How Earmarking Drives Efficiency: A DEA and SFA Approach on Swiss Cantonal Panel Data

Christen Ramon

UNIL | Université de Lausanne

FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES ET D'ADMINISTRATION PUBLIQUE

INSTITUT DE HAUTES ÉTUDES EN ADMINISTRATION PUBLIQUE (IDHEAP)

How Earmarking Drives Efficiency: A DEA and SFA Approach on Swiss Cantonal Panel Data

THÈSE DE DOCTORAT

présentée à la
Faculté de Droit, Des Sciences Criminelles Et D'administration Publique
de l'Université de Lausanne
pour l'obtention du grade de

Docteur en Administration Publique

par
Ramon Christen

Directeur de thèse
Prof. Dr. Nils C. Soguel

Jury
Prof. Dr. Reiner Eichenberger
Prof Dr. Mehdi Farsi
Prof Dr. Christian Thöni

LAUSANNE

2018

UNIL | Université de Lausanne
IDHEAP
Institut de hautes études
en administration publique

IMPRIMATUR

Le Décanat de la Faculté de droit, des sciences criminelles et d'administration publique, sur proposition d'un jury formé des professeurs Nils Soguel, Christian Thoeni, Reiner Eichenberger et Mehdi Farsi, sans se prononcer sur les opinions du candidat, autorise l'impression de la thèse de Ramon Christen intitulée :

How Earmarking Drives Efficiency:

A DEA and SFA Approach on Swiss Cantonal Panel Data

Lausanne, le 13 septembre 2018.

p.o.

Prof. Andreas Ladner
Vice-Doyen de la Faculté de droit,
des sciences criminelles
et d'administration publique

One of the things we have to be thankful for is

that we don't get as much government as we pay for.

(Charles F. Kettering)

Acknowledgements

Certainly, writing a thesis demands a lot of self-motivation and stamina. It would, however, be presumptuous to assume that such a work would be possible without the support of many surrounding persons. First and foremost, I want to express my deep gratitude to my supervisor Prof. Dr. Nils Soguel, who not only backed my ideas from the very beginning but also fostered my other projects. When language barriers decelerated and complicated discussions, he was always patient, and during generously sponsored trips with the team he became more a colleague than a boss.

Further precious and substantial contributions came from the other jury members: Prof. Dr. Reiner Eichenberger, Prof. Dr. Mehdi Farsi, and Prof. Dr. Christian Thöni. I appreciate their effort and the time they have invested to enhance my project.

The newly initiated fund at the IDHEAP encourages PhD students to broaden their horizon and exchange with other universities. Benefitting from this opportunity, I had the chance to spend two months with Prof. Dr. Subal Kumbhakar at the State University New York at Binghamton. Despite his bad health conditions, he always found the time for discussions for which I am very grateful.

Writing a quantitative thesis also means collecting a lot of data, which usually requires obliging contributors. Most of the queried cantons and federal bureaus were very helpful and the staff buckled down to provide me with the necessary data. When requested, the upper management also took the time to listen to my ideas and to propose potential improvements. In particular, I want to express my gratitude to Basilius Scheidegger, who not only challenged my theories but also taught me a lot in the way of working before starting my PhD.

Since the thesis also contains a qualitative part, I relied on the accessibility of non-public information of the canton of Solothurn. Therefore, I thank Dr. Andreas Bühlmann, Beat Loosli, Fritz Brechbühl and Dr. Michael Strebel for the trust they placed in me.

My sincere thanks also go to Michelle Bailat-Jones who put a lot of effort into proofreading.

The working environment is one thing. The private context is another. My family and friends have been constantly supportive – whether by listening to and commenting on heavy subjects or by continually urging me to finally start working. Most importantly, they kept a relaxing atmosphere at home even when I spent all the week in Lausanne. Clearly, Angela is the person to whom I owe the most for this. David and Bene also advised me on mathematical questions.

Contents

Figures

Tables

Acronyms of cantons

AG	Aargau	NW	Nidwalden
AI	Appenzell Innerrhoden	OW	Obwalden
AR	Appenzell Ausserrhoden	SG	St. Gallen
BE	Bern	SH	Schaffhausen
BL	Basel Landschaft	SO	Solothurn
BS	Basel Stadt	SZ	Schwyz
FR	Fribourg	TG	Thurgau
GE	Geneva	TI	Ticino
GL	Glarus	UR	Uri
GR	Graubünden	VD	Vaud
JU	Jura	VS	Valais
LU	Lucerne	ZG	Zug
NE	Neuchâtel	ZH	Zurich

Other acronyms

art.	article
BCC	Banker, Charnes and Cooper (the variable returns to scale DEA model)
CCR	Charnes, Cooper and Rhodes (the constant returns to scale DEA model)
CD	Cobb-Douglas
CHF	Swiss francs
COLS	Corrected Ordinary Least Squares
CRS	constant returns to scale
DEA	Data Envelopment Analysis
DMU	decision making unit
FDH	Free Disposal Hull
FE	fixed effect
GSW	road and route law of May 30, 1996 (*Gesetz über Strassen und Wege des Kantons Zug*)
ha	hectare
HAM	Harmonized Accounting Model
IAFP	Four-year financial blueprint, financial plan (*Integrierter Aufgaben und Finanzplan des Kantons Solothurn*)
IPSAS	International Public Sector Accounting Standards
LSDV	least squares dummy variables
MFHG	model financial budget law of October 10, 2015 (*Musterfinanzhaushaltsgesetz*)
MLE	Maximum Likelihood Estimation
MOLS	Modified Ordinary Least Squares
n.a.	not applicable
NDRS	non-decreasing returns to scale
NIRS	non-increasing returns to scale
no.	number
NPM	new public management
OECD	Organization of Economic Cooperation and Development
OLS	Ordinary Least Squares
RE	random-effects
SFA	Stochastic Frontier Analysis
SVAG	federal heavy vehicle fee law of December 19th 1997 (*Bundesgesetz über die leistungsabhängige Schwerverkehrsabgabe*)
U.S.	United States
TL	translog
VRS	variable returns to scale
WoV-G SO	impact oriented administration management law of January 1, 2015 (*Gesetz über die wirkungsorientierte Verwaltungsführung des Kantons Solothurn*)
WoV-Vo SO	impact oriented administration management decree of July 1, 2015 (*Verordnung zum Gesetz über die wirkungsorientierte Verwaltungsführung des Kantons Solothurn*)

1. Introduction

Earmarking means dedicating specific public revenue streams to a specific purpose. These revenues then accumulate in earmarked funds. Politicians utilize earmarks to serve their clientele while voters are pleased to trace the taxes they pay (Wilkinson 1994). In an exemplary case, a recent ballot on the federal level in Switzerland cemented a road fund which joined its previously installed counterpart for the railroads. [1] Both funds benefit from various earmarked revenues, such as the motor vehicle tax, the petroleum tax, or the fees for the motorway sticker. Revenues from these sources are either totally or partly appropriated to the funds without going through the ordinary budgeting progress. With respect to the road fund, the Swiss Federal Council (2015, p. 2085) argues that "earmarked funds ensure that most of the money accumulated from road traffic fees are allocated to the road infrastructure", which should warrant the user-pays-principle.[2] Counting the motorist as their clientele, the Swiss People's Party (2014) supporting the bill complained about the intended alienation of appropriated funds. They feared that the earmarked revenues could be used for a broader target group beyond their electoral base. On the other side of the political spectrum, the Social Democratic Party (2014), likewise in favor of the bill, wanted to extend the scope of the earmark to fund projects that were useful to the pedestrians and cyclists who are their clientele.

Despite the enthusiasm, politicians also acknowledge the drawbacks that earmarking might entail. When contemplating the railroad fund, the Swiss Federal Council (2012, p. 1610) admits that "from a fiscal policy point of view, fund solutions entail disadvantages as they potentially limit the budget controllability and transparency and they can undermine economic solutions." [2] From this point of view, earmarking prevents an optimal budget allocation. Accordingly, the expenditures for and, consequently, the amount of a public good, do not correspond to the public demand (Deran 1965). For this argument to hold, exponents of public administration literature assume the bureaucrat's nature to be benevolent (Schönenberger 2013). On the contrary, earmarking can also be seen as a mechanism that enforces the government to spend a certain share of the total budget for the intended purpose. In the words of public choice scholars, earmarking is a remedy to tame the Leviathan. As a result, earmarking enhances the budget allocation according to the taxpayers preferences (Buchanan 1963). The two opinions have provoked a rich discussion in the literature, which has not yet come to a conclusion.

The two opposing views show the general disagreement as to whether earmarking forces or impedes the budget allocation to correspond to the taxpayers' preferences. To consider the predicted impacts of earmarking in more detail, some clarification of terminology is vital. Budget allocation means the allotment of public resources to the different public services in response to taxpayer preferences. Moreover, a government is considered technically efficient compared to its peers if it provides the maximum number of public services (i.e. outputs) given its bundle of expenditures (i.e. inputs). In contrast, a government provides its services inefficiently if there is overspending for excessively large offices, unproductive programs, et cetera. Furthermore, allocative efficiency denotes the optimal combination of inputs in consideration of their prices.

---

[1] The vote about the fund for national roads and agglomeration traffic (*Fonds für die Nationalstrassen und den Agglomerationsverkehr*) took place on February 12th 2017 and obtained an approval rate of 61.9 percent. Its rail related pendant, the fund for railroad infrastructure (*Bahninfratrukturfonds*), was accepted by 62.0 percent of voters on February 9th 2014.

[2] Translated from German by the author.

The definition of these terms underline that the technical and allocative efficiency on the one hand differ sharply from the concept of budget allocation on the other hand. While the former two are at the heart of the analysis, the latter is outside its scope. The methods to assess an optimal budget allocation are fundamentally different from an efficiency analysis and a corresponding assessment would therefore focus on a different research area. Accordingly, turning back to the empirical discussion of earmarking and efficiency, the further analysis focusses on the technical efficiency.

Interestingly enough, when the question of whether to earmark transfers turns up, the politics and the scientific literature concur. They agree that earmarked grants negatively affect technical efficiency. Particularly, in 2008, the inefficient use of financial resources led to the reform of the fiscal equalization between the federal state and the cantons. One fundamental aim of the reform was to drastically shrink the share of earmarked transfers.[3] The Swiss Federal Council (2001, p. 2365) considers the cantonal use of earmarked transfers as less efficient than their non-earmarked counterpart. Following this logic, one might wonder why earmarking only unfolds its negative effect on efficiency with grants while other earmarked revenues are unsuspicious. So far, very few scientific papers have discussed the effect that earmarking has on efficiency, and in the practical debate the subject is basically absent.

The neglected connection between earmarking and technical inefficiency in the political debates is surprising. Not only do earmarked revenues amount to significant fractions of the total budget (e.g. the cultural domain benefits on average from 35 percent of earmarked revenues), but also the politicians constantly seek for tools to induce an efficient administration. During the debate on the 2018 federal budget for instance, the terms efficiency and efficient were mentioned more than 40 times covering different parties, departments, and government bodies. While the parliament usually demands a more efficient provision of the public services, the executive claims to act accordingly. A right-wing parliament member expresses his concern about the newly reformed federal management system (*Neues Führungsmodell des Bundes*) that links financial resources with outputs and should then give the parliament more possibilities to control the administration. He doubts the benefits of the new system and he still considers the executive responsible as it "has the necessary expertise to set priorities and to stand up to the administration".[4] The statement exposes the fundamental intricacy involved when the aim is to make an administration more efficient. The external standpoint not only prevents identifying the part of the administration with potential for improvement, but it also challenges any overall determination of efficiency.

A broad literature already works to approach and explain the efficiency of the public sector. All these studies share a common empirical structure. They relate some inputs to some outputs to attain an efficiency measure and they link them with certain environmental variables which potentially affect efficiency. Narbón-Perpiñá & De Witte (2017) provide a systematic overview of the efficiency studies focusing on the local public sector. They demonstrate the immense scope of the studies in terms of geographical coverage, methodological approaches, or the considered inputs and outputs. Moreover, while some authors concentrate on the public sector altogether, others assess the efficiency of some specific domains such as education, health, transportation, et cetera.

[3] While the regime change of 2008 seemingly provides a promising setting to study the effect of earmarking, at least two counterarguments speak against such an identification strategy. First, the reform can be viewed neither as exogenous nor as an unexpected intervention. It responded to critiques of the old inefficiency-promoting system and its implementation had a long lead time. Second, the regime change concerned only the transfers but not all the remaining earmarked revenues. If anything, the reform would only enable testing the effect of earmarked transfers on efficiency.

[4] Hösli, W., November 2017-11-28, winter session of Council of States, Berne; translated from German by the author.

As explaining factors, Narbón-Perpiñà & De Witte (2007) list several categories like social, economic, political, institutional, or financial determinants of efficiency. Among the financial determinants, they mention transfers, debt, or the tax system. However, no study has so far considered earmarked revenues as revenue characteristic, which likewise might affect efficiency. Therefore, this study aims to determine whether earmarking has a positive or negative effect on the efficiency of public service provision and what factors condition this relationship.

While the other financial related efficiency determinants generally arise from the fiscal illusion theory, no comparable line of argument exists for earmarked revenues yet. Arguing with a public choice logic, taxpayers are subject to a fiscal illusion if they fail to correctly perceive the cost of the public service (Buchanan 1967). Underestimated costs seduce bureaucrats to exploit voter generosity and to appropriate personal rents which eventually leads to inefficiency (Wyckoff 1990). The misperception might arise, among others, from the revenue characteristics mentioned above (i.e. debt, transfers etc.)(Oates 1985). Likewise based on the public choice framework, this study develops a formal argument, claiming that ministers prevent the debate during the budgeting process by hiding behind revenues that are earmarked for their department. They are thereby less monitored by their colleagues and thus able to appropriate personal rents. As a result, the government provides its service again less efficiently. This chain of reasoning, henceforth called debate prevention, hypothesizes that higher shares of earmarked revenues lead to a less efficient public service provision.

As an empirical foundation to test the hypothesis, the 26 Swiss cantons provide an appropriate data source. Given their shared legal context paired with ample fiscal autonomy, the Swiss second state layer has already led to various empirical applications in economics, public administration, and political science literature. Particularly in public finance, the data prove valuable as the Federal Finance Administration (2016) harmonizes the figures of the annual accounts, which improves comparability. In addition, a common functional classification splits the revenues and expenditures into different domains of public activities (Conference of cantonal finance directors 2008). This enables testing the hypotheses for specific domains. Since it turns out that earmarked revenues are most dominant in the two domains culture and transportation, the study mainly focuses on these two government functions. In order to account for canton-specific effects and to increase the number of observations, the study rests on panel data from 2000 to 2014.

While the official statistics provide a tremendous amount of quantitative information about the cantons, a statistic regarding earmarking is still lacking. One reason for this gap is the proliferation of different applications of earmarking among cantons. Despite the common chart of accounts, which also outlines the application of earmarking, the actual practice is still rather heterogeneous. This study sets a milestone by defining earmarking in detail and by giving an overview of its relevance in the Swiss cantonal context.

Following numerous previous publications (e.g. Kalb, 2010a; Widmer & Zweifel, 2012; Afonso & Fernandes, 2008; Seifert & Nieswand, 2014), two methods assess the theory and challenge the hypotheses. The non-parametric Data Envelopment Analysis (DEA), on one side, first estimates cantonal efficiency and, in a second step, regresses the revenue characteristics on these efficiency scores. While largely avoiding specification problems of the production frontier, DEA struggles with integrating a stochastic error. The parametric Stochastic Frontier Analysis (SFA), on the other side, incorporates both steps in one estimator and roughly turns the drawbacks of the DEA to advantages and vice versa.

The main results show that earmarked revenues affect efficiency differently depending on which government function is considered. Using data from the culture domain, earmarking indeed impedes an efficient public service provision, which supports the debate prevention hypothesis. On the contrary, earmarking in the transportation domain has a positive effect on efficiency. The principal explanation for this difference comes from the existing literature, which sees three conditions for an earmark to work: first, a strong link between the taxed and the benefitting good; second, the possibility of differentiating the benefitting good from other public goods or services; third, a close match of people utilizing the taxed good and the beneficiary good and an excludability of those who do not contribute (Brazer 1984; Lee & Wagner 1991; McCleary 1991).

However, the empirical evidence found is insufficient to exclude alternative explanations. A second possible reason for the opposing results of the two domains is the inherently differing intentions of the earmarks. While the transportation domain indeed seeks to mimic the user-pays principle of a private good, the earmarks in the cultural domain might instead be driven by a redistribution idea, where efficiency is not a priority. Yet another cause for the unexpected positive effect in the transportation domain even challenges the debate prevention fundamentally. Preventing the debate also means that the administration can focus on effective service provision instead of putting effort into political discussions; in consequence the departments benefitting from earmarked revenue could provide their services more efficiently. A further explanation is a methodological one, since both approaches applied here cannot handle possible endogeneity problems. Briefly speaking, the results show that earmarking is not good or bad per se, but the insights cannot make a final assessment on the true mechanisms leading to a positive or a negative effect on efficiency.

In sum, the study involves three major contributions. First, it formalizes a theory to link earmarked revenues and the efficiency of public service provision. Second, it proposes a measure to quantify the practice of earmarking in the Swiss cantons and provides the respective statistical information for 2000 to 2014. Third, it demonstrates how to apply the two-stage Data Envelopment Analysis in a panel context without contradictory ad-hoc assumptions. In addition, several minor but novel approaches provide alternative practices for old questions. For instance, instead of treating plain transfers, they are split into three tiers according to their intensity of appropriation. Likewise, a single measure to approximate the progressivity of the tax system is developed in order to estimate the effect of the tax illusion in the Swiss cantons. Lastly, a widening of the data sources (to the association of Swiss museums, the federal office of sport, or to the federal office of police) makes it possible to more accurately estimate the cantonal efficiencies.

The study is structured as follows: After the introduction, the literature review presents the previous discussion in the relevant research areas. Hereupon, the theoretical models develop the hypotheses before the subsequent fourth chapter demarcates the empirical context. Next, the revenue characteristics are operationalized in Chapter five and the sixth chapter proposes measures of efficiency. The seventh chapter combines the efficiency estimates with the revenue characteristics and tests the hypotheses. To introduce a further methodological perspective, chapter eight examines the theory qualitatively and, finally, the last chapter concludes. To provide a non-technical summary, an introduction and a conclusion enclose each chapter.

2. Literature review

The current theoretical literature mainly treats two related questions with respect to earmarking: What motivation do governments have to earmark revenues? What influence does earmarking have on public support for new taxes? Regarding the first question, several authors argue that the government utilizes earmarks to bind future administrations to their policies (Brett & Keen 2000; Jackson 2013). At the same time, the incumbent administration might use earmarking to commit to its policy program and thereby to increase its reelection probability (Anesi 2006; Stratmann 2013). Alternatively, some departments might use earmarking as a signal for their competence. According to Glazer and Proost (2007), competent departments bundle a less transparent public service with a more observable one if they provide the latter of high quality. Earmarking thus serves as a credible bundling mechanism. Addressing the second question, Baranzini and Carattini (2017) find in their survey-based study that the acceptance of additional taxes is higher if they are earmarked. They explain the willingness and higher trust the taxpayer puts into the government if she can trace the financial resources. In contrast to these two prominent questions, the aspect treated in this paper has not been developed for a long time, although the research community did not come to a conclusion yet. The further literature review therefore rests largely on older studies that assess the impact of earmarked revenues on efficiency.

The theoretical literature on earmarking in relation to efficiency mainly separates into two camps. On one side, there are the public finance scholars who reject earmarking due to its inflexibility in terms of budget allocation (Lee & Wagner 1991). On the other side, public choice scholars see earmarking as a solution to tame the Leviathan (Buchanan 1963). Ultimately, it is the fundamentally different image of humanity that sunder the two camps, which possibly impedes a final agreement. And yet, critique of the public administration scholars is indicated as most of them insufficiently substantiate their argument, why earmarking should have a negative impact on technical efficiency. Only a few papers unify earmarking and technical efficiency in a formalized theory but they either disregard the supposed underlying mechanism of the causal effect, or the theory covers only a specific field (Martínez 2014; Brennan & Buchanan 1978; Bös 2000). Just as in the theoretical literature, the empirical papers assessing the effect of earmarking on efficiency are very rare.

When it comes to further revenue features that potentially affect efficiency, the theoretical literature is much more developed. Since the seminal paper by Oates (1985), the idea of fiscal illusion disseminated into different directions and encouraged researchers to advance the theory. Today, five sources of fiscal illusion are broadly accepted and three of them have also been tested empirically with technical efficiency as a dependent variable. However, neither grants nor indebtedness furnishes unambiguous evidence for a negative or positive effect on efficiency. Only one study exists with respect to tax complexity, and this hardly suffices for a substantial statement.

2.1. Two views on earmarking

Unlike the still meager literature would suggest, the earmarking of revenue is a widespread practice. Roberts (2002) lists several examples for the US in various policy areas such as security, infrastructure, health, telecommunication, education, environment, and social security. McCleary (1991) provides a similar enumeration for other countries. Particularly in Switzerland, a report of the Swiss Federal Council (2013) names a large scope where public revenues are earmarked: agriculture, air traffic, infrastructure, environment, et cetera. Additionally, the inventive vocabulary used to frame earmarking indicates the popularity of this kind of funding mechanisms. The creativity ranges from ring-fenced fund over recycle revenues, segregated funding, autonomous revenue streams, pork-barrel spending, tax hypothecation, or hypothecation to lockboxing (Roberts 2002; Jackson 2013).

From a political point of view, politicians and citizens may simultaneously relish earmarking revenues. Frequently mentioned among practitioners is the ease to sell new taxes to the citizen in return for a new program, that is earmarking taxes lowers the resistance for increasing them. The argument works on both sides of the political spectrum. Leftwing voters appreciate earmarked revenue as they feel they are paying directly for more or better services. To the libertarian right, earmarking seemingly demonstrates the cost of state services in a more transparent fashion (Wilkinson 1994). In addition, earmarking stabilizes the revenue for a certain service or at least guarantees a minimal financial foundation desired by specific interest groups from either political camp (Brazer 1984). For further and more profound political arguments see Bird & Jun (2005).

Despite the political acceptance of earmarked taxes among citizens, one possibility why policy-makers keep on looking for new terms to describe the same practice is that they feel urged to exculpate themselves when introducing a new fund. Given the mainstream view of classic public administration literature, this feeling is comprehensible. In the Weberian tradition, where the administration is omniscient and follows a logic of a benevolent despotism, earmarking revenues does not compute (Ostrom 2008). A predefined budget allocation unnecessarily limits the administration's scope to maximize social welfare (Lee & Wagner 1991). In general, the argument of misallocation due to predefined appropriation is the most prominent reasoning of opposing earmarks (Deran 1965; Oakland 1984; Jankowski 1984). Schönenberger (2013) underlines the confining aspect towards the administration and he dismisses earmarked revenues as being politically motivated. Jankowski (1984) sees in earmarked funds even an encouragement for special program advocates who seek to prevent a debate about limited financial resources. In sum, theorists following the logic of public administration conclude that earmarking revenue cuts the efficiency of public service provision.

With the advent of public choice literature, the practice of earmarking revenues increasingly attracted attention among researchers. It became seen as a disciplinary tool hindering politicians from abusing their influential position in the budgetary process. To put it more positively, earmarking makes it possible to redress tie-in sales offered by the government under general fund financing (Buchanan 1963; Goetz 1968). The necessity to curb the administration reveals the underlying skepticism against its benevolent nature, which is typical for supporters of the public choice stream (Tullock 2008; Niskanen 1975; Migué et al. 1974; Wyckoff 1990; Persson & Tabellini

2002) and which is writ large in the view of the state as a Leviathan (Brennan & Buchanan 1978). From this perspective, earmarking can even help to fight against rigid budgeting procedures under certain circumstances (Bird 1982). Moreover, linking a levy with a certain public service seems to be a fairer tax system because the beneficiary pays for his own utility (McCleary 1991). Also, it provides the administration with information about the true demand for a public good (Bird & Jun 2005). In this sense, earmarking per se is not the first best solution but it substitutes a direct user charge if the latter is not feasible (Lee & Wagner 1991). Besides the fairness argument, user charges induce no or less distortions in contrast to income taxes for instance and they are therefore more efficient (Brazer 1984). Briefly speaking, in the view of – prior – public choice scholars, earmarked revenues force the administration to provide their services more efficiently (Buchanan 1963).

The positive attitude towards earmarking among adherents of the public choice school has again provoked critique. Two arguments challenge the charge-characteristic of earmarks. On one hand, the buyer of a good whose tax is earmarked for a specific service cannot decide individually upon the amount of the service financed by the earmarked tax; only the median voter has this privilege (Bös 2000). On the other hand, according to Musgrave & Musgrave (1973), earmarking fails to meet the intended benefit (i.e. user-pays) principle because the link between the taxed complement and the benefitting public good is usually rather weak; Lee & Wagner (1991) acknowledge this weakness. In fact, the unavailability of a complement to guarantee a user charging mechanism is often the reason why the state, instead of a private supplier, performs a certain task in the first place (Oakland 1984). McMahon & Sprenkle (1970) detail the argument in pointing towards the different price elasticities of the taxed good and the public service financed by earmarked taxes. Assuming nevertheless that the good to be taxed and the benefitting public service are perfect complements and have the same elasticities, the tax rate still remains arbitrary. Besides many others, Goel & Nelson (1999) and Lee & Tollison (2013) show both theoretically and empirically that the tax rate imposed on a good in order to fund a particular public service depends on political considerations. Moreover, from a distributional point of view a beneficiary financing is doubtful after all because the service might be publicly provided in order to make it available for poorer classes of the population; according to Oakland (1984) beneficiary financing is just as little desirable if the public good in question has important merits. In case of a pure public good, the free riding problem incentivizes the beneficiaries to hide their true preferences anyway (Teja 1988). In sum, the dispute seems to bear three recurrent conditions under which earmarking might lead to more efficient public service provision: First, a high correlation between the taxed good and the benefitting public service or good; second, a significant differentiation from other public goods or services in order to prevent hidden cross-financing; third, a strong homogeneity among payers of the taxed good and the beneficiaries of the good financed by the earmarked revenue and an excludability of those who do not pay (Brazer 1984; Lee & Wagner 1991; McCleary 1991).

Drawing premature conclusions from the theoretical literature bears some risks though. First, it is often unclear what the scholars mean by efficiency. For instance, Buchanan (1963) mostly puts the word efficiency in quotation marks and his examples of tie-in sales indicate that he means a budget allocation that corresponds to the preferences of the taxpayer; this understanding of 'efficiency' is far from how efficiency is defined in our analysis.[5] At the same time, Roberts (2002), among others, refers to technical efficiency citing the expression of 'getting more for less'. Second, the rationale behind the opposition against earmarked revenue is poorly substantiated. The opponents simply criticize the rigidity that earmarking imposes and they defeat theories that emphasize any positive aspects of earmarking. However, they mostly fail to explain how the inefficiency effectively comes about.

Nonetheless, a few critical authors autonomously develop and formalize a line of argument that links earmarking with low technical efficiency. Possibly the first authors who married the two concepts were Brennan & Buchanan (1978). The authors emanate from the Leviathan, in other words from the state as a monopolist that enables the bureaucrat to maximize his utility. Assuming that bureaucrat's goal is to maximize the surplus he can use as slack, earmarking can prevent some inefficiencies. If the public good and the taxed private good are complementary, the bureaucrat has an incentive to provide some of the public good in order to be able to raise taxes in the first place. Perfect complementary goods even lay the ground to create a tax law that ensures a fully efficient public good provision, i.e. there is no surplus left the bureaucrat can pinch as slack. Martínez (2014) chooses a different approach. He relies on the assumption that a voter is better informed about general tax revenue than about earmarked revenue and in consequence, the bureaucrat puts more effort toward an efficient public good provision. This assumption is plausible in his case, because he only considers royalties from the extraction of oil as earmarked revenue, which limits the scope of his model. Bös (2000) makes the decision about earmarking endogenous and concludes that even a welfare-maximizing parliament must collect too much tax in order to make sure they meet the earmarking constraint, i.e. guaranteeing that the earmarked tax covers the cost of the benefiting public service. Because of the distortionary effect of taxes, earmarking indirectly engenders inefficiency.

Albeit all three authors theoretically explain why earmarking decreases the technical efficiency, they stay at the macro level without penetrating the underlying functional chain. Put differently, the earmarking literature did not succeed in exploring the conceptual underpinnings with enough care and rigor. The lack of a substantiated theory is probably the reason why authors of public choice literature usually refer to some vague classic literature on public budgeting, public administration, et cetera when designating their opponents rather than explicitly citing specific articles (see for instance Dye & McGuire, 1992; Gwilliam & Shalizi, 1999; Bös, 2000; Bird & Jun, 2005). Section 3.1 yields an attack surface for the defenders of earmarking. Instead of simply criticizing existing ideas that promote earmarking, the debate prevention theory is a formalized model that explains why earmarking generally impairs the technical efficiency of public service provision when a collegial government is in charge.

