normal tone)	
E - *Legal Support*	No support (Does not mention lawyer)	Support (Mentions lawyer)	

*"(**D**: [Dear Sir / Madam] / [Hello])*
*My name is (**A/B**: [Michael Schäfer] / [Anna Schäfer] / [Mustafa Yilmaz] / [Ayse Yilmaz] / [Mihai Ionescu] / [Ana Ionescu]) and I have a question about Hartz 4. I worked a few years as an independent (**C**: [cleaning person] / [physiotherapist]). Now, I need to close my business because I do not have enough customers and want to apply for Hartz 4. But a friend (**E**: [who is a lawyer] / []) has said that this could be complicated because I live with friends in an apartment. Before I go to your office, I therefore want to ask if I need to bring my roommates' documents, too? Or are my own papers sufficient? And which papers do I need to bring exactly?*
It would appreciate your response via email.
Many thanks"

The fourth treatment, **D**, varies the formality of the request. To this end, we sent out two versions of the email. The first email contained no mistakes and was written in a relatively accurate

18 Note that we were unable to obtain a reliable source of information documenting the exact frequency of these names in Germany.

formal and distant tone. Our intention was to represent the style and structure that an applicant with an average educational background would use. The second version decreased the level of formality by adding grammatical mistakes, as well as spelling and capitalization errors, and reducing the formality of the greeting. [19]

For the final treatment, **E**, we experimented with casually mentioning a lawyer. The treatment was motivated by the fact that most applicants have very little understanding of the application process and their legal rights, weakening their position vis-à-vis the bureaucracy. Moreover, there is anecdotal evidence that bureaucratic behavior in German welfare offices changes in the physical presence of external counsel.

3.4. Hypotheses

We spelled out our main hypotheses in the pre-registration document which we briefly summarize here. The hypotheses were inspired by the expectation that bureaucrats would treat non-Germans less favorably, but that this effect would be mitigated for certain groups. We also registered hypotheses for our other treatments and some observational correlations, the latter of which are not discussed in this paper.

Besides a general tendency of discrimination against non-Germans, we expected foreigners to be treated differently based on their origin: since recent public discourse made immigration from Romania highly salient and connected it explicitly to welfare use, we expected Romanians to be treated less favorably than Turks (BBC (2014)).

19 The different versions of the instrument are reproduced in Appendix section 7.3.

- Hypothesis 1: Bureaucrats treat German aliases more favorably than Turkish aliases, and Turkish aliases more favorably than Romanian aliases

Given discrimination against females in the German labor market (Diekmann, Engelhardt and Hartmann (1993)), labor market integration for women might be more difficult. We therefore hypothesized that bureaucrats would focus their efforts on males, leading to higher discrimination for females.

- Hypothesis 2: Bureaucrats treat male applicants more favorably than female applicants

We expected the skills treatment to have an effect because skills are directly related to the likelihood of success in the integration programs mandated by Hartz IV, and hence to the length of participation in the welfare program. Since welfare offices are assessed based on their success in re-integrating applicants into the job market, it is typically rational for bureaucrats to focus their efforts on the segment of recipients that are more skilled and educated (Heckman, Heinrich and Smith (1997)).

- Hypothesis 3: Bureaucrats treat skilled applicants more favorably than unskilled

Besides occupational skills, the level of formality of the request is also a strong signal of basic language and writing skills, which we expected to affect responses in a similar manner.

- Hypothesis 4: Bureaucrats treat applicants with formal requests more favorably than those with informal requests

Although the casual mentioning of a lawyer is only a very slight hint that an applicant might be more informed about the application process and knowledgeable about their rights, we hypothesized that this intervention would have a positive effect on response quality.

- Hypothesis 5: Bureaucrats treat applicants that mention a lawyer more favorably than those that do not

In addition, exploiting the conjoint nature of the trial, we study how discrimination against non-Germans relates to other non-ethnic characteristics of the requests. This is important because it allows us to investigate whether bureaucrats discriminate simply because they dislike non-Germans (taste-based discrimination), or because they use ethnicity cues to infer other information about applicants (statistical discrimination) (Becker (1957), Phelps (1972)). If discrimination against non-Germans were attenuated for those requests displaying higher skills and knowledge relative to other requests, this would be considered evidence of statistical discrimination.

- Hypothesis 6: Discrimination against foreign-sounding names will be less pronounced for respondents which are more skilled, mention a lawyer, or write more formal requests

As described above, a particular feature of the welfare offices is that some of them are run entirely by local governments (*Optionskommunen*), while the rest are under the umbrella of the national welfare agency and run in conjunction with local governments. The choice of organizational form has been at the discretion of local governments since 2005. About a quarter of agencies are now fully independent. As previous literature suggests, decentralization increases the room for discretion of local bureaucrats, heightening the probability that any potential biases among bureaucrats influence service delivery (Fording, Soss and Schram (2011)).

- Hypothesis 7: Discrimination against foreign-sounding names will be larger in locally-organized welfare offices, and the variance of substantive quality will be larger in them.

Finally, we hypothesized that the formality of requests would affect the formality of responses. However, as we discuss in the pre-registration document, formality is not a useful indicator of response quality since it is unclear whether formal language is a desideratum for applicants; it may both express respect and project professionalism, but also work as a deterrent. Therefore, we were primarily interested in whether responses mirrored the requests that were sent in terms of their formality.

- Hypothesis 8: Correspondence formality will be higher in the "formal" condition. Non-Germans will receive less formal responses.

3.5 Sample and protocol

Our units of observation are welfare office districts, of which there were 408 in early 2015. Within a welfare office district, there is typically one main office and multiple subbranches spread throughout the district. All email addresses were manually collected directly from the districts' websites. Since 2005, local governments can choose to run the welfare bureaucracy independently (Hassel and Schiller (2008)) — an option 105 districts made use of in early 2015. As can be seen in Figure 1, these centralized and independent agencies are fairly evenly spread across Germany though there are some differences between federal states.

Before rolling out the experiment, a pre-test was conducted to ensure that the requests elicited meaningful responses. The pre-test was performed using 24 randomly sampled offices, which received the email in November 2014. The pre-test induced one small change in the instrument; the Turkish female name was changed from Aylin to Ayse because two bureaucrats had mistaken the alias for a male.[20] The main wave of the experiment was implemented in January 2015, when the requests where sent to the remaining 384 welfare offices. Randomization was performed blocking on whether agencies are part of the national bureaucracy or managed independently to reduce variability.

Originally, a smaller second wave of the experiment was scheduled to be implemented a few weeks later. It would have involved emails to the subbranches in the 61 districts where multiple emails were available. However, upon reading the responses of the main wave, we noted that emails had occasionally been internally forwarded among bureaucrats within districts. This meant that sending a similar request would have considerably increased the risk of detection. The

20 Table 8 shows that the results are not affected when excluding the pre-test sample.

protocol was therefore changed — a decision made before any outcomes were coded or analyzed.

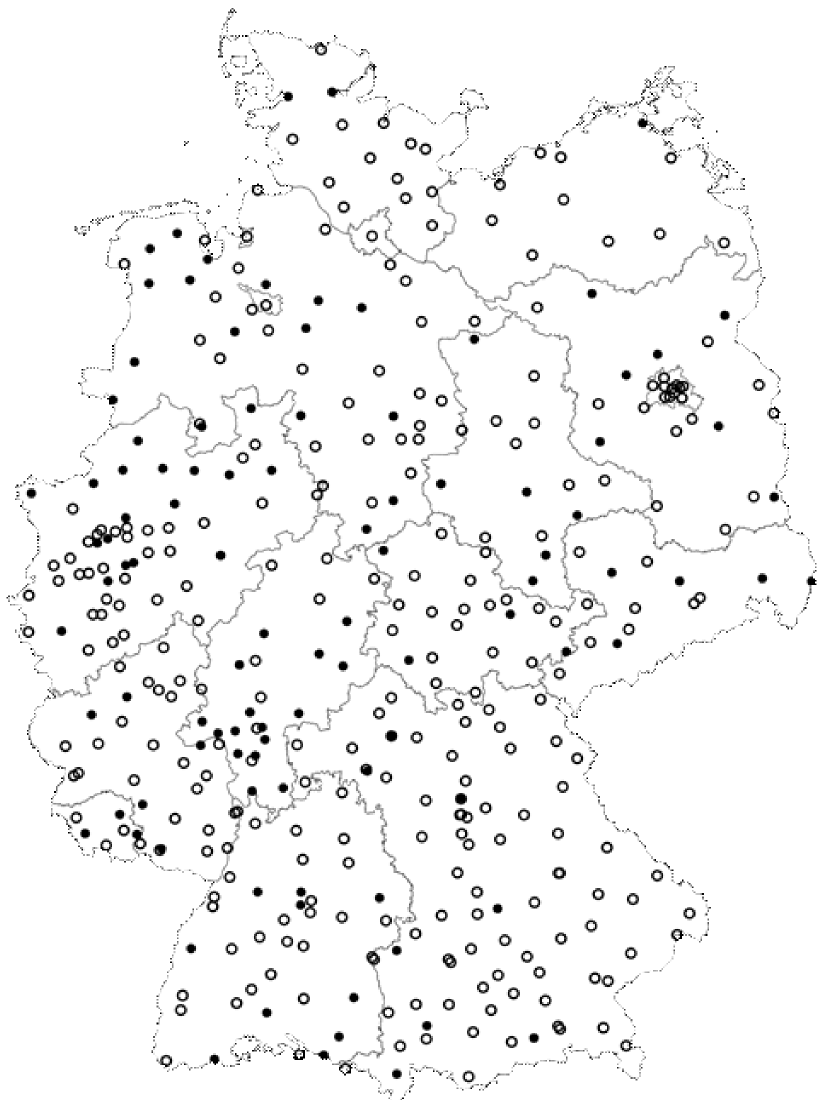


*Figure 1: Location of the 408 employment agencies (*Jobcenter*). Hollow circles mark offices integrated in the national bureaucracy, solid circles mark offices run locally.*

3.6. Ethical considerations

At this point, two ethical concerns merit discussion. First is the study's use of deception vis-à-vis welfare offices. While we believe that the use of deception should be avoided whenever possible, working with real requests did not appear feasible in this context: The German/foreigner cue is difficult to manipulate, and, more importantly, manipulating real cases may have deleterious consequences for actual applicants.

The second issue is the study's diversionary effect on the bureaucrats in the welfare offices: instead of responding to fictional queries, officials could have helped actual applicants. As explained below, we tried to find the shortest and simplest request that would still elicit meaningful variation in responses. The correct responses to the questions we asked are posted in the FAQ section of the national welfare agency's website, highlighting the simplicity of the request.

3.7. Coding of outcome variables

As outlined in the pre-registration document, we recorded a variety of discrimination measures grouped into main outcomes and other outcomes, described in Section 3.8. The three main outcome measures are the substantive quality, the response dummy, and the friendliness of a given response, which we discuss in turn.

First and most importantly, we constructed a substantive quality variable based on whether the two questions contained in the request were answered appropriately (*Answer question 1*, and *Answer question 2*). The coding of this measure requires some discussion given its significance for the study. As detailed in the pre-registration document, for each of the two questions, we gave either no points for no answer or no acknowledgement of the question, one point for a partial response, and two points for the correct response. The substantive quality measure is the sum of these two measures.

Regarding the question on the list of documents, two points were given if the bureaucrat listed at least three *specific* documents of those outlined above. One point was given if the bureaucrat listed one or two documents; this was typically to acknowledge the need for an ID document when visiting the welfare office. No point was given if the question was not addressed or if no single document was indicated.

Regarding the question about co-tenants' documentation, two points were given if the question was answered correctly, informing the applicant that no personal documents are required of the co-tenants. One point was given if the question was acknowledged and partially responded to. No points were given if the question was not addressed. One point was deducted in case the question was answered incorrectly, for example, by stating that documents of co-tenants are generally necessary (this occurred four times).

The two authors coded the emails independently and blindly. To this end, before any coding began, all emails were completely anonymized, removing any information that would have made it possible to infer an agency's treatment status. Overall, the two sets of codings yield a Pearson correlation coefficient of 0.974 for the question about the list of documents, and 0.989 for the question about co-tenants' documentation. There were eight instances of disagreement on the former question and seven on the latter. In the empirical section, we use the average of both coders.

The second main outcome is a simple response dummy, where automatic emails from a server are excluded. The third main outcome is the friendliness of emails, coded subjectively by both coders, on a 7-point scale. On these measures, owing to the flexibility of the concept, inter-coder reliability was considerably lower, with Pearson's R of 0.22.

Besides these three main outcomes, we also recorded secondary outcome measures including whether the response contained a formal greeting or goodbye, or both; the number of spelling mistakes; the number of grammatical and punctuation mistakes; whether the emails had a formal tone (coded subjectively by both coders on a 7-point scale, correlated at only 0.16); the length of the email; and the time it took bureaucrats to respond.

3.8. Descriptive statistics

Table 2 displays the descriptive statistics for the study. The overall response rate was 78 percent — an exceptionally high rate for this type of study.[21] Importantly, this allows us to perform our analyses of response quality on a large sample of responses. The average response time was 39 hours and responses were an average of 616 characters in length, excluding signatures.

Regarding the substantive questions, the average coding of the first question was 0.95 points (*Answer question 1 average*), while the average for the second question was 0.69 (*Answer question 2 average*). We combined the scores for both questions to a comprehensive response quality index (*Response quality average*), which is our primary outcome of interest.[22]

21 For comparison, Butler and Broockman (2011) received a response in 57 percent of all cases, and Distelhorst and Hou (2014) found a 37 percent response rate.

22 Note that in the EGAP pre-registration document, we defined a quality variable that would take the value 0 if no response was received, value 1 if any response was received, and extra points for substantive answers to the

Table 2: Pre-treatment covariates and outcomes.

	N	Mean	SD
Outcome Variables			
Any response	***408***	***78.4***	***41.2***
Response duration (hours)	321	39.3	58.4
Response length (characters)	321	615.8	388.9
Answer question 1 coder 1 (-1-2)	321	0.947	0.978
Answer question 1 coder 2 (-1-2)	321	0.950	0.980
Answer question 1 average (-1-2)	321	0.949	0.976
Answer question 2 coder 1 (0-2)	321	0.689	0.687
Answer question 2 coder 2 (0-2)	321	0.701	0.692
Answer question 2 average (0-2)	321	0.695	0.685
Response quality coder 1 (-1-4)	321	1.636	1.342
Response quality coder 2 (-1-4)	321	1.651	1.347
Response quality average	***321***	***1.643***	***1.340***
Formal greeting	321	93.8	24.2
Formal goodbye	321	94.4	23.0
Any formal	321	100.0	0.0
Typos (#)	321	0.2	0.5
Mistakes (#)	321	1.2	1.5
Friendliness coder 1 (1-7)	321	4.097	0.542
Friendliness coder 2 (1-7)	321	4.047	0.638
Friendliness average (1-7)	***321***	***4.072***	***0.462***
Formality coder 1 (1-7)	321	3.897	0.417
Formality coder 2 (1-7)	321	4.203	0.894
Formality average (1-7)	321	4.050	0.524
Covariates			
Unemployment	408	6.7	3.2
State (#)	408	7.2	4.8
Independent	408	25.7	2.2
Migrants	408	16.7	9.4
Appeals	349	26.1	10.7
Addresses (#)	408	1.5	1.6
Region (#)	408	4.6	2.9
East	408	19.4	39.6

Notes: measured in percent unless stated otherwise. Headline outcomes in bold.

questions. This variable would allow one to make comparisons across treatment conditions even if both response rates and quality varied. However, since response rates were highly similar across treatment conditions, differential non-response is not a concern. The analyses thus exclusively focus on the responses received. This has the advantage of not imposing any form of weighting of response rates versus quality. As detailed in Table 17, implementing our pre-registered coding does not change results appreciably.

Of all responses, 94 percent contained a formal greeting (*Formal greeting*) and 94 percent contained a formal goodbye (*Formal goodbye*). On average, responses contained 0.2 typos (*Typos*) and 1.2 grammatical mistakes including punctuation (*Mistakes*).

Our subjective coding of a given email's tone was done on a scale from 1 (very unfriendly) to 7 (very friendly), captured in the variables *Friendliness coder 1* and *Friendliness coder 2*. The average score of both coders was 4.1 (i.e., the emails were friendly on average; *Friendliness average*). However, as mentioned above, correlations between coders were quite low. Finally, regarding response formality, the average score given was also 4.1 (*Formality average*).

In addition to our outcome variables, Table 2 also reports descriptive statistics for the five pre-treatment covariates at the agency level. The variables above the dotted line were pre-registered, the ones below were not pre-registered and are thus not used in the empirical analysis. First, we obtained the unemployment rate at the welfare office district level from official records (*Unemployment*; average of 7 percent). Second, we indicate which of the 16 German federal states the agency is located in to estimate state fixed effects (*State*). Third, we report whether the agency is independently organized (*Independent*), which 26 percent of the sample are. Fourth, Migrant measures the percentage of citizens with immigration backgrounds in each district according to the 2011 Census.[23] Since these citizens are concentrated in more populous districts, their average share among the residents is lower at the level of districts (17 percent) than at the population level. Fifth, we recorded the fraction of successful appeals and lawsuits against

23 In merging these data to our dataset, we averaged over counties for the few welfare districts containing multiple counties. We were also forced to impute the share of residents with an immigration background in the state of Mecklenburg-Vorpommern, where there was no data on migrant shares in 9 districts. We imputed the missing data with the (known) mean of the state, which is very low at 3.7 percent (lowest decile in the national distribution).

agency decisions (*Appeals*) from national welfare agency databases. We wanted to use this variable as a proxy for the legal quality of decision-making but did not realize at the time of pre-registration that it contains missing values in 59 districts (15 percent of observations). We therefore decided not to include this variable in the benchmark empirical analyses, though we conduct a detailed analysis including this variable in Appendix Tables 15 and 16.

Aside from these five pre-registered control variables, we report statistics for three additional agency-level variables: The number of email addresses posted online (*Addresses*, 1.5), the regional direction under which the office is organized (*Region*; 10 in total), and whether the office is in East Germany, i.e. the former German Democratic Republic (*East*).

3.9. Balance

If randomization was faithfully executed, the distribution of covariates should be similar across treatment conditions. To assess balance, in Appendix Tables 5 through 7, we split the samples across the five treatment conditions, again reporting the sample size, mean and standard deviation of the eight pretreatment covariates. In Table 5, we split the sample along the ethnicity treatment conditions, demonstrating that there are no discernible differences across the three treatments. The final three columns report p-values from t-tests, testing whether the means of the variables are different. Of these, one test for the difference between the share of migrants, between the offices receiving putatively Turkish and Romanian requests, yields a marginally significant difference. In Table 6, we split the sample along the gender and skill treatments. Offices in the male and female conditions exhibit significant differences in their share of migrant population and their likelihood of being located in East Germany. Finally, in Table 7, we assess the balance across the legal backing (endorsed) and formality treatments, finding one significant deviation between informal and formal requests. We conclude that randomization produced a well-balanced sample, and report estimations with and without adjustment for covariates below.[24]

4. Empirical analysis

4.1. Estimation

Our benchmark analyses estimate the effects of the five main treatments on our three main outcome measures.

We estimate the following ordinary least squares (OLS) regression equation:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 \text{Foreign}_i + \beta_2 \text{Female}_i + \beta_3 \text{Unskilled}_i + \beta_4 \text{Unendorsed}_i + \beta_5 \text{Informal}_i \\ + \beta_6 \text{Unemployment}_i + \beta_7 \text{Independent}_i + \beta_8 \text{Migrants}_i + \epsilon_i \end{aligned} \tag{1}$$

where Y_i represents the outcome of agency i and the following variables represent the five treatment indicators each agency was assigned to. ε_i represents the error term. The remaining variables are pre-treatment pre-registered control variables, which are included in some models to assess robustness and reduce estimate variability.[25] In line with Freedman (2008), we do not estimate logistic regressions for the binary outcome (i.e., the response dummy), but fit a linear probability model using OLS.

24 An even more comprehensive overview of covariate balance can be obtained by testing for differences of all 8 covariates across all 55 unique treatment comparisons (e.g., Turkish vs. Female). Doing so results in 440 tests of which 5 yield significant differences — less than one would predict on the basis of chance alone. Figure 10 in the Appendix plots the distribution of p-values of these tests.

25 As described above, we do not include the pre-registered *Appeals* variable as a control given that 15 percent of observations are missing. We address this issue in the Robustness section (4.5.).

Given the conjoint design of the experiment, we estimate the average marginal component effect (ACME; Hainmüller, Hopkins and Yamamoto (2014)). In essence, the AMCE scrutinizes the difference in outcomes by comparing two different attributes (e.g., skilled vs. unskilled), holding all other attributes constant by averaging over them. When comparing two attributes, the random assignment of all five treatments ensures that the remaining attributes are identically distributed. The approach also allows us to compare detailed profiles such as Turkish women vs. Romanian women — a feature we make use of below.

4.2. Response rates

Table 3 displays the main results. To ease interpretation, Figure 2 plots the coefficients from the same six models. No treatment condition has a statistically significant effect on the response dummy (*Any response*). In particular, a foreign alias is only 0.4 percent less likely to receive a response. Female applicants are 2.6 percent less likely, while unskilled applicants are 2.3 percent more likely to receive a response. Applicants that do not mention a lawyer are 3.0 percent less likely to receive a response, while informally written emails are 5.4 percent less likely to be answered. None of these differences, however, are statistically significant (or even marginally so).

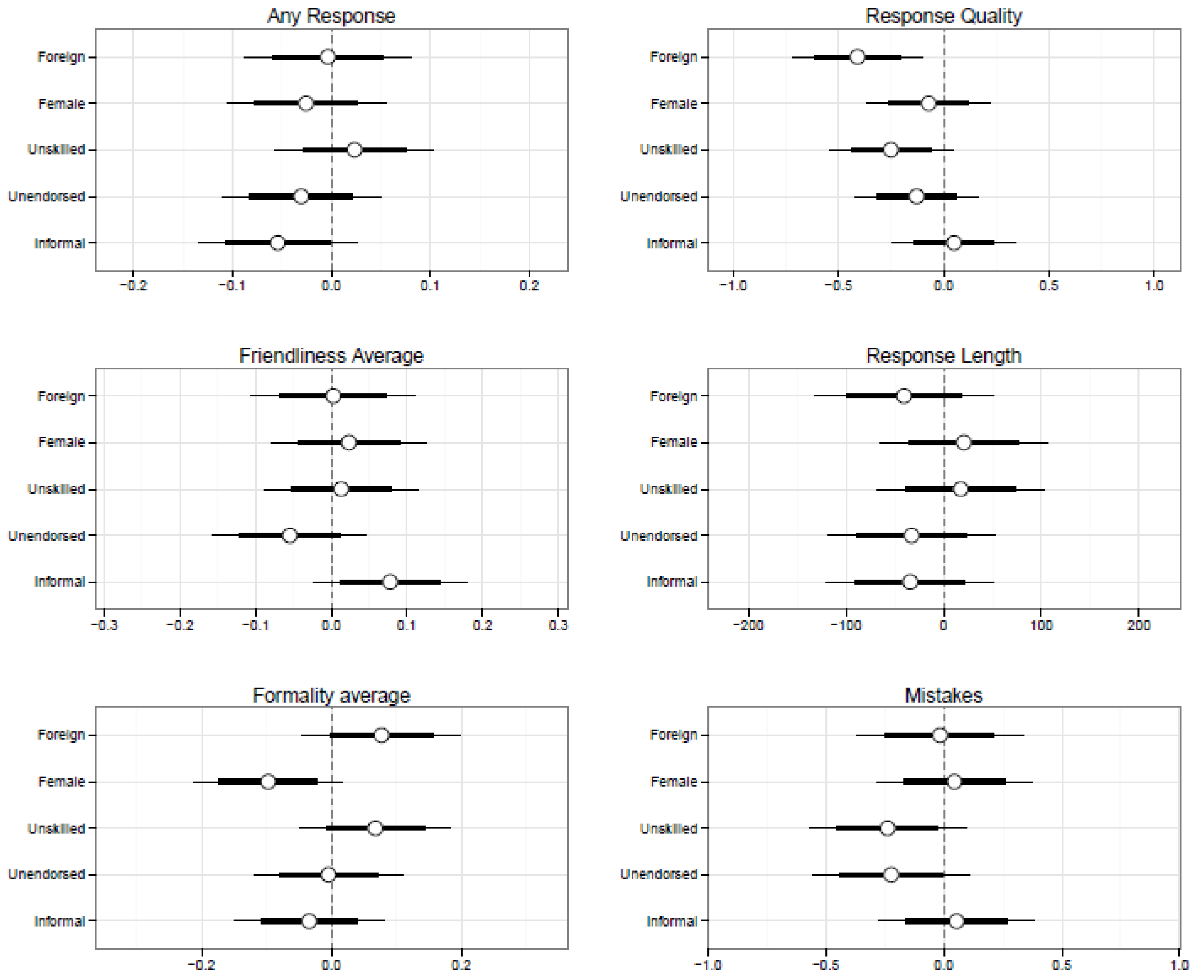


Figure 2: Coefficients of main results, taken from Table 3.

Importantly, the absence of significant response differences demonstrates that non-response is not systematically linked to treatment status. To buttress this finding, we additionally assess whether there are any treatment combinations that experience significantly lower response rates. To do so, we note that the conjoint design of the study produces 155 unique trait counterfactuals one can compare. Specifically, these combinations represent treatment counterfactuals (e.g., female vs. male), where the remaining treatments are either randomly distributed — as is the case in the models discussed thus far — or fixed at one or several specific traits (e.g., skilled male). Overall, there are 31 comparisons per treatment, which adds up to 155 unique comparisons given the five overall treatments. [26]

For all of the 155 combinations, we ran regressions of the unique treatment indicator on the response dummy. In Figure 3, we report p-values from all 155 models. Only 7 models yield significant differences in the response dummy. Overall, the p-values are almost uniformly distributed. This is strong evidence that response rates are unrelated to treatment status. The following analyses, which rely on the subset of the 321 responsive offices, are hence causally identified in the sense that all requests had similar probabilities of being received by any given office and being responded to.

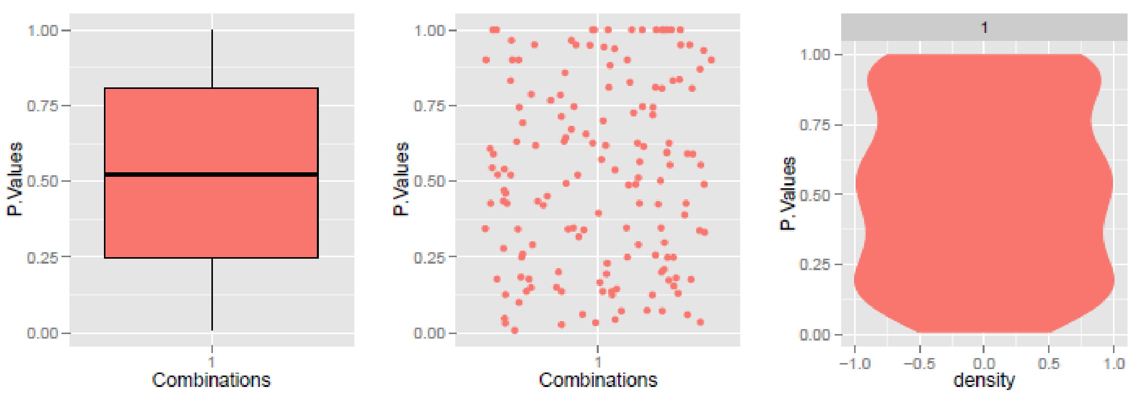


Figure 3: P-values from regressions of the response dummy on all 155 unique trait indicators.

26 Note that the Turkish and Romanian names were grouped into one "foreign" treatment. Examples of the 155 comparisons include: male vs. female; skilled male vs. skilled female; skilled formal male vs. skilled formal female, etc. Note also that these samples have different sample sizes, which renders p-values an imperfect unit of comparison.

4.3. Response quality

Having established that response rates are unrelated to treatment status, we turn to our main quality outcome. As can be seen in column 2 of Table 3, putative non-Germans receive responses of substantively and statistically significantly lower quality. Their responses score an average of 0.41 points lower on the 5-point quality scale. This is equivalent to a 21 percent (0.3 standard deviation) decline.

The female, unskilled and unendorsed treatments also experience lower quality responses (0.07, 0.25, and 0.13, respectively). These differences, however, are not significant, though they are in line with our hypotheses (H1-H3). The informality treatment receives higher quality responses. The finding is contrary to Hypothesis 4, yet the effect is substantively small and not significant.

Table 3: Treatment effects on main outcomes.

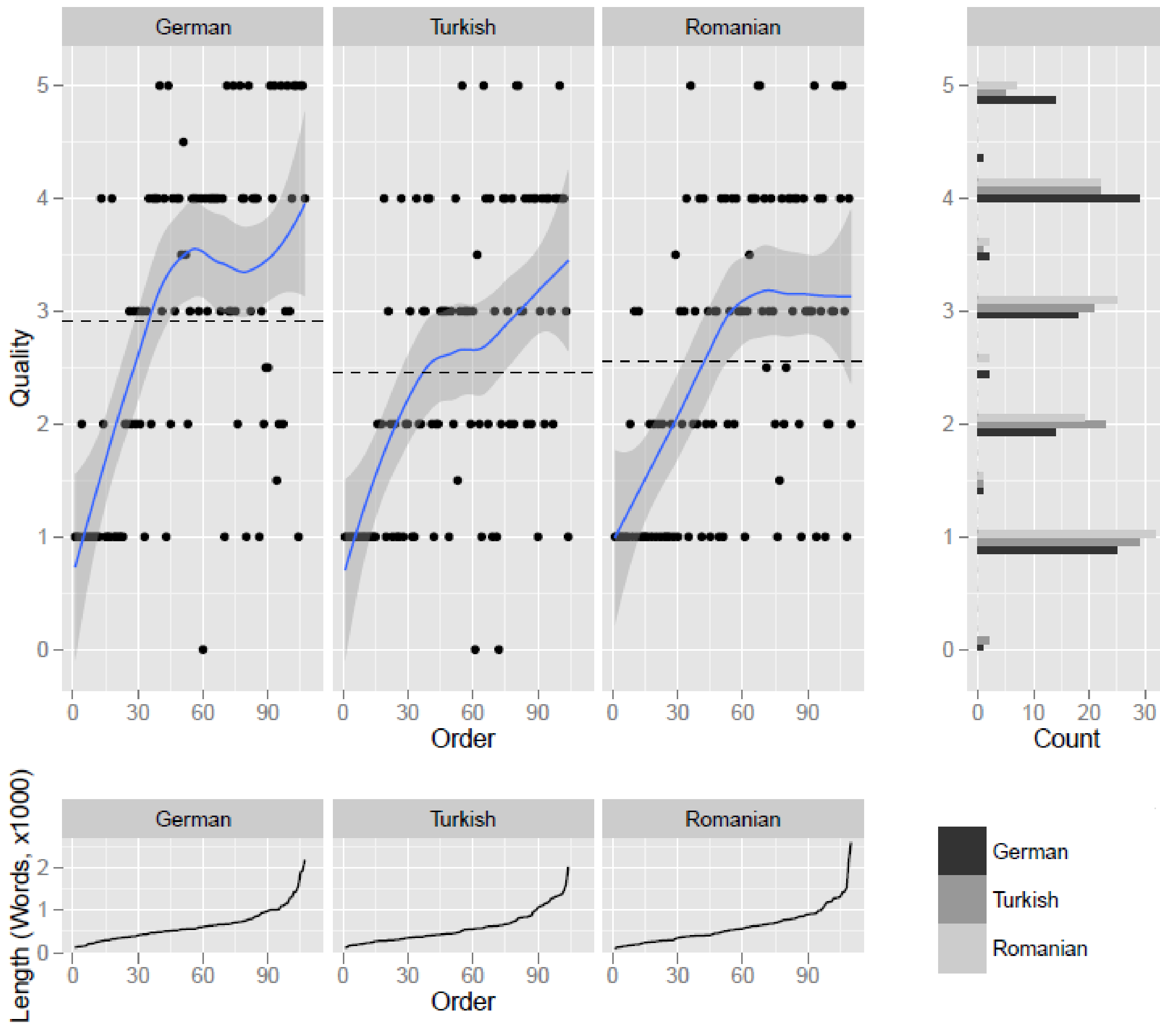
	Headline outcomes			Additional outcomes		
	(1)	(2)	(3)	(4)	(5)	(6)
	Any response	Response quality avg	Friendliness average	Formality average	Response length	Mistakes
Foreign	-0.004	-0.410**	0.003	0.077	-40.655	0.021
	(0.043)	(0.158)	(0.055)	(0.062)	(46.399)	(0.061)
Female	-0.026	-0.072	0.023	-0.097	20.651	-0.015
	(0.041)	(0.149)	(0.052)	(0.058)	(43.672)	(0.057)
Unskilled	0.023	-0.251	0.013	0.068	17.450	-0.088
	(0.041)	(0.149)	(0.052)	(0.059)	(43.705)	(0.057)
Unendorsed	-0.030	-0.129	-0.055	-0.004	-33.014	-0.077
	(0.041)	(0.148)	(0.052)	(0.058)	(43.629)	(0.057)
Informal	-0.054	0.049	0.078	-0.034	-34.447	-0.105
	(0.041)	(0.148)	(0.052)	(0.058)	(43.607)	(0.057)
Observations	408	321	321	321	321	321

Notes: OLS models. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

In column 5 we find that non-German aliases do not receive significantly shorter emails. The estimated length reduction is a mere 41 characters. The other treatment statuses do not elicit significantly longer or shorter responses, either. Indeed, the coefficients for the gender and skill treatment are in the opposite direction of our hypotheses. This result is noteworthy given that response length, being easy to measure, has previously been used as a measure for response quality (Distelhorst and Hou (2014)).

To further assess the relationship between length and quality, in Figure 4 we plot response quality on the y-axis and the response length on the x-axis.[27] To reduce the role of outliers, we plot the order of lengths. The bottom chart motivates this choice, demonstrating that the relationship between length and rank is similar across treatment statuses. Importantly, across ethnic treatment statuses, emails of shorter length are of poorer quality on average. However, as email lengths increase, putative Germans receive higher quality responses than putative Turks and Romanians. This highlights the benefits of coding response quality manually as opposed to relying on simpler measures.

27 The overall correlation between substantive quality and response length in our sample is 0.41.



Notes: The main figure plots the quality outcomes by ethnic treatment status across the email length order. The bottom figure plots the length by order, demonstrating a highly similar trend across treatments. The right figure presents the raw histogram of the quality variable.

Figure 4: Quality of responses by length of email.

4.4. Response tone

Next, we test our hypotheses regarding friendliness and formality. Columns 3 and 4 in Table 3 demonstrate that there are no significant differences across the five main treatments on either outcome measure. All estimates are remarkably close to 0. In Column 6 of Table 3, we show that the mistakes per responses are also statistically indistinguishable across treatments, with differences remarkably close to 0 across all traits.

Taken together, we interpret these results as evidence that German welfare offices discriminate against non-German applicants, and that they do so along the margin of response quality, not response rates. When interpreting these results, it should be kept in mind that our requests do not correspond to negative stereotypes of welfare use: all putative applicants claim to have previously worked in Germany, and therefore do not at all resemble "poverty migrants" as portrayed by some media outlets. Effects would likely be more pronounced if this were not the case.

4.5. Robustness

In the Appendix, we assess the robustness of these findings by estimating the same models controlling for covariates and including fixed effects (Table 8). The estimates are highly similar in size and significance to those reported in Table 3. Moreover, in the same table, we also run models where we drop the 24 pre-test observations that were gathered two months prior to the main wave, which used a different Turkish female given name. In Table 10, we drop all observations from East Germany, as we had pre-registered. In both cases, the estimates are essentially unchanged.

Next, we revisit the issue of the pre-registered *Appeals* control variable which, unbeknownst to us, exhibits missingness in 15 percent of observations. First, we note that the variable is balanced across treatment conditions (see Appendix Tables 5-7). Second, in Appendix Table 15, we demonstrate that, regarding the quality outcome, including the Appeals variable reduces the point estimates for the foreign treatment by 0.15, while slightly increasing the estimates for the other treatments. However, this is driven by the smaller sample induced by the missingness in this variable, not by its inclusion as a control: repeating the analysis for the subset of observations without missingness, without including Appeals as a control, results in essentially the same estimate. Third, in Appendix Table 16 we apply mean and multiple imputation to the missing observations of the *Appeals* variable. For the multiple imputation procedure we use the five treatment indicators and three remaining pre-registered control variables as predictors. Both methods recover the original estimate of around 0.4 for the foreign treatment.

In Table 17 of the Appendix, following our pre-analysis plan, we estimate the same benchmark regressions replacing missing outcome data in the quality variable that stem from non-response with zeroes. This approach, made unnecessary by the fact that response rates are unrelated to treatment status in our trial, also leaves estimates largely unaffected.

Finally, we demonstrate that all headline estimates are also obtained when using randomization inference (Appendix section 7.2., Figures 5 through 9).

4.6. Disaggregating quality

The analyses thus far have shown that foreign-sounding aliases receive qualitatively inferior responses. Which components of the quality measure and which aliases drive this result? To answer this question, Table 4 separately analyzes responses to the co-tenant question (*Question 1*) and the paperwork question (*Question 2*), simultaneously splitting up the foreign variable into the Turkish and Romanian statuses. In the first two columns, we report the simple OLS regression without covariate adjustment; column 3 and 4 add covariates, and columns 5 and 6 add covariates and state-level fixed effects.

First, note that all point estimates are negative across all specifications, highlighting that non-German aliases receive inferior responses relative to German aliases. Second, comparing across questions, coefficients are more pronounced for the first question than for the second question. To the extent that these questions can be thought of as tapping bureaucrats' trust (Question 1) and effort (Question 2), the result appears consistent with discrimination based primarily on lower trust towards foreigners. Third, across both questions the response quality disadvantage is slightly larger for Turks than for Romanians, contradicting our expectation that Romanians would be more disadvantaged (Hypothesis 1). However, note that none of the differences in estimates are statistically significant (Gelman and Stern (2006)).