The empirical evidence is mixed. First and foremost, no more than one article actually empirically studies the efficiency of the public service provision as the dependent variable, whereas all the others examine the effect of earmarked revenues on public expenditure, which is another question. The exception is Martínez (2014) who examines the influence of earmarked revenue on local public good provision in Colombian municipalities. The Colombian law appropriates the

[5] Remember the definition above: A government is considered technically efficient compared to its peers if it provides the maximum number of public services (i.e. outputs) given its bundle of expenditures (i.e. inputs).

revenue stemming from oil royalties to basic public services. Martínez operationalizes efficiency with specific outcome variables depending on the service in question. With the net school enrolment rate of six- to fourteen-year-olds he approximates the efficiency in education. In the health domain, he rests on the infant mortality rate and the percentage of the poor population with access to subsidized health insurance. Finally, the water quality should approximate the quality of the infrastructure. His results show that increases of the general budget boost the public goods indicators at least ten times more than an equivalent increase in earmarked revenues.

Empirical studies explaining the expenditures by earmarked revenue largely pivot on the state level of the U.S. and on a limited scope in terms of government functions. Regarding expenditures for highways, education, and welfare, Deran (1965) finds only for the welfare a significant positive effect of earmarked revenue on expenditures. A later study differentiates between absolute levels of earmarked revenues and the earmarked share of total revenues (Dye & McGuire 1992). Only the latter specification yields a significantly negative effect in the case of education spending. In none of the models does earmarking significantly affect highway expenditures; Crowley and Hoffer (2012) explain this phenomena by crowding-out effects. They claim that the government uses the earmarked revenue to replace unbound funds which in sum leads to a growth of the total expenditures. Likewise, Nesbit & Kreft (2009) assess an almost proportional positive effect. On the contrary, Goel & Nelson (2003) report a 2.5 dollar decrease of highway expenditures for every dollar of motor-fuel taxes diverted for general fund purposes. Afonso's (2015) results point in the same direction. Jung (2002), who focuses on data from Georgian counties, provides some evidence for a positive link. The earmarked revenues originate from sales taxes and are designated to finance local capital projects ranging from infrastructure for transportation (roads, streets, airport facilities), security (police, jails, courthouses, correctional facilities), to health (hospitals), et cetera. According to him, an extra dollar of earmarked revenue leads to a fifty percent increase in the envisaged spending area. Navarro (2002) and Evans & Zhang (2007) even estimate an increase of about 80 percent when lottery revenues are linked with educational spending. In sum, earmarking seems rather to have a crowding-out effect that results in lower spending compared to general fund financing.

The outlined overview should not give the impression of an unrestricted comparability of the results. Particularly when it comes to any international comparison, earmarks and their funds are shaped so differently that a mixing of the results could lead to false conclusions (Bird & Jun 2005). Nevertheless, the literature review shows that there is neither theoretical nor empirical agreement among researchers as to whether earmarking has a positive or a negative effect on the efficiency of public service provision.

2.2. Fiscal illusion

The early literature, from which the idea of fiscal illusion emanates, dates back to Puviani (1903). He already mentions revenue characteristics such as debt financing or tax diversification that later drew a lot of attention in the literature. However, it took more than sixty years until the research community jumped on the bandwagon.[6] While Buchanan (1967) by and large seizes on the earlier ideas and repeats archaic superficial theories, Oates (1985) scrapes out the most important aspects, research-wise for today, and concisely defines fiscal illusion as the "systematic misperception of key fiscal parameters [that] significantly distort fiscal choices by the electorate". The trigger for such a misperception may be various and Oates summarizes them under the following characteristics: 1. the complexity of the revenue structure, 2. the renter illusion, 3. the income elasticity of the tax structure, 4. the debt illusion and 5. the flypaper effect. In the meantime, Wagner (1976) was the first to test fiscal illusion empirically, whereas in the same paper, he developed a formal model which explains public overspending based on the complexity of the revenue system. After a bunch of further empirical studies assessing other revenue characteristics, Dollery & Worthington (1996) embedded the other four revenue characteristics in the formal model. Concentrating on the specificities of the flypaper effect alone, Wyckoff (1990) proposes a similar model which links overspending with intergovernmental transfers. In contrast to the Dollery & Worthington (1996) model though, Wyckoff (1990) shows that the voter's misperception does not only lead to overspending but also to a stronger extraction of rents on the side of the bureaucrats. In other words, Wyckoff (1990) links an inefficient public service provision with intergovernmental grants. The theory chapter of the present work seizes on this idea (section 3.3) and adds the other revenue characteristics to it (section 3.4).

Besides the before-mentioned relatively old papers, fiscal illusion was also tested empirically in more recent works. The most prominent variable among the revenue characteristics is the amount of grants and how this affects efficiency. Mostly, the papers report a negative effect (De Borger & Kerstens 1996; Balaguer-Coll et al. 2007; Kalb 2010b; Widmer & Zweifel 2012; Pérez-López et al. 2015), while there are also ambiguous results (Loikkanen & Susiluoto 2005) and even counterexamples (Geys & Moesen 2009; Bönisch et al. 2011; Da Cruz & Marques 2014; Ashworth et al. 2014). Bischoff et al. (2017) put the latter results into question arguing that state entities with a lower fiscal capacity, on the one hand, are under stronger financial pressure and thus are forced to be more efficient, while on the other hand, these entities receive the most grants. Without taking into account some measure of the pre-grant capacity, the estimates are biased due to the omitted variable. The study of Porcelli (2014) contrasts the other papers in terms of the design. The author focuses on a tax reform in 1998, where the Italian regional governments started to gather their own earmarked taxes to finance their health domain instead of depending on earmarked grants. The results show that efficiency increased after the reform, which the author explains by the higher local autonomy. In sum, most empirical papers find evidence for the flypaper effect.

Turning to the other sources of fiscal illusion, little has been done in relation to efficiency. In fact, the only aspect of fiscal illusion that has been tested besides the flypaper effect is the indebtedness. Benito et al. (2010) find, contrary to the theoretical prediction, an insignificant positive correlation between debt and efficiency. They argue that the current indebtedness is the result of former capital expenditures that allow greater efficiency today. Likewise providing insignificant results, Pérez-López et al. (2015) contribute little knowledge in this manner. Bischoff

[6] Buchanan (1967) presumes that rather than negligence, a lack of knowledge is what prevented the fiscal illusion idea from sparking a broader discussion. Until its translation in 1960 into German, Puviani's (1903) book existed only in Italian.

et al. (2017) claim that governments with high debt face fiscal stress and they are therefore forced to provide the services more efficiently. Indeed, they find a positive effect of the debt on efficiency for municipalities of Saxony-Anhalt. Using Portuguese municipal data, Da Cruz & Marques (2014) and Cordero et al. (2017) provide some evidence for debt illusion without specifically discussing it. Also Ashworth et al. (2014) find a significant negative relationship between indebtedness and efficiency in Flemish municipalities, but they do not refer to the fiscal illusion literature. Altogether, it seems that the researchers working in the field of efficiency analysis are not interested in or aware of the full scope of the fiscal illusion theory. Yet, the few studies taking the debt into account do report illusionary tendencies.

Among the tax related variables (i.e. tax complexity, progressivity, and diversification), only one study provides some empirical insights on their effect on efficiency. Boetti et al. (2012) finds Italian municipalities work less efficiently if their share of non-tax revenues (i.e. fees and charges) is high.

2.3. Interim conclusion

Today's limited interest in the theoretical literature on earmarking in relation to efficiency makes it arduous to recognize that the existing literature actually left a gap. On one side, there are the public finance scholars who reject earmarking due to its inflexibility in terms of budget allocation (Lee & Wagner 1991). On the other side, public choice scholars see earmarking as a solution to tame the Leviathan (Buchanan 1963). Ultimately, it is the fundamentally different views of humanity that sunder the two camps, which possibly impedes a final agreement. And yet, it is possible to criticize public administration scholars as most of them insufficiently substantiate their argument as to why earmarking should have a negative impact on technical efficiency. Only a few papers unify earmarking and technical efficiency in a formalized theory but they disregard the supposed underlying mechanism of the causal effect, or the theory covers only one specific field (Martínez 2014; Brennan & Buchanan 1978; Bös 2000). Just as in the theoretical literature, the empirical papers assessing the effect of earmarking on efficiency are very rare.

When it comes to further revenue features that potentially affect efficiency, the theoretical literature is much more developed. Since the seminal paper by Oates (1985), the idea of fiscal illusion disseminated into different directions and encouraged researchers to advance the theory. Today, five sources of fiscal illusion are broadly accepted and three of them have also been tested empirically with technical efficiency as a dependent variable. However, neither grants nor indebtedness furnishes unambiguous evidence for a negative or positive effect on efficiency. Only one study exists with respect to tax complexity and this hardly suffices for a substantial statement.

3. Theoretical models

This chapter aims at theoretically predicting in what direction and why the different revenue characteristics influence the efficiency of public service provision. In order to reach that objective, a discussion of debate prevention and the flypaper effect both constitute a separate section that develops a theoretical model based on the public choice framework. They thereby cover earmarked revenues and transfers as explaining variables of inefficiency. Because the development of the debate prevention theory turns out to be rather technical, a separate section first presents the intuition behind the theory and stipulates its limits. The fourth subsection embeds the remaining revenue characteristics in the mutual theory of fiscal illusion.

As to the dependent variable, all models understand inefficiency as a positive monetary rent that politicians or bureaucrats use for means other than public service provision (see Le Maux 2009 for an equivalent interpretation). In other words inefficiency means "to hire more staff than necessary, to hire more capital than necessary, to pay workers more than necessary, or to fund perquisites of office such as lavish offices, large expense accounts, short working hours, etc." (Wyckoff 1988: 272).[7] Therefore, a positive rent either induces excessive expenditures or insufficient public good provision, which are both attributes of technical inefficiency. Eventually, each model predicts the expected direction of the causality between the revenue characteristics and the rent, i.e. efficiency. While the models formalizing fiscal illusion and the flypaper effect are established in the current literature, debate prevention is a novel theory.

Beginning with a standard public choice framework, debate prevention takes up the result of an established voting model with uncertainty and an ideological bias (Persson & Tabellini 2002). Because of the latter two factors, politicians in office manage to appropriate a personal rent even under electoral competition. The extent of the rent depends on some transaction cost, which this established model considers as exogenously given. At this point, the new aspect - i.e. the herein developed theory called debate prevention – comes into play, arguing that the transaction cost depends on the monitoring activities within the governing council. The monitoring activities, in turn, are a function of the earmarked part of the total budget linked via ministerial incentives. Hence, the debate prevention theory eventually creates a connection between earmarking and a monetary rent which translates into inefficiency. This chain of reasoning grounds the central hypothesis (H_1): A higher share of earmarked revenues leads to less efficient public service provision, other things equal.

The hypotheses concerning the further revenue characteristics derive from the fiscal illusion idea (Oates 1985). That is, the voter underestimates the price for the public service which opens up the leeway for bureaucrats to appropriate rents. The voter's misperception originates from different revenue characteristics which all engender their own hypothesis. The most prominent among them stems from intergovernmental transfers, which often appear under the label of the flypaper effect (Courant et al. 1979). It states that a higher share of transfers implies poor efficiency levels (H_2) (Wyckoff 1990). Debt illusion, for its part, predicts a lower efficiency for highly indebted states (H_3). Likewise, the voters may be misled by the sheer amount of revenue sources, which substantiates the fourth hypothesis: A more complex tax system, ceteris paribus, results in lower efficiency levels (H_4). Finally, the voters might misjudge the price for public services if it rises due to the non-linear tax curve. Hence the prediction that a rising income in combination with a stronger tax progressivity provokes poor efficiency (H_5).

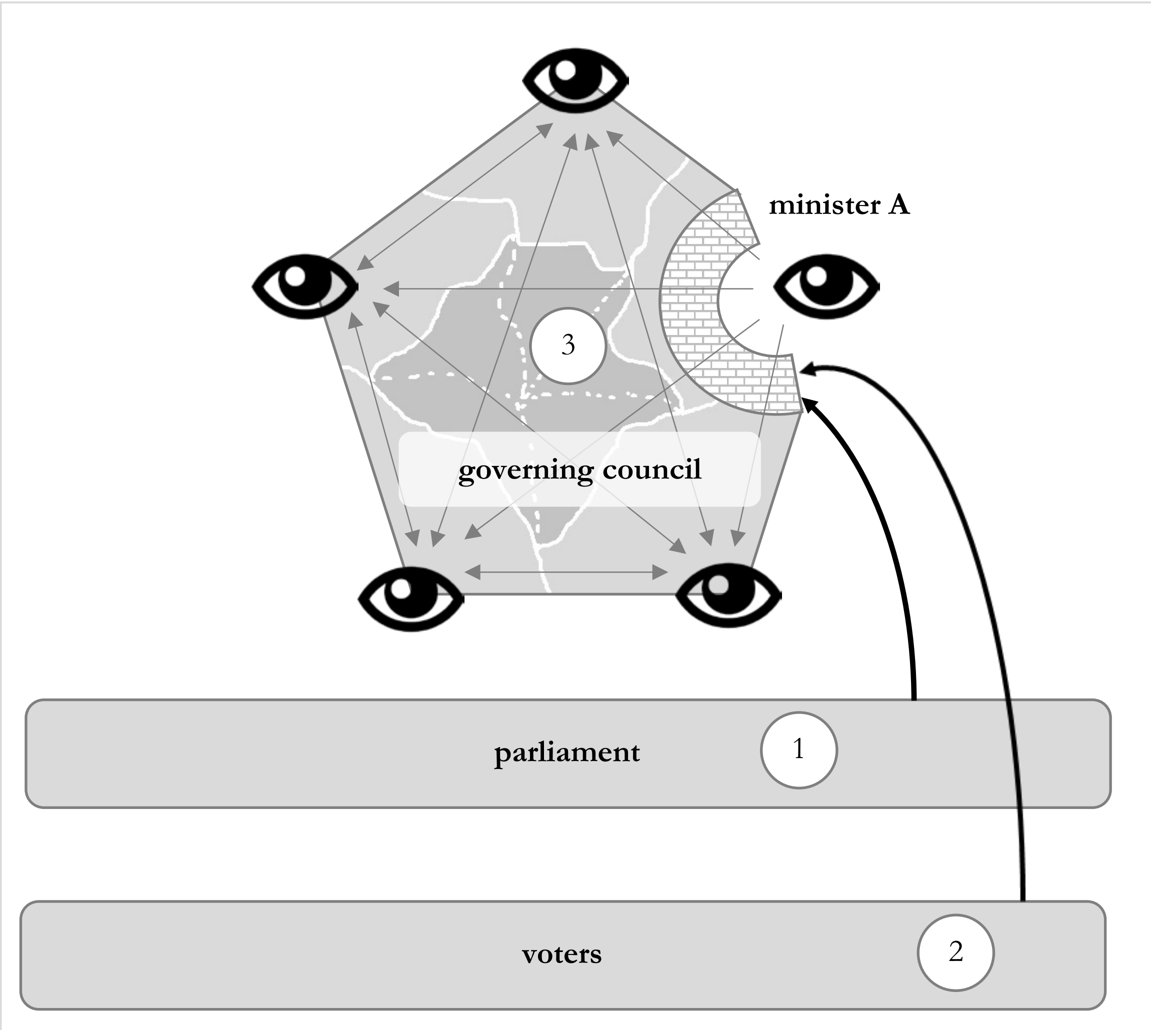
---

[7] The literature also uses terms such as discretionary slack or fiscal residuum as synonyms for the monetary rent when referring to the difference between the total revenue and total cost of production (Le Maux 2009).

3.1. The intuition behind debate prevention and its limits

The subsequent fairly extensive and detailed development of the debate prevention theory calls for a brief intuitive introduction. At the same time this section contrasts the theory with existing public choice models related to earmarked revenue and thereby clarifies its limits. To this end, Figure 1 shows the main actors involved in the dynamics of earmarked revenues. On top, the ministers illustrated as eyes form a pentagon representing the governing council's debate about the budget. In Niskanen's (1994) terminology, the eyes are equivalent to bureaus while the governing council together with the parliament can be viewed as the sponsor.[8] Within the executive branch of government, the ministers monitor each other; this is shown by the grey arrows. While four ministers both actively participate in monitoring activities and passively face their colleague's pressure, minister A hides behind a wall protecting him from the respective surveillance. The wall signifies a special financing. On the bottom, the voters establish the foundation of the parliament and of the executive government branch.

Figure 1 – Interrelationship of the debate prevention with the public choice models



[8] Niskanen (1994, p.28) recognizes the budget evolving from a two-stage process with the executive reviewing it in the first stage and the legislature in the second. Accordingly, the tandem executive-legislative (i.e. the government) positions as a composite sponsor.

Source: Own illustration

The parliament (1) is the first actor motivated to install a wall around minister A. As mentioned in the literature review (section 2.1), the parliament, well aware of its ephemerality, attempts to lock future budgets according to its current preferences (Jackson 2013). While budget restrictions on future governments affect the budget allocation, they have no effect, per se, on the technical implementation, cost, nor the allocative efficiency.[9] Instead, they might reduce the welfare since the earmark prevents politicians from adjusting the budget according to the voters' preferences (Lee & Wagner 1991). The scope of the paper at hand encompasses only the efficiency aspect and leaves aside the welfare considerations.

The second arrow takes up the voter's (2) desire to earmark certain revenues in order to strengthen their preferences (Buchanan 1963). In the same vein as above, the earmarked revenues should ensure a certain budget allocation, which again links to welfare rather than to efficiency. Accordingly, the parliamentary incentives and interests to establish new special financings falls likewise outside the scope of this analysis.

The third relationship refers to the dynamics within the governing council (3) and it illustrates debate prevention. In the first budget round, the five ministers present their policy programs along with the corresponding financial needs (the fraction of the light and dark grey areas attached to each corner). When designing their policy programs, the ministers choose their degree of engagement to reach a certain quality, which simultaneously entails a cost (e.g. longer working hours, etc.). A higher quality ensures that the monitoring activities of the other ministers (i.e. the grey arrow's ends pointing towards each minister) are less effective and that the proposals pass the budgetary discussions without substantial cuts; the combination of the own quality and the other ministers' effort yields the absolute bargaining power. At the same time, the ministers try to convert the other minister's budget share to their own use by monitoring their colleagues (i.e. the arrow's ends pointing away from each minister). If the monitoring is successful and the ministers find potential savings in their colleagues' programs, the free-up funds remain on the negotiation table. After the first budget round, each minister defends the light grey fraction of the budget. In the subsequent budgetary rounds, the ministers allocate the rest of the budget (i.e. the dark grey area) amongst each other according to their relative bargaining power.[10]

Yet, during all these debates, the budgetary fraction behind minister A's special financing wall remains undisputable. Out of A's demanded budget, there is only a small part up for discussion and this alleviates any incentive to monitor A's policy programs in the first place. Being less monitored, minister A enjoys a larger leeway to generate budgetary slack and consequently he offers the services within his domain less efficiently. In general, the larger the share behind earmark-walls, the less incentive the ministers have to monitor each other and the less efficient the public service will be provided.

---

[9] Remember the earlier definition, according to which budget allocation refers to the allotment of public resources to the different public services (outputs); technical efficiency describes a production process that converts a given its bundle of inputs into the maximum number of outputs. Finally, allocative efficiency denotes the optimal combination of inputs taking their prices into account.

[10] Note that debate prevention entails no Nash equilibrium but a Nash bargaining solution making implicitly use of the axiomatic approach as proposed by Nash (1950). In other words, the ministers define some conditions that the outcome of the bargaining game has to satisfy.

The rather specific setup of the model raises the question of its applicability to an environment outside of Switzerland. Answering this question requires some contemplation about the central elements and whether they are exclusively present in the Swiss case. The most critical of which is the formally equal distribution of power among the ministers debating about the budget allocation. Such a system is unique in the world. Yet, it is likely that the dynamics are similar in a system with a president or a prime minister who takes the executive branch's final decision about the budget. The subordinate ministers still debate about their share of the budget and the tactics to protect one's own budget share or to snag a part of the other minister's budget share remain the same. Although, the minister's incentives to monitor their colleagues are less clear, because – depending on the concrete budgeting process – the president or the prime minister can decide upon the budget allocation independently of the possibly found weaknesses in some policies. For instance, if the prime minister has an ideological preference for a specific public service, she might provide generous financial resources in spite of considerable flaws in the respective policy proposals. Ministers who face such a prime minister have less incentive to monitor their colleagues, which disperses the negative effect that earmarking has on monitoring. In sum, debate prevention might affect the budgeting process in other countries too, but its effect is possibly less distinct.

3.2. A formal model of debate prevention

Primal median voter models with electoral competition à la Lindbeck & Weibull (1987) came to the gratifying conclusion that politicians cannot reap rents in equilibrium. If the politicians do so, another candidate would waive his own rent in order to increase the public service provision or decrease taxes. With a more efficient combination of public services and taxes, the candidate can kick the incumbent out of office. The anticipatory incumbent therefore reduces his rent to zero in the first place, guaranteeing the voters an optimal bundle. Referring to Polo (1998), Persson & Tabellini (2002) introduce in their probabilistic voting model not only an uncertainty about the election outcome but also ideological preferences and thereby they formulate the conditions under which politicians can obtain positive rents, even under electoral competition. An extension of Persson & Tabellini's model reveals the effect earmarking has on the extent to which politicians can retain rents. Equally, the modification of the model accounts for the specific composition of collegial governments. In contrast to the usual probabilistic voting model, here elected politicians cannot rule the government by themselves as a monolithic planner and executer, but they make part of a governing council, where they take over a department that provides specific public goods. The literature calls this type of government consensus democracies with executive power-sharing (Vatter & Freitag 2007). When running for office, the candidates commit to an overall policy platform, since, before the election, they do not know which department they will undertake.[11]

In the following, capital letters in equations denote functions while lowercase letters signify parameters. Upper bars on letters indicate externally given parameters whereas a tilde marks a guess

[11] Admittedly, there is usually more than one candidate running for office. As Persson & Tabellini (2002) show though, additional candidates lead to a lower extraction of rents, but as long as the number of candidates does not go to infinity, there are still some rents left. It is therefore appropriate to only consider two candidates for the sake of simplicity.

based on incomplete information. Consider a large number of voters-consumers-taxpayers who have the quasi-linear preferences w depending on their consumption of two market-delivered goods c and x as well as of two types of publicly-provided goods g_e and g_g:

$$w = c + H(g_e, g_g, x) \tag{1}$$

The concave increasing function $H(\bullet)$ describes the voter's appreciation for the respective goods. The way the voter's preferences are modelled goes deeper into detail than the original formulation in two aspects. First, there are two types of publicly-provided goods that differ in the manner they are financed. While the source of funding of g_g is the general fund from ordinary taxes, g_e obtains its resources from taxes or charges imposed on a related market good x. Typical market goods x with earmarked taxes are fuel, the lottery, et cetera. Second, g_g is not a particular publicly provided good but a bundle (for a similar interpretation see Goetz, 1968). Concretely, g_g typically contains goods that do not allow excluding anyone from their use: security, foreign policy, et cetera. The opposite applies to g_e, which represents health, roads, or education. It would be easy to extend the model by splitting the publicly-provided goods g_e and g_g into their individual components and to regard each separately. However, such an extension would only complicate the model while leading to the same result.[12] The possibility of splitting is useful in a later stage of this analysis though, once the candidates are in office. The government budget must balance out and is restricted by

$$\tau\bar{y} + \varphi x = g_e + g_g + r_e + r_g \tag{2}$$

The governmental revenue on the left-hand side of equation (2) stems from the taxes τ imposed on the given income $\bar{y}$ and the taxes φ on the good x. The total expenditures consist of spending for the politician's rents r_e and r_g and for the two publicly-provided goods g_e and g_g. Both g_e and g_g have constant production costs (i.e. shadow prices) normalized to unity. Since the revenue generated through the taxation of the market good x is earmarked, it must not be used to produce g_g. Indeed, the revenue can either flow into the production of g_e or into rents r_e that politicians can convert to their own use (McMahon & Sprenkle 1970). In formal terms the earmarking constraint is:

$$g_e + r_e = \varphi x \tag{3}$$

With the parameters defined so far, the voter's budget constraint is

$$c + (1 + \varphi)x = (1 - \tau)\bar{y} \tag{4}$$

The voters can consume as much from the private goods c as they have left after paying taxes on their income and after having consumed the market good x taxed at rate φ. The equal sign reflects the assumption that voters cannot save anything from their income and they cannot consume on credit. For simplicity, define for now the total operational public expenditures as $g \equiv g_e + g_g$ and the total rent as $r \equiv r_e + r_g$. Combining (1), (2), (3) and (4) yields the voter's policy preferences

[12] Note that, in contrast to the median voter model, the probabilistic voting model eludes cyclic majorities when aggregating g_e and g_g (Blankart 2012).

$$W(\boldsymbol{p}) = \bar{y} - g - r - x + H(g_e, g_g, x) \qquad \text{where } \boldsymbol{p} \equiv (g_e, g_g, r_e, r_g, \varphi) \tag{5}$$

The vector $\boldsymbol{p}$ subsumes the policy platform, that is to say the variables that politicians can adjust. Note that φ does not appear in (5), which implies that the voter does not care about the level of the tax on the market good x. To see why, solve the government's budget restriction (2) for τ and replace the respective variable in (4). Then, solve the resulting expression for *c* whereby φx cancels out. How can the disappearance of φx be interpreted? On the one hand, the tax paid on the market good φx enters positively in the voter's budget constraint as it reduces the income tax required to reach the same public budget level. On the other hand, φx decreases the voter's budget constraint as it limits the scope to consume c. In essence, the two effects neutralize each other. Consequently, the politician can arbitrarily alter φ such that the earmark constraint (3) holds however he sets g_e and r_e (Bös 2000). The earmark constraint is therefore in fact non-restrictive. The disappearance of the earmarking constraint is only possible because the voter can directly express her preferences about the amount of each public good individually.

Now turn to the politician's objective. The politician benefits from being in office, which manifests in ego rents $\tilde{R}(\tilde{N}(g) + \tilde{b})$. In the ego rents, politicians aggregate expected non-monetary benefits $\tilde{b}$ such as power, prestige, et cetera and Niskanen-type utility $\tilde{N}(\bullet)$ (Niskanen 1975; Bös 2000). The latter depends on the operational governmental expenditures g. While g is part of the policy platform, $\tilde{b}$ is non-discretionary and equal for each politician before the election. Also, before the election, the politicians do not know exactly how to achieve the ego rents; that is, the functional form of $\tilde{N}(\bullet)$ and which of its parameters become discretionary once the politicians are in office is unknown. They know, though, that $\tilde{R}(\bullet)$ is an increasing function of $\tilde{N}(\bullet)$ and $\tilde{b}$, while $\tilde{N}(\bullet)$ increases in g. The ego rents do not appear in the governmental budget as, on the one hand, they partly coincide with government spending g and, on the other hand, they reflect non-monetary benefits $\tilde{b}$.

Besides ego rents, politicians benefit from endogenous monetary rents r. To what extent politicians can achieve them depends on the transaction costs $\tilde{\gamma}$. The higher $\tilde{\gamma}$, the lower the transaction costs and the easier politicians can internalize the monetary rents. As a concrete example of transaction costs, Persson & Tabellini (2002) name the transparency of the budget or administrative procedures. In the pre-election stage, the politicians cannot manipulate the transaction cost and they have to take it as externally given. The politicians can appropriate the ego and the monetary rents only if they are elected, whereas the probability of being elected is denoted by f_p. As the voters, politicians have a quasi-linear objective too. Politician P's objective function reads as

$$\mathbb{E}[v_p] = f_p\big(\tilde{R}(\tilde{N}(g) + \tilde{b}) + \tilde{\gamma} r\big) \tag{6}$$

where v_p is the politician P's stochastic utility. In line with Persson & Tabellini (2002), assume that voters have an individual ideological bias σ^i towards candidate B. The reason why σ^i might be different from zero are attributes of the politician that each voter esteems differently and additionally to the economic policy platform. Concrete examples are the politicians' educational background, the origin, et cetera. The bias can take positive or negative values, whereas the latter case indicates a preference for the candidate. Among the entire population, the bias is uniformly distributed with density ϕ and mean zero:

$$\sigma^i \sim U\left[-\frac{1}{2\phi}, \frac{1}{2\phi}\right] \tag{7}$$

As a second stochastic term, δ measures the average relative popularity of candidate B. This parameter can also take values above and below zero with negative values implying that candidate enjoys relatively more popularity in the entire population. Note that δ is the same for all voters and therewith might be interpreted as attributes of the politician such as, for instance, scandals. The politician cannot observe his relative popularity before election, but indeed, he knows its distribution. It is also assumed to be uniform with density ψ and mean zero:[13]

$$\delta \sim U\left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right] \tag{8}$$

Having outlined the model's setting, consider the timing of events:

[1] The incumbent and a challenging candidate learn the distributions of σ^i and ψ as well as the voters' preferences $W(\boldsymbol{p})$. Based on this information, they proclaim simultaneously and non-cooperatively their policy platforms $\boldsymbol{p}$ that maximize their objective $\mathbb{E}[v_p]$.