Table 4: Treatment effects across both questions.

	(1) Answer Question 1	(2) Answer Question 2	(3) Answer Question 1	(4) Answer Question 2	(5) Answer Question 1	(6) Answer Question 2
Turkish	-0.303*	-0.161	-0.306*	-0.175	-0.318*	-0.164
	(0.134)	(0.095)	(0.135)	(0.094)	(0.137)	(0.096)
Romanian	-0.192	-0.168	-0.188	-0.152	-0.216	-0.126
	(0.132)	(0.093)	(0.132)	(0.092)	(0.136)	(0.096)
Female	-0.069	-0.002	-0.073	-0.007	-0.062	-0.007
	(0.108)	(0.077)	(0.110)	(0.077)	(0.113)	(0.079)
Unskilled	-0.236*	-0.015	-0.238*	-0.013	-0.216	-0.029
	(0.108)	(0.077)	(0.109)	(0.076)	(0.111)	(0.078)
Unendorsed	-0.125	-0.005	-0.126	-0.005	-0.116	0.021
	(0.108)	(0.077)	(0.109)	(0.076)	(0.110)	(0.077)
Informal	0.055	-0.003	0.057	0.005	0.021	-0.024
	(0.108)	(0.077)	(0.109)	(0.076)	(0.112)	(0.079)
Controls	No	No	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No	Yes	Yes
Observations	321	321	321	321	321	321

Notes: OLS models. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

Note also that unskilled applicants receive lower-quality responses to question 1. This effect is roughly similar in magnitude to the ethnicity effect but not detectable in question 2 and not significant in all models.

4.7. Interactions

Next, we utilize the conjoint design of the study to parse out the interactions between the individual-level traits. Here, we restrict our analysis to the hypotheses spelled out in the pre-registration document.

First, in Appendix Table 11, we separate formal from informal requests. The results demonstrate that the response quality effect is higher for putatively foreign applicants with formal requests than for those with informal requests. For the latter group, the quality disadvantage is small and statistically indistinguishable from zero, while for the former group requests receive a substantial reduction by 0.7 points (a statistically significant difference). This finding contradicts hypothesis 6, which states that discrimination against foreigners is mitigated by more formal writing. Instead, the observed pattern might be explained by more formal requests reducing perceptions of applicants' need.

Second, in Appendix Table 12 we split up the sample along applicants mentioning (endo) and not mentioning lawyers (unendo). Here, our hypothesis that unendorsed applicants' emails receive lower quality responses than endorsed applicants is confirmed, though the estimates are not significantly different. There are two additional findings. First, endorsed and unskilled applicants receive qualitatively inferior emails (a 0.45 drop, though not significant). Second, informal but unendorsed applicants seem to receive more friendly emails.

Finally, in Appendix Table 13, we look at differences between skilled and unskilled applicants. Several findings result from these comparisons. We find one of the very few significant differences in response rates: skilled foreign applicants appear to be more likely to receive a response relative to unskilled ones. Both estimates are significantly different. Yet, among the applicants which received answers, substantive quality is somewhat lower for skilled foreigners than for unskilled ones, though this difference is not significant. In addition, we find that unendorsed applicants receive less qualitative emails when skilled and less friendly emails when unskilled.[28]

4.8. Bureaucratic organization

We now turn our attention to the bureaucratic organization of the offices under study. In particular, we hypothesized that independent welfare offices run by local governments are more likely to discriminate than those embedded in the national bureaucracy (H7). In Appendix Table 9, the sample is split into "independent" and "centralized" offices. Recall that in centralized offices, work is jointly administered by the national welfare agency and local governments. Independent offices, on the other hand, are run entirely by the local government in a given district.

28 For completeness, Appendix Table 14 splits the sample into male and female applicants, for which we did not specify any hypotheses.

Although estimate variability is high due to the small number of observations of independent offices (76), the estimate of the quality disadvantage for foreigners in independent offices is about twice as large as in centralized offices. Two words of caution, however, are in order. First, bureaucratic organization was not randomly assigned. Even if the difference between independent and centralized offices is not due to chance, it may well be due to selection of districts into the independent status. The map in Figure 1, however, shows that, geographically speaking, independent agencies are fairly evenly spread across Germany. Second, the difference in the estimates, though sizeable, is not significant.

Having emphasized these caveats, the finding is consistent with a literature on the effects of centralized welfare administrations. Fording, Soss and Schram (2011), for example, argue that more centralized bureaucratic arrangements are less conducive to biased behavior by officials in the United States. If bureaucrats are subject to increased monitoring and have less discretion in centralized versus decentralized arrangements, this could help explain differences in discriminatory behavior even if individual-level bias is regionally uniform.

Previous research on the organization of the German welfare offices indicates that bureaucrats in independent offices do indeed have more leeway in decision-making than their colleagues in centralized offices. The official evaluation of the different organizational forms of welfare agencies, written in 2008, states: “Independent offices exhibit decentralized structures which, based on the judgment of staff, is reflected in more flexibility and freedom of decision, but also has the disadvantage of lower transparency in terms of regional and national controlling and monitoring” (Deutscher Bundestag (2008), p. 17). Thus, while more research on the effects of organizational forms is warranted, it is plausible that non-Germans are more likely to experience discrimination in independent offices.

5. Discussion and conclusion

This study used a conjoint experiment to assess whether German welfare offices treat requests from German and non-German applicants equally. Relying on a request designed to elicit substantively meaningful variation in responses and a pre-registered measure of response quality, the trial provided evidence that prospective benefit applicants with foreign-sounding aliases are no less likely to receive responses in general. However, putative foreigners receive emails of significantly lower quality. We believe that the finding has several implications for the study of discrimination in the public sphere.

First, the trial overcomes what we think is an empirical shortcoming in the existing literature. By restricting attention to easily observable dimensions of interactions — above all, response rates or lengths — researchers run the risk of making potentially faulty inferences about substantive discrimination. Had the present study only measured response rates and lengths, one would have inferred that German welfare offices treat all applicants equally. However, as was shown, such a reading would have been wrong: responses sent to putative foreigners were, on the whole, less helpful to applicants in a way that is substantively important. The reduction in quality we report plausibly affects applicants' perceptions of the burdens involved in applying for benefits, potentially impacting their decision on whether or not to apply.[29] Arguably, there is no straightforward and sensible way of determining the substantive importance of response rates

29 A large body of research has demonstrated that information plays an important role in decision-making about benefit applications (Duflo and Saez (2002), Daponte, Sanders and Taylor (1999)).

relative to the information value contained in responses. But at a minimum, we believe that future studies that use information requests should make explicit attempts, defined ex ante if possible, to measure the quality of responses, in order to avoid the pitfalls we have highlighted. The use of outcome variables that are directly substantively meaningful may be an even more promising alternative.

Second, the findings give rise to an interesting substantive question, namely, why discrimination occurs along the quality and not the response dimension. We see two potential explanations.

On the one hand, it may be psychologically easier for biased bureaucrats to respond to requests in a relatively superficial way rather than ignoring them outright — especially in the presence of strong antiprejudice norms. Blinder, Ford and Ivarsflaten (2013) suggest that citizens try to control their prejudiced thoughts and actions in order to comply with such norms. However, their ability to do so varies across situations. If not responding to a query is more cognitively salient as a prejudiced act than drafting a less helpful response, this might explain the divergence between the two dimensions.

On the other hand, the effect may be due to our case selection: to the extent that front-line bureaucrats are subject to more stringent monitoring by their hierarchy in Germany than elsewhere, discriminating along the quality margin instead of the response margin might be a way of obscuring discrimination vis-à-vis superiors. Since response rates to official emails are plausibly easier to monitor than their content, this could explain why we do not find discrimination along the response margin — in stark contrast with previous studies.

Third, we find tentative evidence suggesting that the prevalence of discrimination is negatively related to the bureaucratic centralization of welfare offices. While the study cannot conclusively answer this question, the effects of administrative decentralization for equality and impartiality in public service provision constitute a fruitful area for future research.

6. References

Applebaum, L. (2001). The Influence of Perceived Deservingness on Policy Decisions Regarding Aid to the Poor. *Political Psychology,* 22(3), 419–442.

Art, D. (2005). *The Politics of the Nazi Past in Germany and Austria*. Cambridge: Cambridge University Press.

Atkeson, L., Bryant, L. Hall, T., Saunders, K. and Alvarez, M. (2010). A New Barrier to Participation: Heterogeneous Application of Voter Identification Policies. *Electoral Studies*, 29(1), 66–73.

BBC (2014). *Romania and Bulgaria Migration Issue Divides Germans*. URL: http://www.bbc.com/news/world-europe-25635864

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* , 94(4), 991–1013.

Castronova, E., Kayser, H., Frick, J. and Wagner, G. (2001). Immigrants, Natives and Social Assistance: Comparable Take-Up Under Comparable Circumstances. *International Migration Review*, 35(3), 726-748.

Blinder, S., Ford, R. and Ivarsflaten, E. (2013). The Better Angels of Our Nature: How the Antiprejudice Norm Affects Policy and Party Preferences in Great Britain and Germany. *American Journal of Political Science*, 57(4), 841–857.

Broockman, D. (2013). Black Politicians Are More Intrinsically Motivated to Advance Blacks Interests: A Field Experiment Manipulating Political Incentives. *American Journal of Political Science*, 57(3), 521– 536.

Bruckmeier, K. and Wiemers, J. (2012). A New Targeting: A New Take-Up? *Empirical Economics,* 43(2), 565–580.

Deutscher Bundestag. (2008). *Unterrichtung durch die Bundesregierung: Bericht zur Evaluation der Experimentierklausel nach 6c des Zweiten Buches Sozialgesetzbuch. Drucksache 16/11488.* [Information by the Federal Government: Report on the evaluation of the experimentation clause following 6c of the Second Book of German Social Law]. Public Report.
URL: http://dipbt.bundestag.de/dip21/btd/16/114/1611488.pdf

Butler, D. and Broockman, D. (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science*, 55(3), 463–477.

Cobb, R., Greiner, J. and Quinn, K. (2010). Can Voter ID Laws Be Administered in a Race-Neutral Manner? Evidence from the City of Boston in 2008. *Quarterly Journal of Political Science*, 7(1), 1–33.

Currie, J. (2004). *The Take Up of Social Benefits*. National Bureau of Economic Research Working Paper 10488. URL: http://www.nber.org/papers/w10488.pdf

Daponte, B., Sanders, S. and Taylor, L. (1999). Why Do Low-Income Households Not Use Food Stamps? Evidence from an Experiment. *Journal of Human Resources,* 34(3), 612–628.

Diekmann, A., Engelhardt, H. and Hartmann, P. (1993). Einkommensungleichheit in der Bundesrepublik Deutschland: Diskriminierung von Frauen und Ausländern. [Income Inequality in the Federal Republic of Germany: Discrimination Against Women and Foreigners] *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 26(3), 386–398.

Distelhorst, G. and Hou, Y. (2014). Ingroup Bias in Official Behavior: A National Field Experiment in China. *Quarterly Journal of Political Science,* 9(2), 203–230.

Duflo, E. and Saez, E. (2002). Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues Choices. *Journal of Public Economics,* 85(1), 121–148.

Fix, M. and Struyk, R. (1993). *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C: The Urban Institute.

Fording, R., Soss, J. and Schram, S. (2011). Race and the Local Politics of Punishment in the New World of Welfare. *American Journal of Sociology,* 116(5), 1610–1657.

Freedman, D. (2008). Randomization Does Not Justify Logistic Regression. *Statistical Science*, 23(2), 237–249.

Gelman, A. and Stern, H. (2006). The Difference Between "Significant" and "Not Significant" is not itself Statistically Significant. *The American Statistician,* 60(4), 328–331.

Giulietti, C., Tonin, M. and Vlassopoulos, M. (2015). *Racial Discrimination in Local Public Services: A Field Experiment in the US*. . IZA Working Paper No. 9290.
URL: http://www.iza.org/de/webcontent/publications/papers/viewAbstract?dp_id=9290

Deutsche Welle (2014). *Experte für Migrationsberatung: Abwehrhaltung in Behörden*. [Expert on Immigration Counseling: Defensive Posture in Offices]. URL: http://www.dw.com/de/experte-fuer-migrationsberatung-abwehrhaltung-in-behoerden/a-17425680l

Hainmüller, J., Hopkins, D. and Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, 22(1), 1– 30.

Hassel, A. and Schiller, C. (2008). *Die Politische Dynamik von Arbeitsmarktreformen in Deutschland am Beispiel der Hartz IV-Reform.* [The Political Dynamics of Labor Market Reforms in Germany: The Example of Hartz IV]. Unpublished Manuscript. URL: http://www.boeckler.de/pdf_fof/S-2007-996-4-6.pdf

Heckman, J., Heinrich, C. and Smith, J. (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review: Papers and Proceedings*, 87(2), 389–395.

Hewstone, M., Rubin, M. and Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology*, 53(1), 575–604.

Humphreys, M., Fang, A. and Guess, A. (2014). *Messaging Strategies to Combat Housing Discrimination in New York City: Evidence from a Field Experiment.* NYC Commission for Human Rights Report. URL: http://cu-csds.org/wp-content/uploads/2012/02/NYC-Housing-Discrimination-Study-Final-Report-2014-06-07-NO-APPENDICES.pdf

Jolls, C. and Sunstein, C. (2006). The Law of Implicit Bias. *California Law Review*, 94(4), 969–996.

Kaas, L. and Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1–20.

Keiser, L., Mueser, P. and Choi, S. (2004). Race, bureaucratic discretion, and the implementation of welfare reform. *American Journal of Political Science,* 48(2), 314–327.

Krumpal, I. (2012). Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning. *Social Science Research*, 41(6), 1387–1403.

Lipsky, M. (1980). *Street-level Bureaucracy: Dilemmas of the Individual in Public Services*. New York, NY: Russell Sage Foundation, 1980.

Mayntz, R. (1965). Max Webers Idealtypus der Bürokratie und die Organisationssoziologie. [Max Weber's Ideal Type of Bureaucracy and Organizational Sociology] In J. Fijalkowski (Ed.). *Politologie und Soziologie* (pp. 91–100) [Political Science and Sociology]. Wiesbaden, Germany: VS.

McClendon, G. (forthcoming). Race and Responsiveness: A Field Experiment with South African Politicians. *Journal of Experimental Political Science.*

Mustard, D. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the US Federal Courts. *Journal of Law and Economics*, 44(1), 285–314.

Pager, D. and Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34, 181-209.

Pager, D. Quillian, L. (2005). Walking the Talk? What Employers Say Versus What They Do. *American Sociological Review,* 70(3), 355–380.

Phelps, E. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, 62(4), 659–661.

Rachlinski, J., Johnson, S., Wistrich, A. and Guthrie, C. (2009). Does Unconscious Racial Bias Affect Trial Judges? *Notre Dame Law Review,* 84(3), 1195-1246.

Rosanvallon, P. (2011). *Democratic Legitimacy: Impartiality, Reflexivity, Proximity*. Princeton, NJ: Princeton University Press.

Rothstein, B. and Teorell, J. (2008). What is Quality of Government? A Theory of Impartial Government Institutions. *Governance*, 21(2), 165–190.

Schmid, L. (2015). *Ethnische Diskriminierung bei der Wohnungssuche: Feldexperimente in sechs deutschen Großstädten* [Ethnic Discrimination In Housing Search: Field Experiments in Six German Metropolitan Areas]. PhD thesis. University of Konstanz, Germany.
URL: http://kops.uni-konstanz.de/bitstream/handle/123456789/31349/Schmid_0-295831.pdf?Sequence= 3&

Schönwälder, K. (2013). Immigrant Representation in Germany's Regional States: The Puzzle of Uneven Dynamics. *West European Politics,* 36(3), 634–651.

Spiegel (2014). *Seehofer verteidigt Kurs der CSU im Streit um Armutsmigration* [Seehofer defends position of the CSU in the debate around poverty migration]. URL: http://www.spiegel.de/politik/deutschland/seehofer-verteidigt-kurs-der-csu-im-streit-umarmutsmigration-a-941433.html

Stern, T., Wecking, C. and Reinecke, M. (2008). *Expertise zum Thema Interkulturelle Kompetenz der Job-Center* [Report on the Intercultural Competence of Jobcenters]. Report by Rambøll Management for the Senate of Berlin (Senatsverwaltung für Integration, Arbeit und Soziales). URL:
http://www.ramboll-management.de/ news/~/media/03BA5036745D4D8087F0453C6B4ECA79

Thränhardt, D. (1995). The Political Uses of Xenophobia in England, France and Germany. *Party Politics,* 1(3), 323–345.

van Oorschot, W. (2006). Making the Difference in Social Europe: Deservingness Perceptions Among Citizens of European Welfare States. *Journal of European Social Policy,* 16(1), 23–42.

Weber, M. (2009). *The Theory of Social and Economic Organization*. New York, NY: Simon and Schuster.

7. Appendix

7.1. Tables

	Sample			German			Turkish			Romanian			Ger - Tur	Ger - Rom	Tur - Rom
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	P-Value	P-Value	P-Value
Unemployment	408	6.6	3.1	136	6.7	3.4	136	6.2	3.0	136	6.9	2.9	0.231	0.639	0.074
State (#)	408	7.2	4.8	136	7.3	5.0	136	6.8	4.8	136	7.4	4.6	0.411	0.764	0.244
Independent	408	25.7	43.8	136	25.7	43.9	136	25	43.5	136	26.5	44.3	0.889	0.890	0.782
Migrants	408	16.7	9.4	136	16.3	9.5	136	18.1	9.3	136	15.8	9.3	0.121	0.671	0.046
Appeals	349	26.1	10.7	113	25.6	10.2	120	25.9	10.7	116	26.6	11.3	0.820	0.477	0.625
Addresses (#)	408	1.4	1.5	136	1.3	0.8	136	1.5	2.0	136	1.5	1.5	0.118	0.115	0.814
Region (#)	408	4.8	2.8	136	4.8	2.9	136	4.7	2.9	136	5.0	2.7	0.728	0.623	0.397
East	408	19.4	39.6	136	21.3	41.1	136	16.2	37	136	20.6	40.6	0.278	0.882	0.349

Table 5: Balance of covariates across ethnic treatments

	Male			Female			Male - Female	Skilled			Unskilled			Skill - Unskill
	N	Mean	SD	N	Mean	SD	P-Value	N	Mean	SD	N	Mean	SD	P-Value
Unemployment	202	6.5	3.1	206	6.7	3.2	0.411	207	6.7	3.1	201	6.6	3.2	0.749
State (#)	202	7.1	4.9	206	7.2	4.7	0.917	207	7.4	4.7	201	7.0	4.9	0.404
Independent	202	26.7	44.4	206	24.8	43.3	0.649	207	26.6	44.3	201	24.9	43.3	0.696
Migrants	202	15.7	9.4	206	17.8	9.3	0.023	207	16.9	9.6	201	16.6	9.2	0.719
Appeals	170	26.2	10.4	179	25.9	11.0	0.771	180	25.5	10.5	169	26.6	11.0	0.324
Addresses (#)	202	1.4	1.3	206	1.5	1.7	0.564	207	1.5	1.8	201	1.4	1.2	0.352
Region (#)	202	4.8	2.9	206	4.8	2.7	0.871	207	4.9	2.8	201	4.7	2.9	0.364
East	202	24.3	43.0	206	14.6	35.4	0.013	207	18.8	39.2	201	19.9	40.0	0.787

Table 6: Balance of covariates across gender and skill treatments

	Endorsed			Unendorsed			Endo - Unendo	Formal			Informal			Form - Inform
	N	Mean	SD	N	Mean	SD	P-Value	N	Mean	SD	N	Mean	SD	P-Value
Unemployment	200	6.7	3.2	208	6.6	3.0	0.769	202	6.6	3.0	206	6.6	3.2	0.839
State (#)	200	7.3	4.9	208	7.0	4.7	0.471	202	7.6	4.6	206	6.7	4.9	0.072
Independent	200	25.5	43.7	208	26.0	43.9	0.915	202	25.7	43.8	206	25.7	43.8	0.997
Migrants	200	16.3	9.6	208	17.2	9.2	0.303	202	16.5	9.0	206	17.0	9.8	0.553
Appeals	173	26.9	10.5	176	25.2	10.9	0.140	171	26.4	10.6	178	25.7	10.9	0.540
Addresses (#)	200	1.5	1.8	208	1.4	1.2	0.756	202	1.4	1.2	206	1.5	1.8	0.337
Region (#)	200	4.9	2.9	208	4.7	2.8	0.463	202	5.1	2.8	206	4.5	2.8	0.031
East	200	21.5	41.2	208	17.3	37.9	0.285	202	19.8	39.9	206	18.9	39.3	0.824

Notes: all variables in percent unless stated. P-values taken from t-tests.