[2] On election day, their actual relative average popularity δ becomes publicly known.

[3] Elections are held.

[4] The elected politicians learn about the practice in the governing council and become acquainted with their council colleagues. Based on that, they specify their objective ξ.

[5] The governing council debates the details of the policy program by dispersing the total budget.

[6] The governing council implements the detailed policy program.

Since politicians are aware of the electoral procedure, they use backwards induction to maximize their objective function (6). As they know the voters' preferences, they can anticipate their probability of winning the election. They therefore need to identify the condition under which the voters actually prefer them. From politician A's point of view this condition is

$$W(\boldsymbol{p}_A) > W(\boldsymbol{p}_B) + \sigma^i + \delta \qquad \Leftrightarrow \qquad \sigma^i < W(\boldsymbol{p}_A) - W(\boldsymbol{p}_B) - \delta \tag{9}$$

The rearranged equation (9) shows that all the voters whose ideological bias towards candidate B is smaller than the term on the right-hand side prefer candidate A. By the distributional assumption on σ^i, A's vote share therefore is

$$\pi_A = \phi\left(\sigma^i + \frac{1}{2\phi}\right) = \phi(W(\boldsymbol{p}_A) - W(\boldsymbol{p}_B) - \delta) + \frac{1}{2} \tag{10}$$

The probability that A wins the election equals the probability of acquiring more than half of the votes:

$$f_A = Pr\left(\pi_A > \frac{1}{2}\right) = Pr(W(\boldsymbol{p_A}) - W(\boldsymbol{p_B}) > \delta) = \frac{1}{2} + \psi\big(W(\boldsymbol{p_A}) - W(\boldsymbol{p_B})\big) \tag{11}$$

[13] These distributions might appear rather particular. As Persson & Tabellini (2002) state though, non-uniform distributions would lead to equal final results, while they complicate the calculations considerably.

Exploiting the distributional assumption on δ yields the third equality of (11). With the knowledge about his probability of winning, politician A maximizes his objective function. Since the focus is on rents, the first derivative of the politician's objective with respect to r promises interesting insights:

$$\frac{\partial \mathbb{E}[v_A]}{\partial r} = \frac{\partial f_A}{\partial r}\left(\tilde{R}(\tilde{N}(g) + \tilde{b}) + \tilde{\gamma} r\right) + \tilde{\gamma} f_A = -\psi\left(\tilde{R}\left(\tilde{N}(g), \tilde{b}\right) + \tilde{\gamma} r\right) + \frac{1}{2}\tilde{\gamma} = 0 \qquad (12)$$

Since the objectives of the two politicians are the same, in equilibrium their probability of winning is one-half, which justifies the second equality. The same reasoning determines the content of g, which is exactly the combination of publicly-provided goods that the median voter wants. Solving for r yields the equilibrium rents:

$$r = \max\left[0, \frac{1}{2\psi} - \frac{R\left(\tilde{N}(g) + \tilde{b}\right)}{\tilde{\gamma}}\right] \qquad (13)$$

Primarily, the result (13) shows that, in certain circumstances, politicians can nail down rents in equilibrium even under competition. The circumstances depend first and foremost on the density of the politician's relative popularity, which is the base for the monetary rent. The more uncertain the outcome of the election, meaning the lower the popularity density, the larger the possible monetary rents. Second, higher ego rents, scaled by the transaction cost, reduce the scope of the monetary rents. Remember that the higher the transaction costs, the lower the $\tilde{\gamma}$. Thus, high transaction costs unsurprisingly diminish the monetary rents. The further analysis about the politicians in office reveals what factors determine the transaction cost. Up to this point, when the elections are held, the model followed by and large what Persson & Tabellini (2002) outlined earlier. The novel part follows with the politicians debating about the detailed governmental budget (indicated as stage [4] above). Nevertheless, the repetition of the base model proves useful, as its results remain stable even under partial earmarked funding.

As soon as the politician is elected, he does not have to care about the probability of winning anymore as f_p equals one by definition. Due to the politician's commitment during the campaign, he has to stick with his proposed policy platform $\boldsymbol{p}^*$.[14] Consequently, the total government spending $\bar{g}$ and the monetary rent $\bar{r}$ are externally given parameters in the upcoming budgetary debate. The claim at the opening of this section was that splitting the two types of publicly-provided goods g_e and g_g into their individual components would lead to the same results. For the further analysis, remember that $\bar{g}$ actually contains several separate publicly provided goods of each type. Once the politicians are in office, the imaginary partition of $\bar{g}$ into more than only two parts makes sense, because otherwise, in the debate, they could not allocate the budget in detail to the different departments. Moreover, the politicians in office do not have to form expectations anymore as they learn about the functional form of the Niskanen-type utility $N(\bullet)$, the non-monetary benefits $\bar{b}$ and the transaction cost $\tilde{\gamma}$. Once in office, the minister A's objective (6) becomes

[14] Admittedly, the assumption that politicians have to stick with their promises is a rather strong one. In fact, some models extend the probabilistic voting model by allowing for post-election alterations of the policy platform. However, the debate prevention model uses the policy platform more as a guideline for further budgetary debate, which makes the assumption less far-fetched after all. Because the platforms respond to citizens' demand in the first place these promises have a constraining character to politicians (Rubin 1990).

$$\xi_A = R(N(\bar{g}) + \bar{b}) + \tilde{\gamma}\bar{r} \tag{14}$$

The elected politician learns about the practice in the governing council, that is he learns how the budgeting and spending processes, discussed empirically in section 4.1 below, proceed. In either process, the governing council plays an important role in allocating the resources to the departments, as they debate about the share each department receives of the total government budget $\bar{g}$. Note that a budgetary debate within the government is a major difference to classic bureaucracy models because it breaks the executive's monopoly power to some extent (i.e. its Leviathan status) (Brennan & Buchanan 1978; Niskanen 1975).

Consider D as all departments different from A henceforth, i.e. $D \equiv \{1, \dots, \dots, n\}$ and D^{+A} respectively includes the department A, that is $D^{+A} \equiv \{A, 1, \dots, n\}$. In the debate, each minister starts with a certain share of the budget $\bar{\lambda}_{D^{+A}} \equiv \frac{\bar{g}_A}{\bar{g}}$ depending on the promises about $\bar{g}$ the politicians made during the campaign. As such the ex-post share $\bar{\lambda}_{D^{+A}}$ is externally given. Naturally, $\lambda_{D^{+A}}$ takes only values between zero and one and the shares of all ministers must sum up to one.

As regards the minister of department A, his goal in the debate is to maximize his own share of the total budget that gives him the Niskanen-type utility $N(\bullet)$. In order to reach that objective, he has an offensive and a defensive possibility. In the defensive case, minister A seeks to protect his own budgetary claims and sets a certain level of quality q_A in his policy programs. At the same time, he faces the other ministers' monitoring effort $\bar{e}_D$. His absolute bargaining power then turns out to be $\Omega_A(q) = \Omega(\bar{e}_D, q_A)$. A higher dossier quality makes it hard to challenge his proposals even if extensively screened by the other ministers, i.e. $\Omega(\bullet)$ increases in q. More intense screening reveals further weaknesses, which is why $\Omega(\bullet)$ decreases with respect to $\bar{e}_D$. The bargaining power is normalized to one, where one is the highest bargaining power and zero the lowest. Assume that the second derivative of $\Omega(\bullet)$ with respect to e and q is negative, which is plausible since politicians pick the low hanging fruits first when engaging in monitoring effort. The concavity with respect to the quality connotes a decreasing marginal effect.

Given the earmarked budget share $\bar{\eta} \equiv \frac{\bar{g}_e}{\bar{g}}$ and the total budget $\bar{g}$, minister A can claim the following part of the budget based on his own defensive activities in the debate:

$$\bar{\lambda}_A \Omega(\bar{e}_D, q_A)(1-\eta)\bar{g} = \bar{\lambda}_A \Omega_A(q)(1-\eta)\bar{g} \tag{15}$$

In the offensive case, minister A interferes with the other departments' bargaining power by monitoring them with an effort e_A. They in turn repel the monitor-attack by ensuring a high quality of their policy programs $\bar{q}_D$. From A's viewpoint, the resulting bargaining power parameters of ministers D are $\Omega_D(e) = \Omega(e_A, \bar{q}_D)$. Unless all ministers manage to reach a bargaining power of one, some resources remain undistributed after the first budgetary round. In fact, all resources that the ministers cannot claim for their department accumulate to an unbound residual, that is

$$\Phi(e,q) \equiv \frac{1+\bar{n}-\sum_D \Omega(e_A,\bar{q}_D)-\Omega(\bar{e}_D,q_A)}{1+\bar{n}} = \frac{1+\bar{n}-\sum_D \Omega_D(e)-\Omega_A(q)}{1+\bar{n}} \tag{16}$$

The nominator indicates simply the total number of departments including A minus the sum of the shares each department successfully claimed during the debate. To obtain a share, the nominator is divided by the number of departments. Based on the same logic as in the first round, the ministers use their bargaining power to demand a part of the residual, which again incentivizes the other ministers to make a monitoring effort. Since all ministers stick to their initial quality q and monitoring effort e, the entire budget eventually divides among all departments according to the relative bargaining power of each minister. The relative bargaining power is therefore

$$\Psi_A(e,q) \equiv \frac{\Omega(\bar{e}_D\bar{\vartheta}_D, q_A\bar{\vartheta}_A)}{\sum_D \Omega(e_A\bar{\vartheta}_A, \bar{q}_D\bar{\vartheta}_D) + \Omega(\bar{e}_D\bar{\vartheta}_D, q_A\bar{\vartheta}_A)} = \frac{\Omega_A(q)}{\sum_D \Omega_D(e) + \Omega_A(q)} \tag{17}$$

The total government budget partly consists of earmarked resources that must flow into a specific department. Each department receives a fix and externally given share $\bar{\alpha}_{D+A} \equiv \frac{\bar{g}_{e,D+A}}{\bar{g}_e}$ of the earmarked revenue $\bar{g}_e$. Therefore, the earmarked share of the budget $\bar{\eta}$ cannot be allocated discretionarily and is not up for discussion. Finally, making an effort as well as increasing the quality induces non-monetary costs $C(e,q)$ to the ministers. $C(\bullet)$ is an increasing convex function in both its arguments and the marginal cost is zero when the effort is zero, i.e. $\frac{\partial C(e=0,q)}{\partial e} = 0$. In sum, combining (20), (21), (22) and adding the cost leads to the following functional form of the Niskanen-type utility:

$$N(\bar{g}) = N(e,q) = (1-\bar{\eta})\bar{g}\left[\bar{\lambda}_A\Omega_A(q) + \Psi_A(e,q)\Phi(e,q)\right] + \bar{\alpha}_A\bar{\eta}\bar{g} - C(e,q) \tag{18}$$

The Niskanen-type utility (18) reveals the two roles of the bargaining power. First, the absolute bargaining power determines the residual share $\Phi(e,q)$ that remains after allocating the budget according to the promised share $\bar{\lambda}$. Second, it sets the relative bargaining power $\Psi_A(e,q)$ and thereby the share each department can claim out of the residual.

In addition to the detailed Niskanen-type utility, minister A learns his true non-monetary benefits:

$$\tilde{b} = \bar{b} \tag{19}$$

Turning to the transaction costs, remember that they indicate to what extent politicians can pocket monetary rents. Thereby, the higher the γ, the more rents politicians can rake in. Following the discussion above, it is straightforward to claim that transaction costs for minister A are high, if his colleagues invest a lot in monitoring activities.[15] Transaction costs are therefore

$$\tilde{\gamma} = \frac{1}{1+\bar{e}} \tag{20}$$

When inserting (18), (19), and (20) in (14), minister A's objective becomes

$$\xi_A(e,q) = \bar{b} + (1-\bar{\eta})\bar{g}\left[\bar{\lambda}_A\Omega_A(q) + \Psi_A(e,q)\Phi(e,q)\right] + \bar{\alpha}_A\bar{\eta}\bar{g} - C(e,q) + \frac{1}{1+\bar{e}}\bar{\lambda}_A\bar{r} \tag{21}$$

In short, minister A's objective contains seven parts. The first term of (21), i.e. $\bar{b}$, denotes the politician's non-monetary benefits. The second term, i.e. $(1-\bar{\eta})\bar{g}$, marks the non-earmarked part of the total budget. The third term, i.e. $\lambda_A\Omega_A(q)$, represents minister A's claim of the non-earmarked budget depending on the promises during the campaign and his absolute bargaining power. The fourth term, i.e. $\Psi_A(e,q)\Phi(e,q)$, captures the budget he can grab from the other minister's budget depending on his relative bargaining power and the residual share of the budget. The fifth term is the budget share earmarked for department A. The sixth term subtracts the cost accruing from the effort put into monitoring and quality. Finally, the last term is A's share of the

[15] For a discussion of how monitoring potentially disciplines the ministers see Nagin et al. (2002).

monetary rents depending on A's ex-post share and the transaction cost. In order to maximize his objective, each minister can only choose the levels of effort e and quality q. Since the emphasis lies on the last term including the monetary rent $\bar{r}$, the relevant variable is e. Minister A finds his optimal monitoring effort level by differentiating (21) with respect to e and setting it equal to zero. Rearranging then yields[16]

$$(1-\bar{\eta})\bar{g}\bar{\vartheta}_A\left(\frac{\partial\Psi_A(e,q)}{\partial e}\Phi(e,q)+\frac{\partial\Phi(e,q)}{\partial e}\Psi_A(e,q)\right)=\frac{\partial C(e,q)}{\partial e} \qquad (22)$$

where $\frac{\partial\Psi_A(e,q)}{\partial e}>0$ [residual share after first budgetary round]

$\frac{\partial\Phi(e,q)}{\partial e}>0$ [A's relative offensive bargaining power]

$\frac{\partial\Omega_D(e)}{\partial e}<0$ $\frac{\partial^2\Omega_D(e)}{\partial e^2}<0$ [A's absolute offensive bargaining power]

$\frac{\partial C(e,q)}{\partial e}\geq 0$ $\frac{\partial^2 C(e,q)}{\partial e^2}>0$ [cost function]

Equation (22) contains a basic result from microeconomics: Utility maximization means equalizing the marginal cost and the marginal revenue. Here, the marginal revenue on the left-hand side is additionally weighted by the earmarked revenue share $\bar{\eta}$ and the available budget $\bar{g}$. Note that minister A needs to keep the two sides balanced in order to achieve maximum utility. For our analysis, the relevant question is how a change of the earmarked revenue share affects the minister's decision about his monitoring effort. The standard approach to answer this question would be to solve for e followed by a derivation with respect to $\bar{\eta}$. For the sake of generality, neither $\Omega(\bullet)$ nor $C(\bullet)$ was defined, which makes solving for e impossible.

To find the respective answer in a different way, consider the derived marginal effects listed in (22). If the earmarked share of the budget $\bar{\eta}$ increases, the left-hand side of (22) decreases. That is, changing $\bar{\eta}$ linearly alters the marginal benefit of the monitoring effort e, whereas its cost on the right-hand side remains equal. In order to reach the maximum of his objective (21), minister A equalizes (22) again. Unfortunately, the so far imposed assumptions are not sufficient to show that there is one and only one intersection curve between the marginal cost and the marginal benefit. Appendix A. 1 develops the necessary assumption, which is basically a sufficiently high second order derivative of the offensive bargaining power. Under this weak additional assumption, e and $\bar{\eta}$ are negatively linked. Therefore, minister A will set his monitoring effort e lower the higher the earmarked share η is. An interim conclusion leads to the realization that the monitoring effort is actually a function of the earmarked share of the budget $\bar{\eta}$, i.e.

$$e(\bar{\eta}) \qquad \text{where} \qquad \frac{\partial e(\bar{\eta})}{\partial\bar{\eta}}<0 \qquad (23)$$

That is, the effort depends on the external factor $\bar{\eta}$ and the higher the share of earmarked revenue becomes, the less politicians engage in monitoring activities. An intuitive explanation for this relation is that monitoring activities do not pay off enough anymore if the earmarked revenue share increases. Put differently, a high earmarked revenue share partly prevents ministers from engaging in the budgetary process debate.

[16] See appendix A. 1.

The insights regarding the dependence of the monitoring activities e and the earmarked revenue share $\bar{\eta}$ puts another complexion on the transaction cost $\tilde{\gamma}$ as defined in (20). Since e decreases in $\bar{\eta}$, the transaction costs decrease (i.e. $\tilde{\gamma}$ increases) as the share of earmarked revenue increases. The full specification of the transaction costs is therefore

$$\tilde{\gamma}(\bar{\eta}) = \frac{1}{1+\bar{e}(\bar{\eta})} \qquad \text{where} \quad \frac{\partial \tilde{\gamma}(\bar{\eta})}{\partial \bar{\eta}} > 0 \tag{24}$$

With reference to equation (2), the efficiency is some relation between financial inputs (i.e. $\tau\bar{y} + \varphi x$) and outputs (i.e. $g_e + g_g$) whereas larger rents (i.e. $r_e + r_g$) decrease efficiency by demanding more inputs per output. A precise definition of efficiency and its different forms follows in chapter 6. Concretely, the efficiency measure applied later focusses on input efficiency, which means considering the outputs fixed while inputs are flexible. With respect to the theory, the fixed outputs hypothesis translates into fix g_e and g_g, leaving politicians to propose a policy platform containing their rents that imply the input. Facing lower transaction costs $\tilde{\gamma}$, politicians appropriate rents easier. From the politicians' point of view, the specification of the transaction cost (24) is not relevant, as they cannot manipulate the earmarked revenue share $\bar{\eta}$ once in office – and before their swearing in they cannot tell the functional form of the ego rents $\tilde{R}(\bullet)$. Taking the long run perspective from a policy advisor though, η can possibly be altered with some nudges. One of them is to set the hurdle for an earmark high. Hence, it shall be one of the main objectives of our analysis to empirically test whether the predicted relation between the earmarked revenue share and the efficiency of public service provision exists.

3.3. Flypaper effect

In contrast to the previous section, the two that follow draw on the existing models rather than developing new ones. Applying the traditional median voter model on a government that posts revenue from taxes and from transfers of other state entities, Bradford & Oates (1971) arrive at a relatively disenchanting conclusion. Whether the government's budget constraint increases due to the higher income of its citizens or because of lump-sum grants from other governments has no effect on the amount of additional public goods supplied.[17] The irrelevance comes from the unchanged preference curves of the voter-taxpayer and the steady relative prices of the public and the private good. In economic terms, both regimes only have an income but no substitution effect (Bailey & Connolly 1998). Interpreted differently, the source of revenue divided into lump-sum transfers and taxes has no influence on the quantity of public goods provided. Among the first to do so, Gramlich et al. (1973) tested the theoretical prediction with United States data for the period 1954 to 1972 and found contradictory results. An additional dollar of transfer triggers an additional public spending of 0.43 dollars, while an income increase of citizens only stimulates expenditures by 0.1 dollar. The difference between the two effects suggests that the government does not pass

[17] With a lump-sum grant the government can allocate the resources arbitrarily among the public goods and even, by reducing taxes, provide its citizens with the possibility of using them for more private goods. In contrast, a matching grant is earmarked and requires the government use it for the designated task or program.

the grants to the citizens in the form of lower taxes. The puzzling phenomena that money sticks where it hits came to be known as the flypaper effect (Courant et al. 1979).

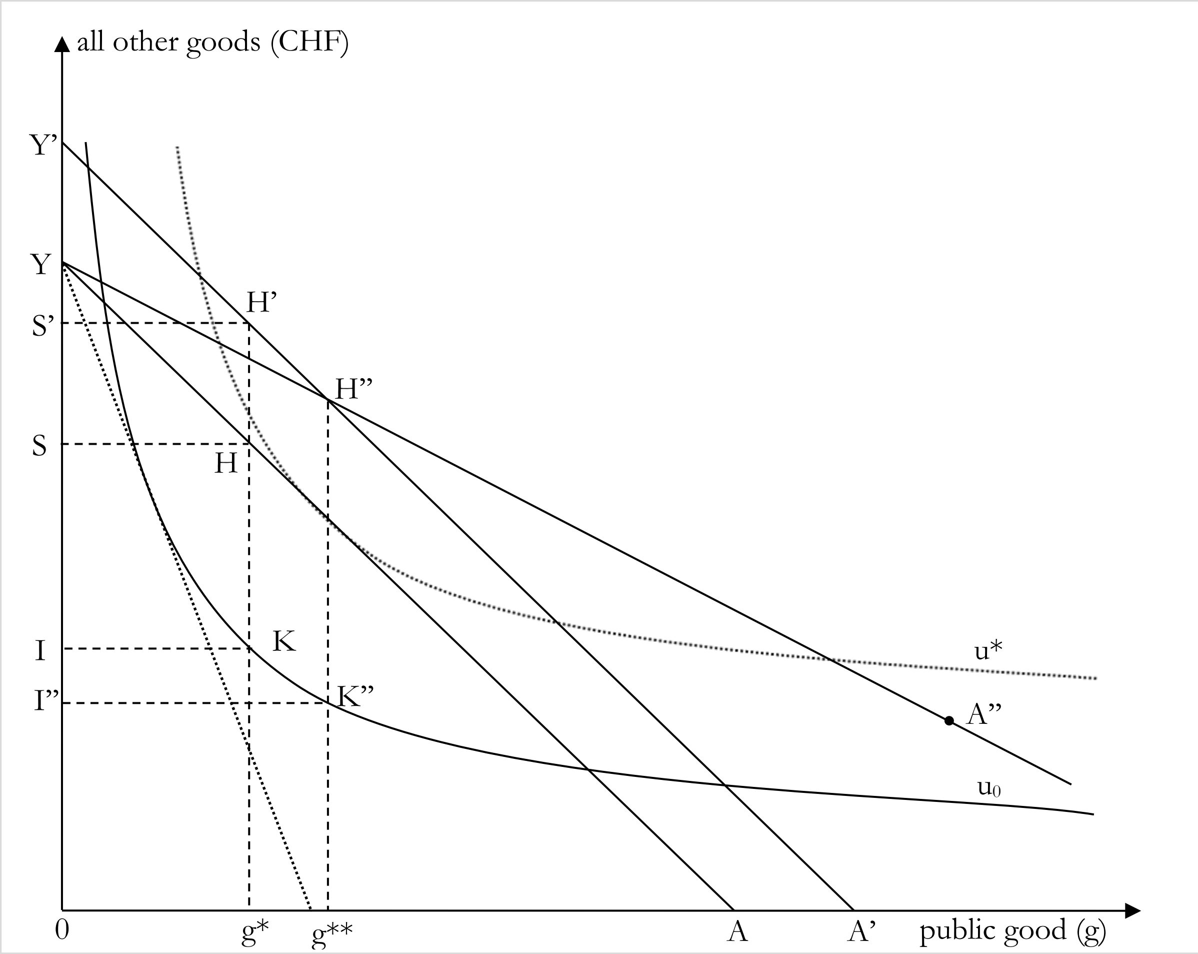
Since the detection of this anomaly, researchers have pursued different paths to find theoretical explanations. While the fiscal illusion is one possibility (see section 3.4), slack maximizing bureaucrats might also trigger the effect.[18] Even after four decades of discussion, there remains no consensus in the literature regarding the explanations of the flypaper effect nor whether it indeed exists. At this juncture, our analysis rests on the model that includes inefficiency as a link in its chain of reasoning. Like the next section, no modifications—apart from some specific interpretations—on the model are needed to address the purpose of the analysis. Therefore, the explanation serves to prime the elements.

If a Migué et al. (1974) type bureaucrat, who maximizes his discretionary slack (i.e. the rent), attains some grants from other state entities, his reaction turns out to partly explain the flypaper effect (Wyckoff 1990).[19] More importantly for the current analysis though, lump-sum and matching grants allow the bureaucrat to increase his discretionary slack. The slack is the difference between the revenue, i.e. the tax the bureaucrat levies as a reward for the public good, and the cost of its production. When the bureaucrat maximizes his slack, he compiles, from the taxpayer's viewpoint, the worst possible combination of taxes and public goods. The larger the slack, the less technically efficient the bureaucrat's provided public service. Therefore, if a grant helps the bureaucrat to increase his slack, the public service provision becomes less efficient the more grants the bureaucrat receives. Figure 2 illustrates that grants indeed support the bureaucrat to expand his slack.

[18] For a dated but still excellent overview of the different attempts to explain the flypaper effect see Bailey & Connolly (1998).

[19] Musgrave & Musgrave (1989: 461 et seqq.) show with a similar model that the effect also results under the assumption of a benevolent administration who uses the lump-sum grant partly to reduce taxes. Clearly, a Migué et al. (1974) type bureaucrat has no incentives to do so.

Figure 2 – Bureaucrat's reaction on lump-sum and earmarked transfers



Source: Adapted from Wyckoff (1990)

The pre-grant budget constraint $\overline{AY}$ sets the limit of each possible combination of goods that the citizens can buy. $\overline{0Y}$ represents the citizens' income. They can choose between the public good *g* or all other goods. As the latter is measured in one Swiss franc unit, the slope of the budget constraint indicates the negative constant per unit costs of the public good *g* as the citizens face it. The per-unit-cost does not necessarily coincide with the true cost of production. In fact, the slack maximizing bureaucrat exploits his information advantage to impose charges to the citizens that exceed the true cost. The citizens have an indifference curve u_0, which sets the utility they obtain without *public* provision of the good g. Imagine u_0 would intersect with the 'all other goods axis' at Y, that is, its position is slightly on the left of the true u_0. The indifference curve would then stop at Y because negative amounts of the public good g do not make sense. Such an indifference curve would depict any combination of g and the other goods that provide the same utility to the citizens, as if no private good was available. As Figure 2 is drawn though, u_0 does not intersect with the 'all other goods axis' at Y which means that private (but imperfect) substitutes for the public good exist.[20]

It is worthwhile to further examine what determines the position of the indifference curve – an aspect Wyckoff (1990) discusses briefly in his paper. Imagine that there is a perfect substitute for the public good for which a competitive market exists. The competitive prices implicitly inform voters about the true production cost of the good g, because every supplier has to offer g at the price of marginal costs (Frank 2006, p.385). Based on information about the cost, voters know their true budget constraint and they end up with the indifference curve u^*. Therefore, the illustrated (solid) budget constraint $\overline{AY}$ represents the voters' budget as if the public good would be produced at its true production cost. Now, the government has two instruments to deceive the voter. First, since the state usually provides services for which a competitive market does not exist, the government can claim to bear higher production costs, such that they coincide with the price of the voters' next best solution; i.e. that is the substitute of the publicly provided good. This means that in reality the voter perceives the imaginary (dotted) budget constraint that is tangent to u_0. Accordingly, the less there are appropriate possibilities to substitute g and the less informed voters are about the true cost of producing g, the more the government can force voters to set their indifference curve closer to the origin by turning the imaginary budget constraint counterclockwise around Y. Second, by disguising the true governmental budget, the state can make the voter believe that her own budget restriction seems smaller than it actually is. Again, the virtual reduction of the governmental budget turns the voter's budget constraint counterclockwise around Y. In a nutshell, the government imposes the minimal indifference curve u_0 on the voter by turning the budget constraint as far counterclockwise as possible.

Based on the knowledge of what u_0 actually denotes, it is clear that a slack maximizing bureaucrat must provide the citizens with at least the utility u_0. At the same time, the bureaucrat needs to acknowledge the citizens' true budget constraint $\overline{AY}$. Therefore, the area between the minimum indifference curve u_0 and the true budget constraint $\overline{AY}$ is the government's room for maneuver as a monopolist; any point is feasible in terms of production and will be accepted from the voter (Wyckoff 1990). As a slack maximizer, the government seeks the longest distance between these two restrictions which is $\overline{HK}$. In K the slope of the indifference curve equals the slope of the budget line. The bureaucrat then pretends the cost of producing g^* is $\overline{IY}$ (whereas it truly is $\overline{SY}$), which leaves the voter $\overline{I0}$ to consume all other goods. The slack the bureaucrat withdraws is $\overline{SI}$ – the distance he intended to maximize.

Starting from the equilibrium derived above, consider another state entity transfers a lump-sum grant of $\overline{Y'Y}$ to the bureaucrat. The citizens' budget constraint therefore moves parallel further away of the origin to $\overline{A'Y'}$. At the same time, the median voter's indifference curve u_0 remains where it was, because the lump-sum grant does not change the price of the next best alternative, i.e. the imaginary (dotted) budget constraint. How does the bureaucrat react to the new situation?

[20] Note that spending the entire available budget for all the other goods yields an inferior utility than u_0, as such a point lies left of the indifference curve. Since u_0 does not go through Y, the citizen cannot entirely relinquish the public good (such as security, for instance) and she instead seeks a private substitute (such as a private security firm).