Table 7: Balance of covariates across form and endorsed treatments

Table 8: Treatment effects across main outcomes with covariates and fixed effects

	Headline outcomes								**Additional outcomes**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any response		Response quality average		Friendliness average		Formality average		Response length		Mistakes	
Sample: PT	Full	PT	Full	PT	Full	PT	Full	PT	Full	PT	Full	
Foreign	-0.005	-0.007	-0.411*	-0.418*	-0.002	-0.005	0.083	0.091	-43.828	-52.166	-0.021	-0.017
	(0.044)	(0.045)	(0.162)	(0.166)	(0.056)	(0.058)	(0.063)	(0.066)	(47.129)	(48.676)	(0.187)	(0.192)
Female	-0.021	-0.008	-0.072	-0.064	0.044	0.040	-0.055	-0.064	55.097	57.207	0.061	0.107
	(0.042)	(0.042)	(0.155)	(0.158)	(0.054)	(0.056)	(0.060)	(0.063)	(44.988)	(46.301)	(0.178)	(0.182)
Unskilled	0.027	0.035	-0.244	-0.277	0.020	0.033	0.051	0.055	-2.813	3.463	-0.221	-0.265
	(0.041)	(0.042)	(0.152)	(0.156)	(0.053)	(0.055)	(0.059)	(0.062)	(44.327)	(45.747)	(0.175)	(0.180)
Unendorsed	-0.030	-0.017	-0.095	-0.100	-0.059	-0.057	0.013	0.005	-22.960	-7.694	-0.200	-0.139
	(0.041)	(0.042)	(0.151)	(0.155)	(0.052)	(0.054)	(0.059)	(0.061)	(43.997)	(45.328)	(0.174)	(0.179)
Informal	-0.062	-0.058	-0.006	-0.048	0.095	0.104	-0.031	-0.021	-34.991	-38.796	0.063	0.022
	(0.042)	(0.042)	(0.154)	(0.158)	(0.053)	(0.056)	(0.060)	(0.063)	(44.727)	(46.294)	(0.177)	(0.182)
Unemployment	0.018	0.020	-0.022	-0.031	-0.024	-0.025	-0.012	-0.010	2.470	2.411	-0.024	-0.019
	(0.012)	(0.012)	(0.045)	(0.046)	(0.016)	(0.016)	(0.018)	(0.018)	(13.198)	(13.548)	(0.052)	(0.053)
Independent	-0.090	-0.093	-0.201	-0.138	0.009	-0.019	-0.062	-0.081	5.199	-7.971	-0.153	-0.072
	(0.050)	(0.052)	(0.188)	(0.194)	(0.065)	(0.068)	(0.073)	(0.077)	(54.726)	(56.921)	(0.217)	(0.224)
Migrants	0.330	0.230	-0.662	-0.734	-0.170	-0.166	0.620	0.765	-442.393	-620.446	-1.297	-2.315
	(0.376)	(0.381)	(1.385)	(1.424)	(0.479)	(0.501)	(0.540)	(0.564)	(403.129)	(417.488)	(1.596)	(1.645)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	408	384	321	306	321	306	321	306	321	306	321	306

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Full**: is the entire sample. **PT** is the entire sample excluding pre-test observations.

Table 9: Treatment effects in independent and centralized agencies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	Ind	Ind	Cent	Cent	Ind	Ind	Cent	Cent	Ind	Ind	Cent	Cent
Foreign	0.014	0.042	-0.009	-0.001	-0.666*	-0.593	-0.333*	-0.316*	-0.096	-0.108	0.033	0.034
	(0.094)	(0.095)	(0.049)	(0.050)	(0.352)	(0.398)	(0.177)	(0.183)	(0.116)	(0.127)	(0.063)	(0.065)
Female	-0.114	-0.161*	0.002	0.010	-0.339	-0.397	-0.016	0.031	0.002	0.046	0.021	0.042
	(0.089)	(0.094)	(0.046)	(0.047)	(0.336)	(0.393)	(0.167)	(0.173)	(0.111)	(0.125)	(0.059)	(0.062)
Unskilled	-0.002	-0.009	0.031	0.040	-0.312	-0.365	-0.252	-0.243	-0.129	-0.083	0.053	0.048
	(0.089)	(0.094)	(0.046)	(0.047)	(0.333)	(0.396)	(0.167)	(0.173)	(0.110)	(0.126)	(0.059)	(0.061)
Unendorsed	-0.083	-0.121	-0.013	-0.019	-0.338	-0.347	-0.078	-0.039	-0.199*	-0.280**	-0.015	-0.008
	(0.089)	(0.093)	(0.046)	(0.047)	(0.335)	(0.419)	(0.166)	(0.171)	(0.111)	(0.134)	(0.059)	(0.061)
Informal	-0.048	-0.120	-0.055	-0.059	0.078	0.206	0.029	-0.112	0.053	-0.008	0.080	0.101
	(0.089)	(0.092)	(0.046)	(0.048)	(0.331)	(0.385)	(0.166)	(0.176)	(0.109)	(0.123)	(0.059)	(0.062)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	105	105	303	303	76	76	245	245	76	76	245	245

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Ind**: is the sample of independent agencies. **Cent** is the sample of centralized agencies.

Table 10: Treatment effects in East and West German agencies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	East	East	West	West	East	East	West	West	East	East	West	West
Foreign	0.032	0.065	-0.017	-0.011	-0.050	-0.070	-0.471**	-0.441**	0.028	0.017	0.003	0.006
	(0.102)	(0.102)	(0.047)	(0.048)	(0.384)	(0.392)	(0.175)	(0.180)	(0.116)	(0.122)	(0.063)	(0.064)
Female	-0.154	-0.147	-0.018	-0.003	-0.107	-0.169	-0.132	-0.080	-0.002	-0.043	0.033	0.051
	(0.100)	(0.098)	(0.044)	(0.045)	(0.390)	(0.391)	(0.164)	(0.171)	(0.118)	(0.122)	(0.059)	(0.061)
Unskilled	0.295**	0.326**	-0.040	-0.033	0.337	0.474	-0.340**	-0.369**	0.132	0.204	-0.019	-0.009
	(0.097)	(0.096)	(0.044)	(0.044)	(0.378)	(0.395)	(0.164)	(0.168)	(0.114)	(0.123)	(0.059)	(0.060)
Unendorsed	0.011	-0.023	-0.050	-0.047	-0.155	-0.114	-0.128	-0.085	-0.062	-0.100	-0.048	-0.046
	(0.097)	(0.096)	(0.044)	(0.045)	(0.357)	(0.352)	(0.164)	(0.168)	(0.108)	(0.109)	(0.059)	(0.060)
Informal	0.112	0.097	-0.100**	-0.099**	-0.238	-0.348	0.109	0.076	0.099	0.146	0.069	0.090
	(0.098)	(0.101)	(0.044)	(0.045)	(0.356)	(0.370)	(0.164)	(0.170)	(0.108)	(0.115)	(0.059)	(0.060)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	79	79	329	329	58	58	263	263	58	58	263	263

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **East** is the sample of East German agencies. **West** is the sample of West German agencies.

Table 11: Treatment effects across informal and formal applicants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	Informal	Informal	Formal	Formal	Informal	Informal	Formal	Formal	Informal	Informal	Formal	Formal
Foreign	-0.033	-0.041	0.027	0.025	-0.156	-0.030	-0.651**	-0.675**	0.051	0.065	-0.040	-0.041
	(0.064)	(0.065)	(0.059)	(0.060)	(0.224)	(0.226)	(0.221)	(0.236)	(0.074)	(0.076)	(0.081)	(0.084)
Female	-0.035	-0.011	-0.017	0.003	-0.184	-0.156	0.034	0.093	0.108	0.124*	-0.062	-0.082
	(0.060)	(0.061)	(0.056)	(0.059)	(0.211)	(0.213)	(0.208)	(0.230)	(0.070)	(0.071)	(0.076)	(0.082)
Unskilled	0.010	0.003	0.037	0.049	-0.162	-0.117	-0.330	-0.479**	0.001	0.017	0.033	0.050
	(0.060)	(0.061)	(0.056)	(0.057)	(0.212)	(0.215)	(0.208)	(0.225)	(0.070)	(0.072)	(0.076)	(0.080)
Unendorsed	-0.064	-0.064	0.004	0.018	0.133	0.146	-0.366*	-0.360	0.023	0.031	-0.133*	-0.156*
	(0.060)	(0.060)	(0.056)	(0.057)	(0.211)	(0.210)	(0.208)	(0.224)	(0.070)	(0.070)	(0.076)	(0.080)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	206	206	202	202	157	157	164	164	157	157	164	164

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Informal** is the sample of informal applicants. **Formal**: is the sample of formal applicants.

Table 12: Treatment effects across endorsed and unendorsed applicants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	Unendo	Unendo	Endo	Endo	Unendo	Unendo	Endo	Endo	Unendo	Unendo	Endo	Endo
Foreign	-0.002	-0.000	-0.003	-0.018	-0.528**	-0.488**	-0.275	-0.298	-0.096	-0.074	0.101	0.083
	(0.062)	(0.065)	(0.061)	(0.062)	(0.217)	(0.234)	(0.228)	(0.234)	(0.082)	(0.089)	(0.072)	(0.074)
Female	-0.093	-0.090	0.044	0.071	-0.173	-0.241	0.027	0.093	0.071	0.090	-0.023	-0.036
	(0.058)	(0.061)	(0.057)	(0.060)	(0.205)	(0.217)	(0.214)	(0.232)	(0.077)	(0.082)	(0.068)	(0.074)
Unskilled	0.016	0.033	0.032	0.044	-0.015	0.004	-0.478**	-0.444*	-0.078	-0.050	0.107	0.085
	(0.058)	(0.059)	(0.057)	(0.059)	(0.205)	(0.214)	(0.214)	(0.225)	(0.077)	(0.081)	(0.068)	(0.071)
Informal	-0.088	-0.100	-0.020	-0.025	0.307	0.287	-0.200	-0.284	0.146*	0.170**	0.000	0.029
	(0.058)	(0.061)	(0.057)	(0.059)	(0.204)	(0.221)	(0.214)	(0.225)	(0.077)	(0.084)	(0.068)	(0.071)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	208	208	200	200	161	161	160	160	161	161	160	160

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Unendo**: is the sample without a lawyer. **Endo**: is the sample with a lawyer.

Table 13: Treatment effects among skilled and unskilled applicants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	Unskill	Unskill	Skill	Skill	Unskill	Unskill	Skill	Skill	Unskill	Unskill	Skill	Skill
Foreign	-0.125**	-0.122**	0.115*	0.115*	-0.335	-0.271	-0.495**	-0.526**	0.072	0.070	-0.071	-0.084
	(0.060)	(0.060)	(0.062)	(0.063)	(0.223)	(0.232)	(0.224)	(0.233)	(0.074)	(0.079)	(0.081)	(0.084)
Female	-0.040	-0.016	-0.010	-0.040	-0.019	0.051	-0.123	-0.196	-0.054	-0.049	0.096	0.114
	(0.057)	(0.058)	(0.058)	(0.061)	(0.215)	(0.224)	(0.206)	(0.224)	(0.072)	(0.077)	(0.074)	(0.080)
Unendorsed	-0.038	-0.030	-0.020	-0.019	0.112	0.222	-0.367*	-0.423**	-0.144**	-0.130*	0.034	0.022
	(0.057)	(0.057)	(0.058)	(0.059)	(0.215)	(0.223)	(0.206)	(0.212)	(0.072)	(0.076)	(0.074)	(0.076)
Informal	-0.066	-0.097*	-0.041	-0.034	0.126	0.053	-0.015	-0.070	0.055	0.059	0.092	0.120
	(0.057)	(0.058)	(0.058)	(0.060)	(0.215)	(0.232)	(0.206)	(0.216)	(0.072)	(0.080)	(0.074)	(0.077)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	201	201	207	207	161	161	160	160	161	161	160	160

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Unskill:** is the sample of unskilled applicants. **Skill:** is the sample of skilled applicants.

Table 14: Treatment effects across female and male applicants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Any Response				Response quality average				Friendliness average			
Sample	Female	Female	Male	Male	Female	Female	Male	Male	Female	Female	Male	Male
Foreign	0.072	0.070	-0.079	-0.086	-0.514**	-0.456*	-0.311	-0.387*	0.066	0.055	-0.054	-0.040
	(0.062)	(0.062)	(0.060)	(0.064)	(0.228)	(0.238)	(0.221)	(0.231)	(0.085)	(0.091)	(0.069)	(0.072)
Unskilled	0.009	0.030	0.037	0.049	-0.214	-0.191	-0.290	-0.239	-0.050	-0.058	0.077	0.111*
	(0.058)	(0.059)	(0.057)	(0.059)	(0.212)	(0.230)	(0.211)	(0.215)	(0.079)	(0.088)	(0.066)	(0.067)
Unendorsed	-0.098*	-0.124**	0.037	0.070	-0.228	-0.235	-0.019	0.066	-0.016	-0.004	-0.101	-0.121*
	(0.058)	(0.059)	(0.057)	(0.060)	(0.211)	(0.217)	(0.211)	(0.224)	(0.079)	(0.083)	(0.066)	(0.070)
Informal	-0.065	-0.040	-0.047	-0.072	-0.058	-0.084	0.161	-0.030	0.160**	0.166*	-0.009	0.005
	(0.058)	(0.059)	(0.057)	(0.061)	(0.211)	(0.218)	(0.210)	(0.228)	(0.079)	(0.084)	(0.066)	(0.071)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	206	206	202	202	160	160	161	161	160	160	161	161

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. **Female** is the sample of female applicants. **Male** is the sample of male applicants.

Table 15: Treatment effects when dealing with missingness in Appeals variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any Response			Response quality average			Friendliness average		
Sample	Full	App I	App II	Full	App I	App II	Full	App I	App II
Foreign	-0.006	0.012	0.012	-0.408**	-0.252	-0.269	0.005	0.037	0.033
	(0.043)	(0.048)	(0.048)	(0.157)	(0.172)	(0.172)	(0.055)	(0.061)	(0.061)
Female	-0.035	-0.018	-0.018	-0.082	-0.143	-0.137	0.039	0.030	0.031
	(0.041)	(0.045)	(0.045)	(0.150)	(0.161)	(0.162)	(0.052)	(0.057)	(0.057)
Unskilled	0.023	0.033	0.033	-0.251	-0.260	-0.270	0.006	0.026	0.024
	(0.041)	(0.045)	(0.045)	(0.149)	(0.161)	(0.162)	(0.052)	(0.057)	(0.057)
Unendorsed	-0.034	-0.013	-0.013	-0.130	-0.180	-0.153	-0.050	-0.070	-0.064
	(0.041)	(0.045)	(0.045)	(0.148)	(0.161)	(0.161)	(0.052)	(0.057)	(0.057)
Informal	-0.056	-0.062	-0.062	0.057	-0.056	-0.030	0.081	0.058	0.064
	(0.041)	(0.045)	(0.045)	(0.148)	(0.161)	(0.161)	(0.052)	(0.057)	(0.057)
Unemployment	-0.001	0.004	0.004	-0.045	-0.043	-0.049	0.003	0.007	0.006
	(0.007)	(0.007)	(0.007)	(0.024)	(0.026)	(0.026)	(0.008)	(0.009)	(0.009)
Independent	-0.080	-0.083	-0.083	-0.245	-0.254	-0.260	0.040	0.056	0.055
	(0.047)	(0.052)	(0.052)	(0.174)	(0.192)	(0.193)	(0.061)	(0.068)	(0.068)
Migrants	0.380	0.258	0.263	0.148	0.510	0.185	-0.553*	-0.389	-0.458
	(0.221)	(0.247)	(0.240)	(0.801)	(0.862)	(0.849)	(0.280)	(0.307)	(0.301)
Appeals		0.017			-1.459			-0.311	
		(0.217)			(0.776)			(0.276)	
Observations	408	349	349	321	272	272	321	272	272

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$. ***Full***: is the benchmark sample.
App I benchmark with controlling for Appeals.
App II benchmark without Appeals, but restricted to observations where it is available.

Table 16: Treatment effects using imputations for missingness in Appeals variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Response quality average				Friendliness average			
Sample	without	with	mean	multiple	without	with	mean	multiple
Foreign	-0.408**	-0.252	-0.398*	-0.391*	0.005	0.037	0.007	0.008
	(0.157)	(0.172)	(0.157)	(0.157)	(0.055)	(0.061)	(0.055)	0.055
Female	-0.082	-0.143	-0.088	-0.092	0.039	0.030	0.038	0.038
	(0.150)	(0.161)	(0.149)	0.149	(0.052)	(0.057)	(0.052)	0.052
Unskilled	-0.251	-0.260	-0.247	-0.243	0.006	0.026	0.007	0.008
	(0.149)	(0.161)	(0.148)	0.148	(0.052)	(0.057)	(0.052)	0.052
Unendorsed	-0.130	-0.180	-0.148	-0.159	-0.050	-0.070	-0.053	-0.055
	(0.148)	(0.161)	(0.148)	0.148	(0.052)	(0.057)	(0.052)	0.052
Informal	0.057	-0.056	0.041	0.029	0.081	0.058	0.078	0.076
	(0.148)	(0.161)	(0.148)	0.149	(0.052)	(0.057)	(0.052)	0.052
Unemployment	-0.045	-0.043	-0.041	-0.040	0.003	0.007	0.003	0.004
	(0.024)	(0.026)	(0.024)	0.024	(0.008)	(0.009)	(0.008)	0.008
Independent	-0.245	-0.254	-0.244	-0.243	0.040	0.056	0.040	0.041
	(0.174)	(0.192)	(0.174)	0.174	(0.061)	(0.068)	(0.061)	0.061
Migrants	0.148	0.510	0.409	0.480	-0.553*	-0.389	-0.502	-0.490
	(0.801)	(0.862)	(0.811)	(0.819)	(0.280)	(0.307)	(0.284)	0.286
Appels		-1.459	-1.387	-1.489		-0.311	-0.271	-0.281
		(0.776)	(0.776)	(0.834)		(0.276)	(0.272)	0.278
Observations	321	272	321	321	321	272	321	321

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$.
Without: models without Appeals variable. ***With***: models with Appeals variable.
Mean: models where missing Appeals observations are imputed with mean imputation.
Multiple: models where missing Appeals observations are imputed using multiple imputations.