The new budget constraint $\overline{A'Y'}$ has the same slope and it is the slope that determines the amount of the public good the bureaucrat provides. Thus, the supplied public good is still g^*. With the new budget constraint, though, the bureaucrat can extract even more slack. The governments expenditures increase exactly by the same amount it received as an intergovernmental lump-sum grant and they now sum up to $\overline{0Y'}$. Still $\overline{0I}$ flows into the consumption of all other goods and $\overline{S'Y'} = \overline{SY}$ is the true production cost of the public good. It remains $\overline{IS'}$ as slack for the bureaucrat. Thus, in terms of efficiency, a grant only increases the expenditures while the public good provision freezes. In terms of the flypaper effect, the model predicts public expenditures to grow 100 percent of the lump-sum grant and the bureaucrat does not refund the citizen through lower taxes.

Wyckoff (1990) examined another situation, where the grant is not lump-sum but earmarked for the public good under investigation. Figure 2 also reveals the bureaucrat's reaction to this kind of matching grant. While earmarked grants have no effect on all other goods, they decrease the price of the public good which graphically translates into a flatter budget constraint that crosses Y. Remember that the donor transferred $\overline{Y'Y} = \overline{H'H}$ as lump-sum aid in the example above. Anticipating the receiving government's reaction, the donor's aim is to assign exactly the same amount as the matching grant. Thus, the budget restriction under an earmarked grant regime is $\overline{YA''}$. As before, the slack maximizing bureaucrat supplies the amount of public good that maximizes the distance between the budget constraint and the indifference curve. The maximization problem leads him to g^{**} where he can extract a slack of $\overline{K''H''}$. This slack under earmarked grants is smaller than the one under the lump-sum regime (i.e. $\overline{IS'}$). Thus, in terms of efficiency, earmarking grants have a less negative effect on efficiency compared to lump-sum grants. In terms of the flypaper effect, the model predicts no effect of earmarked grants on public expenditures.

In sum, whether the donor government transfers lump-sum or earmarked matching grants, the bureaucrat's slack and therewith the inefficiency of the public service provision increases with higher grants. A comparison of the two regimes in relation to the slack shows meaningful differences. Under earmarked grants, the bureaucrat withdraws a smaller slack than under lump-sum grants. The magnitude of the difference between the two regimes depends on the cost of the public good and on the marginal rate of the substitution.

3.4. Other sources of fiscal illusion

The refined discussion about the position of the reservation indifference curve u_0 in Figure 2 lays the ground for some simple extensions of the formal model. Indeed, voter misperception of the true production cost or the true governmental budget suffices to explain not only higher public expenditures but also a lower efficiency. Therefore, any further revenue characteristic that potentially leads to a misperception, eventually impairs the efficiency of the public service provision. The ongoing research on 'fiscal illusion' has generated various factors that mislead the voters in telling the true production costs of public services or the true state budget.

In particular, Oates (1985) names four sources of fiscal illusion besides transfers, and each source has its own mechanism in terms of how it misleads the voter; Dollery & Worthington (1996) add some examples of empirical papers to each of the four sources of fiscal illusion. First, the tax composition entails the tax complexity and the tax diversification. A highly complex tax structure challenges the voter to keep the overview of the total amount payed to the government. In contrast, the more fragmented and diversified the tax system is, the less the government depends on one single tax source. This reasonable goal from the finance department's point of view is often accompanied by other agendas seeking to distribute the tax burden more equally or to attain easier public acceptability (Carroll 2009). An administration, reasoning with an equal distribution, the public acceptability, and financial stability potentially distorts less the perception of the true cost of the public service provision. Thus, revenue diversification is primarily a strategy to prevent severe slumps in revenue, while revenue complexity supposedly reduces efficiency (Dollery & Worthington 1996).

Second, the income elasticity of the revenue system refers to the progressivity of the tax system (Oates 1985). The idea that the progressivity of the tax system allegedly veils the true cost of the public service provision is linked to the voter's reference when she pays her taxes. The theory posits that people are more aware of the base tax rate than of their actual tax bill. If their income grows and they in turn pay more taxes due to the progressivity, voters perceive the tax bill to be lower than if the base tax rate increases. Put differently, people care more about the base tax rate than the tax bill (Oates 1975). As long as public expenditures are fundable without raising the tax rate, people will not oppose (Dollery & Worthington 1996). These statements are already provocative, but even if such a mechanism exists, there is an immediate counter argument. Changing the tax rate involves a high transaction cost because an adjustment usually needs some democratic legitimation. The legitimation may come from the parliamentary process or even a ballot, and the latter makes a change even more costly and difficult. This means that people might desire a progressive tax system in order to prevent the government from constantly changing the tax rate, which finds reflection in a less efficient government (Lee & Wagner 1991). In this case though, if any, a positive effect of the progressivity on efficiency would result.

Third, the voter may experience debt illusion. This happens because individuals actually observe the true cost of a public service if they have to pay for it immediately. Alternatively, if the government defers the cost by borrowing the money and inflates its debt, individuals will underestimate the true cost. This myopic view challenges the traditional Ricardo-Barrow theorem. Barro (1988) denies a difference between tax and debt financing as indebtedness simply results in a higher present value of discounted future tax liabilities. Assuming there is myopia instead of a rational equivalence, the voter effectively perceives the two financing sources differently and higher debt triggers lower efficiency levels.

Fourth, the renter illusion targets the different taxation of property and income which makes the tenant underestimate the governmental cost because she is not subject to the property tax. The landlord who faces the tax directly is aware of these costs and probably even forwards the taxation partly to the tenant through higher rents. The tenant, however, cannot differentiate between the part of the rent that originates from the property value and the premium due to the tax. Hence, a lower proportion of property owners presumably leads to less efficiently provided public services (Oates 1985; Dollery & Worthington 1996).

It is worth mentioning that Martínez (2014) could have simply added earmarks to the sources of fiscal illusion, since his model is based on the idea that the administration has an information advantage towards its citizens. He assumes that the latter are less informed about earmarked revenues as they stem from royalties. Given the funding sources of earmarked revenues, rather the contrary is true. Namely, citizens are likely better informed about earmarked revenues.

3.5. Interim conclusion

Based on the public choice framework, this chapter theoretically develops the link between various revenue characteristics and a monetary rent that the bureaucrat can appropriate for his own use. Concretely, the rent can be observed, for instance, through oversized bureaus in terms of physical and personal resources, which, eventually, is tantamount to inefficient public service provision. Reasoning with a diminishing incentive between ministers to monitor each other, the debate prevention theory founds the central hypothesis (H_1): A higher share of earmarked revenues leads to less efficient public service provision, other things equal.

The hypotheses concerning further revenue characteristics derive from the fiscal illusion idea (Oates 1985). That is, the voter underestimates the price for the public service which then opens up leeway for bureaucrats to appropriate rents. The voter's misperception originates from different revenue characteristics which all engender their own hypothesis. The most prominent among them stems from intergovernmental transfers, which often appears under the label of the flypaper effect (Courant et al. 1979). It states that a higher share of transfers implies poor efficiency levels (H_2) (Wyckoff 1990). Debt illusion, for its part, predicts a lower efficiency for highly indebted states (H_3). Likewise, the voters may be misled by the sheer amount of revenue sources, which substantiates the fourth hypothesis: A more complex tax system, ceteris paribus, results in lower efficiency levels (H_4). Finally, the voters might misjudge the price for public services if it rises because of the non-linear tax curve. Hence the prediction that a rising income in combination with greater tax progressivity provokes poor efficiency (H_5).

In succession, all the revenue characteristics and the notion of efficiency require an adequate operationalization, before the hypotheses face the empirical testing. In the first instance, the subsequent chapter provides some background information about the empirical context in which the hypotheses will be tested.

4. Empirical context

In order to test the previously outlined hypotheses, the empirical foundation must not only be comparable in terms of the institutional and legal settings, but it should also provide enough variation in the relevant variables. The 26 Swiss cantons meet these conditions and thus they are an adequate laboratory for the estimations. This chapter deals with the first requirement and goes through the budgeting and spending process whose understanding is key to evaluate the theories' suitability to the Swiss case. Subsequently, it follows a brief overview of the public domains under investigation. Operationalizing and summarizing the variables are dealt with in the next chapter.

Debate prevention possibly intervenes in the cantonal budgeting and spending process in two stages. If ministers want to undertake a larger project, they need to pass the governing council and convince their colleagues of the adequacy and necessity of their plan. If the project exceeds a certain financial threshold, it goes through the government council twice. The governing council first decides upon therewith contingent appropriations (i.e. the credit) and in a second stage the project requires the annual budgetary appropriations (i.e. the budget). In both stages, debate prevention may affect the dynamics of the negotiations.

To account for a possible heterogeneity of the effect in question, the analysis distinguishes between ten different government functions. The division goes along the line used in the official financial statistics. A look on the expenditure share of each government function identifies the education, social security, and health as the top cantonal domains. Later on, the focus is on the two domains transportation and culture that occupy ranks six and nine, respectively.

4.1. Budgeting and spending process

The financial figures and stats of public policies reported in the public annual accounts are the result of relatively complex dynamics that take place within the administration. Understanding the interplay that leads to public spending requires knowledge of the budgeting and the spending process. An explanation of how these two processes work in the Swiss cantons is forthcoming. Stemming from the fairly high autonomy of the Swiss cantons, there is no law determining how to design budgeting and spending. However, there is a non-binding recommendation regarding the organization of fiscal policy and the accounting standards, namely the second Swiss harmonized accounting model for the cantons and municipalities (HAM2) (Conference of cantonal finance directors 2008).[21] The according manual contains an annotated model law concerning financial interests which sets the scene for the budgeting process (art. 14 to 20 of the model financial budget law, MFHG) and the spending process (art. 36 to 48 MFHG). Because these are only recommendations, the manner in which the budget is designed and implemented differs from one canton to another and consequently, this section traces only one possibility. Still, once the exemplary processes presented here are understood, the knowledge can easily be transferred to the other ones. There is less divergence between the cantons with respect to the spending process because the HAM2 treats it in more detail. As a second reference, particularly to reveal the budgeting process, the canton of Solothurn fills the gaps which the HAM2 lefts over and serves as an illustrative example.[22] Table 1 helps to follow the subsequent explanations.

[21] The HAM2 was developed by the conference of the Swiss cantonal directors of finance and follows to some extent the international public accounting standards (IPSAS). The most recent version is published online; see: http://www.srs-cspcp.ch/.

[22] Note the link to chapter 7.8, where the canton of Solothurn also serves as a case study. The canton of Solothurn regulates the budgetary and spending process in the impact-oriented administration management law (WoV-G SO) and its respective decree (WoV-Vo SO). Additional information stems from an interview conducted with the finance manager of the canton of Solothurn and one of his controllers on August 29th 2016.

The budgeting process starts with two parallel actions. On one side, each spending office gives its input into the electronic accounting system [1.1].[23] Each accounting area specifies their expected revenues and requests expenditures which results in most cases in negative balances. The aggregation from these inputs results in a budget with an institutional perspective (Pfäffli 2011). At the same time, the parliamentary finance commission (corresponds approximatively to the appropriation committee in the U.S. system) decides upon the budgetary targets for the attention of the government council [1.2]. The targets are based on a rolling 4-years-ahead financial blueprint (§ 22 WoV-G SO). The finance commission's view is, in contrast to the spending offices, rather macroeconomically driven with a strategic perspective. The finance office then receives the commission's budgetary targets and prepares the government council's answer to it [1.3]. Concretely, in the case of the canton of Solothurn, after some informal talks between the finance minister and his financial manager, the finance office writes the comments for the government council who then simply passes them on [1.5]. The answer of the government council goes afterwards back to the finance commission, which discusses and comments upon it [1.7]. Once the spending offices have placed their input into the accounting system, the finance office collects the tentative information on behalf of the government council [1.4]. The government council then instructs each department on how much they have to cut their budgeted expenditures in order to meet the target of the finance commission or at least to come close to it [1.6]. The government council's instructions mark the end of the first version of the budget.

For the purpose of complying with the financial target set by the finance commission, the financial manager, together with one of his controllers, meets a delegation of the spending offices to discuss possible budget adjustments [2.1]. Once the spending ministers have adjusted the inputs in the accounting system, the finance office prepares the government council's budget report [2.2]. In fact, some cantons (the canton of Aargau, for instance) even pass the prior consultation of the finance commission and start their budgeting process with the government council setting the budgetary targets and instructing the spending offices. As a final step in the administration stage, the government council transfers the compiled budget to the parliament [2.3].

Having passed the executive, the budget proceeds to the parliamentary stage. There, the functional commissions (e.g. social and health commission, education and culture commission etc.) contemplate the budget within their occupation and propose adjustments for the attention of the finance commission [3.1]. The latter then discusses the budget as a whole, taking into account the adjustment proposals of the functional commissions [3.2]. In particular, the finance commission is concerned about financial ratios and accordingly agrees on or rejects the budget adjustments of the various functional commissions. The latter then reconcile the differences together with the finance commission, if there are any [3.3]. The adjusted budget goes back to the finance office [3.4] that integrates the adjustments into the accounting system and prepares the government council's comments on it [3.5]. The government council then passes on the comments to the attention of the parliament without modifications in the budget [3.6]. Finally, the parliament discusses the budget and decides on it [3.7].

---

[23] The number in brackets refers to the budgeting stage in Table 1.

Table 1 – Budgeting process

<table>
<tr><th colspan="3"></th><th colspan="7">year t</th></tr>
<tr><td rowspan="3">executive</td><td rowspan="2">administration</td><td>finance office</td><td></td><td>[1.3] prepares the govern-ment council 's answer to the budgetary targets</td><td>[1.4] collects tentative information from accounting system</td><td></td><td rowspan="2">[2.1] one or several representatives of each spending office meet with the finance office to discuss adjustments of expected revenues and expenditures; spending offices adjust the inputs in accounting system</td><td>[2.2] compiles budget report</td><td></td></tr>
<tr><td>spending offices</td><td colspan="2">[1.1] give input on expected revenues and expenditures into the accounting system</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">governing council</td><td></td><td></td><td>[1.5] comments upon the budgetary targets and passes the preliminary decisions on the budget based on a 4-year financial blueprint</td><td>[1.6] instructs the spending offices to adjust their expected revenues and expenditures in order to meet the targets</td><td></td><td></td><td>[2.3] passes the budget</td></tr>
<tr><td rowspan="3">legislative</td><td colspan="2">finance commission</td><td>[1.2] decides upon the budgetary targets based on a 4-year financial blueprint</td><td></td><td></td><td>[1.7] comments upon the government council's answer to the budgetary targets</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">functional commissions</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">parliament</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Table 1 – Budgeting process (continued)

		year t						
Executive – administration	finance office					[3.5] integrates budget adjustments into the accounting system		
Executive – administration	spending offices							
Executive	governing council						[3.6] comments upon the proposed budget adjustments	
legislative	finance commission		[3.2] discusses the budget overall (particularly financial ratios etc.); proposes budget adjustments to functional commissions	[3.3] reconcile differences	[3.4] passes on the proposals of the budget adjustments agreed between the various commissions			
legislative	functional commissions	[3.1] discusses the part of the budget concerning their function		[3.3] reconcile differences				
legislative	parliament							[3.7] discusses the budget and decides on it

Source: own composition

Once the budget is effective, the executive is, in principle, authorized to actually spend the money. This is only true for expenses that fulfill certain conditions though. In the words of the model financial budget law, an expense is a financial liability that in any case needs a certain type of credit before it is incurred (art. 36 MFHG). The model law distinguishes between four types of credit:

- the ordinary budgetary credit is decided upon and approved automatically through the accounting balances of budget and is thus valuable for one year (art. 44 MFHG);
- the supplementary credit brings to bear if the approved ordinary budget credit is insufficient or if there is none (art. 46 MFHG);
- the contingent credit is used for projects whose expenses exceed a certain threshold or endure several years (art. 37 MFHG);
- the ancillary credit is needed if the approved contingent credit is insufficient (Art. 43 MFHG).

The spending process comes only into play for projects that need more than the ordinary budget credit. Consider some spending office that intends to carry out a new project, which requires a non-ordinary credit, i.e. a supplementary, contingent, or ancillary credit [4.1]. The office prepares a report justifying the need for the project and discussing the financial liabilities expected [4.2]. In order to avoid overly large discrepancies in the government council, the spending office then opens a co-report procedure where it invites the possibly interested offices. As soon as a project requires financial resources, inviting the finance office to comment upon the report is mandatory anyway (§ 26 WoV-Vo SO). In its co-report the finance office evaluates the project with respect to the question of whether the funds are applied economically [4.3]. In the next step the government council debates the project, taking into account the report as well as the co-reports. Depending on the level of expenses, the government council can definitively decide, or it must pass the report on to the attention of the parliament (art. 37 MFHG) [4.4, 4.5]. After a prior consultation of the functional or the finance commission [4.7, 4.6], the parliament discusses the project together with the credit [4.8]. Again, depending on the level of expenses, the parliament definitively decides or a referendum applies. If the last resort, be it the government council, the parliament or the voters, approved the credit, the spending office may spend the money within the frame defined in the credit request [4.9].

The two parallel processes serve as a reference for the previously presented debate prevention theory. A special focus lies on the actions between [1.1] and [2.3] for the budgeting process and between [4.2] and [4.5] for the expenditure process. The interplay between the executive and earmarking is key in these two phases, and this matches with the debate prevention theory. Note that the debate prevention theory did not differentiate between the budgeting and the spending process. Indeed, both processes are relevant for what the researcher finally observes in terms of public expenditures (inputs) and public service (outputs). The mutual consideration seems therefore appropriate.

Table 2 – Spending process for non-ordinary credits

		year t						year t + 1
executive	administration: finance office			[4.3] comments upon the project as to whether the funds are applied economically				
executive	administration: spending offices	[4.1] intends to carry out a new project such that it requires a non-ordinary credit	[4.2] prepares the report that justifies the need for the project and contains the credit request to approve the expenses					[4.9] spends the money within the frame defined in the credit
executive	government council				<u>credit above certain threshold</u>: [4.4] passes project and hence credit <u>credit below certain threshold</u>: [4.5] decides on project and hence on credit			
legislative	finance commission					<u>credit above certain threshold</u>: [4.7] discusses the credit		
legislative	functional commissions					<u>credit above certain threshold</u>: [4.6] discusses the credit		
legislative	parliament						[4.8] discusses the project together with the credit and decides on it (above a certain threshold, a referendum applies)	

Source: own composition

4.2. Public domains under investigation

The functional chart of accounts of the HAM2 guarantees a certain comparability between the cantons and years with reference to the various categories of government function. The HAM2 discusses ten functional categories in which the cantons classify their expenditures. Each of these categories has several sub and sub-sub categories (Conference of cantonal finance directors 2008). Table 3 lists the ten superior categories together with some prominent sub categories as examples and the respective expenditures in absolute and relative terms in 2014. The figures provide a feeling for the absolute and relative sizes of the government functions. The last column anticipates the operationalization of the variable related to earmarked revenues. Concretely, it shows the fraction of expenditures that is funded through earmarked revenues.

Table 3 – Cantonal expenditures and earmarked funding in 2014 by government function

Government function	Prominent sub and sub-sub categories	Expenditures (in 1'000 CHF)	Expenditures (in percent)	Earmarked revenues[1]
General administration	General services, executive and legislative organs	65'34'036	7.67	0.28
Public order and security, defense	Police, justice, corrections, army	77'02'816	9.04	1.71
Education	Obligatory schooling, universities, research	239'21'269	28.07	0.37
Culture, church, sport and leisure	Theaters, cultural heritage, libraries, sport events	16'82'716	1.97	32.80
Health	Hospitals, preventive care	120'17'287	14.10	0.53
Social Security	Old age and survivors, disability, social welfare	172'50'345	20.24	3.24
Transportation and communication	Construction and maintenance of roads, public transport	61'98'567	7.27	28.35
Protection of the environment and spatial planning	Sewage disposal, waste management	14'54'930	1.71	11.73
National economy	Agriculture, forestry, hunting and fishing	44'73'515	5.25	3.81
Financing and taxes	Asset and debt management, fiscal equalization	39'84'959	4.68	n.a.

[1] in percent of expenditures

Source: Federal Finance Administration (2016);

The three largest expenditure categories are education, social security, and health. Together they account for almost three quarters of the total expenditures of all cantons in 2014. Even if defense falls completely under the responsibility of the confederation, the domain of public order and security, defense (hereafter public order and security) still ranks fourth. Note that general administration appears rather high on the list. One reason is that this domain captures not only the parliaments and executives, but also a large part of the financial administration and the tax authority. Transportation and communication (hereafter transportation) takes a middle position and, on the cantonal level, is reduced to tasks related to transportation, as communication is by and large settled on the national level. The national economy is positioned seventh on the list, representing the administration, regulations, and – to a limited extent – the support of different industries and sectors. The function of financing and taxes does not provide any public service per se, but it does encompass interest payments, financial flows of the fiscal equalization, et cetera. The last two functions each account for less than two percent of the expenditures. In both domains, in protection of the environment and special planning as well as in culture, church, sport and leisure (hereafter culture), the municipalities as the lowest state layers play a significant role, which explains the limited cantonal expenditures.

In contrast to its meager weight with respect to expenditures, earmarking plays the most important role in the government function of culture as the later analysis shows. Around one third of all expenditures for culture were funded through earmarked revenue in 2014; the share remains about the same when considering the entire timespan from 2000 to 2014 (see Table 27 in the appendix). With almost thirty percent earmarked funding, transportation comes second (over the entire timespan the share is slightly smaller). In recent years, the domain of protection of the environment and spatial planning also reaches shares of more than ten percent of expenditures funded through special financings. In the remaining functions, earmarking is almost negligible with percentage figures in the low one-digit range. Consequently, this analysis concentrates on the two domains of culture and transportation, while all the data and the estimations of the other functions largely appears in the appendix.

4.3. Interim conclusion

Although a theory can describe reality only in an abstract manner, it still needs to account for the central elements of the relationship it approaches. Concretely, the debate prevention theory predicts less monitoring between ministers if they benefit from earmarked revenue, which eventually leads to a less efficient public service provision. This chapter identifies the stages in the budgeting and spending process, where these configurations typically occur and are thus observable. If ministers want to achieve a larger project, they need to pass the governing council and convince their colleagues of the adequacy and necessity of their plan. If the project exceeds a certain financial threshold, it goes through the government council twice. First the governing council decides upon therewith accruing liabilities (i.e. the credit) and in a second stage the project requires the annual financial allotments (i.e. the budget). In both stages, debate prevention may affect the dynamics of the negotiations.

To account for a possible heterogeneity of the effect in question, this analysis distinguishes between ten different government functions. The division follows what is used in the official financial statistics. A look at the expenditure share of each government function identifies education, social security, and health as the top cantonal domains. Later on, the focus is on the two domains of transportation and culture that occupy ranks six and nine, respectively.

5. Operationalization of revenue characteristics

A glance on the existing literature in section 2.1 demonstrates the prevalence of earmarking practices. While the idea of earmarking in general is to dedicate certain revenues to a specific purpose, its understanding in practice varies vastly. Anecdotal evidence comes from an episode in the United States Bush Administration era. In 2007, the Congress passed the Honest Leadership and Open Government Act which, inter alia, demands a public disclosure of each earmarking request that made it to the bill.[24] During the Senate debate, the question of how to define earmarks arose. While the Senate's original version of the act only covered appropriations contained in the bill's text, a finally passed amendment extended the definition to earmarks listed in the report to the bill. This subtle difference increased the requests in question nearly twentyfold (Finnigan 2007).[25] Similar issues on how to define earmarks arise in Switzerland, where the cantons use them, usually under the term special financing, very heterogeneously.

Thanks to the HAM1 and HAM2, the definitions of debt, tax complexity and transfers turn out more spontaneously. Having said this, the transfers are broken down into an earmarked and a non-earmarked subgroup. Although this refinement complicates the operationalization to some extent, some existing operationalizations in this matter give some guidance. Tax progressivity as a last revenue characteristic imposes further operationalization issues. A novel approach yields interesting insights here.

Once the terms are defined and operationalized, this chapter provides an overview of the practical relevance of the revenue characteristics. There are two cantonal domains in which earmarks play a significant role. On average, the cantons fund 35 percent of their expenditures via earmarked revenues in the cultural domain. In the transportation domain this share still adds up to 24 percent. The remaining domains yield amounts between 0.5 percent (education) and 12.1 percent (environment and spatial planning) and they therefore are not analyzed in closer detail (summary statistics are given in Table 27 in the appendix).

5.1. Earmarked revenue

Before going into the details of the operationalization of earmarked revenues using data of the Swiss cantons, a more general discussion embeds the term 'earmarked fund' into a broader context. Figure 3 comprises the possible financial flows related to earmarked funds. The filled boxes and arrows represent elements that are directly linked through the chain of financial flows, while the hatched arrows connote in- or outflows. Theoretically, in- and outflows are possible, but if inflows are present, the respective earmarked fund is, according to the definition used here, not considered as such anymore.[26] Note that the exclusion of inflows prevents any crowding out of general fund financing of a specific purpose by earmarked revenues. In addition, there is obviously no outflow from the earmarked fund to the general budget or to any other specific purpose, which is the core of the earmarking idea. For this to hold in reality, the benefitting good must be defined in a sufficiently narrow way.

The diagram of the financial flows starts with a tax for a private good (e.g. gasoline) or a charge for a public service (e.g. fee for hunting license). The revenue stemming from either source then

---

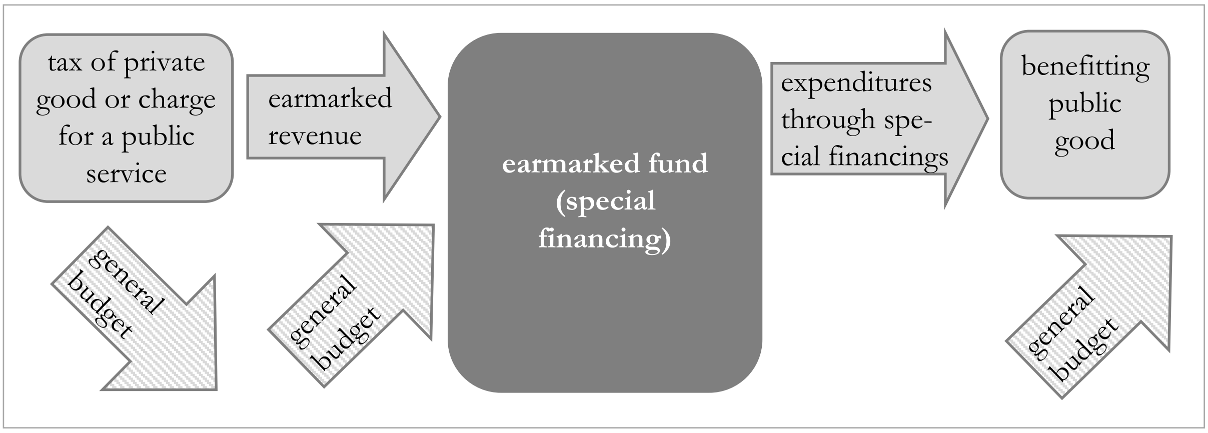
[24] In 2005, the House Appropriations Committee received almost 35'000 requests out of which about a third effectively found their way into the bill (Porter & Walsh 2008).

[25] Ultimately, in 2008, President Bush issued an executive order forcing the administration only to expend on the basis of the text of the laws. Since only the bill itself but not the corresponding report entails the characteristic of a law, this order implicitly forced Congress to include all earmarks in the bill (Porter & Walsh 2008).

[26] Of course, different understandings are possible. For a classification of different types of earmarked funds see Bird & Jun (2005). According to their typology, earmarked revenue as it is grasped here, would be considered substantive earmarking, in which an increase of earmarked revenues has a direct effect on the special fund or the linked expenditures.

flows either entirely – equipped with an earmark – into the respective fund, or a part of it contributes to the general budget. In addition to the earmarked revenue, the parliament could theoretically add additional financial resources of the general budget to the earmarked fund. In this case though, the fund is excluded from any further analysis. The stock of financial resources in the earmarked funds then flows (labelled 'expenditures through special financings') to a previously specified public good. This good also possibly collects further resources from the general budget. In this case, again, the earmarked fund is not taken into account as such anymore.