Table 17: Treatment effects when assigning 0s to non-responses

	Headline outcomes									Additional outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Response quality average			Friendliness average			Formality average			Response length			Mistakes		
Sample:	Bench	Non-Missing		Bench	Non-Missing		Bench	Non-Missing		Bench	Non-Missing		Bench	Non-Missing	
Foreign	-0.410**	-0.312*	-0.322*	0.003	0.002	-0.007	0.077	0.054	0.044	-40.655	-31.760	-30.937	-0.016	0.002	-0.001
	(0.158)	(0.143)	(0.142)	(0.055)	(0.181)	(0.181)	(0.062)	(0.181)	(0.181)	(46.399)	(44.950)	(44.980)	(0.179)	(0.150)	(0.150)
Female	-0.072	-0.100	-0.112	0.023	-0.065	-0.095	-0.097	-0.158	-0.193	20.651	3.054	6.618	0.045	0.018	0.032
	(0.149)	(0.135)	(0.135)	(0.052)	(0.171)	(0.172)	(0.058)	(0.171)	(0.172)	(43.672)	(42.384)	(42.736)	(0.169)	(0.141)	(0.142)
Unskilled	-0.251	-0.129	-0.136	0.013	0.125	0.123	0.068	0.164	0.162	17.450	33.357	31.081	-0.238	-0.153	-0.162
	(0.149)	(0.135)	(0.134)	(0.052)	(0.171)	(0.171)	(0.059)	(0.171)	(0.170)	(43.705)	(42.387)	(42.412)	(0.169)	(0.141)	(0.141)
Unendorsed	-0.129	-0.142	-0.151	-0.055	-0.150	-0.160	-0.004	-0.104	-0.115	-33.014	-42.038	-40.107	-0.223	-0.208	-0.207
	(0.148)	(0.135)	(0.134)	(0.052)	(0.171)	(0.171)	(0.058)	(0.171)	(0.170)	(43.629)	(42.389)	(42.446)	(0.168)	(0.141)	(0.141)
Informal	0.049	-0.035	-0.041	0.078	-0.139	-0.146	-0.034	-0.228	-0.235	-34.447	-56.627	-55.794	0.054	-0.012	-0.013
	(0.148)	(0.135)	(0.134)	(0.052)	(0.171)	(0.171)	(0.058)	(0.171)	(0.170)	(43.607)	(42.382)	(42.400)	(0.168)	(0.141)	(0.141)
Controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	321	408	408	321	408	408	321	408	408	321	408	408	321	408	408

SEs in parentheses. * $p < 0.05$, ** $p < 0.01$.
Bench: refers to the model of Table 3.
Non-missing: refers to models where non-responses are assigned 0, instead of a dot for missing data

7.2. Figures

Figure 5: Effect of treatments on response dummy using randomization inference

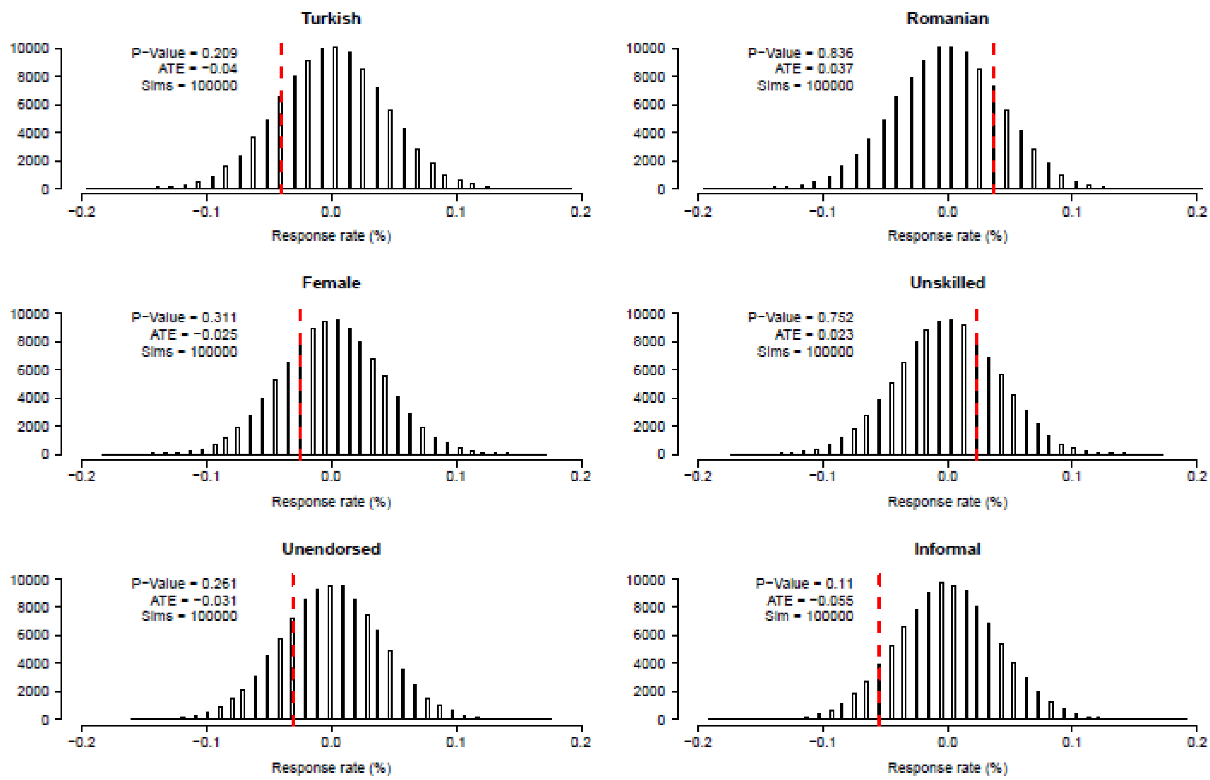


Figure 6: Effect of foreign treatment on quality measure using randomization inference

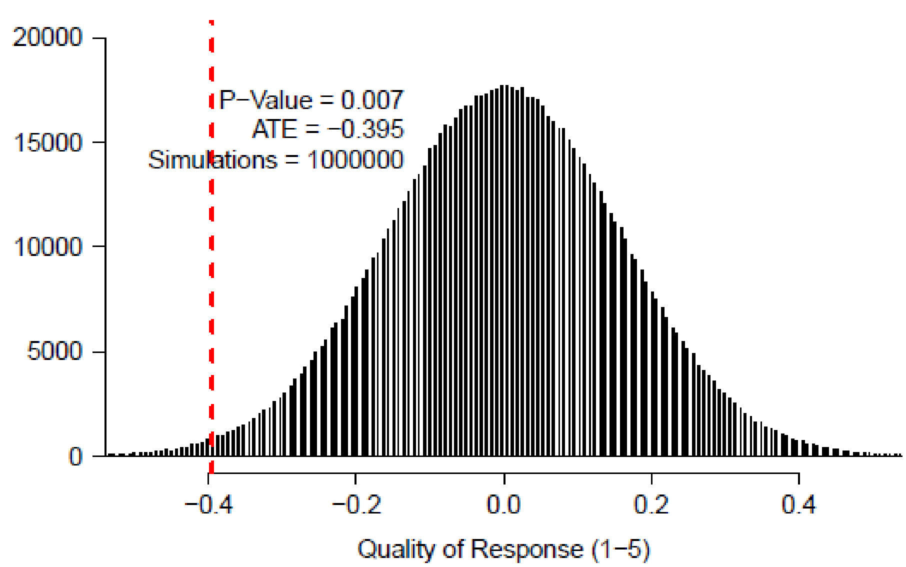


Figure 7: Effect of treatments on response quality using randomization inference

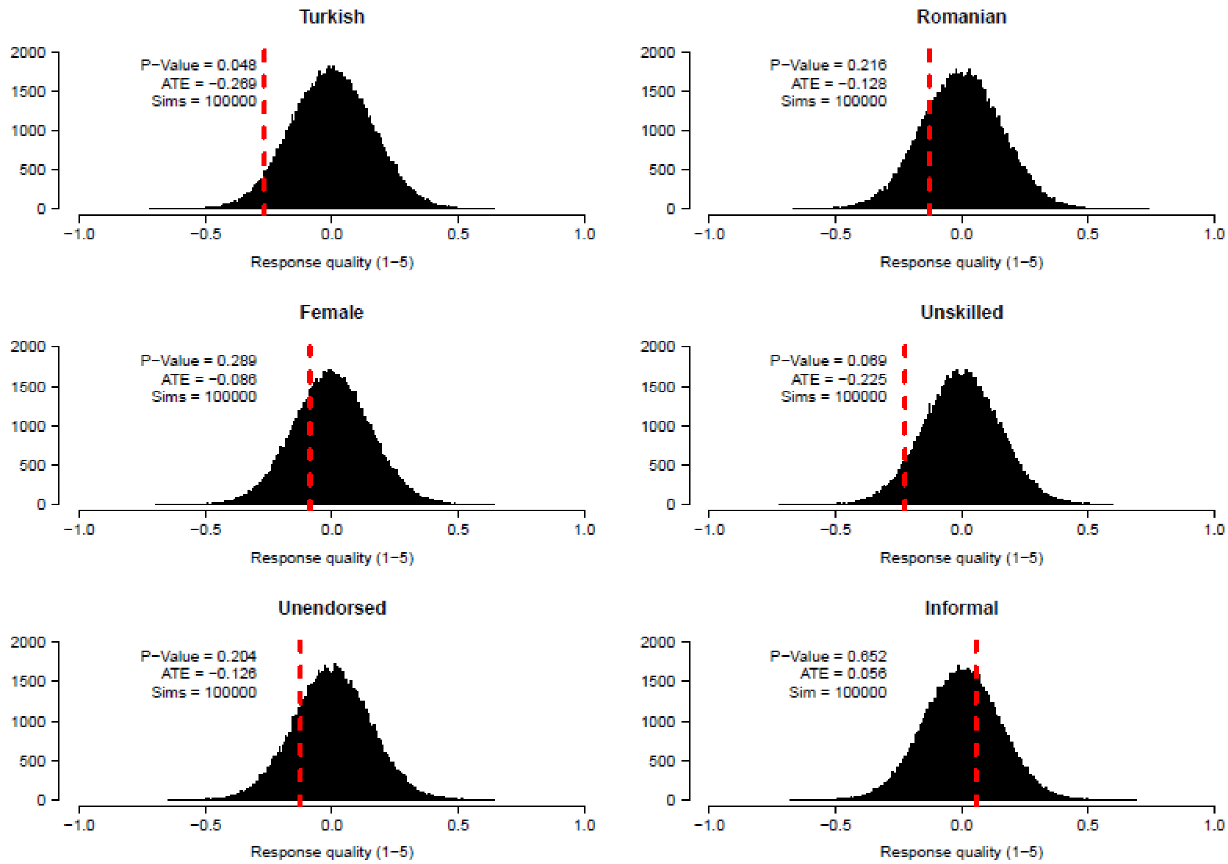


Figure 8: Effect of treatments on friendliness measure using randomization inference

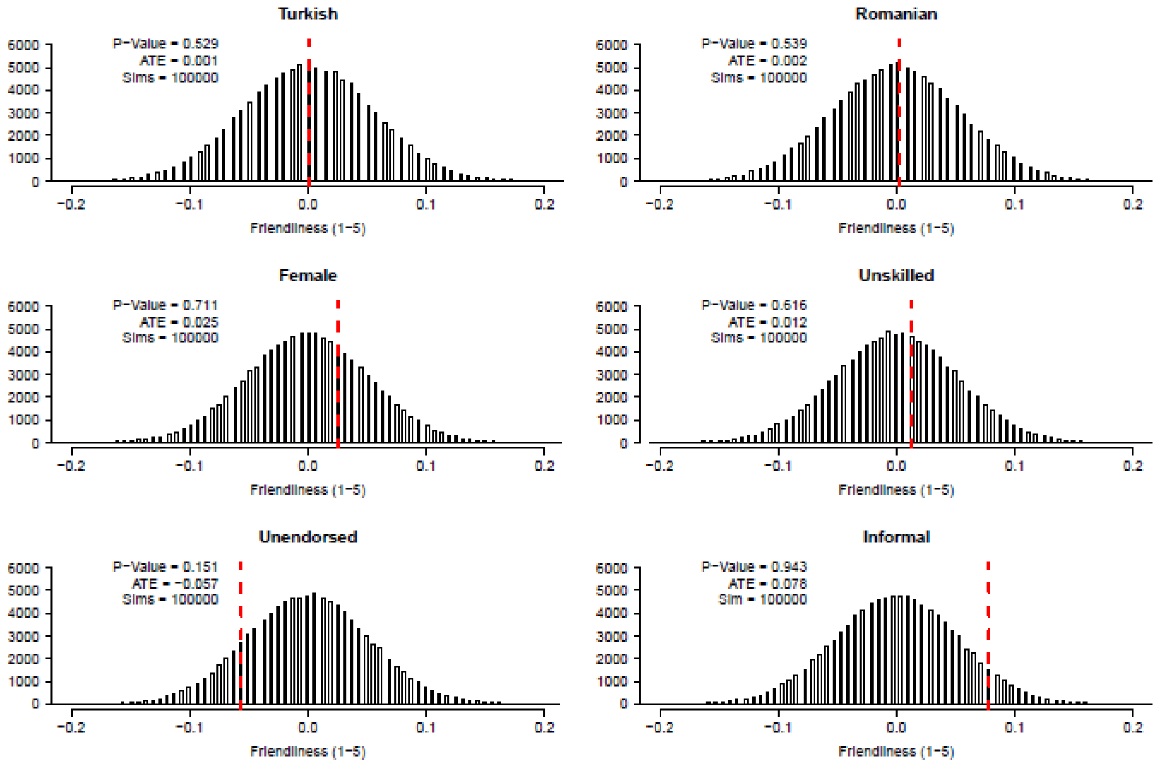


Figure 9: Effect of treatments on formality measure using randomization inference

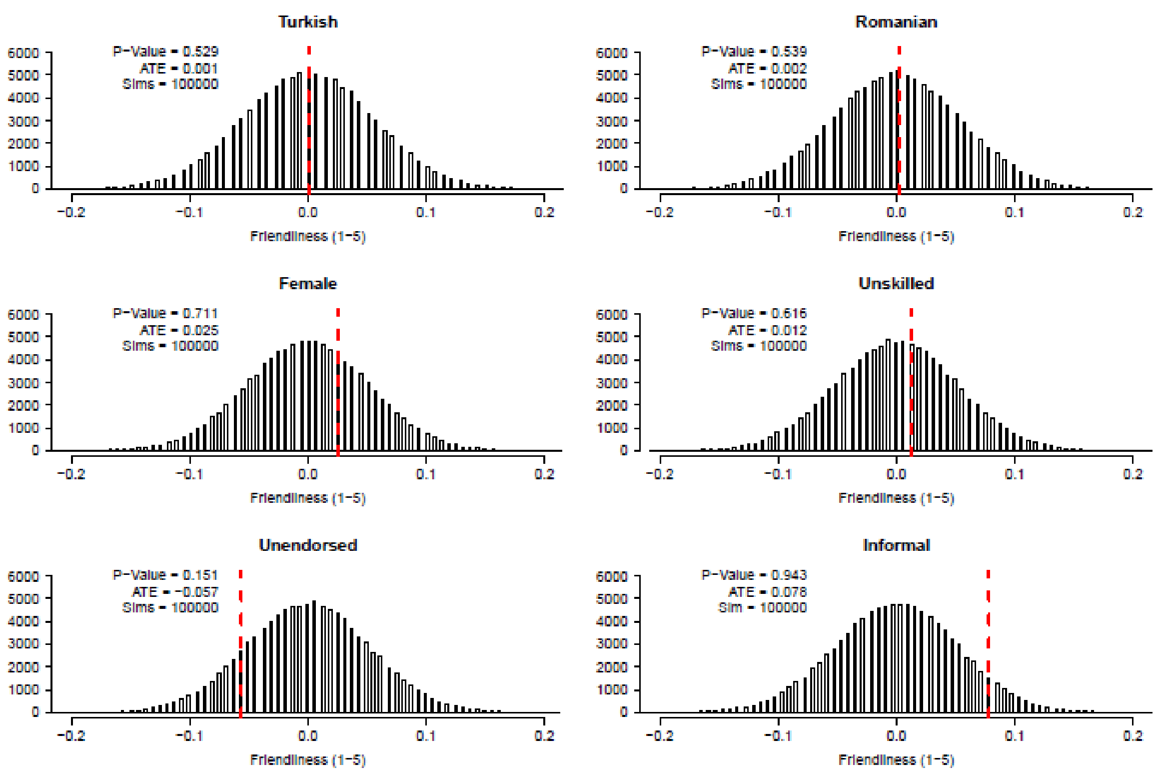
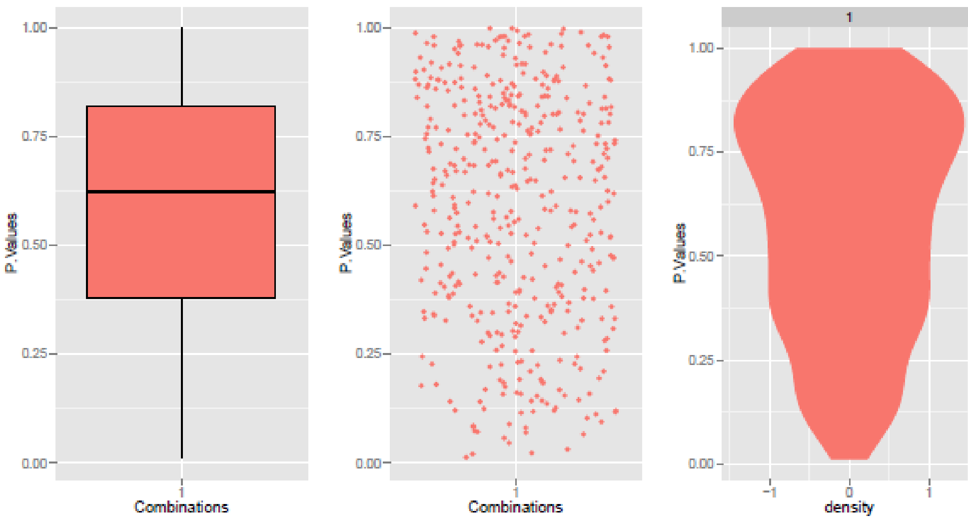


Figure 10: P-values of t-tests comparing means of eight covariates across all 55 unique treatment comparisons.



7.3. Instrument

Formal Instrument English

"Dear Sir / Madam,
My name is [A/B] and I have a question about Hartz 4. I worked a few years as an independent [C]. Now, I need to close my business because I do not have enough customers and want to apply for Hartz 4. But a friend [D] has said that this could be complicated because I live with friends in an apartment. Before I go to your office, I therefore want to ask if I need to bring my roommates documents, too? Or are my own papers sufficient? And what papers do I need to bring exactly?
I would appreciate your response via email.
Many thanks, [A/B]"

Informal Instrument English

Hello,
My names [A/B] and I have a question about harz4. I worked a few years as an independent [C].. Now, I need to close my business because I do not have enough customers! And I want to apply for harz4. But a friend [D] said that may be complicated because I live with friends in an apartment? Before I go to your office I therefore want to ask if I need to bring roommates documents? Or are my papers sufficient?
And what papers do I need to bring exactly??
Please write me an email. Thanks [A/B]

Formal Instrument German

Sehr geehrte Damen und Herren, ich heisse [A/B] und habe eine Frage zu Hartz 4. Und zwar war ich einige Jahre als [C] selbstständig. Jetzt muss ich mein Geschäft wegen zu wenig Kunden schliessen und will Hartz 4 beantragen. Aber ein Freund [D] hat gesagt dass das kompliziert werden könnte, weil ich mit Freunden in einer Wohngemeinschaft wohne. Bevor ich zum Amt komme wollte ich deshalb fragen, ob ich über meine Mitbewohner auch Dokumente mitbringen muss. Oder reichen die Papiere über mich? Und welche Papiere brauche ich genau?
Über eine Antwort per Email würde ich mich sehr freuen.
Vielen Dank, [A/B]

Informal Instrument German

Hallo, ich heisse [A/B] und habe eine frage zu harz 4. Und zwar war ich einige jahre als [C] selbstständig.. Jetzt muss ich mein Geschäft wegen zu wenig Kunden schliesen! Und ich will harz4 beantragen. Aber ein freund [D] hat gesagt das ist vieleicht kompliziert weil ich nämlich mit freunden in eine wohngemeinschaft wohne? Bevor ich zum amt gehe will ich sie deswegen fragen ob sie von meinen mitbewohnern auch dokumente brauchen. Oder reichen die papiere über mich? Und welche papiere brauche ich genau??
Bitte schreiben sie mir eine email. Danke [A/B]

IV. TAX CAPS AND POLITICAL PARTICIPATION IN THE U.S.[30]

Abstract

A growing literature argues that citizen's political participation is motivated at least in part by the state's power to levy taxes on them. In line with that argument, it has been suggested that caps on local taxation decrease turnout in local elections. This in turn implies that there might be a causal relationship between the spread of tax caps in the U.S. states and decreasing turnout in local elections in the past decades. This paper uses county-level tax restrictions on municipalities in Illinois and micro-level turnout data from over 7 million voters to investigate this hypothesis systematically. Using panel regression models with and without matching adjustment, I find no statistically discernible effect of tax caps on turnout. This might be due to the fact that the fiscal effects of the Illinois tax cap only accrued slowly over time, potentially dampening behavioral responses.

30 I would like to thank Olle Folke, Yotam Margalit and Justin Phillips for their advice and guidance, and Bruce Cain, Donald Green, Jens Hainmüller, David Nickerson, Dick Simpson, and participants at the 2014 Meeting of the Midwest Political Science Association and the 2014 Alexander Hamilton Center NYU Graduate Student Conference on Political Economy for helpful comments and suggestions. Special thanks are due to Wolfgang Silbermann for helping me start this project. I am also grateful to staff at the Illinois State Board of Elections, the Illinois Municipal League, the Illinois Department of Revenue, and Stanford University's libraries, who were very helpful and generous with their time.

1. Introduction

Taxation is one of the most significant links between individuals and the state. A sizable literature in political economy has investigated aspects of this relationship in democratic polities. Bates and Lien (1985) developed a seminal theory of representative democracy based on a quid-pro-quo between citizens as a tax base and a government in need of revenue offering them some decision-making power over policy in exchange for taxation. More recently, several studies have argued that government taxing power is an important motive for political participation. Kasara and Suryanayaran (2014) argue based on cross-country evidence that the patterns of turnout differences between affluent and poor citizens are explained by the taxing capacities of governments. In their argument, affluent citizens turn out to vote if the government can tax them in order to avoid an increasing tax burden, but they leave voting to poor citizens if the threat of taxation is not imminent. In a similar vein, experimental work by Paler (2013) shows that citizens are more willing to monitor their governments if government revenue is based on taxation as opposed to windfalls. If government's taxing power motivates citizens to participate in the political process, then reductions in or restrictions of this taxing power should result in lower political participation.

This logic suggests that there might be a causal connection between two distinct phenomena that have occurred in the United States over the past decades: the spread of tax caps in U.S. states on the one hand, and the decline of turnout in local elections on the other. The rapid increase in tax caps, or tax and expenditure limitations (TELs) more generally, began with the passage of Proposition 13 in California in 1978. These rules generally restrict the ability of local governments to raise revenue, often targeting property tax collection which is the most important local revenue source. Currently, 46 states have some kind of TEL constraining local governments in place (Mullins and Wallin (2004)).

Concurrently, turnout in U.S. local elections has been declining from its already-low levels. While the lack of a central data repository complicates the empirical description of this phenomenon, Caren's (2007) study of turnout in big U.S. cities suggests that average turnout declined by 7 percentage points between the late 1970s and the early 2000s. Some of this decline seems to be accounted for by institutional factors such as the form of local government and variation in campaign efforts.

Might the spread of tax caps across the U.S. have played a causal role in local turnout decline? Some anecdotal evidence suggests so. In a qualitative study of the effects of Proposition 13 in California, Sokolow (1998) finds that capping property taxation visibly affected participation in local politics. As he argues, "severely limiting local control over the property tax tends to diminish representative democracy at the community level" (ibid., p. 182 ff.).

In this study, I measure the effects of an effective tax cap reform in Illinois on turnout in local elections in order to assess this hypothesis more systematically. Using county-level variation in tax cap status induced by the reform and administrative turnout data from the Illinois voter file, I find that the tax cap did not reduce local turnout appreciably. This result is highly similar across subsets of locales and voters and unaffected by matching adjustments to the sample of control observations. This suggests that the Illinois tax cap did not contribute to local turnout decline.

The remainder of the paper is organized as follows. Section 2 discusses previous research on tax caps, local turnout and mechanisms that could relate the two. Section 3 introduces the Illinois tax cap while section 4 gives an overview of the data and descriptive statistics. Results are presented in Section 5. Section 6 concludes.

2. Tax caps and local turnout

Electoral participation at the local level is typically very low in the United States. Exactly how low it is in the average municipality is unknown due to severe data constraints. However, samples from large cities indicate an average of only about 27% (Caren (2007)) in recent decades, with the bulk of the distribution between 20 and 40% and a negative development over time. This is significantly lower than the average estimated in the first major study of local turnout by Alford and Lee (1968), who found an average of 43.5% turnout in local elections that were held concurrently with state and national races, and 31.2% in elections that were not.

Turnout in American local elections is so low that it actually matters for outcomes. This is not obvious, since a literature on elections to national offices has mostly found turnout not to matter much for election outcomes. For example, Citrin Schickler and Sides (2003) show that U.S. Senate election outcomes would not change very much if turnout was higher in total or equal across groups. However, as Hajnal and Trounstine (2005) show using the example of ethnic minorities in large cities, the partisan and ethnic distribution of representation in local offices would be changed substantially if all groups voted at similar levels. This is because in local elections with lower turnout, there is much more room for skew, and because given sorting patterns, minority groups are much more likely to be numerically important in smaller geographic units.

Scholarship attempting to explain variation in local turnout has focused on broadly institutional factors. Most notably, Progressive Era institutions designed to reduce the power of local political machines, including non-partisan local elections and council-manager government, have been frequently been found to correlate negatively with turnout in many studies including Alford and Lee (1968), Karnig and Walter (1983) and Caren (2007).

An important study by Hajnal and Lewis (2003) corroborates this finding, but also makes an important and more general argument: as the authors argue, “institutional changes that tend to raise the stakes of local elections also increase turnout.” (ibid., p. 645) Using data from a survey of California county clerks, they show that besides the timing of elections and the form of city government, institutional factors such as the privatization of city services and the prevalence of direct democracy are also strongly related to local turnout: outsourced service provision is associated with lower turnout, and direct democracy is associated with higher turnout.

Following this logic, reducing the room for fiscal decision-making at the local level by instituting a binding tax cap should also be expected to decrease local election turnout. As Mullins and Wallin (2004) document, tax caps of various forms have become ubiquitous across the U.S. states in the past decades. A large body of economic research has investigated the effects of these caps, and found them to be generally effective in reducing revenue and expenditure, although effectiveness varies as a function of their design and the economic environment (Poterba and Rueben (1995), Figlio (1998), Cutler, Elmendorf and Zeckhauser (1999), Skidmore (1999)).

Thus, the co-occurence of the spread of tax caps and decreasing local turnout could be more than a coincidence: increasing constraints on local fiscal policy brought about by tax caps might have causally contributed to turnout decline.

Two types of mechanisms could be responsible for such a relationship. First, one could imagine a "stakes mechanism" driven by reduced utility differentials between alternative candidates or programs. Second, there could be a "salience mechanism" driven by the reduced visibility of the tax burden and the increasing complexity of local public finance following tax restrictions.

To clarify the "stakes mechanism", suppose there are two candidates, whose platforms are characterized by promised levels of local public goods provision, in a local election. Public goods provision is fully funded by a lump sum tax on residents, and citizens' utility is a function only of public goods consumption and private consumption. Since spending on public goods can not exceed taxes collected, there is a trade-off between public and private consumption. In a very simple calculus-of-voting framework, voters turn out at the election if their benefits from doing so exceed their costs, where their benefits are the probability of being decisive in the election times the utility difference between the two platforms (Downs (1957), Riker and Ordeshook (1968)). This implies that turnout increases in a voter's chances of being pivotal, and in the voter's utility difference between the alternative candidates. Now, a ceiling (TEL) is imposed on the maximum tax that can be collected. If both proposals imply taxes below the maximum, nothing changes. If one of the proposals was in excess of the new maximum, the difference between the two proposals decreases. If both proposals were in excess of the maximum, the difference becomes zero, and therefore (weakly) decreases. Therefore, under a ceiling that constrains at least one of the platforms, one would expect voters to turn out less following the instatement of a tax cap (see also Andersen, Fiva and Natvik (2014)).

Besides reducing the stakes of elections, TELs may also affect participation through a salience mechanism. Qualitative research by Sokolow (1998) on the politics of local tax restrictions suggests that this was the case for Proposition 13 in California. Based on interviews with local officials in that state, he finds that the strong constraints that Proposition 13 imposed on local fiscal policy-making fundamentally changed local politics. It is worth quoting his hypothesis at some length here:

> *"The search in this state for alternative revenues has transformed a fairly simple local fiscal system into one much more complicated and little understood by elected officials and citizens. In place of the process before Proposition 13 that revolved around annual decisions on property-tax rates by local governments, the revenue side of budgeting now has no central focus. Instead, it is an amalgamation of considerations about numerous smaller revenue sources (user fees, building revenues, sales tax, motor vehicle license fees, and so on) and yearly state fiscal actions. With ongoing uncertainty and less control over revenues, elected governing boards engage less in comprehensive priority-setting and long-range planning than before Proposition 13. Local budgets also receive less public scrutiny. With no controllable target like the property-tax rate to retain their interest and activity, taxpayers seldom turn out for local budget hearings, nor do they communicate directly with their elected representatives on fiscal matters as frequently as in the old days." (ibid., p. 184)*

While he does not utilize the term "salience", Sokolow's hypothesis clearly concerns the visibility of a given tax burden and the complexity of local public finance, and not the stakes of elections in an objective sense. His claim dovetails with more recent behavioural public finance research by Cabral and Hoxby (2013) on the high salience of the property tax. They show that the property tax is the most salient (and least popular) tax in the United States because unlike most taxes it is not paid through withholding, but through relatively rare property tax bills with relatively high bill amounts that require many individuals to save in advance or go into short-term debt to pay. Using arbitrary variation in property tax salience due to different payment methods, they are able to show that higher tax salience has substantive political effects, causing local property tax rates to be lower and more likely to be politically restricted. By the same token, the "salience mechanism" would lead one to expect individuals to participate in local politics more when a given tax burden is made more salient, even if the total tax burden and the stakes of elections remain constant.

Although this hypothesis is plausible, little direct evidence of such a causal relationship is offered in Sokolow's work on tax caps, and it remains difficult of course to systematically disentangle changes brought about by the tax cap from other developments in local politics.

However, research on tax caps in other contexts offers an indication that tax caps may indeed reduce turnout. A recent study by Revelli (2013) investigates the turnout effects of a tax limitation in Italian local elections. Using freezes of a local income tax surcharge rate, he shows that tax freezes are associated with modest turnout declines in a panel analysis. Moreover, looking at a specific freeze that only affected a subset of municipalities, he shows that this result holds up even when using national elections as a comparison group.

3. The Illinois tax cap

Whether U.S. tax caps have similarly detrimental effects on local politics is difficult to measure for several reasons. First, U.S. tax caps often arise endogenously from direct democratic institutions and anti-tax movements, making it hard to identify their own effect. For example, an active and combative anti-tax movement may bring a referendum on a TEL, but may also in itself change the electoral incentives of politicians to change the size of the public sector.

Second, although most of the variation in tax caps is across states, tax caps are highly heterogeneous. There are caps on revenue growth, caps on assessments, caps on tax rates, and caps on expenditures. Some caps are indexed to inflation, and others to a fixed growth factor. Moreover, the ability of local governments to circumvent TELs also diffes based on factors specific to a state's legal environment and the wording of the TEL. This means that any cross-state study of caps would necessarily lump together estimates of the effects of very heterogeneous policies, making it difficult to interpret results.

Thirdly, even if caps were reasonably homogenous, the voting data necessary for a large-N cross-state study would be near-impossible to obtain: as mentioned above, extant studies of local turnout are typically forced to send send out surveys to county clerks in order to obtain local turnout results.

To circumvent these problems, I estimate the effect of the Property Tax Extension Limitation Act (PTELA) in Illinois, which was passed in the summer of 1991. Although PTELA is legally an Illinois Public Law, it never applied to the entire state, but has only applied to so-called non-home-rule municipalities in select counties over time. The tax cap was first applied to the five "collar counties" around Cook County in 1991 by the State Legislature, then to Cook County itself in 1994, and then to other counties following county-level referenda in the years after 1996. I use this within-state variation in tax cap status to estimate the effects of the cap on political participation. As regards the outcome measure, Illinois local elections are held in the spring of odd-numbered years, while state-wide elections are held in even-numbered years. This allows me to use the Illinois voter file to measure local turnout in each municipality. In the remainder of this section, I briefly describe the history and design of the PTELA tax cap in more detail .

The Illinois tax cap was one of the cornerstones of the gubernatorial campaign of Jim Edgar, a Republican who entered the 1990 race to replace James Thompson, who had been governor for 14 years. He had tried to advance it in the General Assembly before the election, but it had not been reported out of committee because of Democratic opposition (Chicago Tribune (2/13/1991)). Upon taking office, Edgar still faced the same democratic majority in the Illinois General assembly. In his first "state of the state" address, he outlined his plans for a tax cap that would limit the growth of property tax collections in all local entities with taxing power. That same day, he also called for a special legislative session of the Assembly, which did not act on the proposal. (ibid.)

In a series of interviews given for an oral history project for the Abraham Lincoln Presidential Library in 2009, 10 years after he decided not to run for reelection in 1999, Edgar described the situation in the following way: "The Democrats didn't laugh at me, but they ignored me. They said, "That isn't going to happen." And the lobby groups like the teachers' unions, they said, "That isn't going to happen."" (Edgar (2010), p. 557) When the Assembly did not react to bills proposed by Republican representatives, Edgar called additional special sessions.

> "We'd call these special sessions, and they really got a little testy because I think one time I made them stay in an extra day or something like that. This was after about the third time they hadn't responded, so they really thought I was grandstanding, because they weren't going to do this. Of course, I kept trying to put pressure on them to do something, or at least looked like I was trying. I didn't want people to say, "Well, he didn't really mean this; he didn't really push hard enough." I began to think, we may not get this. Republican legislators kind of liked it because it was something they could talk about back home. I don't know if they cared about passing it, but it was a great issue to blame the Democrats for holding up" (ibid., 558).

Edgar himself considered the tax cap proposal he made "draconian" because it capped tax collections growth at the *lesser* of 5% or the rate of inflation. According to him, this was a strategically exaggerated demand that he hoped to use in a deal later by allowing 5% growth. In his words, "I thought, we'll compromise on this, but this is a place to start. I'm not going to start with my compromise" (ibid.).