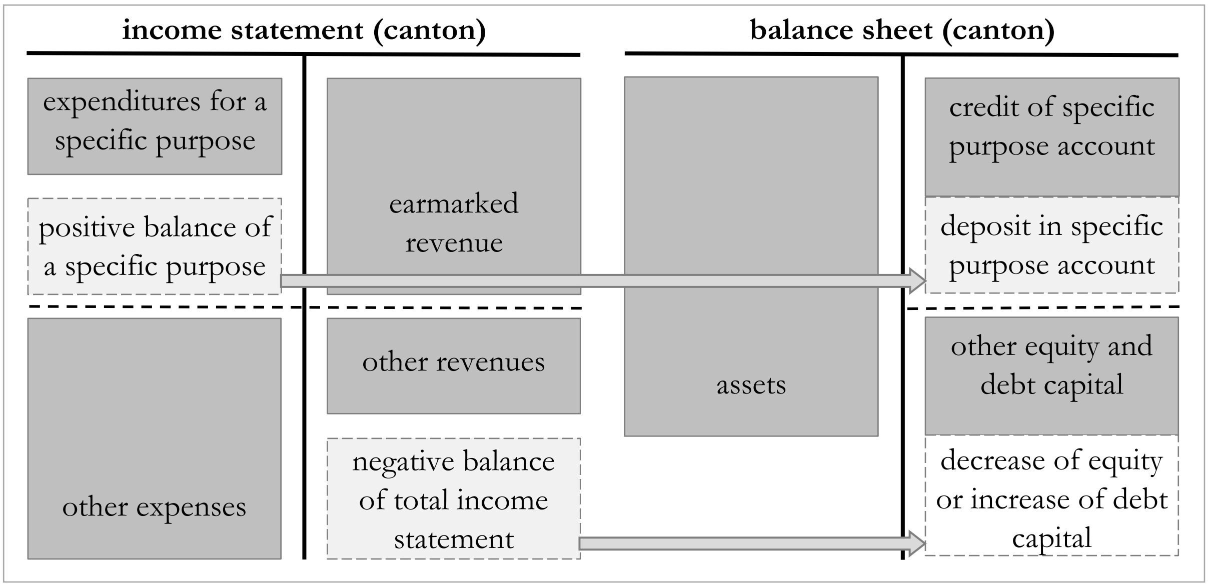
Figure 3 – Financial flows of earmarked funds



Source: Own illustration

Using the aforementioned definition of earmarked funds, it is clear that the spirit and purpose of earmarking revenues is to ensure that they are used solely for a specific predefined application. From a bookkeeping point of view, a separately reported account guarantees the claimed exclusive use while keeping the disclosure transparent. In such a system, earmarked revenues still enter the general financial statement and they are subsumed under the total revenues. Likewise, the corresponding expenditures represent a part of the total expenses of the annual financial statement. However, by the end of the year, the specific purpose account (called earmarked fund above) nets the earmarked revenue and the corresponding expenditures independently of all other transactions. If the balance is positive, then a credit note increases the respective account in the balance sheet. Otherwise, the account decreases by the amount of the deficit. The process is explicated using the diagram in Figure 4.

Figure 4 – Accounting diagram of earmarked revenues



Source: Own illustration

The example in Figure 4 describes the case in which the specific purpose account (i.e. above the dotted line) in the income statement closes positively. The surplus increases the specific purpose account of the balance sheet. At the same time, the total income statement shows a deficit, which in turn decreases the equity, or, if the equity becomes negative, requires additional borrowings. In this sense, the specific purpose accounts sit in between equity and debt capital. On the one hand, they can be viewed as debt capital, which the government owes to the respective purpose. On the other hand, they share characteristics with equity as their revenue stems from saved surpluses. The same arguments hold if the specific purpose account is in deficit. The HAM1 acknowledged only the first reasoning and assigned specific purpose accounts to debt capital (recommendation 454 of the HAM1; Conference of cantonal finance directors, 1981).[27] The HAM2 solves the ambiguity by defining criteria as to the leeway in decision-making over the account. Depending on whether the superior law or the state entity itself determines the earmarking and details the use, the specific purpose account is part of the debt capital or the equity (recommendation 8 of the HAM2; Conference of cantonal finance directors, 2008).

Since the two domains of transportation and culture are central in the forthcoming analysis, it is worth looking closer at their concrete earmarked revenues and, accordingly, at their legal base. For transportation, the most prevalent specific purpose accounts hold resources for roads. Comparing their legal bases reveals that the revenue sources are similar in most cantons. The generic revenues are taxes and fees for motor vehicles, federal transfers stemming from the petroleum tax and from the heavy vehicle fee (*Leistungsabhängige Schwerverkehrsabgabe*, LSVA), and fees for excessive or special use.[28] Despite the federal law urging cantons to use the before-mentioned transfers specifically for the roads,[29] they can nevertheless decide on their own whether they want to set up an affiliated specific purpose account. Cantons can additionally and

[27] At the same time, the chart of accounts of HAM1 assigns the specific purpose accounts to distinct category.
[28] See, for instance, § 24 road law of the canton of Solothurn as of 24 September 2000, § 35 road and route law of the canton of Zug as of 30 Mai 1996.
[29] See, for instance, 85(2) Swiss Federal Constitution as of 18 April 1999.

autonomously define whether the transfers exclusively fund road maintenance and construction or if the means can also be used for the periphery (wildlife corridors for instance). Hence, the cantonal parliaments decide independently of the federal law on the installation of a specific purpose account in the transportation domain.

In comparison, in the cultural domain, the usage of specific purpose accounts differs more from one canton to another. Following the inter-cantonal agreement on the joint implementation of lotteries as of 26 Mai 1937 (*Interkantonale Vereinbarung betreffend die gemeinsame Durchführung von Lotterien*), every canton maintains a specific purpose account for accumulating lottery profits. Consequently, the amount of earmarked revenues deposited in these accounts depends on the players' appetite to bet. Besides the lottery fund, most cantons support culture, sport and leisure with other specific purpose accounts that are funded with a yearly fixed amount of the general budget, with revenues from specific divestments, fines for misdemeanors or with pecuniary compensations.[30] Also in the cultural domain, it is the cantonal parliament that decides upon the specific purpose account.

The term 'specific purpose account' defines a family of accounts whose resources are earmarked. The HAM1 and HAM2 name several types that differ significantly in kind:

- Special financings: According to HAM1 „[t]he term special financing denotes […] the fully or partial assignment of revenue sources to specific functions (earmarked revenues)" (Conference of cantonal finance directors 1981: 27)[31]. The revised HAM2 stayed with the same definition (recommendation 8). The two standards specify the purpose of special financings in restricting their scope to domains where there is a real causality between the charge and the service. Since such a causality hardly exists between ordinary (income, property, profit and capital) taxes and any government function, the recommendation is to desist from earmarking them. Because special financings potentially and rather restrictively narrow financial leeway, they need a legal basis. Special financings can be part of the equity or the debt capital under the HAM2. The HAM1 creates a third category within the liabilities only for special financings.
- (Special) funds: When developing the HAM1, the authors recognized that the term 'fund' signifies different things in financial practice. They therefore explicitly avoided using the term (item 453 HAM1). With the revision, the HAM2 reintroduced funds as a light version of special financings. Funds also require a legal basis. In contrast to special financings, funds get along without a causality between the revenue and their intended use. Therefore, even taxes can be earmarked and allocated to a fund (see the description of the account 1091 'receivables from funds within the debt capital'; Conference of cantonal finance directors, 2008). Funds can be part of the equity or the debt capital.
- Legacies and foundations: If third parties want to contribute financially to closely related government functions, they usually donate in form of legacies or foundations. The canton administers these donations and allocates the resources according to the testament or to the foundation goal. The HAM1 recommends displaying neither legacies

[30] See, for instance, 2(1) law regarding the fund for contemporaneous art of the canton of Geneva as of 7 May 2010 or 16 nature and heritage protection law of the canton of the canton of Glarus as of 2 May 1971.
[31] Translated from the German version.

nor foundations in the balance sheet because they are not funded by general public revenues (item 1023 and 1043 et seq.). The HAM2 reverses the old recommendation by listing specific accounts for legacies and donations in the balance sheet. Both can be part of the equity or the debt capital.

- Leftover accounts with earmarked resources: Because the HAM1 stretched the special financing further than the HAM2, the former did not need a leftover category. HAM2 names research contributions as a concrete example for the leftover accounts (account 2093). The leftover accounts exist only within the debt capital. Apart from that, the manual provides no further additional information or a recommendation concerning these accounts.
- Prefinancings: Particularly for larger projects, prefinancings serve to spread the expenditures over several preceding years before the government achieves a not yet legally decided project (item 445 HAM1 and recommendation 8 HAM2; Conference of cantonal finance directors, 2008). HAM1 did not provide a specific account but recommended to use special financings instead (item 453 HAM1). As such, they are always part of the debt capital. The revision of the accounting standard relaxed the recommendation concerning prefinancings. HAM2 views them as a reserve and creates a separate account for them which consequently is part of the equity (see the description of the account 293 'prefinancings'; Conference of cantonal finance directors, 2008).
- Reserves from domains managed through global budgets: Global budgets are an instrument from the new public management (NPM) approach. Instead of governing the administration by granting a certain amount of resources for every task (i.e. input control), NPM suggests targeting indicator goals (i.e. output control) and grant global budgets for an entire department (Pfäffli 2011). In addition, global budgets do not necessarily expire by the end of the year, but the departments can deposit surpluses in reserves in order to tap them later on. Because the parliament approves global budgets for a specific department, it is the only one that can have access to the respective reserves. In this sense, the resources are earmarked too (art. 48(3) MFHG and its corresponding comment; Conference of cantonal finance directors, 2008). As the name suggests, reserves from domains managed through global budgets are always part of the equity. As far as the old standard is concerned, global budgets did not exist at the time the HAM1 was developed.
- Provisions: If at the end of the year, the exact amount and/or the due date for a yet unpaid but justified, probable, legal or factual liability is unknown, then a provision applies. In the subsequent year, the provision serves to settle the invoice. In comparison to the HAM1 (item 425), the HAM2 (recommendation 9) only expands on the number of accounts of specific provisions but otherwise the respective recommendation remained equal. In particular, they both decree that a provision must be used for the purpose it was formed for. Provisions are always part of the debt capital.
- Deferments: In contrast to provisions, deferments apply for the short term and if the amount to be paid is fairly sure. In all other aspects, deferments are equal to provisions (see comment to recommendation 9 HAM2 and item 426 HAM1).

- General reserves: The HAM1 simply qualifies ‘false provisions’ to cover future liabilities for deliveries and services as general reserves and bars them (item 425). Neither does the HAM2 provide an accurate definition. Instead, it provides the scope of the general reserves. Two accounts are the result of purely accounting transactions: restatements and reevaluations. The aim of the third, i.e. fiscal reserve (account number 294), involves reducing the deficit of the income statement or financing future capital expenditures. It is also known as the economic cycle or the compensation reserve. General reserves are always part of the equity.

With their different characteristics, not all of the previously listed candidates for specific purpose accounts comply with the debate prevention theory in section 3.1. This only applies to specific purpose accounts that satisfy the conditions described above as comment to Figure 3 and Figure 4. Simply referring to the list above in order to determine which account to exclude from the analysis is impossible, however, because the HAM1 as well as the HAM2 are only non-binding recommendations, something which leaves the cantons the leeway to interpret them somewhat differently. The practical applications do not necessarily coincide with the example description above and the cantons may use different accounts for the same transaction. The subsequent paragraphs examine whether the example case of each specific purpose account meets the relevant assumptions from the theory section. In case it exists, a counterexample is included to demonstrate that the cantons indeed use the respective account in different ways. Table 4 gives an overview of the accounts and states whether they take part in the analysis:

Table 4 – Specific purpose accounts of HAM1 and HAM2 and their consideration

Account	To consider in principle	Counterexample where (not) to consider the account anyway
Special financings	yes	special financing for road construction, canton of Solothurn
(Special) funds	yes	fund for disabled people; canton of Aargau
Legacy / Foundation	no	N/A
Leftover accounts with earmarked resources	yes	N/A
Prefinancings	HAM1: yes HAM2: no	N/A
Reserves from domains managed through global budgets	no	N/A
Provisions	no	N/A
Deferments	no	N/A
General reserves	no	alcohol-tenth; canton of Aargau

Source: own composition

The first theoretical assumption is about the appropriation. Indeed, the revenue has to be earmarked in the sense that the scope is sufficiently narrow in order to allocate the revenue to a specific department. If this assumption does not hold, then there might still emerge a debate about the allocation and the incentives to monitor revert. In the words of Wilkinson (1994), who classified different types of earmarking, this assumption refers to narrow earmarks. General reserves leave considerable room for manoeuver and provoke potentially tough debates when it comes to their distribution among departments. Therefore, there is no real, or at least no restricting, earmarking ascertainable and the general reserves in principle drop out of the analysis.[32]

The second assumption the theory makes concerns the externality of the earmarked share. Concretely, the law predetermines the allocation of earmarked revenue to certain public services. In the budgetary process, each department receives a given fixed proportion of earmarked revenue (denoted as α_{D+A} above). This assumption rules out the consideration of reserves from domains managed through global budgets, provisions, deferments, and potentially prefinancings. Albeit the credits of all these accounts are earmarked for specific purposes, their revenue source is not. For instance, a department whose domain is managed through a global budget must debate on the part it receives from the general budget just as all the other departments do. Only if it manages to acquire a surplus does the latter flow into the earmarked account. The same reasoning holds true for provisions and deferments.[33] In terms of prefinancings, their principle exclusion depends on the accounting standard. HAM1 assigns them to special financings that comply with the theory. Under HAM2, prefinancings embody a discrete account whose resources are allocated either by law or during the budgetary process to certain public services. In the former case, the law effectively prevents the debate, either by stipulating a fixed amount or by allowing the apportioning of a potential surplus to the prefinancing at the end of the year.[34] In the latter case, the first assumption rules out prefinancings from being considered because they do not prevent the debate.[35]

Legacies and foundations do not necessarily conflict with the two assumptions (i.e. appropriation and externality) above. However, they differ fundamentally from the other accounts in one aspect—legacies and foundations are always a voluntary donation to the state and not a charge or a tax imposed on the citizens. With regard to the theory, they should be classed among the market delivered goods (denoted c above) rather than among the public goods. People can freely decide whether they want to contribute to the charitable good. The state only acts as a service supplier who fiduciary manages the donation. In addition, some cantons even have rules listing the conditions under which they accept donations in the form of legacies and foundations. Consequently, legacies and foundations are disregarded from this analysis without any exceptions.

Finally, three special purpose accounts remain. Special financings, funds, and leftover accounts with earmarked resources are paragons for the stocks of earmarked revenue, as detailed by the

32 The canton of Aargau provides a counterexample to the previously described typical usage of general reserves. Whereas most of the cantons use a special financing to register the so-called alcohol-tenth, the canton of Aargau enters it as general reserve (i.e. 'Rücklagen') in its books. The alcohol-tenth is the part of the federal tax on alcoholic beverages passed to the cantons. The alcohol law obliges the cantons to use their share for combating alcoholism and other addictive substances (art. 45(2) Alcohol Law as of 21 June 1932).

33 There are no counterexamples for deferments, provisions, and reserves from domains managed through global budgets. They do not need to be considered in the analysis.

34 The canton of Fribourg established a prefinancing funded with surpluses. According to the Financial Budget Law of the canton of Fribourg, the government can appropriate a part of the possible surplus at the end of the year to the prefinancing for infrastructures (art. 42a[bis] of the Financial Budget Law as of 25 November 1994). The canton of Appenzell Ausserrhoden provides an example of a law that stipulates a fixed amount, originating from a revaluation surplus of its hospitals, and deposits it in a prefinancing (art. 31 of the Hospital Network Law as of 19 September 2011, *'Spitalverbundgesetz'*).

35 For instance, the canton of Nidwald maintains diverse prefinancings. Their purposes range from compensating shortfalls from tax revisions, to revenue shortfalls due to decreased profit distribution of the central bank, to pre-finance the hospital (see the appendix of the annual financial statement of the canton of Nidwald for the year 2014). To build the capital stock of these prefinancings, the cantonal Financial Budget Law requests a respective decision during the budgetary process (art. 10(2) of the Financial Budget Law as of 21 April 2010).

theory explanation in section 3.1. In principle, they combine the beforementioned assumptions about the appropriation and the externality of the revenue share. Likewise, their source of funding is public, i.e. taxes or charges. Despite these characteristics, there still exist counterexamples.[36]

Special financings deserve a more detailed discussion according to the example of the special financing for road construction in the canton of Solothurn. When the canton advanced to the HAM2, it decided to depreciate its roads linearly over their useful life instead of immediately and it registered the recovery value within the special financing.[37] As illustrated in Figure 4, the deposits in or the withdrawals from special financings balance out the income statement of the specific purpose account. Therefore, under immediate depreciation all transactions of specific purpose accounts are neutral with respect to the overall balance of the income statement. Under non-immediate depreciation, the depreciations in the income statement differ from the net capital expenditures. This difference, together with the deposits and withdrawals, then determines the stock change of the special financing. Put differently, the stock change of the special financing not solely represents the balance of the income statement of the specific purpose account, but also the difference between net capital expenditures and depreciations. In addition, under non-immediate depreciation the special financing inflates the balance sheet because it contains the recovery value, i.e. the administrative property. If the transactions of the specific purpose account are no longer neutral with respect to the overall income statement, then the respective domain is exposed to the budgetary debate again. Consequently, special financings with such a system must drop out of the analysis. Having said that, most cantons decided to register potential recovery values outside of the special financings when they changed to a non-immediate depreciation, which makes the transactions of the specific purpose account neutral. In this case, the respective domain can prevent the debate, which is in line with the theory.

The bullet points above list stock-accounts, whereas the hypothesis utilizes profit and loss accounts. More precisely, the theory hypothesizes that a higher proportion of the expenditures for a specific purpose in comparison to the total expenditures leads to less efficiency. With reference to Figure 4, the required data are located above the dotted line of the income statement. Although neither the HAM1 nor the HAM2 explicitly recommends it, some cantons individually provide income statements and balance sheets of the specific purpose accounts in the appendix of the annual financial statements, thus furnishing the information in the most transparent way. The second-best solution is to run cost centers financed entirely through one specific purpose account and to report these cost centers individually in the institutional classification of the annual financial statement. Lastly, there are cantons in which the information is not publicly accessible. Most of the cantons of the last category are cooperative and provide the information upon request.

[36] The canton of Aargau partly uses funds for accumulating donations for a specific purpose (§ 32 of the Decree about the Impact Oriented Management of Tasks and Financial Resources as of 5 June 2012). As such, they are a third possibility besides legacies and foundations to stock donated earmarked revenue. Hence, they are not considered in this analysis. There are no counterexamples for leftover accounts with earmarked resources and they are therefore always considered in this analysis.

[37] Figure 23 in the appendix schematically illustrates the mechanism of an accounting regime with non-immediate depreciations. As comparison, Figure 24 draws on the same example while depreciating immediately.

5.2. Transfers

In contrast to the earmarked revenue, HAM2 proves more useful in directly identifying the transfers the cantons receive. The federal finance administration publishes the cantonal financial data in a harmonized and hence comparable form in accordance with these very standards. Accordingly, it suffices to determine the relevant accounts of the HAM2 with respect to transfers.

The obvious account category to consider is transfer receipts (item 46 HAM2). The category subsumes receipt shares (item 460 HAM2), compensations (item 461 HAM2), fiscal equalization and cost compensation (item 462 HAM2), contributions from public authorities and third parties (item 463 HAM2), and a residual account (item 469 HAM2). Also for capital expenditures, the cantons receive additional funding from other state entities. Namely, the category contributions on capital expenditures (item 67 HAM2) pools receipts from different sources: the confederation, cantons, concordats, et cetera. Together, the transfer revenue and the transfer receipt constitute the total transfers.

The operationalization of the earmarked part of the transfers is less simple. The de-appropriation of the transfers was one of the objectives of the fiscal equalization reform implemented in 2008 (Swiss Federal Council 2001). As a consequence, the federal council reviews how successfully the reform achieves this goal in the quadrennial efficacy report (Swiss Federal Council 2014). The report defines the earmarked transfers as the sum of the cantonal shares of the heavy vehicle fee (*Leistungsabhängige Schwerverkehrsabgabe*, LSVA) (item 46008 HAM2), the petroleum tax (item 46005 HAM2), and the alcohol tax (item 46003 HAM2). Contributions from public authorities and third parties (item 463 HAM2) and contributions on capital expenditures (item 67 HAM2) are included alike; see Table 5.01 of the report (Swiss Federal Council 2014). From the confederation's point of view, this definition seems reasonable as it reflects the legal requirements. For instance, according to art. 85(2) Swiss Federal Constitution as of 18 April 1999 the net receipts of the heavy vehicle fee must be used to cover road traffic related costs. Yet, the reality in the cantons looks different as a report of the Federal Office for Spatial Development (2009) reveals. As per their own declaration, six cantons enter the respective resources into the general budget, while others comply more strictly with the law or even establish a specific purpose account. This example justifies a closer focus on the effective implementation in the cantons.

In order to capture the earmarking of the transfers more appropriately, the money inflows are categorized into three tiers. The first tier contains all transfers, for which the canton itself elaborated a law to earmark them. This category overlaps with the definition of earmarked transfers from above (Swiss Federal Council 2014) but is more restrictive. Also, such a definition allows for variation between the cantons instead of simply lumping entire accounts together. The second tier encompasses transfers that are carried in the books of departments with specific tasks that are different from those of the financial department. Typical examples of this category are compensations (item 462 HAM2) or contributions (item 463 HAM2). The third tier includes only transfers that enter the books of the financial department that is responsible for the general budget; example transfers here are those of the fiscal equalization system after its reform in 2008. A differentiation between tier two and three is important because of the different obstacles to get at the money. If a transfer is affiliated with the general budget, it becomes part of the general budget debate, while transfers flowing directly to a department are no longer up for discussion. Admittedly, the departments that benefit largely from such direct transfers might face more difficulties in withdrawing means from the general budget. In this sense, transfers of the second tier entail a crowding out effect. Tier one and two can be seen together as quasi-earmarked transfers.

5.3. Further variables related to fiscal illusion

To measure the complexity and the diversification of the tax system, it is useful to look more at the fiscal possibilities of the cantons. The HAM2 differentiates between the following taxes that apply to the cantons: income tax (natural persons), wealth tax (natural persons), withholding tax (natural persons), personal tax (natural persons), other direct taxes (natural persons), profit tax (legal entities), tax on capital (legal entities), other direct taxes (legal entities), property tax, capital gains tax, capital transfer tax, inheritance and gift tax, casino and slot machine tax, motor vehicle tax, boat tax, entertainment tax, dog license, and other property and expenditure taxes (Conference of cantonal finance directors 2008). The list shows that the cantons are able to exploit a fairly broad tax base which also underlines the importance of the federal system in Switzerland (Feld & Reulier 2009). At the same time, some elements of the list raise the question as to why the analysis is limited to tax sources instead of also considering other types of revenue. There are at least two reasons for this. First, the coverage is already exceptionally large as some items are in fact charges and not taxes. Second, as Oates (1985) states, other revenue sources have other properties with respect to their visibility and are therefore incomparable to the tax sources listed above.

Measuring tax complexity and diversity is straightforward, given the concrete tax sources and their respective importance with respect to the revenue they generate for the cantons. According to Carroll (2009), the tax diversity refers to the number of tax sources a canton actually exploits. Tax complexity takes the relative importance of each source into account and is based on the normalized Herfindahl-Hirschmann index (nHHI).[38] Concretely, the index builds on the proportion of each tax source s_i and is calculated as $HHI = \sum_{i=1}^{N} \left(s_i / \sum_{i=1}^{N} s_i \right)^2$ while its normalized version is $nHHI = (HHI - 1/N) * (1 - 1/N)^{-1}$. Normalizing facilitates its interpretation as the nHHI ranges from zero to one instead of $1/N$ to one. Yet, the normalization dumps the information about the number of tax sources. In the concrete case of the tax sources of the Swiss cantons, the results of the two indices prove to be very similar, which speaks in favor of the nHHI due to its more intuitive interpretability.

Operationalizing the progressivity of the tax system is tricky for several reasons. First, the different state layers usually have diverse tax bases with varying rates. It is unclear though, whether and how this heterogeneity can be combined into a single variable (Oates 1985). Second, progressivity often manifests in a staircase form rather than a smoothed function. Likewise, this circumstance complicates finding a single parameter to describe the curvature. Third, often only aggregated data is available, something which poorly approximates individual revenue elasticity. All

[38] See Pommerehne & Schneider (1978) who exploited the Swiss data in a similar way in an early empirical paper on the subject.

these deficiencies leave the research community somewhat frustrated about the proposed ways of a possible operationalization (Dollery & Worthington 1996). Oates (1975) uses the ratio of income tax revenues to total tax revenues. He thereby ignores that tax revenues other than the income tax are subject to some – admittedly much smaller – progressivity too. Further possibilities include the ratio of the tax burdens of the highest quintile of a state's income distribution to that of the lowest quintile (Chernick 2005) or a measure based on the idea of the Gini-ratio[39] (Suits 1977). While Oates (1975) approaches progressivity roughly, the other two methods require information about the distribution of tax revenue among income classes. This information is not publicly available on the cantonal level for the cantonal tax in Switzerland.[40] In sum, none of these options seem appropriate or feasible in the present context.

A viable and yet exhaustive way to measure the progressivity of the cantonal tax system is a logarithmic approximation of the tax curve. Concretely, the method exploits the fact that most of the tax rates as a function of the income follow a logarithmic curve, which in turn depends on two parameters. The first parameter b indicates the curve's position, while the second parameter a stands for the curvature. Together they characterize the logarithmic function $t(y_i) = a * \ln(y_i) + b$, whereas $t(y_i)$ represents the tax rate as a function of the income y_i. Particularly, the tax rate corresponds to the rate a married couple with two children has to pay at the cantonal chief city. Since the tax rate is a steadily increasing function of the income in every canton, the parameter $a > 0$. Naturally, if $a = 0$ then $t(y_i) = b$ and the curve is flat without any progressivity. As a increases, so does the progressivity of the tax system and, technically, there is no upper limit. In reaction to the curvature parameter a, the position parameter b automatically adjusts in order to guarantee the curve's intersection around the origin. Usually, the function intersects the (horizontal) income axis to the right of the origin, meaning that the tax burden only takes effect at some positive income. The simple manner of measuring progressivity has its drawbacks though. Particularly, the interpretation of the curvature parameter a is not intuitive and only permits qualitative statements. In addition, the reliability of the approximation significantly correlates with the curvature parameter itself. The higher the curvature parameter a, the better the logarithmic function approximates the observed tax rate as a function of the income. Moreover, the misspecification is systematic in the sense that the curvature parameter is downwards biased. Using the curvature parameter as a measure of progressivity therefore underestimates any true effect. However, given the first deficiency that only qualitative statements are possible anyway, the systematic error does not impose a further problem.

The operationalization of the debt happens optimally in net terms in order to prevent misleading interpretations if the debt decreases due to sold assets (Brennan 2012). However, empirical analyses reveal the problem of the valuation of the assets. If the same asset has a different value depending on the government that holds it, the gross debt approximates the concept more appropriately after

---

[39] Concretely, the $progressivity = \sum_{i=1}^{N} 0.5\left(T(Y_i) + T(Y_{i-1})\right)(Y_i - Y_{i-1})$, where $T(Y_i)$ is the accumulated percent of the tax burden as a function of the accumulated percent income y and the running indices i to N are the known stepladders of the tax structure.

[40] Indeed, Schmidheiny (2006) exploits such data in a study covering the two demi-cantons Basel Landschaft and Basel Stadt. However, the fact that he actually uses individual level data reveals that he has direct access to information from the cantonal tax departments.

all. In the case of the Swiss cantons, the harmonized accounting standards do not necessarily guarantee a comparability of the asset valuation, but they suggest a composition of the gross debt (recommendation 18 of the HAM2). According to this definition, the gross debt is the sum of the current liabilities (item 200 HAM2), the short-term financial liabilities (item 201 HAM2) less the derivative financial instruments (item 2016 HAM2), and the long-term liabilities (item 206 HAM2) less the passivated contributions on capital expenditures (item 2068 HAM2).

To operationalize the renter illusion, a common measure is the percentage of housing owners (Dollery & Worthington 1996). Unfortunately, the respective statistics for the Swiss cantons are only available as of 2010. Accounting for the renter illusion would therefore dramatically reduce the sample which is why the concept is dropped for the further analysis.

5.4. Practical relevance of the revenue characteristics

This section gives an overview of the manifestation of the different revenue characteristics in the Swiss cantons. Depending on the variable, a temporal (i.e. along the time axis) or a cross-sectional (i.e. among cantons) illustration promises more fruitful insights.[41] Table 5 contains the standard descriptive statistics together with the source of the data. Note that the descriptive statistics cover all cantons and all years provided the data is available. The four variables that differ from one public domain to another are shown for the functions of culture and transportation. It is in these two domains, where the cantons fund the most expenditures through special financings as measured by the total expenditures. The first and second moment per revenue characteristic are shown separately for the other domains in Table 26 in the appendix.

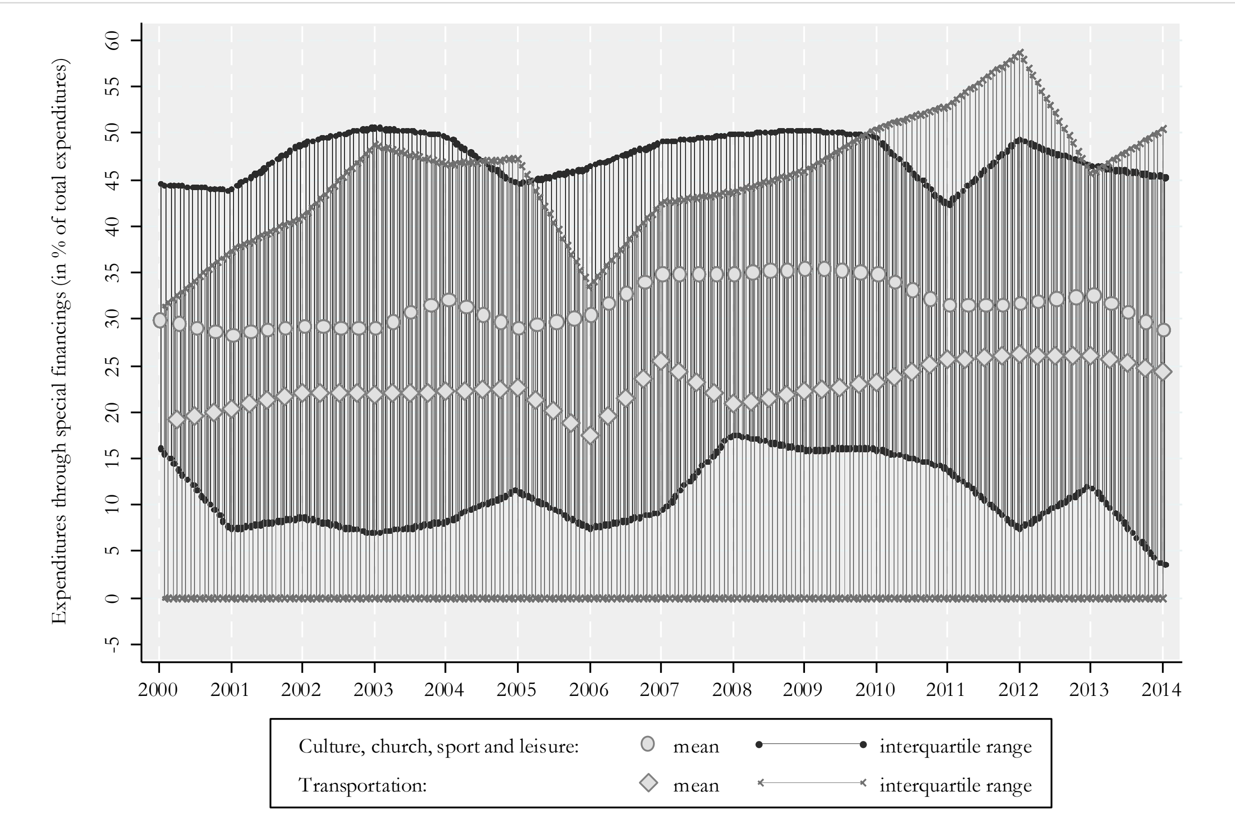
[41] If the graph shows the data canton-wise, the corresponding graph illustrating the development over the years can be found in the appendix.

Table 5 – Descriptive statistics of the revenue characteristics

Variable	Mean	Standard deviation	Minimum	Maximum	Data source [f]
Tax complexity (normalized HHI in %)	40.374	7.071	21.966	58.849	federal finance administration
Tax diversification (number of tax sources)	9.040	1.529	6	12	federal finance administration
Tax progressivity (curvature parameter)	5.588	1.343	2.522	7.753	federal tax administration
Gross debt per capita (in 1'000 Swiss francs)	6.502	6.595	1.051	37.097	federal finance administration
Domain: Culture					
Expenditures through special financings [a c]	35.431	20.296	0.000	116.698	cantonal annual financial statements
Volume of the special financing [b]	1.104	0.776	0.000	3.849	cantonal annual financial statements
Transfers [c]	26.783	21.609	0.331	86.925	federal finance administration
Pseudo-earmarked transfers [d]	96.324	11.946	12.806	100	federal finance administration
Domain: Transportation					
Expenditures through special financings [a c]	23.628	33.455	0.000	200.197	cantonal annual financial statements
Volume of the special financing [b]	1.131	3.253	-14.512	12.456	cantonal annual financial statements
Transfers [c]	3.58	12.892	0.000	88.428	federal finance administration
Pseudo-earmarked transfers [d]	97.284	11.622	35.956	100	federal finance administration

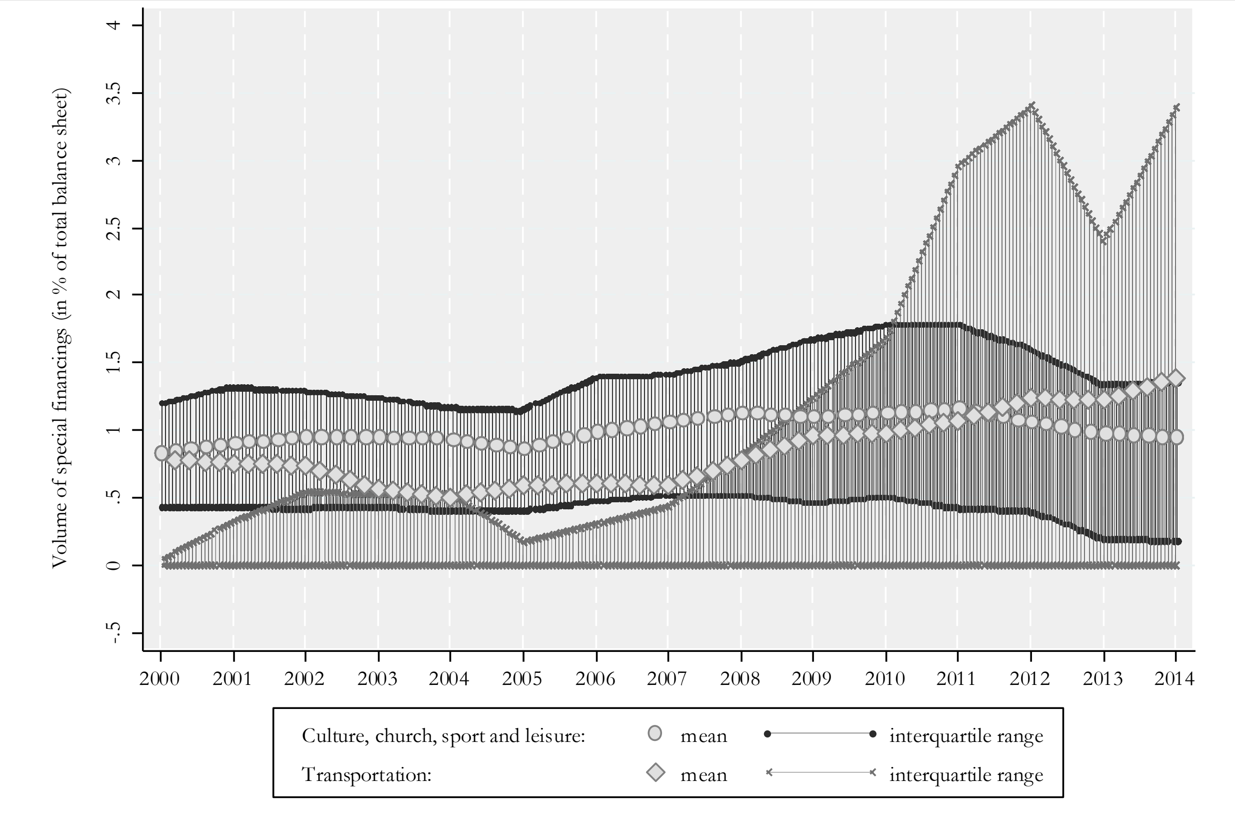
[a] In percent of total expenditures
[b] In percent of the balance sheet's total
[c] In percent of total revenues
[d] In percent of total transfers
[e] Values above 100% possible due to different data sources of the nominator (cantonal annual financial statements) and the denominator (federal finance administration)
[f] All data are subject to own calculations based on the data source indicated in the last column
Notes: Without missing values there should be 390 observations (26 cantons and 15 years); the true N = 301; missing cantons (and years) are BS (if year < 2013), VD (all years), TG (all years), LU (all years), GR (all years), GL (all years), GE (2014); the N can be slightly smaller in some regressions in section 7.5 due to excluded outliers.
Source: own composition

Figure 5 – Expenditures through special financings 2000 to 2014



Source: Own illustration based on data from the cantonal annual financial statements

Figure 6 – Volume of special financings 2000 to 2014

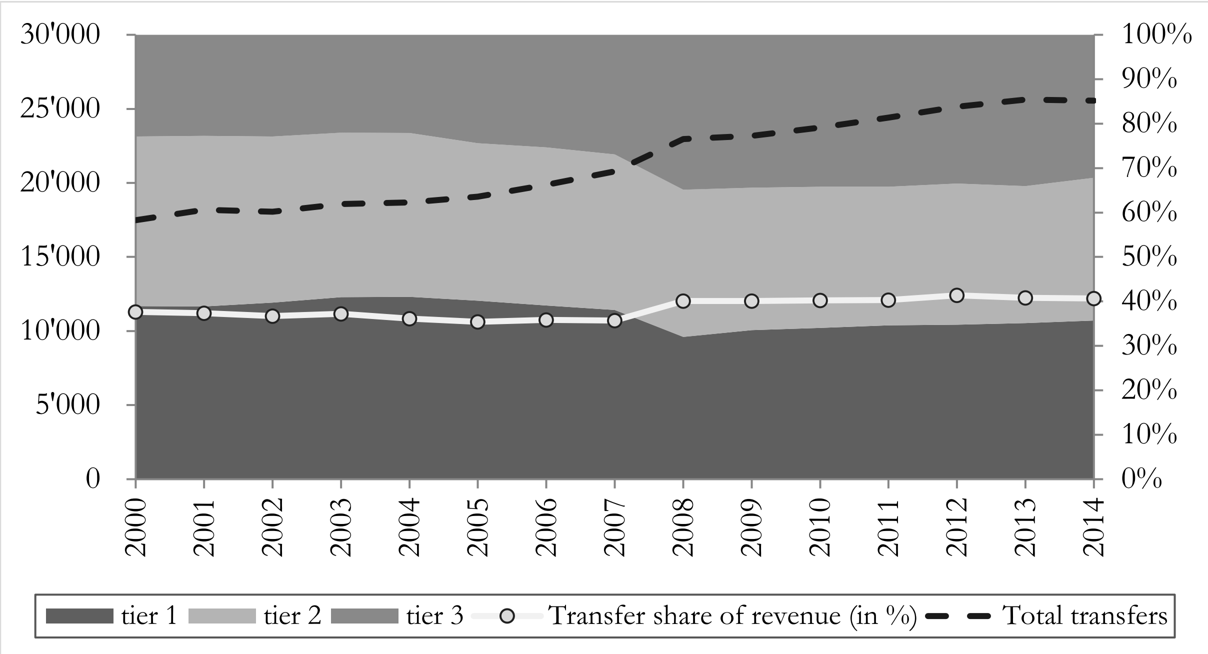


Source: Own illustration based on data from the cantonal annual financial statements

In a temporal comparison of the expenditure shares that are funded with earmarked revenues (Figure 5), the mean value stays roughly stable over time. On average, the cantons fund their expenditures by 35 percent (culture) and 24 percent (transportation) using the aid of special financings. The interquartile range among the cantons is relatively large compared to the temporal variation. The small changes over time are surprising particularly around the year 2008 when the new fiscal equalization scheme took effect. It would have been a window of opportunity for the cantons to rethink their earmarks, yet from an expenditure point of view little has changed. Focusing on the volumes (Figure 6) shows a completely different picture. Starting from 2007, the mean volume of special financings and particularly the 75th percentile of the distribution rose sharply in the transportation domain. Taken together, the two graphs suggest a drastic increase of earmarked revenues for transportation, while the expenditures remained roughly constant. The culture domain seems not to have experienced a similar break in 2008, and if anything, the expenditures and the volume shrink.

The relation between the volume's interquartile range and its mean (Figure 6) illustrate an interesting phenomenon. Given the highly right-skewed distribution, the mean value lies above the 75th percentile from 2000 to 2007. Thus, most cantons have very poorly funded special financings, while in a few, the stocks of earmarked revenues represent a considerable part of their balance sheets. Also note that the interquartile range of the volume does not contain any negative values while the minimum indeed falls below zero, suggesting that only a scarce number of cantons are indebted vis-à-vis their special financing.

Figure 7 – Total (in million CHF) and percentage of different tiers of earmarked transfers 2000 to 2014



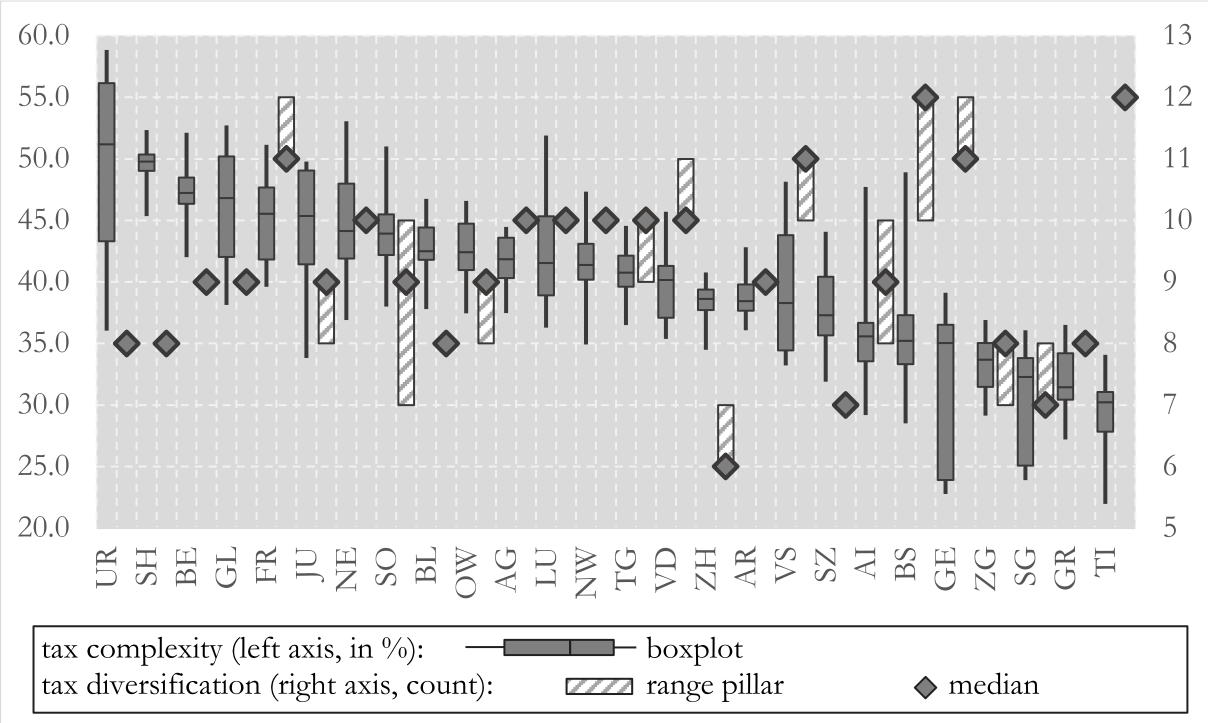
Source: Own illustration based on data from the Federal Finance Administration (2016)

Figure 7 illustrates the cantonal dependency on earmarks in different aspects. The dashed line refers to the left axis and traces the total transfers (revenue and receipts) taken in by the cantons. Its upward slope corresponds largely to the general increasing trend of the total revenues. Consequently, the transfer share stagnates at a level of around 40 percent, as the solid line (right axis) emphasizes. Overall, the transfer share varies widely between 11.48 percent (Geneva in 2000) and 53.63 percent (Jura in 2009). From 2007 to 2008, the average share jumps from 36 percent to 40 percent, which, again, may be due to the fiscal equalization reform. Another rather technical explanation for the jump is a revision of the manner in which the federal finance administration harmonizes the financial data (Federal Finance Administration 2016).

Zooming in reveals the destination of the transfers. On average, 66.77 percent of them are either earmarked or flow directly into specific departments rather than into the general budget of the financial department. In the canton of Jura, the pseudo-earmarked share goes up to 83.53 percent in 2004, whereas the canton of Zug appropriates only 31.01 percent directly or indirectly in 2008. Note that these values vary across governmental domain, which, as mentioned above, justifies the more detailed array in Table 27 in the appendix.

The different tiers of earmarked transfers also record a discontinuity. The first tier, covering truly earmarked transfers, decreases from 38 percent in 2007 to 32 percent in 2008. At the same time, the third tier (i.e. the non-appropriated transfers, right axis) climbs from 27 percent to 35 percent. In combination with Figure 5, this shift from earmarked to non-earmarked transfers is striking given that the expenditures through special financings do not replicate such a pattern. For the transfers, this shift signifies that the aim of the reform, namely to increase cantonal autonomy by reducing earmarked transfers, was successful (Swiss Federal Council 2001; Swiss Federal Council 2014). However, there appears to be a slight backlash as the first tier's share reaches 36 percent again in 2014. The second tier's share (transfers in favor of a specific department without explicit earmark) declined almost constantly over the interval, moving from 38 percent in 2000 to 32 percent in 2014.

Figure 8 – Tax complexity and diversification by canton 2000 to 2014

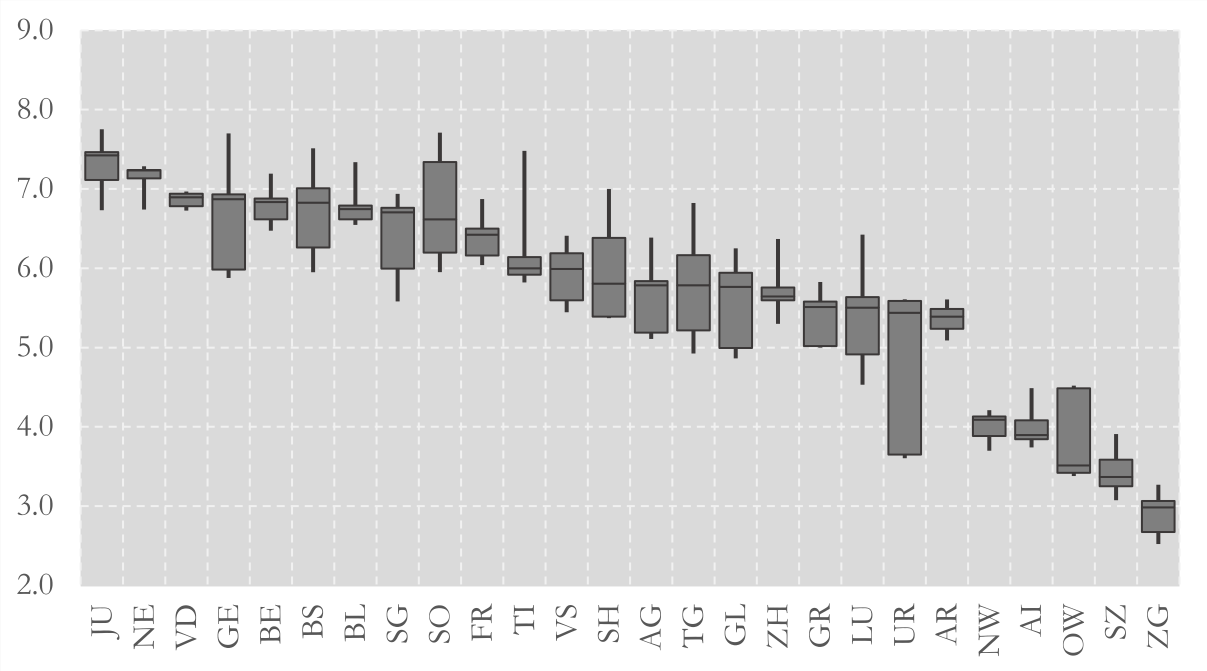


Source: Own illustration

Since the two concepts are closely related, tax diversification and complexity are plotted together in Figure 8. The mean of the normalized HHI (boxplot, left axis) that measures the tax complexity is relatively high at 40.37 percent, meaning that about one half of the tax revenue is based on a single source (the income tax). The variation is fairly small with the canton of Uri representing the upper end of the scale (58.50%) in 2003 and the canton of Ticino reaching the smallest value in 2014 (21.97%). The low value, indicating a somewhat higher complexity, stems from the geographic location of the canton of Ticino and hence its comparably high revenue from the withholding tax (tax at source). Overall, the cantons follow a trend towards a less complex tax system that particularly originates in a revenue concentration on a few tax sources rather than in an effective reduction of tax sources (see Figure 25 in the appendix).

Clearly, tax complexity is linked with tax diversification (range pillar, right axis in Figure 8) because only cantons with a large number of tax sources can spread their base and increase complexity. It is therefore reasonable that tax diversification also does not vary strongly. In fact, half of the cantons kept the same number of tax sources over the observed time span. Due to the small variation, a classical boxplot illustration is futile. Thus, the crosshatched pillars mark the entire range of the tax diversification of a canton throughout the years 2000 to 2014 with the diamond figuring the median. On average, the cantons have nine tax sources while some cantons expanded this to twelve (Ticino over the entire time range and Geneva between 2011 and 2013) and another gets along with six (Zurich starting from 2002 and ongoing).

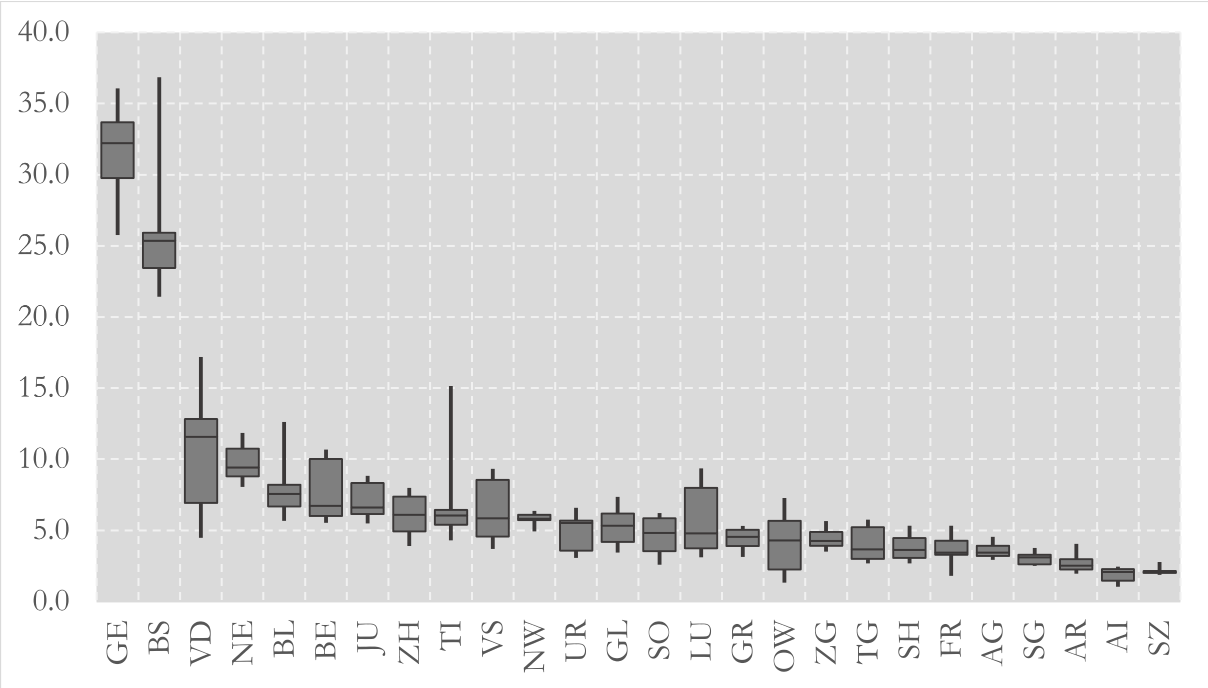
Figure 9 – Progressivity of the tax system by canton 2000 to 2014



Notes: The ordinate shows the estimated a-value of the logarithmic function $t(y_i) = a * \ln(y_i) + b$ fitted into data on the tax rate (t) and income (y_i) of a household consisting of a married couple with two children.
Source: Own illustration

The boxplot in Figure 9 shows the differences between the cantons with respect to the exploitation of the tax progressivity parameter (see section 5.3 for an explanation of the parameter). Even if the parameter itself is hard to interpret, the variation still permits qualitative statements. As the classification by canton reveals, the variation between the cantons is usually more important than the difference between the years within a single canton. In most of the years under investigation, the tax system of the canton of Zug was the least progressive one (in 2014 the parameter is 2.522). That is, the curve tracing the percental tax burden per income is flat. In contrast, the most progressive tax systems are found in the canton of Jura, Solothurn, and Geneva (each in 2000). In these cases, the tax burden still increases sharply even in higher income classes. As regards the variation within, the canton of Uri underwent a massive and rapid tax reform that became effective in 2009 (Governing Council of the Canton of Uri 2008). The lack of whiskers, which cover here the distance from the minimal (maximal) observed value to the first (third) quartile, point out that there is almost no variation before and after the reform; that is, all the variation comes from the one reform. The same is true for the canton of Obwalden, whose tax system reform occurred in 2006 (Governing Council of the Canton of Obwalden 2005). The canton of Ticino is different with its long upper whisker that originates in an early tax reform (taking effect in 2001). Since that reform, the canton of Ticino has maintained roughly the same progressivity. On the contrary, the cantons with two long whiskers, as the canton of Lucerne for instance, smoothly adjusted their tax over the years. Taken together, all cantons flattened their tax system between 2000 and 2014, something which can be condensed into a decreasing trend in the progressivity parameter (see Figure 26 in the appendix).

Figure 10 – Gross debt (in 1'000 CHF per capita) by canton 2000 to 2014



Source: Own illustration

When comparing cantonal debt per capita, the canton of Basel Stadt and the Canton of Geneva immediately attract attention; the situation in Basel Stadt is somewhat special as the fiscal data cover the canton and the capital city. Even if the variation within the two cantons is remarkable, their whiskers do not overlap with those of the other cantons. Again, here the whiskers cover the distance from the minimal (maximal) observed value to the first (third) quartile. In defiance of the two extreme cantons, the within as well as the between variation of the remaining cantons cover a relatively broad range. Among these cantons, the canton of Ticino's upper whisker leaps to the eye. While its debt amounted to 9'241 Swiss francs per capita in 2013, the canton moved up to the third rank in 2014 with an indebtedness of 15'132 Swiss francs per capita in 2014. On average, the cantonal gross debt fluctuates on a low level of around 6'519 Swiss francs per capita. The least indebted canton was Appenzell Innerrhoden in 2010, with only 1'051 Swiss francs per capita. As regards the overall development of the debt among the cantons, the mean remains roughly stable around 5'000 Swiss francs per capita (see Figure 27 in the appendix).

5.5. Interim conclusion

The practical relevance of the different revenue characteristics delivers on the promise of the previous chapter. With considerable variation in terms of earmarked revenue, transfers, tax complexity, tax progressivity, and indebtedness, the Swiss cantons provide a welcome data source to assess potential determinants of efficiency. Among the different revenue characteristics, this analysis mainly focuses on the earmarked revenue as an explaining variable of efficiency. The descriptive statistics therefore justify concentrating on the domains of culture and transportation as the most prominent beneficiaries of earmarks.

6. Approaching efficiency

While the revenue characteristics as right-hand side variables are observable, the efficiency as dependent variable only manifests when combining different factors. In his seminal paper, Farrell (1957, p.254) notices that "[w]hen one talks about the efficiency of a firm one usually means its success in producing as large as possible an output from a given set of inputs". This statement inspires two questions. What exactly are inputs and outputs? How to combine several inputs and outputs? Regarding the first question, little help comes from the theoretical literature to support the researcher trying to identify the relevant inputs and outputs. With respect to inputs, at any rate, the OECD (2001) has established the KLEMS categories referring to capital, labor, energy, and material. In terms of outputs, the literature actually differentiates between outputs, outcomes, and impacts. While outputs measure what is actually produced within the scope of a policy, outcomes cover the effects of a policy on a broader social group, and the impact describes the reaction of the policy's target group (Jann & Wegrich 2007). [42] The identification of the relevant inputs and outputs pave the way for the actual estimation of efficiencies. Due to the lack of data, this analysis mostly relies here on outputs.

In regard to the second question, this chapter extensively develops the two most prominent methods that aggregate inputs and outputs and that finally produce a measure of efficiency. Among researchers in the field of efficiency, there is a fairly sharp line between proponents of the Data Envelopment Analysis (DEA) and its parametric counterpart, the Stochastic Frontier Analysis (SFA). The separation stems from a fundamental disagreement in some basic assumptions of both methods. Consequently, it seems appropriate to apply both and test the robustness of the results. However, the emphasis is on the DEA which eventually provides more distinct results.

Throughout this and the subsequent chapter, several terms appear repeatedly and should be defined. A 'reference set' describes a vector of input and output values of all observed peers (benchmarks could be used as a synonym for peers) taken into account to estimate DEA efficiency scores. Following the distinction of Tulkens & Eeckenhaut (1995), a 'contemporaneous reference set' considers only those peers' values that are observed in the same year. By contrast, the 'intertemporal reference set' covers the peer's values of all years and, hence, it pools the data. Accordingly, 'intertemporal efficiency scores' stem from a DEA estimation using an intertemporal reference set, whereas 'contemporaneous efficiency scores' are based on a contemporaneous reference set.