The proposal faced stiff opposition in the House, and Democratic leaders labeled it as "Reaganomics" because of its overall regressive effects (Chicago Tribune (6/2/1991)). Beyond the Democratic party, however, some interest groups also came out strongly in opposition to the plan. Teacher's unions mobilized against the proposal, and the Illinois Municipal League described it as "an encroachment upon the powers of local governments" (Chicago Tribune (2/14/1991)). Municipalities argued they would have to increase non-tax user fees, and reduce public goods provision: "In Kane County, Elgin City Manager Larry Rice said that if the property-tax cap becomes law, the fast-growing suburb would be unable to begin a project to improve several major roads, including Dundee, Highland and Big Timber Roads." (Chicago Tribune (7/16/1991))

While the tax cap was popular with voters and some Democrats in the Senate, it could not have passed against the opposition of the Speaker of the House. However, it also became entangled in the broader budget fight of 1991: towards the end of the fiscal year, there was still no budget, and a lot of uncertainty surrounding revenue because of a temporary income tax surcharge that many representatives wanted to make permanent. Democrats introduced a $1 billion spending bill to fund AFDC, nutrition programs for children and and the indigent even if the budget were to be adopted late. Edgar, demanding a budget solution "in one piece", vetoed it, and Democrats failed to override his veto in the Senate. After that defeat, Democrats sought a compromise on the property tax cap. According to Edgar, Democratic leaders approached him with a suggestion to limit the property tax caps to collar counties surrounding Chicago, under which condition they would back it in the House. To Edgar's own surprise, they did not take issue with the 5%-or-inflation rule, and so for the counties affected, the effect was even more pronounced than Edgar had expected when proposing the cap. (Edgar (2010), p. 560)

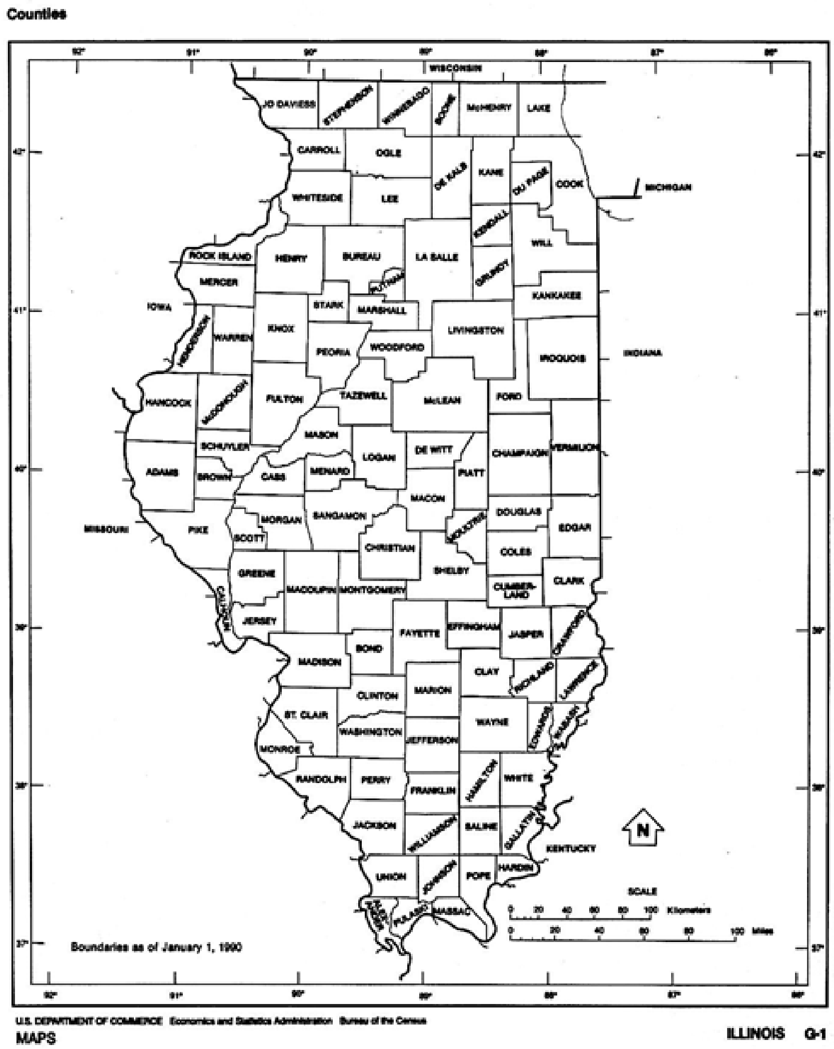


Figure 1: Illinois counties (Census Bureau).

The reasons behind the Democrat's proposal to limit the tax cap to some counties are not entirely clear. One plausible explanation is that the five collar counties (Will, Kane, Lake, McHenry and DuPage, see Figure 1), containing many wealthy suburbs, were (and are) leaning more towards the Republican Party than Chicago and Cook County. Democrats might have cared more about teacher's and municipal employee's unions in Cook county, which were more clearly on their side politically. The focus on the area around Chicago could also be related to the fact that this is where property prices went up the most in the 1990s, thus increasing the value of the property tax base and boosting property tax collections.

After publication of the deal, the Illinois Library Association, AFSCME (The American Federation of State, County and Municipal Employees) the Illinois Municipal League and some other organizations sought to challenge the property tax cap legally, arguing that it would render municipalities unable to provide services they were mandated by law to provide. However, funding for the lawsuit was insufficient due to classic collective action problems (Chicago Tribune (10/15/1991)), and the property tax cap went into effect as Illinois Public Law 87-17 on October 1st, 1991, limiting the growth of total property tax collections (not rates or assessments) to the lesser of 5% or the rate of inflation in the collar counties. Partly because the Democrats had surprised Edgar with their offer, it was one of the more restrictive laws of this kind.

Due in part to the popularity of the tax cap even in Democratic constituencies, the property tax cap was then extended in two waves. First, Cook County came under the tax cap in 1994 after a non-binding referendum in that county had shown overwhelming support. Then, in 1996 the Illinois General Assembly allowed voters in all counties, on proposals from county boards, to hold referenda on whether or not to adopt the tax cap. Successively, 39 of 102 Illinois counties adopted the tax cap between 1994 and 2004, the last year of turnout data in my voter file. 9 counties rejected the tax cap in referenda (Illinois Library Association (2013)).

However, even within capped counties, not all municipalities' tax rates are restricted by PTELA: the cap only applies to non-home rule taxing districts (Illinois Department of Revenue (2012)).

Home rule is a status of local jurisdictions that is regulated in different ways in each state. The Illinois constitution of 1970 affords home-rule municipalities more latitude over local issues ranging from taxation to local debt and investment, unless the General Assembly explicitly restricts these. Municipalities with population above 25,000 are automatically home-rule, while those with lower population can choose to have referenda to make them home-rule. Municipalities with population above 25,000 can also hold referenda to make them non-home-rule. As a result, the tax cap applies to a larger share of municipalities than Illinois residents. Figure 2 displays the share of municipalities under the tax cap over time.

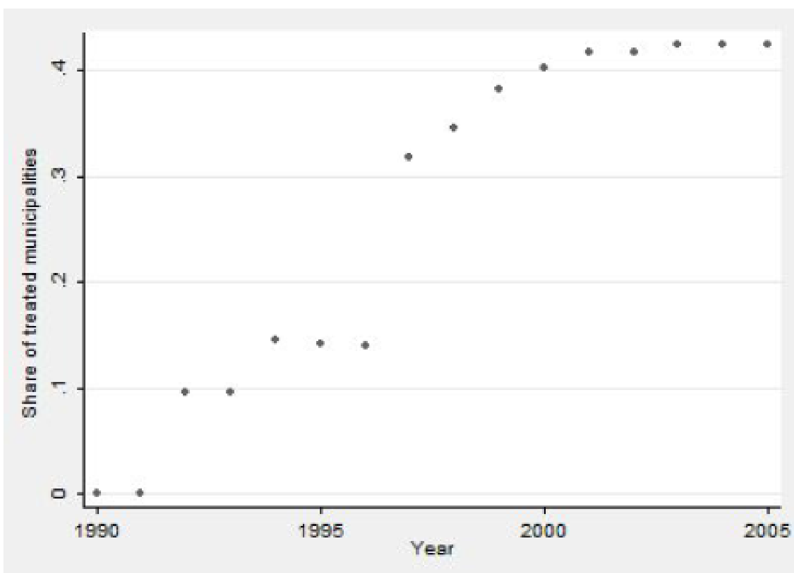


Figure 2: Share of municipalities under the Illinois tax cap between 1990 and 2005.

The Illinois tax cap is widely considered to have been very effective at reducing the growth of property tax revenue in the affected counties. Research has also shown that, relative to the counties not affected, property tax revenue decreased after the reform both for school districts and for municipalities (Dye and McGuire (1997)). Later research by Dye, McGuire and McMillen (2005) also indicates that the county-level caps enacted after 1996 were just as effective as the earlier ones, and notably became more effective over time.

These results accord with the statements and behavior of the relevant actors involved at the time. The tax cap's initiator Jim Edgar was, perhaps unsurprisingly, convinced of its effectiveness, saying “if you look at the numbers, it’s pretty spectacular” (Edgar (2010), p. 561). Moreover, the fervent opposition of relevant interest groups even before passage of the tax cap suggests that its fiscal effects were not impossible to anticipate.

4. Data and descriptive statistics

Throughout the study, I use the 2004 Illinois voter file to measure turnout. It contains information on every voter that was registered to vote in Illinois in 2004, about 7.1 million voters in total. Most importantly, and with equally important exceptions, it contains turnout in primary and general elections for every year (that the voter was registered) between 1990 and 2004. Beyond that, it contains geographic information, the date of voter registration, and voter’s age and gender. It also contains a variable on a voter’s party registration. While Illinois does not require voters to identify their party when they register, voters are coded as Republican or Democratic when they choose either partisan ballot in the primary election. Voters that never vote in primaries, or only vote on referenda and non-partisan races using the nonpartisan ballot, are coded as Independents.

The voter file contains all registered voters in 2004, but it does not necessarily contain the voting history of *all* voters that voted in elections in the early 1990s: Voters that died, moved or were imprisoned could have been purged from the rolls. Voter registration lists are maintained by counties, who are required by Illinois law to verify voter addresses every two years. This is done by sending out new voter ID cards, and tagging those returned as undeliverable by USPS. Those voters tagged are then suspended from voting, but not deleted from the list. Instead, county clerks try to to send a letter to a forwarding address, if available, to verify a move and then delete a voter from the list. If the ID card is undeliverable but the forwarding does not work and the county clerk does not hear from a voter confirming he has moved out of the state, it takes 5 years for a voter to actually be deleted from the list. Therefore, the data give an almost complete, but not entirely complete picture of turnout because of out-migration.

Some of the registration dates in the data are problematic. I generally code voters as registered in an election if their registration date was before the election date (grace period registration allows voters to register up to 3 days prior to an election in Illinois). However, first of all, just under 100,000 observations (out of 7.1 million) have missing registration dates. Among those observations with registration dates, some are coded as having turned out in an election which took place before their initial registration date. For example, in the 1991 election, about 180,000 voters are coded as having voted although they are not coded as registered in 1991. Around 35,000 of these have a missing registration date, but the remaining 145,000 observations (around 2% of total observations) have non-missing but logically inconsistent registration dates. These cases need not be actual data errors, but likely represent in-state movers whose registration date was changed when they moved between counties. Because I do not know whether and to what extent voter's registrations are connected in the voter file, I code voters that voted in any election as having been registered at the time of that election, and keep all registered voters registered in all subsequent elections; if their registration had become invalid, they would be unobserved.

Consolidated local elections in Illinois are generally held on the first Tuesday in April in odd-numbered years, while local primaries are held on the third Tuesday in March. Most but not all municipal elections in Illinois are non-partisan, while township and county elections are partisan. The first local elections recorded in the voter file were held in March and April of 1991, which was the spring before the first wave of the tax cap. In primary elections, voters determine who runs in the general elections, which then determine who holds the local offices (aldermen, village presidents/mayors, clerks, trustees, etc.). This is true both for partisan and non-partisan local elections. In the latter, having more than two nominations for an office triggered a non-partisan primary during the observation period.

For the analyses below, I link voters to jurisdictions using their census block and discard voters that live in unincorporated areas (outside of any municipality). I then aggregate turnout, the fraction of registered voters who turned out in any given election, at the census block level. The final dataset contains 4392 census blocks with an average of about 280 registered voters per block. Besides overall turnout, the data also allows me to calculate turnout for subgroups such as registered Democrats or Republicans, males and females, and old and young voters.

This choice of census blocks as level of analysis is motivated primarily by a desire to correctly code when no election took place in a locale in a given year. For example, when a deceased or retiring council member is replaced, some voters in a municipality may get to vote in the local election while in a neighboring ward in the same municipality, there are no offices to be filled. Aggregating turnout at the census block level avoids averaging across the two wards, which would give the mistaken impression of low city-wide turnout. However, results are very similar when aggregating at the municipality or county level.

Another advantage of using census blocks is that they are most fine-grained level at which covariates are available. I merged Census data on housing and income to the voter file dataset. Table 1 displays descriptive statistics for the outcome variables and covariates.

Table 1: Descriptive statistics.

VARIABLES	(1) N	(2) Mean	(3) S.D.	(4) min	(5) max
Share of owners	35,128	77.57	15.70	0	100
Median income	35,136	50,677	24,951	0	200,000
Population density	35,136	0.000790	0.00103	0	0.00983
Median house value	35,088	81,826	72,176	14,999	500,001
Share of reg. Dems	33,744	0.531	0.130	0.169	1
Turnout - general	31,984	0.336	0.189	0.00243	1
Turnout - primary	8,681	0.111	0.139	0.000334	1
Dem turnout - general	30,045	0.481	0.230	0.00279	1
Dem turnout - primary	6,560	0.210	0.202	0.00117	1
Rep turnout - general	30,147	0.507	0.231	0.00515	1
Rep turnout - primary	6,627	0.227	0.208	0.00134	1
Tax cap	35,136	0.459	0.498	0	1

5. Empirical strategy and results

The most straightforward way to estimate the effect of the tax cap on turnout is to estimate linear regression models with unit and year fixed effects. The identifying assumption of this approach is that in the absence of the tax cap, changes in turnout would have evolved similarly in capped counties and uncapped counties. However, this assumption could plausibly be violated: for example, the tax cap may have been more likely to apply to jurisdictions with positive or negative trends in turnout. This is especially relevant given that assignment to tax cap status was intransparent in the first years of the reform and counties could select into the tax cap beginning in 1996.

Under the assumption that selection into tax cap status was not independent of potential turnout but driven exclusively by observable covariates (selection on observables), one can improve over the fixed-effects approach by using matching methods to adjust the sample of control observations (Ho, Imai, King and Stuart (2007)) . I implement this in two different ways.

First, I use propensity score matching to discard control observations that were very observationally different from treated observations in the baseline year 1991. Since turnout data begins in 1991, this allows for including one year of pre-treatment turnout as a predictor of future tax cap status. I then run the fixed-effects model on the pre-processed sample.

Second, in order to be able to include a richer pre-treatment series of turnout outcomes as predictors of tax status, I restrict attention to the extension of the tax cap through referenda after 1996. Disregarding the six counties that were treated before 1996, I use covariates and pre-treatment turnout in all years prior to 1996 to estimate the propensity score. I then discard unmatched control observations and estimate the fixed-effects model. This strategy is likely to considerably reduce selection bias since any remaining bias would have to be unrelated to pre-treatment turnout, but affect turnout after the tax cap. Details on the matching procedures and balance tables are available in the Appendix.

Figure 3 gives an overview of the estimated effects of tax caps on general election turnout in the aggregate and among registered Democrats and Republicans. Results from four specifications are shown: the standard fixed-effects model with (1) and without (2) weighting of observations by the number of registered voters, and the two matching procedures outlined above. Since tax cap status varies at the county level, standard errors are clustered at the county level in all analyses.

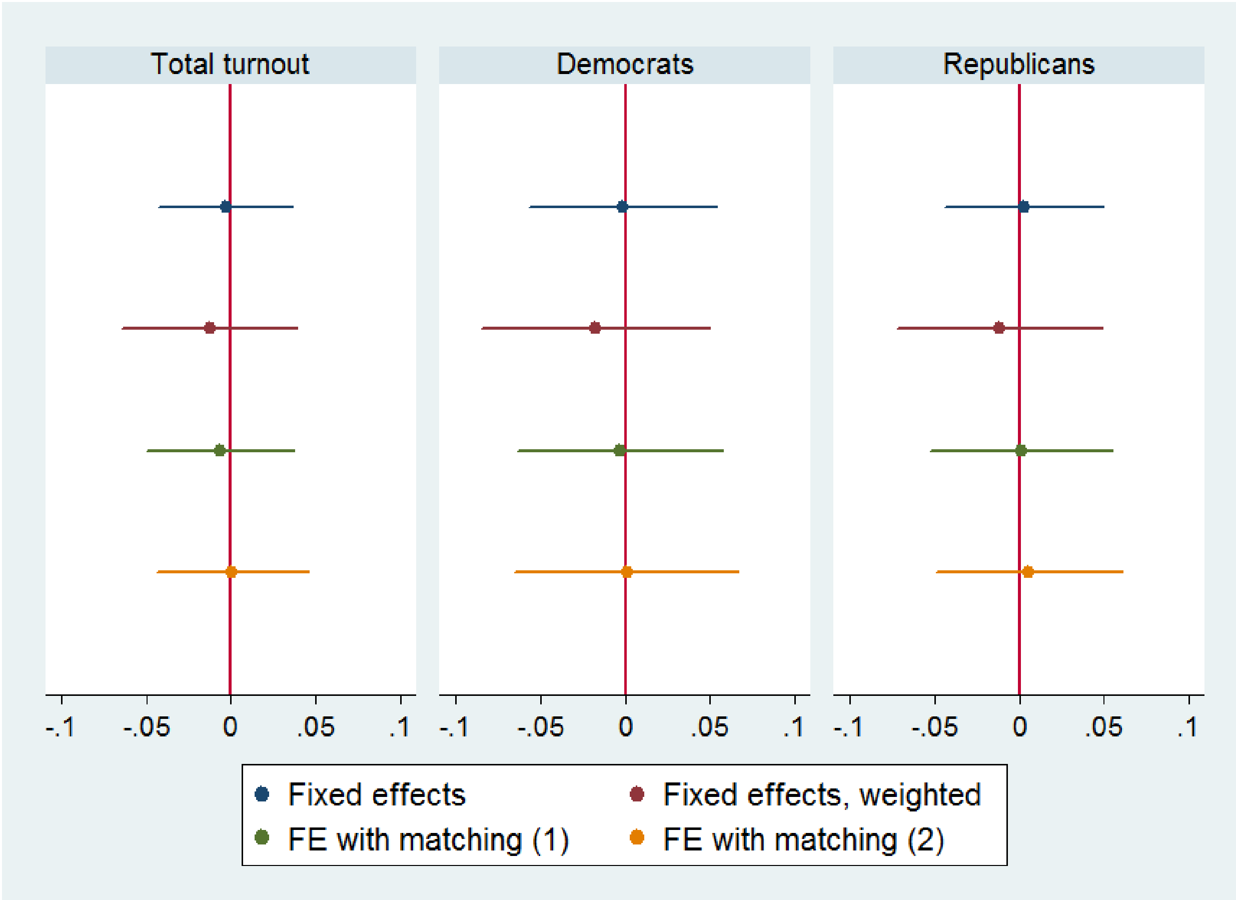


Figure 3: Estimated effects of the tax cap on local turnout in general elections among all registered voters, Democratic voters and Republican voters, according to four main specifications.

The estimated effect of tax caps on turnout in local elections is very close to zero across all specifications. The largest estimate in absolute terms is a mere negative 1.3 percentage points in the fixed-effects specification with weights, and is not statistically distinguishable from zero.

Full results are also reproduced in Tables 2a, 2b, 2c and 2d.

Table 2a: Estimated effects of the tax cap on local turnout on turnout, fixed effects regression specification.

	Total turnout – General election	Democrats – General	Republicans – General	Total turnout – Primary	Democrats – Primary	Republicans – Primary
Tax cap	-0.003	-0.002	0.003	0.004	0.004	0.004
	(0.020)	(0.028)	(0.024)	(0.016)	(0.022)	(0.023)
R^2	0.08	0.10	0.10	0.01	0.02	0.02
N	31,984	30,045	30,147	8,681	6,560	6627

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 2b: Estimated effects of the tax cap on local turnout on turnout, fixed effects regression specification, weighting census blocks by the number of registered voters.

	Total turnout – General election	Democrats – General	Republicans – General	Total turnout – Primary	Democrats – Primary	Republicans – Primary
Tax cap	-0.013	-0.018	-0.011	-0.000	-0.007	-0.007
	(0.026)	(0.034)	(0.030)	(0.016)	(0.024)	(0.024)
R^2	0.06	0.13	0.13	0.01	0.03	0.02
N	31,984	30,045	30,147	8,681	6,560	6,627

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table 2c: Estimated effects of the tax cap on local turnout on turnout, fixed effects regression specification on matching-adjusted sample (see Appendix for details).

	Total turnout – General election	Democrats – General	Republicans – General	Total turnout – Primary	Democrats – Primary	Republicans – Primary
Tax cap	-0.006	-0.003	0.001	0.006	0.011	0.004
	(0.022)	(0.030)	(0.027)	(0.017)	(0.023)	(0.024)
R^2	0.07	0.12	0.12	0.02	0.02	0.02
N	21,653	20,835	20,903	6,719	5,095	5,153

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

This estimate is stable across general across and primary elections, as well as across the voter subgroups among which I can measure turnout using the voter file. As depicted in Table 3 for the benchmark specification, independents, voters younger or older than the median age of registered voters, and male and female voters also do not appear to turn out more or less in response to tax caps.

Table 2d: Estimated effects of the tax cap on local turnout on turnout, fixed effects regression specification on matching-adjusted sample, post-1996 tax cap switches (see Appendix for details).

	Total turnout – General election	Democrats – General	Republicans – General	Total turnout – Primary	Democrats – Primary	Republicans – Primary
Tax cap	0.001	0.001	0.006	0.029	0.033	0.044
	(0.022)	(0.033)	(0.028)	(0.019)	(0.028)	(0.030)
R^2	0.19	0.21	0.20	0.07	0.08	0.08
N	11,850	11,554	11,471	3,594	2,979	2,944

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

One might hypothesize potential turnout effects of tax caps to be larger in locales with a higher share of owners, higher incomes or higher housing values: individuals who own their housing are more likely to be aware of the property tax burden than renters, and richer owners of more expensive housing might care more about property taxes than other people. However, analyses of subsamples obtained by dividing samples at the median of these variables show that estimates are very similar to zero in all subsamples, as shown in Table 4.

Table 3: Estimated effects of the tax cap on local turnout in general elections for sub-samples of registered voters, fixed effects regression specification without weighting.

	Independents	Old voters	Young voters	Male voters	Female voters
Tax cap	-0.002	-0.002	0.002	-0.003	-0.002
	(0.018)	(0.022)	(0.013)	(0.020)	(0.020)
R^2	0.12	0.08	0.05	0.08	0.07
N	28,678	31,702	27,961	31,345	31,415

* p<0.1; ** p<0.05; *** p<0.01

Table 4: Estimated effects of the tax cap on local turnout in general elections for sub-samples of census blocks, fixed effects regression specification without weighting.

	Share of owners above median	Share of owners below median	Income above median	Income below median	Home values above median	Home values below median
Tax cap	-0.001	-0.005	-0.002	-0.003	0.000	-0.001
	(0.018)	(0.024)	(0.024)	(0.022)	(0.030)	(0.019)
R^2	0.07	0.09	0.04	0.12	0.04	0.13
N	15,658	16,326	15,872	16,112	15,916	16,068

* p<0.1; ** p<0.05; *** p<0.01

These null results are stable, but not extremely precise. To appreciate the size of effects that can be ruled out, consider the benchmark point estimate in Table 2a, the 95% confidence interval of which ends at an effect of about 4 percentage points. One recurrent finding in the literature on local turnout is that council-manager government is associated with turnout reductions of about 8% (Caren (2007), Hajnal and Lewis (2003)). Thus, based on the benchmark results in Table 2a, it appears highly unlikely that the true effect of the tax cap on local turnout was more than half as large than the effect of council-manager government in absolute terms.

Additional analyses indicate that the tax cap also seems to have had little effect on the salience of the property tax in local politics. To measure salience, I used an online database ("Access World

News") to count the number of articles and news items mentioning the term "property tax", but not the tax cap itself, in Illinois publications. Sources were assigned to counties based on the address of publication, and yearly article counts were summed up over sources and z-standardized at the county level. Figure 4 provides a descriptive overview of this measure of property tax salience across counties.

Figure 4: Number of articles mentioning "property taxes", but not "tax cap", by county and year in the "Access World News" database, Z-normalized. Lowess estimates

Unfortunately, since most of the digitized data begins in the late 1990s, there are only three instances of counties adopting the tax cap for which newspaper data are available before and after the switches. In order to visualize the development of newspaper mentions of the property tax in those counties before and after implementation of the tax cap, Figure 5 plots the normalized mentions per year in those counties relative to the time of tax cap adoption along with a lowess estimate of the relationship. At least on the basis of this extremely small sample, the salience of the property tax in the media does not appear to decline following institution of the tax cap.

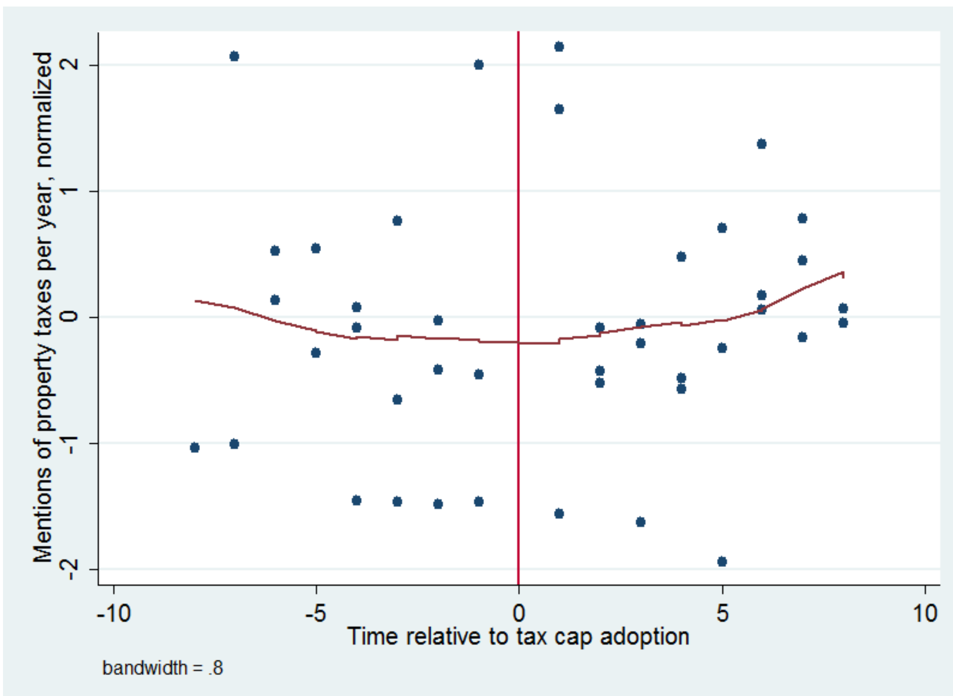


Figure 5: Articles mentioning the property tax per year at the county level (z-normalized), relative to timing of the adoption of the tax cap. Lowess estimate.

6. Conclusion

Research on California's Proposition 13 suggests that the tax cap reduced local political participation. However, this study finds no discernible effects of a major property tax cap in Illinois. What might account for this apparent divergence?

First, there might be no divergence at all; the Illinois tax cap may have had a negative but quantitatively small effect on turnout. Most estimates of its effect presented in the above study are very close to zero, but small effects cannot be ruled out.

Second, differences in the design of the tax cap plausibly matter. The Illinois tax cap may not have had a sufficiently strong and direct economic effect to effectively reduce election stakes or the salience of the property tax. While there is clear evidence that the cap was effective at reducing revenue growth (Dye and McGuire (1997)), and the discussion above shows that relevant actors also perceived it to be effective, this effect was perhaps not as dramatic in terms of overall revenue implications as California's Proposition 13 analyzed by Sokolow (1998).

Moreover, and perhaps crucially, the effects of the Illinois tax cap were spread out over time by design: capping total revenue growth annually means that the total revenue effects of the policy accumulate over the long run. Such incremental changes over a long time horizon, even if quantitatively meaningful, are plausibly more likely to go unnoticed by voters. In contrast, California's Proposition 13 reduced property taxes quite drastically from one year to the next in many jurisdictions, giving voters a clear signal to interpret. The gradually increasing bite of the tax cap may also explain why the salience of the property tax did not decline measurably in those counties for which newspaper data are available.

Third, differences in local institutional context are also likely to condition the effects of tax caps on political participation. For example, tax caps are only likely to reduce local election stakes if local government spending is really a key dimension of electoral competition; the degree to which this is the case likely varies across locales. As another example, property taxes are highly salient and unpopular everywhere (Cabral and Hoxby (2013)), but the degree to which they are more salient than alternative revenue sources varies with the structure of local public finance (e.g. reliance on service fees versus local income taxes). Where other forms of revenue generation are almost as salient as the property tax, the salience mechanism is likely to be muted.

Future research on this subject could assess whether the more abrupt changes engendered by California's Proposition 13 had measurable impacts on turnout in that state. Bringing additional evidence to bear on this question could have important implications for the desirability of tax caps, since lower turnout and citizen monitoring might result in exacerbated agency losses, counteracting one of the key rationales for such policies. (Cutler, Elmendorf and Zeckhauser (1999)).

7. References

Alford, R. and Lee, E. (1968). Voting Turnout in American Cities. *American Political Science Review*, 62(3), 796-813.

Andersen, J. , Fiva, J. and Natvik, G. (2014). Voting When the Stakes are High. *Journal of Public Economics*, 110, 157-166.

Bates, R. and Da-Hsiang D. (1985). A Note on Taxation, Development and Representative Government. *Politics & Society*, 14, 53-70.

Cabral, M. and Hoxby, C. (2013). *The Hated Property Tax: Salience, Tax Rates and Tax Revolts*. National Bureau of Economic Research Working Paper 18514. URL: http://www.nber.org/papers/w18514

Caren, N. (2007). Big city, big turnout? Electoral Participation in American cities. *Journal of Urban Affairs*, 29, 31–46.

Chicago Tribune (1991). *Edgar seeks property-tax cap by March*. February 14th.

Chicago Tribune (1991). *Edgar To Seek Cap on Property Taxes.* February 13th.

Chicago Tribune (1991). *Educators on road to stop Edgar tax cap.* May 7th.

Chicago Tribune (1991). *Key number in property tax reform plans looks like 1992.* June 2nd.

Chicago Tribune (1991). *School officials see `serious cuts' under property tax cap.* May 17th.

Chicago Tribune (1991). *Tax cap challenge is losing ground*. October 15th.

Chicago Tribune (1991). *Tax cap could be burden on services*. July 16th.

Citrin, J., Schickler, E. and Sides, J. (2003). What if Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections. *American Journal of Political Science*, 47, 75–90.

Cutler, D., Elmendorf, D. and Zeckhauser, R. (1999). Restraining the Leviathan: property tax limitation in Massachusetts. *Journal of Public Economics,* 71(3), 313-334.

Downs, A. (1957). *An Economic Theory of Democracy*. New York, NY: Harper.

McGuire, T. and Dye, R. (1997). The Effect of Property Tax Limitation Measures on Local Government Fiscal Behavior. *Journal of Public Economics*, 66(3), 469-487.

Dye, R. McGuire, T. and McMillen, D. (2005). Are Property Tax Limitations More Binding Over Time? *National Tax Journal*, 58(2), 215-225.

Edgar, J. (2010). *Interviews with Jim Edgar.* Abraham Lincoln Presidential Library Illinois Statecraft – Jim Edgar – Oral History Project. URL: http://www2.illinois.gov/alplm/library/collections/oralhistory/illinoisstatecraft/edgar/Pages/EdgerJim.aspx

Figlio, D. (1998). Short-term effects of a 1990s-era property tax limit: panel evidence on Oregon's Measure 5. *National Tax Journal,* 51(1), 55-70.

Hajnal, Z., and Lewis, P. (2003). Municipal institutions and voter turnout in local elections. *Urban Affairs Review,* 38(5), 645-668.

Hajnal, Z. and Trounstine, J. (2005). Where Turnout Matters: The Consequences of Uneven Turnout in City Politics. *Journal of Politics*, 67, 515–535.

Illinois Department of Revenue (2012). *Property Tax Extension Limitation Law: A technical manual*. Mimeo. URL: http://www.revenue.state.il.us/publications/LocalGovernment/PTAX1080.pdf

Illinois Library Association (2013). *Tax Cap Information.* Website. URL: http://www.ila.org/advocacy/tax-cap-information

Karnig, A. and Walter, B. (1983). Decline in municipal voter turnout: A function of changing structure. *American Politics Quarterly*, 11(4), 491–505.

Kasara, K., and Suryanarayan, P. (2014). When do the rich vote less than the poor and why? Explaining turnout inequality across the world. *American Journal of Political Science*, 59(3), 613-627.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199-236.

Mullins, D. and Wallins, B. (2004). Tax and Expenditure Limits: Introduction and Overview. *Public Budgeting and Finance*, 24, 2-15.

Paler, L. (2013). Keeping the public purse: An experiment in windfalls, taxes, and the incentives to restrain government. *American Political Science Review*, 107(4), 706-725.

Poterba, J. and Rueben, K. (1995). The Effect of Property-tax Limits on Wages and Employment in the Local Public Sector. *American Economic Review,* 85(2), 384-389.

Revelli, F. (2013). *Tax Limits and Local Democracy.* Department of Economics and Statistics at the University of Turino Working Paper 36/13.
URL: http://www.est.unito.it/do/home.pl/Download?doc=/allegati/wp2013dip/wp_36_2013.pdf .

Riker, W. A. and Ordeshook, P. (1968). A theory of the calculus of voting. *American Political Science Review,* 62, 25–42.

Skidmore, M. (1999). Tax and expenditure limitations and the fiscal relationships between state and local governments. *Public Choice,* 99, 77-102.

Sokolow, A. D. (1998). The Changing Property Tax and State-Local Relations. *Publius*, 28(1), 165-187.

8. Appendix: Matching procedure

Here I describe the two matching procedures employed in the paper in more detail. In both cases, the strategy is to pre-process the sample by discarding control observations that are very observationally dissimilar to the observations that were affected by the tax cap (Ho, Imai, King and Stuart (2007)).

In the first approach, I use 1991 data to predict whether an observation ever falls under the tax cap. I use turnout in 1991, and the value and square of each of the following variables to predict tax cap status: the median value of housing, median income, the share of owners, the share of registered voters who register as Democrats, and population density. I use nearest-neighbor matching on the propensity-score (with replacement) to identify controls that are observationally similar to observations that experienced the tax cap. I then discard control observations that remain unused and run the benchmark fixed effects regression on this sample. Table 5 displays the balance of the covariates for the matched and unmatched samples. In the unmatched sample, balance is very weak on some key covariates that one might expect to be important, such as the share of registered voters that are Democrats, median income and median housing value. The matching procedure almost universally improves balance, notably on economic covariates and the share of Democrats. However, differences between treated and control observations in terms of pre-treatment turnout, income, housing values and population density remain important even after matching.

Table 5: Balance table, first matching approach.

Variable	Sample	Mean		Standard Deviation	
		Treated	Control	Treated	Control
Turnout – 1991	Unmatched	0.30	0.33	0.20	0.20
	Matched	0.30	0.33	0.20	0.20
Median house value	Unmatched	1.00E+05	42377	83926	15938
	Matched	1.00E+05	88480	83926	41256
Median house value ²	Unmatched	1.70E+10	2.00E+09	3.70E+10	1.70E+09
	Matched	1.70E+10	9.50E+09	3.70E+10	7.20E+09
Median income	Unmatched	56125	38277	28656	10072
	Matched	56125	49370	28656	14078
Median income²	Unmatched	4.00E+09	1.60E+09	5.10E+09	8.40E+08
	Matched	4.00E+09	2.60E+09	5.10E+09	1.40E+09
Share of owners	Unmatched	77.29	76.24	17.7	11
	Matched	77.29	77.34	17.7	14.9
Share of owners²	Unmatched	6287	5934	2415	1543
	Matched	6287	6204	2415	2035
Share of reg. Dems	Unmatched	0.55	0.5	0.1	0.1
	Matched	0.55	0.55	0.1	0.1
Share of reg. Dems²	Unmatched	0.33	0.26	0.2	0.1
	Matched	0.33	0.31	0.2	0.1
Population Density	Unmatched	0.0011	0.0004	0	0
	Matched	0.0011	0.0005	0	0
Population Density²	Unmatched	2.70E-06	5.80E-07	0	0
	Matched	2.70E-06	8.70E-07	0	0

For the second approach, I first discard counties that were already treated before 1995. I then use covariates to predict which observations were treated after 1995; the period when the tax cap was expanded through county-level referenda. For this, I use the same covariates as above, as well as turnout in local elections in the years 1991, 1993 and 1995. These pre-treatment outcomes are arguably the most important covariates, since differences in pre-treatment turnout patterns between capped and uncapped observations are most likely to drive biases in the panel analysis. As Table 6 shows, balance on pre-treatment turnout improves substantially when matching. Moreover, balance on economic covariates is also significantly better than in procedure 1. This is driven primarily by the fact that the observations discarded from counties treated before 1995 are difficult to match, since they are disproportionately wealthy and dense. Just like in procedure 1, I then discard unused control observations and estimate the fixed-effects model on this sample. In interpreting these results, it should be kept in mind that discarding treated observations changes the estimand (ATT), limiting comparability of results.

Table 6: Balance table, second matching approach.

Variable	Sample	Mean		Standard Deviation	
		Treated	Control	Treated	Control
Turnout – 1995	Unmatched	0.25	0.32	0.2	0.2
	Matched	0.25	0.27	0.2	0.2
Turnout – 1993	Unmatched	0.44	0.42	0.2	0.2
	Matched	0.44	0.42	0.2	0.2
Turnout – 1991	Unmatched	0.28	0.32	0.2	0.2
	Matched	0.28	0.27	0.2	0.2
Median house value	Unmatched	52283	43165	24400	16487
	Matched	52283	51935	24400	23134
Median house value ^{2}	Unmatched	3.30E+09	2.10E+09	3.70E+09	1.80E+09
	Matched	3.30E+09	3.20E+09	3.70E+09	3.00E+09
Median income	Unmatched	40945	38608	14323	10277
	Matched	40945	39592	14323	13524
Median income^2	Unmatched	1.90E+09	1.60E+09	1.40E+09	8.70E+08
	Matched	1.90E+09	1.80E+09	1.40E+09	1.20E+09
Share of owners	Unmatched	73.74	76.19	16.5	11.1
	Matched	73.74	71.74	16.5	17.1
Share of owners^2	Unmatched	5709	5927	2107	1546
	Matched	5709	5438	2107	2186
Share of reg. Dems	Unmatched	0.51	0.5	0.1	0.1
	Matched	0.51	0.51	0.1	0.1
Share of reg. Dems^2	Unmatched	0.28	0.26	0.1	0.1
	Matched	0.28	0.26	0.1	0.1
Population Density	Unmatched	0.0007	0.0004	0	0
	Matched	0.0007	0.0006	0	0
Population Density^2	Unmatched	1.20E-06	6.00E-07	0	0
	Matched	1.20E-06	9.70E-07	0	0