Applying the extensively reviewed performance measurement methods and the operationalization of inputs and outputs, this chapter lists mean DEA and SFA estimates across the Swiss cantons from 2000 to 2014. In anticipation of when this analysis will narrow down to focus on the two domains of culture and transportation, the comments are confined to their results. For culture, the DEA attests a mean efficiency of about 57 percent, while the SFA's mean climbs to 98 percent. Their rank correlation amounts to only 19 percent. Accordingly, both results should be interpreted cautiously. For the transportation domain, the mean efficiencies come to 67 (DEA)

[42] Note that the usage of the three terms is not coherent in the literature. Niskanen (1994, p.26), for instance, contends that "[b]ureaucrats and their sponsors do not, in fact, talk much about output – in terms of military capability, the value of educational services, the number and condition of the poor, etc." This entire enumeration actually lists impacts, while his example of 'activity levels' (the number of the poor served by a program) actually describes an output.

and 21 (SFA) percent with a rank correlation of -16 percent, respectively. Additional evaluations revealed that the DEA results are more trustworthy in the further analysis.

6.1. Efficiency measurement in general

There are various models to measure efficiency, all of which share one basic idea. They are all based on a production frontier in which efficiency is expressed as the relative distance to this frontier (Farrell 1957). Efficiency therefore depends on the so-called decision-making units (DMU) or firms under investigation and specifically on the DMU that produces a set of outputs using a set of inputs in the most efficient manner. Most efficient means that given a fixed set of inputs, the DMU maximizes its output set, or, given a set of outputs, it maximizes the input set. The difference between the various models lies in the way to find and describe such a frontier. Two general strands, deterministic and stochastic, have developed in the literature, both with different strengths and weaknesses. The deterministic strand claims to find the exact frontier given the observed data. By imposed this kind of strong claim, deterministic approaches generally do not need, but also do not allow for conducting statistical inference.[43] The stochastic strand acknowledges that the observed data is only a sample, which prevents determining the exact frontier. Instead, stochastic approaches estimate the frontier which thus enables statistical inference (Bogetoft 2013).

Both strands embrace a further categorization with respect to the functional form of the frontier. While the classic stochastic models are parametric and need to assume a specific functional form, recent developments also propose non-parametric approaches without such assumptions. The differentiation between parametric and non-parametric models also exists within the deterministic family. Table 6 gives an overview of the different categories of efficiency measurement methods and names the respective models as examples.

Table 6 – Categorization of efficiency measurement methods

	Deterministic	Stochastic
Parametric	• Corrected Ordinary Least Squares • Modified Ordinary Least Squares • Maximum Likelihood Estimation	• Stochastic Frontier Analysis • Distribution Free Approach • Thick Frontier Approach • True fixed / random effect model
Non-parametric	• Data Envelopment Analysis • Free Disposal Hull	• Stochastic Data Envelopment Analysis

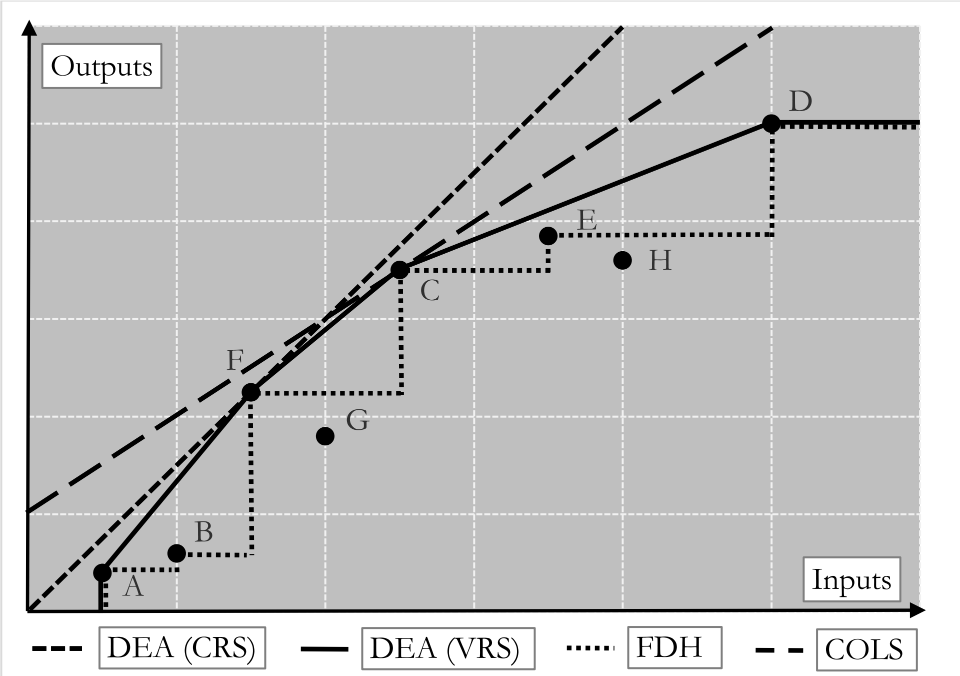
Source: Own illustration

Exponents of the deterministic family are the Data Envelopment Analysis (DEA), the Free Disposal Hull (FDH), and Corrected Ordinary Least Squares (COLS). All these approaches somehow envelop the data points on an input-output plane. Figure 11 illustrates their respective efficiency frontiers. The data points A to H represent input-output combinations of the decision-

[43] The deterministic character, which is actually an assumption, is relaxed later.

making units under investigation. As shown, DEA can further distinguish between a constant returns to scale (CRS) and variable returns to scale (VRS) model (Cooper et al. 2007).[44]

Figure 11 – Deterministic efficient frontiers



Source: Own illustration based on Kalb (2010a)

The two DEA models differ in the form with which they envelop the data. Constant returns to scale DEA draws the efficient frontier as a straight line from the origin to the point closest to the upper left corner (Charnes et al. 1978). The underlying assumption is that all decision-making units, irrespective of their size of production, can theoretically produce on the same production frontier. The only fully efficient decision-making unit in terms of constant return to scale is C. Variable returns to scale DEA accounts for possible scale efficiencies by enveloping the data points convexly (Banker et al. 1984). Now the fully efficient decision-making units are A, F, C, and D. The intersection of the two frontiers, that is decision-making unit C, identifies the input volume up to which a DMU can become more efficient by increasing its input. Above the intersection, the DMU can increase its efficiency by reducing the input. In between the constant and the increasing returns to scale model there is the non-increasing return to scale (NIRS) DEA, which Figure 11 does not explicitly illustrate.[45] As the name suggests, it combines the two models by drawing the efficient frontier from the origin to the intersection, which corresponds to the constant returns to scale DEA. From the intersection upwards, the frontier of the non-decreasing DEA is equal to the frontier of the variable returns to scale DEA. Likewise the non-decreasing returns to scale (NDRS) DEA first follows the variable returns to scale DEA frontier up to the intersection and then corresponds to the constant returns to scale DEA (Charnes et al. 1978).

[44] Often the constant returns to scale DEA model is referred to as the CCR model named after its authors Charnes, Cooper & Rhodes (1978). The variable returns to scale DEA model is also known as BCC model according to its authors Banker, Charnes & Cooper (1984).
[45] For further illustrations of the models see for example Kalb (2010).

Criticism of the DEA models is linked to the direct connections between the data points and argues that these input output combinations are only theoretical and not feasible. To account for that critique, Free Disposal Hull (FDH) envelopes the data only with lines parallel to the axes (Deprins et al. 2006). The resulting frontier ensures that only real input output combinations act as reference. Thereby even more decision-making units become fully efficient. The only inefficient decision-making units left are G and H. A small number of inefficient units is a usual result when applying FDH on small samples. In fact, the convergence rate of the FDH estimator is much smaller compared to the DEA estimator; the rate difference between the two methods is specifically severe, if the number of inputs and outputs exceeds two (Kneip et al. 2015). Therefore, FDH appears to be an invalid method to analyze the rather small sample investigated here.

Turning to the parametric models within the deterministic family leads to Corrected Ordinary Least Squares (COLS) as proposed by Richmond (1974).[46] As its name implies, it originates from Ordinary Least Squares. This econometric technique fits a straight line into the data on the input output plane such that it minimizes the squared distance between the data points and the line. The correction takes place in a parallel upwards shift of the line until it is tangent to the variable returns to scale DEA frontier. The tangency is reached in the decision-making unit F which is fully efficient under COLS. Note that even if COLS is based on an econometric approach, it is still deterministic. The reason is that it interprets the entire distance between the efficient frontier and the data points as inefficiency, leaving no room for statistic noise.

Besides COLS, similar techniques like the Modified Ordinary Least Squares (MOLS) and specific forms of the Maximum Likelihood Estimation (MLE) have emerged. They basically differ in the assumption about the error term. Whereas COLS assumes normally distributed errors, in MOLS the error term follows a one-sided distribution (Afriat 1972; Richmond 1974). If estimated by MLE, for instance, the functional form of the error can be assumed to be gamma distributed (Greene 1980). Note that the different errors only alter the shift and possibly the slope of the regression line, but they do not change the deterministic character of the model. They all assign the entire error to inefficiency.

The first authors who separated efficiency from statistical noise were Meeusen & Van den Broeck (1977), Aigner et al. (1977) and Battese & Corra (1977). They laid the foundation for the family of stochastic models which are mainly based on econometric ideas. Like the COLS model, stochastic models either explain the output by a combination of inputs through a production function or they explain the input by a combination of outputs and the input factor-prices through a cost function. While the stochastic noise has a symmetrical distribution around zero, the efficiency term is strictly positive and possibly follows some other distribution. Marrying the frontier idea and the standard econometric models paved the way for further developments, including the consideration of panel data, heteroscedasticity, endogeneity, et cetera. The respective models are accordingly called true fixed effects model, true random effects model, et cetera. Not least, the parametric approach also makes it possible to test many of the underlying assumptions based on some statistics (Kumbhakar et al. 2015).

[46] Originally the idea of COLS dates back to Winsten (1957) but the actual model was only developed around twenty years later.

In anticipation of the 7th chapter a short note is appropriate. The proceeding paragraphs presented the different families of efficiency measures. As outlined, the DEA belongs to the category of non-parametric deterministic approaches since it does not presume a specific functional form of the production process while neither does it *a priori* allow for statistical noise. The lack of statistical noise is indeed consistent with its classification as a deterministic approach. However, from a statistical point of view, the deterministic character is an assumption rather than a property. This is an important distinction when it comes to the assumptions needed for the second stage. Since the second stage uses statistical methods to explain the previously estimated efficiency scores, the distribution of the scores as a dependent variable needs to be known – or plausibly assumed. Unfortunately the literature has only generated very limited results about the sampling distribution of efficiency estimates (Gijbels et al. 1999). By all means, as Simar & Wilson (2007, p.33) state, "[a] more serious problem in *all* of the two-stage studies that we have found arises from the fact that DEA efficiency estimates are serially correlated".[47] Section 7.1 shows that most of the statistical models used in the second stage rely on implicit underlying assumptions that are not consistently plausible. The chapter also outlines different possibilities of how to deal with such issues.

6.2. Data Envelopment Analysis

This section develops the different measures of Data Envelopment Analysis step by step. The new quasi fixed effects algorithm which follows in section 7.1 builds strongly on the ideas of DEA, which is why it is important to review this method in detail. Although mathematical concepts are reduced to a minimum, some are nonetheless necessary. To understand them, keep in mind the following coherent notation. While row or column vectors appear in small bold letters, $\boldsymbol{x}$, matrices are in capital bold letters, $\boldsymbol{X}$. The superscript T, as in $\boldsymbol{x}^T$ for instance, indexes transposed vectors (i.e. scalars) and matrices.

As mentioned above, the DEA is based on the idea that efficiency is a relationship between inputs and outputs. This fundamental thought is understood most easily by considering a production process that involves only one input, say clerks, and one output, say documents.[48] Take departments A to H, where each treats a certain number documents employing a number of clerks as given in Table 7.

[47] The correlation emerges from the fact that in finite samples every efficient observation expands the frontier and thereby alters the efficiency estimates of all or most of the other observations.

[48] The examples build on those from (Cooper et al. 2007).

Table 7 – Input-oriented DEA example with one input and one output

Department (i)	A	B	C	D	E	F	G	H
I: Clerks	1	2	5	10	7	3	4	8
O: Documents	8	12	70	100	77	45	36	72
Documents/Clerk	8	6	14	10	11	15	9	9
Efficiency	0.53	0.40	0.93	0.67	0.73	1.00	0.60	0.60

Source: Own illustration inspired by Cooper et al. (2007)

The number of documents treated per clerk is called the productivity. The department with the highest productivity (which is F here) is the relatively most efficient and therefore obtains an efficiency score of 1. Department F is thus the peer for the other departments, whose efficiency is the result of dividing their productivity by the productivity of the peer. Formally, in the one-input-one-output case the efficiency is

$$\theta_i = \frac{output_i / input_i}{\max(output_i / input_i)} \quad \text{where} \quad i \in n = \{A, \dots, H\} \tag{25}$$

Note that, because of the compound fraction, the efficiency measure θ is unit invariant, that is, the number of documents could be in the hundreds or in the thousands while the efficiency of all departments remains equal. Furthermore, the efficiency measure is always between one and zero, i.e. $\theta \in (0,1]$ by construction. This property is maintained throughout the subsequent modifications if not stated otherwise.

Besides treating documents, clerks might also answer inquiries. Table 8 introduces the number of inquiries to every department as an additional output variable. As before, divide the number of inquiries by the number of clerks to obtain the productivity of each department. Having two different productivities for each department at hand, efficiency can no longer be obtained by comparing each department with the peer. In fact, there are several peers now. While department F has the highest productivity in treating documents, department B answers the most inquiries per clerk.

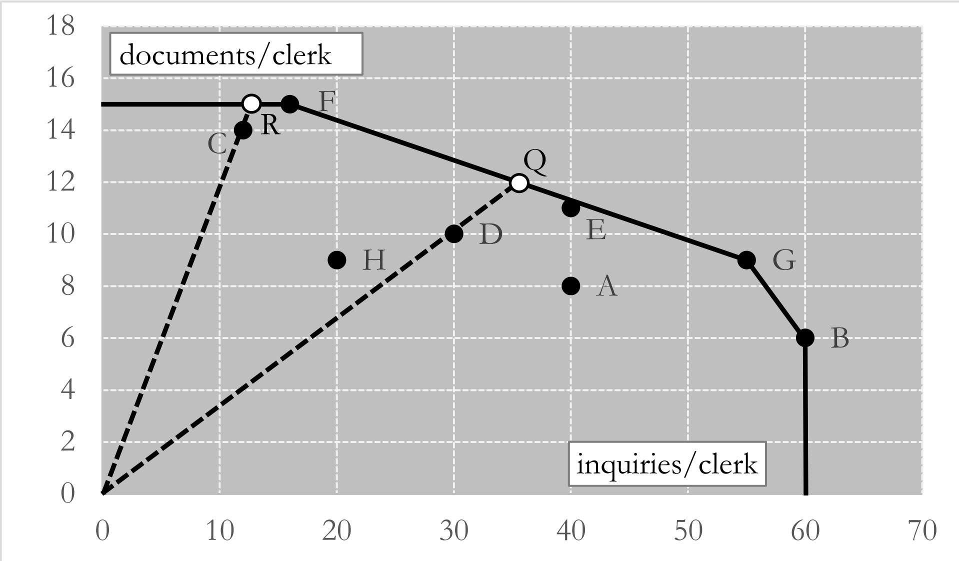
Table 8 – Input-oriented DEA example with one input and two outputs

Department (i)	A	B	C	D	E	F	G	H
I: Clerks	1	2	5	10	7	3	4	8
O: Documents	8	12	70	100	77	45	36	72
O: Inquiries	40	120	60	300	280	48	220	160
Documents/Clerk	8	6	14	10	11	15	9	9
Inquiries/Clerk	40	60	12	30	40	16	55	20
Efficiency	0.81	1	0.93	0.84	0.98	1	1	0.69

Source: Own illustration inspired by Cooper et al. (2007)

In order to identify the inefficient departments and their respective peer(s), Figure 12 plots productivity with respect to documents and with respect to inquiries. The line connecting the data points with the largest distance to the origin, the departments B, F, and G, is called the efficient frontier and serves as a reference to obtain the efficiency scores. Consider, for instance, the department D. The line $\overline{OD}$ going through the origin O and department D's data point intersects the frontier in Q. Note that Q's coordinates (35.84, 11.95) are a direct result from those of F (16, 15), G (55, 9), D (10,30) and O (0,0). The hypothetical department Q is D's efficient analogue while F and G are its real reference set.

Figure 12 – DEA example with one input and two outputs



Source: Own illustration inspired by Cooper et al. (2007)

Geometrically, the part of the distance of $\overline{OQ}$ that is covered by $\overline{OD}$ is D's efficiency. Formally D's efficiency can be obtained using the coordinates of D (30, 10) and Q (35.84, 11.95) to calculate the ratio of the Euclidian distances to the origin, which is called the radial measure:

$$\theta_D = \frac{\overline{OD}}{\overline{OQ}} = \frac{\sqrt{30^2+10^2}}{\sqrt{35.84^2+11.95^2}} = 0.84 \qquad (26)$$

The efficiency score of 84% means that D should only use 84 percent of its input to produce the same output in order to be efficient. Likewise, the efficiency score indicates that D could increase both its outputs by ($\theta_D^{-1} = 1.19$) 19 percent to attain 100 percent efficiency. Calculating efficiency scores by the previously presented procedure is not easy to put into practice when the number of departments increases and not at all feasible if the number of outputs is further increased. For these purposes, Charnes et al. (1978) developed the so called CCR model that measures efficiency by linear programming. To reproduce the CCR model, recall equation (25) where the efficiency resulted from a ratio of ratios of one input and one output per department. As opposed to this, take w inputs (x) and q outputs (y) and add them up after attaching a weight (u and v) to each of them, which yields

$$\theta_i = \frac{\sum_{r=1}^{q} u_r y_{ri} / \sum_{j=1}^{w} v_j x_{ji}}{\sum_{r=1}^{q} u_r y_{rk} / \sum_{j=1}^{w} v_j x_{jk}} \quad \text{where} \quad \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{j=1}^{w} v_j x_{jk}} = \max_{i=1,\ldots,n} \left\{ \frac{\sum_{r=1}^{q} u_r y_{ri}}{\sum_{j=1}^{w} v_j x_{ji}} \right\} \qquad (27)$$

Then, restrict the weights to be semi positive, i.e. $u_r, v_j \geq 0$. Moreover, set the denominator of (27) (i.e. $\sum_{r=1}^{q} u_r y_{rk} \,/ \sum_{j=1}^{w} v_j x_{jk}$), which is the ratio of the most productive department (here department k), equal to one. In order to be in line with the ordinary textbook nomenclature, the departments are hereafter called decision-making units (DMU). The efficiency of each DMU is the result of choosing the weights u_r, v_j of each input and each output that maximizes the value of the ratio of ratios subject to the two imposed restrictions. The optimal weights are a byproduct of the efficiency measurement, which is not relevant here. For the sake of simplicity, define the multipliers $\boldsymbol{u}$ and $\boldsymbol{v}$ as vectors with their corresponding outputs $\boldsymbol{y}_i$ and inputs $\boldsymbol{x}_i$ of department D. The fractional programming problem then reads as

$$\max_{\boldsymbol{u},\boldsymbol{v}} \theta_i = \frac{\boldsymbol{u}^T \boldsymbol{y}_i}{\boldsymbol{v}^T \boldsymbol{x}_i} \tag{28}$$

$$\text{s.t.} \quad \frac{\boldsymbol{u}^T \boldsymbol{y}_i}{\boldsymbol{v}^T \boldsymbol{x}_i} \leq 1 \ \forall\, i \in \{1, \dots, n\}$$

$$\boldsymbol{u}, \boldsymbol{v} \geq \boldsymbol{0}$$

Since a linear programming problem is easier to solve than a fractional one, next impose the restriction that the denominator of (28) ($\boldsymbol{v}^T \boldsymbol{x}_i$), that is the sum of weighted inputs of the DMU i, is equal to one. In addition, under the assumption that $\boldsymbol{x}_j > \boldsymbol{0}$ for all j, rearrange the first restriction to obtain the linear programming problem that leads to the same results as (28):

$$\max_{\boldsymbol{u},\boldsymbol{v}} \theta_i = \boldsymbol{u}^T \boldsymbol{y}_i \tag{29}$$

$$\text{s.t.} \quad \boldsymbol{u}^T \boldsymbol{y}_i \leq \boldsymbol{v} \boldsymbol{x}_i \ \forall\, i \in \{1, \dots, n\}$$

$$\boldsymbol{v}^T \boldsymbol{x}_i = 1$$

$$\boldsymbol{u}, \boldsymbol{v} \geq \boldsymbol{0}$$

To simplify further, define the data matrices for the p inputs, the q outputs (rows), and the n DMUs (columns):

$$\text{input: } \boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1i} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{j1} & \cdots & x_{ji} & \cdots & x_{jn} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{w1} & \cdots & x_{wi} & \cdots & x_{wn} \end{bmatrix} \quad \text{and output: } \boldsymbol{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1i} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ y_{r1} & \cdots & y_{ri} & \cdots & y_{rn} \\ \vdots & & \vdots & \ddots & \vdots \\ y_{q1} & \cdots & y_{qi} & \cdots & y_{qn} \end{bmatrix}$$

Note that the repeatedly mentioned DMU D is one column of $\boldsymbol{X}$ and $\boldsymbol{Y}$ just as i, for instance. To clarify, the weighting vectors are also written out in full here:

input multiplier: $\boldsymbol{u}^T = [u_1, \dots, u_j, \dots, u_w]$ and the output multiplier: $\boldsymbol{v}^T = [v_1, \dots, v_r, \dots, v_q]$

The (primal) linear programming problem in full matrix notation then reads as

$$\max_{\boldsymbol{u},\boldsymbol{v}} \theta_i = \boldsymbol{u}^T \boldsymbol{y}_i \tag{30}$$

$$\text{s.t.} \quad \boldsymbol{Y}\boldsymbol{u} - \boldsymbol{X}\boldsymbol{v} \leq \boldsymbol{0}$$

$$\boldsymbol{v}^T \boldsymbol{x}_i = 1$$

$$\boldsymbol{u}, \boldsymbol{v} \geq \boldsymbol{0}$$

As (30) shows, the number of DMUs (n) boosts the number of restrictions. Thus, since the number of DMUs is generally larger than the number of inputs (p) plus outputs (q), the efficiency measure $\boldsymbol{\theta}_i$ is obtained more efficiently through solving the dual of (30). By turning to the dual, the inputs plus the outputs yield the number of variables and the number of DMUs becomes the number of restrictions.[49] Concretely, the dual of (30) is

$$\min_{\theta,\boldsymbol{\lambda}} \theta_i \tag{31}$$

$$\text{s.t.} \quad \theta_i \boldsymbol{x}_i - \boldsymbol{X}\boldsymbol{\lambda} \geq \boldsymbol{0}$$

$$\boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{y}_i \geq \boldsymbol{0}$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0}$$

with $\boldsymbol{\lambda}$ as column vector containing weights of the reference DMUs. Sometimes the dual (31) is referred to as the envelopment form while the primal (30) is the multiplier form. By the strong duality theorem of linear programming, the optimal value of θ_i is the same for both the primal and the dual. Turning back to the example of departments A to H, define the data matrices of the inputs and the outputs with reference to Table 8 as

$$\boldsymbol{X} = \begin{bmatrix} 18 & 2 & 5 & 10 & 7 & 3 & 4 & 8 \end{bmatrix} \tag{32}$$

$$\boldsymbol{Y} = \begin{bmatrix} 8 & 12 & 70 & 100 & 77 & 45 & 36 & 72 \\ 40 & 120 & 60 & 300 & 280 & 48 & 220 & 160 \end{bmatrix}$$

The application of the defined data matrices $\boldsymbol{Y}$ and $\boldsymbol{X}$ in (32) to the linear programming problem (31) for each of the departments results in the efficiency measures presented in the last row of Table 8.[50] Note that the calculation by the linear programming problem (31) provides the same result as the geometric calculation in (26). However, the linear programming allows additional inputs and outputs.

Before developing this model further, it is worth backing up to reconsider the underlying data, specifically the outputs. All the outputs involved up to this point are desirable, that is, it is the very aim of an administration to process documents and inquiries. However, the communications unit might be exposed to complaints from the different departments, which is clearly not their intention. It is therefore clear that the number of complaints as a further output cannot be treated like the other two outputs; it is an undesirable or bad output. As the overview of Ramli & Munisamy (2013) demonstrates, the research has moved in many different directions trying to account for undesirable outputs. Following Seiford & Zhu (2002), Liang et al. (2009), Wu et al. (2013) and others, the approach is to transform undesirable outputs by a linear monotone function. The upside of this kind of transformation is that the standard DEA models are still applicable, which in the end keeps the estimation complexity on a reasonable level. On the downside, the transformation obscures the interpretation of the results, as it is less clear what the transformed variable actually measures. To address this concern intuitively, the proposed transformation function turns the parameter upside down and squeezes it between two boundaries. The resulting transformed vector is an index between 1 and 100, where the DMU with the lowest output attains an index of 100 and the highest output value equals to an index of 1. Concretely, the function transforms the bad output y^b into the index y^i as follows[51]

[49] For a detailed discussion of linear programming problems see Sydsæter & Hammond (2009).

[50] Sydsæter & Hammond (2009) explain the simplex method that solves linear programming problems. Software solutions like the MS Excel Solver makes it possible to solve this kind of problem using the simplex method.

[51] Note that the usual transformation sets $b = 1$ and $a > \arg\max(\boldsymbol{y}^b)$ in order to ensure that $\boldsymbol{y}^i > 0$ (Seifert & Nieswand 2014; Liang et al. 2009; Wu et al. 2013). The resulting transformed output lacks any intuition though.

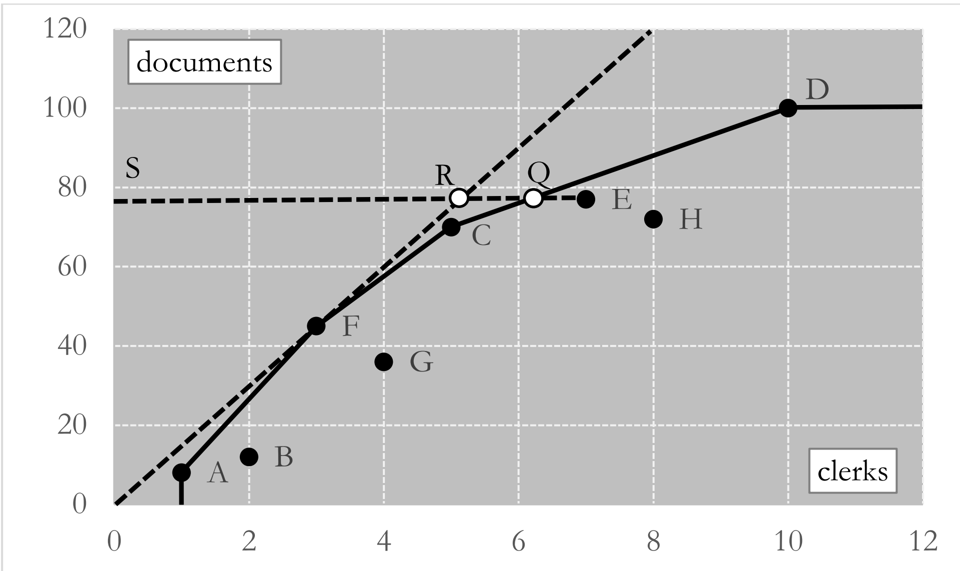
$$y^i = a - b * y^b \qquad \text{where} \qquad a = 100 + \frac{99 * \arg\min(\boldsymbol{y^b})}{\arg\max(\boldsymbol{y^b}) - \arg\min(\boldsymbol{y^b})} \tag{33}$$

$$b = \frac{99}{\arg\max(\boldsymbol{y^b}) - \arg\min(\boldsymbol{y^b})}$$

So far, illustrating the CCR model only shows one side of the situation. Instead of an input orientation that is minimizing the input while holding the output constant, the possibility of the opposite case is apparent. Of course, the presented linear programming problems are convertible to an output orientation where the inputs are kept fixed while maximizing the outputs. As the interpretation of the equation (26) result has shown, in the CCR model the output orientation is nothing other than the invert of the input orientation. It is equivalent whether a DMU decreases its inputs by θ or increases its outputs by θ^{-1} as long as the proportion of the various outputs remains equal. This equivalence does not hold anymore, when the thus far implicitly imposed assumption of constant returns to scale is dropped. The respective theory comes from Banker et al. (1984), who gave the BCC model its name.

Reconsider the very first example of the departments A to H that need only one input (clerks) to produce one output (documents). Table 7 provides the respective data. Plotted on a simple input-output plane, the data looks just like Figure 13. Remember that the CCR model identified department F to be fully efficient because it processes the most documents per number of clerks engaged. In fact, the output-input ratio of each department creates the slope of the line connecting each data point with the origin. The department with the steepest line can claim to be the most efficient. However, microeconomic theory, for instance, advises caution with respect to possible scale efficiencies (Frank 2006). The BCC model attends to this caution by enveloping the data with a convex curve rather than with a straight line to define the efficient frontier. Adapting the data in this way usually leads to more DMUs declared as efficient. While under the constant returns of scale assumption, department F is the only efficient department, under the variable returns to scale assumption the departments A, C, and E become efficient too.

Figure 13 – DEA example with one input and one output



Source: Own illustration inspired by Cooper et al. (2007)

To demonstrate the different efficiency measures and how to decompose the technical efficiency examined above, take department E as an example. The CCR model evaluates E's *overall efficiency* by equation (25), relating the steepness of the line $\overline{OE}$ to that of the line $\overline{OF}$. This is geometrically equal to relating the distance of $\overline{SR}$ to that of $\overline{SE}$. The efficient frontier, as it is defined by the BCC model, intersects the line $\overline{SE}$ at point Q (6.17, 77). The relation between $\overline{SQ}$ and $\overline{SE}$ therefore yields the *technical efficiency*, which evaluates efficiency under consideration of possible disadvantages of the size. Thus, the *scale efficiency* is the part of the *overall efficiency* that the *technical efficiency* cannot explain, as denoted in equation (34).

$$\underbrace{overall\ efficiency}_{\overline{SR}/\overline{SE}} = \underbrace{scale\ efficiency}_{\overline{SR}/\overline{SQ}} \times \underbrace{technical\ efficiency}_{\overline{SQ}/\overline{SE}} \tag{34}$$

Note that the overall efficiency equals the technical efficiency if there are no returns to scale. Banker et al. (1984) built on the CCR model to account for scale effects. For didactic reasons, consider the envelopment form of the CCR model (31). In fact, the BCC model needs only one additional constraint. Let $\boldsymbol{e}$ be a vector of ones. The BBC model (35) then seeks to

$$\min_{\theta,\boldsymbol{\lambda}} \theta_i \tag{35}$$

$$\text{s.t.} \quad \theta_i \boldsymbol{x}_i - \boldsymbol{X\lambda} \geq \boldsymbol{0}$$
$$\boldsymbol{Y\lambda} - \boldsymbol{y}_i \geq \boldsymbol{0}$$
$$\boldsymbol{\lambda} \geq \boldsymbol{0}$$
$$\boldsymbol{e}^T\boldsymbol{\lambda} = \boldsymbol{1}$$

The additional constraint $\boldsymbol{e}^T\boldsymbol{\lambda} = \boldsymbol{1}$ means that the multiplier $\boldsymbol{\lambda}$ contains in fact weights for the DMUs because the sum of all lambdas must be equal to one. While the second constraint holds the weighted average of the optimal output combinations fixed, the first constraint together with the objective minimizes the difference between the similarly weighted average of the input bundles and the input bundle of the DMU under examination (Ray 2004). As a concrete example take department E from the one input one output case above (see Table 7) where $\boldsymbol{X}$ shrinks to a row vector with the inputs of the different departments as elements. Then lambda is $\boldsymbol{\lambda}^T = [0, 0, \lambda_C, \lambda_D, 0, 0, 0, 0]$ because department E has the fully efficient departments C and F as a reference set. The optimal input bundle of E are the opposite relative distances from Q to C and D, respectively; so $\lambda_C = \overline{QD}/\overline{DC}$ and $\lambda_D = \overline{QC}/\overline{DC}$. Holding these weights fixed, the objective seeks to minimize the distance between E and the $\overline{DC}$ line by only modifying the input value. Obviously, the minimum is $\overline{EQ}$ which yields the BCC efficiency measure of pure technical efficiency. Note that without the BCC constraint $\boldsymbol{e}^T\boldsymbol{\lambda} = \boldsymbol{1}$, Q would not lie on the line $\overline{DC}$ but on $\overline{OF}$ which corresponds to the CCR frontier. The expansion to the multiple input multiple output case is straightforward and proceeds parallel to the explanation of the CCR model above. Table 9 reproduces both models in order to contrast them.

Table 9 – Input-oriented BCC vs. CCR model to estimate efficiency

Technical efficiency (CCR model)		Overall technical efficiency (BCC model)	
$\min_{\theta,\boldsymbol{\lambda}} \theta_i^{CCR}$		$\min_{\theta,\boldsymbol{\lambda}} \theta_i^{BCC}$	
s.t.	$\theta_i^{CCR}\boldsymbol{x}_i - \boldsymbol{X\lambda} \geq \boldsymbol{0}$	s.t.	$\theta_i^{BCC}\boldsymbol{x}_i - \boldsymbol{X\lambda} \geq \boldsymbol{0}$
	$\boldsymbol{Y\lambda} - \boldsymbol{y}_i \geq \boldsymbol{0}$		$\boldsymbol{Y\lambda} - \boldsymbol{y}_i \geq \boldsymbol{0}$
	$\boldsymbol{\lambda} \geq \boldsymbol{0}$		$\boldsymbol{\lambda} \geq \boldsymbol{0}$
			$\boldsymbol{e}^T\boldsymbol{\lambda} = \boldsymbol{1}$

Source: Own illustration inspired by Cooper et al. (2007)

For further reference, consider the more compact form of writing the BCC model:

$$\theta_i^{BCC}(\boldsymbol{X},\boldsymbol{Y}) = \min_{\theta,\boldsymbol{\lambda}}\{\theta_i^{BCC}|\theta_i^{BCC}\boldsymbol{x}_i \geq \boldsymbol{X\lambda}, \boldsymbol{Y\lambda} \geq \boldsymbol{y}_i, \boldsymbol{\lambda} \geq \boldsymbol{0}, \boldsymbol{e}^T\boldsymbol{\lambda} = \boldsymbol{1}\} \tag{36}$$

As a comparison, Table 10 reproduces the data and provides the results from both models.

Table 10 – Efficiencies of an input-oriented DEA example with one input and one output

Department (i)	A	B	C	D	E	F	G	H
I: Clerks	1	2	5	10	7	3	4	8
O: Documents	8	12	70	100	77	45	36	72
Techn. efficiency	0.53	0.40	0.93	0.67	0.73	1.00	0.60	0.60
Pure techn. efficiency	1.00	0.61	1.00	1.00	0.88	1.00	0.63	0.67
Scale efficiency $\left(\frac{\theta_i^{CCR}}{\theta_i^{BCC}}\right)$	0.53	0.66	0.93	0.67	0.83	1.00	0.95	0.90

Source: Own illustration inspired by Cooper et al. (2007)

For the practitioner, it is inconvenient to rely on a rather ad hoc assumption about whether one should impose the constant returns to scale restriction or allow variable returns to scale. At the same time, the literature has generated a statistical test to support the decision. Specifically, Simar & Wilson (2002) use the fact that θ_i^{BCC} (the efficiency allowing for variables returns to scale) is always consistent, while θ_i^{CCR} (imposing the constant returns to scale restriction) is not; θ_i^{CCR} conflates technical efficiency with scale efficiency if the true production frontier exhibits variable returns to scale. If the constant returns to scale restriction holds, then the θ_i^{CCR} should be used, since its estimates are more efficient in the econometric sense (Kneip et al. 2016). To decide upon the returns to scale, determine whether $\hat{S}_{2n}^{crs} = \frac{\sum_{i=1}^{n}\hat{\theta}_i^{CCR}(\boldsymbol{x}_i,\boldsymbol{y}_i)}{\sum_{i=1}^{n}\hat{\theta}_i^{BCC}(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ significantly differs from one (i.e. H_0: $\hat{S}_{2n}^{crs} = 1$). The decision of whether $\hat{S}_{2n}^{crs}$ is close to one relies on a distribution that is approximated by the bootstrapping technique. When rejecting H_0, the CRS and VRS frontiers are different and there are scale efficiencies. A similar question arises as to whether, given there are scale efficiencies, the frontier is convex or has the form of a stepladder. The latter would favor the FDH as the appropriate method. Although Kneip et al. (2016) propose a respective statistical test, it is not applicable to the data at hand since it requires a much higher number of observations in order to provide useful results. Consequently, if the $\hat{S}_{2n}^{crs}$-test suggests variable returns to scale, convexity is assumed.

The models discussed so far did not consider the time dimension. The inputs and outputs were assumed to operate within a production process that is not subject to change over time. Although entailing some logistical difficulties, longitudinal data exhibit several advantages. To mention the three most important ones in the present context, consider shocks that dramatically change the production frontier in one year, individual endogenous effects, and the limited number of decision making units for cross-sectional studies.

Tulkens & Eeckenhaut (2006) concisely explain the different ways to treat panel data in the DEA framework. They also show what questions and difficulties arise from such data. One possibility that induced a particular strand in the literature is the Malmquist index (Malmquist 1953). The Malmquist index uses panel data in order to evaluate productivity change over time. Productivity splits in a change in efficiency (the catch up effect) and a frontier shift (the innovation effect) (Cooper et al. 2007). However, since it is the efficiencies but not their changes that are relevant in the present context, the Malmquist index is not an option to choose. A further possibility to account for panel data is the window analysis initially proposed by Klopp (1985).[52] His idea was first to define a time window typically between two and the total number of years available minus one; for illustrative purposes take four years. Klopp (1985) then treats the four observations stemming from the same unit but from different years as individual DMU, just as all the other DMUs in the data set. This provides him with four times the number of DMUs he originally had in the data set to conduct an ordinary DEA. In the last step, he calculates the mean per DMU of all efficiency scores to obtain the final efficiency. As Tulkens & Eeckenhaut (2006) point out, Klopp (1985) rules out regress, that is a negative innovation effect. If all decision-making units jointly fall below the efficient frontier in one period, he interprets it as inefficiency instead of acknowledging the regress. In addition, the window size is rather arbitrary. Therefore, neither window analysis is an option for the present study.

The intertemporal and the contemporaneous reference set are two further alternatives to consider the time dimension. In fact, they are extreme cases of the window analysis. The former defines the window over the entire set of observations, assuming that there is no frontier shift at all. The latter sets the time window to one, letting the efficient frontier shift independently every year (Tulkens & Eeckenhaut 2006). Intuitively, the contemporaneous reference set is relatively simple, needs sparse assumptions and seems therefore to be the most appropriate option. Most importantly though, the thereby obtained efficiency scores yield an important property. Namely that they are bound between zero and one on a yearly basis (in case of an input-orientation). That is, if there is an unobserved fixed effect that varies every year, potentially even endogenously, the relative measure of efficiency absorbs it. This property unfolds visibly when considering the fractional form of the CCR model in equation (25). In order to interpret the result as relative efficiency, the CCR model rests upon the assumption of a constant production frontier (Cooper et al. 2007). A common frontier ensures that the different DMUs are comparable after all. However, if the frontier changes over time due to an altering contingent common factor – which becomes

[52] Tulkens & Eeckenhaut (2006) actually treat the window analysis as an evaluation of time series data. As Cooper et al. (2007) demonstrate though, it is straight forward to use window analysis with panel data.

more likely the longer the time period analyzed – the assumption does not hold anymore. A contingent common factor could be a certain technology, a disaster, the statutory framework, et cetera. All these aspects are time specific but commonly influence either the required inputs or the achieved outputs of a production process. Adding the time to the outputs of the CCR fractional problem yields[53]

$$\theta_{it} = \frac{\tau_t \sum_{r=1}^{q} u_{rt} y_{rit} / \sum_{j=1}^{p} v_{jt} x_{jit}}{\tau_t \sum_{r=1}^{q} u_{rt} y_{rkt} / \sum_{j=1}^{p} v_{jt} x_{jkt}} \quad \text{where} \quad \tau_t \frac{\sum_{r=1}^{q} u_{rt} y_{rkt}}{\sum_{j=1}^{p} v_{jt} x_{jkt}} = \max_{i=1,\dots,n} \left\{ \tau_t \frac{\sum_{r=1}^{q} u_{rt} y_{rit}}{\sum_{j=1}^{p} v_{jt} x_{jit}} \right\} = 1 \qquad (37)$$

Because it is a relative measure, a contingent common factor τ_t drops out of the analysis as long as it is the same over the entire time interval t. In this case the subscript t becomes irrelevant. Therefore, under the assumption of a stable τ_t, a DEA based on the intertemporal reference set is appropriate.[54] However, if there are different points in time with a varying contingent common factor, then the time fixed effect may remain in force. In this case, the year fixed effect τ_t biases the efficiency measure for all points in time t, in which τ_t differs from that of the most efficient peer year. Since the time fixed effect is not directly observable, it is not separable from the inputs and outputs. Thus, liberating the inputs and outputs from an altering time fixed effect is only possible with a contemporaneous reference set. Using the contemporaneous reference set, the input-oriented CCR model writes as

$$\theta_{it}^{CCR}(\boldsymbol{X}_t, \boldsymbol{Y}_t) = \min_{\theta_{it}, \boldsymbol{\lambda}_t} \left\{ \theta_{it}^{CCR} \middle| \theta_{it}^{CCR} \boldsymbol{x}_{it} \geq \boldsymbol{X}_t \boldsymbol{\lambda}_t, \boldsymbol{Y}_t \boldsymbol{\lambda}_t \geq \boldsymbol{y}_{it}, \boldsymbol{\lambda}_t \geq \boldsymbol{0}, \boldsymbol{e}\boldsymbol{\lambda}_t = \boldsymbol{1} \right\} \qquad (38)$$

Compared to the original CCR model (34) the contemporaneous reference set simply adds the index t to indicate the year. Implementing (38) comes with a trade-off though. On the one hand, the resulting efficiency measure is unbiased in the sense that it only measures the actual efficiency but excludes changes, i.e. shifts, of the frontier. In this sense, year per year DEA incorporates the same advantages as de-meaning the data in linear panel data models. On the other hand, the number of DMUs per run shrinks drastically. A small number of DMUs negatively affects the efficiency estimates in two ways. First, it restricts the number of inputs and outputs.[55] Second, as Simar & Wilson (2007) point out, the convergence rate of DEA is low. In small samples, the estimated efficiency scores most likely differ considerably from their true values. As usual, both caveats lose their importance as the number of observations increases.

The last extension of the model introduces input prices. Recall the input-oriented case of the CCR model, i.e. with constant returns to scale. Instead of simply regarding the amount of inputs used in the production process, cost efficiency models add another dimension to the efficiency measure by multiplying each input with its corresponding price.[56] The input price matrix is:

---

[53] It is irrelevant whether the time fixed effect is multiplied with the inputs or with the outputs.

[54] In fact, there is a test statistic based on the Malmquist-Index that makes it possible to evaluate whether there is innovation, i.e. productivity change, over time (Simar & Wilson 1999). If there is no productivity change, conducting a DEA with an intertemporal reference set in the first stage and running the Simar-Wilson procedure without time dummies in the second stage is appropriate. If there is productivity change, the quasi fixed effect estimator applies.

[55] As a rule of thumb, Cooper et al. (2007) propose choosing the number of inputs p and outputs q so that $\max\{p * q, 3 * (p + q)\} \leq n$.

[56] Banker et al. (2004) and others use the term 'aggregated technical and allocative efficiency' as synonym for cost efficiency.

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & \cdots & p_{1i} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ p_{j1} & \cdots & p_{ji} & \cdots & p_{jn} \\ \vdots & & \vdots & \ddots & \vdots \\ p_{w1} & \cdots & p_{wi} & \cdots & p_{wn} \end{bmatrix} \tag{39}$$

The respective linear program minimizes total input costs, i.e. the sum of all inputs multiplied with their price, while restricting the solution to be within the feasible frontier. Analytically, the cost-CCR model is therefore:

$$\begin{aligned} (\boldsymbol{p}_i^T \boldsymbol{x}_i)^* &= \min_{\boldsymbol{X},\boldsymbol{\lambda}} \{\boldsymbol{P}^T \boldsymbol{X} | \boldsymbol{x}_i \geq \boldsymbol{X\lambda}, \boldsymbol{Y\lambda} \geq \boldsymbol{y}_i, \boldsymbol{\lambda} \geq \boldsymbol{0}\} \\ \theta_i^{C-CCR}(\boldsymbol{X},\boldsymbol{Y}) &= \frac{(\boldsymbol{p}_i^T \boldsymbol{x}_i)^*}{\boldsymbol{p}_i^T \boldsymbol{x}_i} \end{aligned} \tag{40}$$

The second line of equation (40) shows how the cost efficiency measure can be obtained based on the results of the linear program. The transformation to the cost-BCC model is done by simply adding the restriction $\boldsymbol{e}^T \boldsymbol{\lambda} = \boldsymbol{1}$. Adapting the time dimension is equally straightforward. To continue with the previous example, the number of computers refine the measurement of the input side and, by assumption, all departments face the same input prices. That is, each clerk receives 10 units of salary, while each computer generates maintenance costs of three units. This yields the price matrix $\boldsymbol{P}$:

$$\boldsymbol{P}^T = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_j, \dots, \boldsymbol{p}_w] = [\boldsymbol{10}, \boldsymbol{3}] \tag{41}$$

The further data corresponds to the previous examples and is given in the upper part of Table 11. The middle part of Table 11 results directly from the number of inputs with their prices and the number of outputs. Clerk cost is the product of the number of clerks and their wage, whereas computer cost is calculated by multiplying the number of computers with their maintenance cost. The next two lines divide the products with the output from the production process. The last three lines report the efficiency scores of the cost-CCR model in equation (54), of the CCR model in (31) and a combination of the two. More specifically, the allocative efficiency in the last line of Table 11 arises from the fact that the cost efficiency combines the technical and the cost efficiency (see equation (56) below).

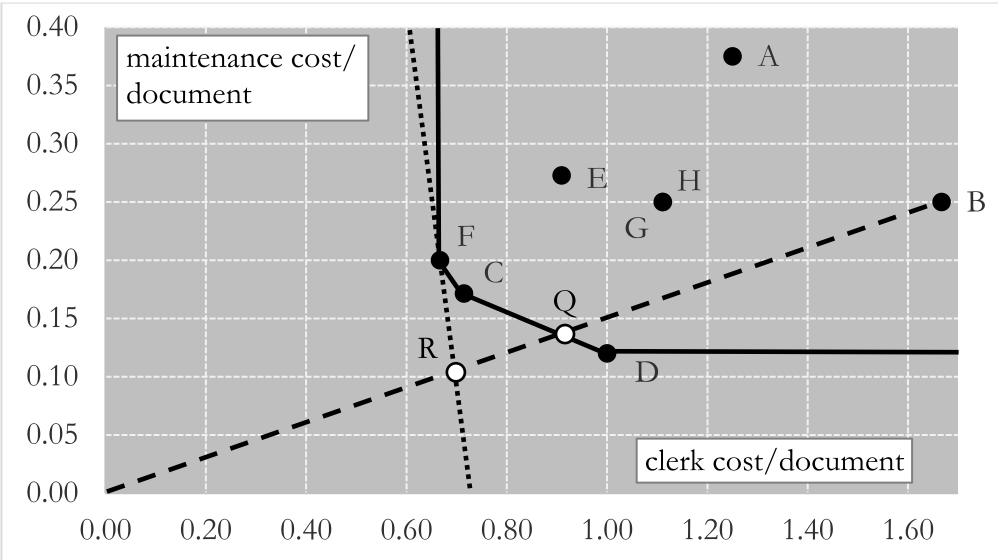
Table 11 – Efficiencies of an input-oriented DEA example with two inputs and one output

Department (i)	A	B	C	D	E	F	G	H
I: Clerks	1	2	5	10	7	3	4	8
I: Computers	1	1	4	4	7	3	3	6
O: Documents	8	12	70	100	77	45	36	72
Clerk cost	10	20	50	100	70	30	40	80
Computer cost	3	3	12	12	21	9	9	18
Clerk cost/Document	1.25	1.67	0.71	1.00	0.91	0.67	1.11	1.11
Computer cost/Document	0.38	0.25	0.17	0.12	0.27	0.20	0.25	0.25
Cost efficiency	0.53	0.45	0.98	0.77	0.73	1.00	0.64	0.64
Technical efficiency	0.53	0.55	1.00	1.00	0.73	1.00	0.67	0.67
Allocative efficiency	1.00	0.83	0.98	0.77	1.00	1.00	0.96	0.96

Source: Own illustration inspired by Cooper et al. (2007)

To complete the example, a graphical illustration supports the previous explanation. Figure 14 plots the clerks' cost and the computer cost per document. As an example, take department B, whose connection line to the origin intersects the solid lined efficiency frontier at Q. As in the one-input-two-output case, the *overall efficiency* follows from $\overline{OQ}/\overline{OB}$. The relative price line $\overline{FR}$ is the tangent to the frontier and has the slope of the two inputs' relative prices, i.e. $-10/3$. The intersection R represents a hypothetical DMU where B would be cost efficient. B's *cost efficiency* is therefore $\overline{OR}/\overline{OB}$.

Figure 14 – Cost frontier example with DEA



Source: Own illustration inspired by Cooper et al. (2007)

Using Figure 14 as a graphical illustration of the data in Table 11, the origin of the last line becomes evident. The three efficiency measures are interrelated as follows:

$$\underbrace{cost\ efficiency}_{\overline{OR}/\overline{OB}} = \underbrace{overall\ efficiency}_{\overline{OQ}/\overline{OB}} \text{ x } \underbrace{allocative\ efficiency}_{\overline{OR}/\overline{OQ}} \quad (42)$$

The last model presented to measure cost efficiency seems to have a rather high demand regarding data availability and it might therefore trigger doubts about the practical usability. However, Banker et al. (2004) summarize the following propositions proven in different papers, where the intuition comes from the unit invariance:

- <u>Cost efficiency with total cost data only</u>: If the individual input prices are equal for all DMUs, then the cost-CCR and the cost-BCC model with individual input costs return the exact same cost efficiency scores as when the total costs apply as input in the CCR and the BCC model, respectively.
- <u>Overall efficiency with cost data only</u>: If all input prices are equal for all DMUs, then the cost-CCR and the cost-BCC model with individual input amounts and input prices returns exactly the same overall efficiency scores as when individual input costs are used as input in the CCR and the BCC model, respectively.

Applied to empirical data, the two propositions prove to be valuable when only cost data, i.e. expenditures on individual inputs, but no individual inputs or input prices, are available. In other words, contrary to the first impression, the cost efficiency model potentially helps to overcome data availability issues, because cost data are usually easier accessible. The word 'potentially' indicates that the propositions rely on the assumption of common prices among DMUs. To make use of the two propositions, first enter the input costs as multiple inputs either in the BCC or in the CCR model. The resulting scores denote the overall efficiency. Next, obtain the cost efficiency scores by using total expenditures, i.e. the sum of all input costs, as single input. Regarding Table 11, whose data is based on uniform input prices for all departments, it suffices to know the clerk cost and the computer cost per DMU to obtain all three efficiencies. Neither the prices of each input nor the amounts of inputs are necessary.

The introductory section to efficiency classified the DEA as a deterministic approach because the method interprets the entire distance to the production frontier as inefficiency. This fairly strict assumption is the reason why most econometricians are averse to this method, as they underline the problems coming with measurement errors and outliers (Tauchmann 2011). The issue of extreme values is particularly severe with the DEA because the observations determining the production frontier influence the efficiency of all the other DMUs. On the one hand, this criticism led to a research strand looking for methods to detect outliers. Two of these methods, i.e. order-α (Aragon et al. 2005) and order-m (Cazals et al. 2002) efficiency, are employed here.[57] On the other hand, the implementation of bootstrapping techniques accounts not only for a possible bias of the efficiency scores but also makes it possible to perform statistical inference. Simar & Wilson (2011a) show how to apply the bootstrap technique to obtain a pseudo-sample of B estimates $\widehat{\boldsymbol{\theta}}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$, with which a bias-corrected efficiency estimate $\widehat{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ results from

$$\widehat{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 2\widehat{\boldsymbol{\theta}}(\boldsymbol{x}_0, \boldsymbol{y}_0) - B^{-1} \sum_{b=1}^{B} \widehat{\boldsymbol{\theta}}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0) \tag{43}$$

where $\widehat{\boldsymbol{\theta}}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ are the original efficiency estimates of each DMU using the linear program in equation (36), for instance. Note that the bias correction is independent of the assumption about the returns to scale. As will be shown below, under certain conditions the technique is also applicable using data from different time periods. Applying the bootstrap to the initial example with the data given in Table 8 yields efficiency estimates that are slightly smaller. The pseudo-sample stems from a smoothed homogenous bootstrap with 2000 replications.[58] As a comparison, Table 12 reports also the original efficiencies without bias correction. Both efficiencies are input oriented and the underlying production function is assumed to exhibit no scale efficiencies.

[57] When calculating the order-m efficiency score, only m randomly chosen observations form the efficient frontier. This leaves some DMUs outside the production possibility set and therefore their efficiency is bigger than one, i.e. they are super-efficient. Repeating the procedure many times theoretically yields robust estimations (Cazals et al. 2002). The order-α efficiency uses the $(100 - \alpha)$th percentile of the peers' inputs (or outputs, depending on the orientation) to define the frontier, which also results in super-efficient DMUs (Aragon et al. 2005). If $\alpha = 100$ then the order-α is equivalent to the FDH.

[58] The appropriate type of the bootstrap depends on the dependency between the set of inputs and the efficiency score (input-oriented case) or the set of outputs and the efficiency score (output-oriented case). If the efficiency is independent, then the simpler homogenous bootstrap produces consistent estimates. Otherwise the more complex and computationally more demanding heterogeneous bootstrap comes into play (Badunenko & Mozharovskyi 2016). The dependency is testable (Wilson 2003).

Table 12 – Original vs. bootstrap-corrected efficiencies

Department (i)	A	B	C	D	E	F	G	H
I: Clerks	1	2	5	10	7	3	4	8
O: Documents	8	12	70	100	77	45	36	72
O: Inquiries	40	120	60	300	280	48	220	160
Original efficiency	0.81	1	0.93	0.84	0.98	1	1	0.69
Bias-corrected efficiency	0.78	0.92	0.89	0.82	0.96	0.94	0.94	0.67

Source: Own illustration

The bootstrap identified DMU B's original efficiency as a small sample artefact and downscaled it to 92 percent. The efficiency score of the DMUs F and G that were initially also fully efficient, are apparently more likely and therefore remain at 94 percent.

6.3. Stochastic Frontier Analysis

In principle, the SFA estimators are based on the same idea as the DEA, or, more generally, the non-parametric estimators (Aigner et al. 1977; Meeusen & Van den Broeck 1977). Resulting from some efficient frontier, the deviation of each observation represents its inefficiency to some extent. However, instead of forming the efficient frontier directly from the observed sample, the Stochastic Frontier Approach (SFA) predefines a specific functional form of the frontier and fits it into the data. Also, in contrast to the two-step approach, where the methods to evaluate panel data are still in a very early stage of development, many different such models emerged within the one-step SFA framework. Hence, since there is no need to adjust the existing models for this analysis, this section reviews the SFA estimators more superficially and refers to the respective literature for more details. Unlike in the non-parametric framework, the usage of total cost instead of specific inputs needs to be modelled from the beginning. A simple adaptation, as in the case of DEA, does not exist. Therefore, this section treats both, the input distance function to estimate the technical efficiency and cost function to estimate the cost efficiency separately.

Starting with the distance function, let $\boldsymbol{y}_i$ be a vector of outputs and $\boldsymbol{x}_i$ a vector of inputs of individual i, as before. Based on this information, define the input-oriented production possibility function as

$$f(\boldsymbol{y}_i, \boldsymbol{x}_i\theta_i) = 0 \qquad \text{where} \qquad \theta_i = \exp(-u_i) \text{ and } u_i \geq 1 \tag{44}$$

The production possibility function specifies how the inputs, deflated by some inefficiency factor u_i, are transformed into a set of given outputs without leaving any spare resources. The efficiency term θ_i is interpreted in the same way it was for the DEA case and is also bound between zero and one. Unless the outputs are fixed, the $f(\cdot)$ is also called the transformation function. Yet another way to write the transformation function in a more general fashion under the assumption of separability of the input and output function, is

$$Ag(\boldsymbol{y}_i)h(\boldsymbol{x}_i\theta_i) = 1 \tag{45}$$

where A embodies observed and unobserved factors affecting the transformation function neutrally (Kumbhakar et al. 2015). For the simplicity of the further modifications, the transformation function is normalized to one. At this stage, an assumption about the functional form of $g(\cdot)$ and $h(\cdot)$ is necessary, which is characteristic for the parametric approaches as mentioned earlier. For simplicity, consider for now the Cobb-Douglas specification for both $g(\cdot)$ and $h(\cdot)$ leading to the following transformation function [59]

$$\ln(x_i) = \ln(x_{1i}) = \alpha_0 + \sum_{r=1}^{q} \alpha_r \ln(y_{ri}) + \sum_{j=2}^{w} \beta_j \ln(x_{ji}) + u_i$$
$$\text{where} \quad \alpha_0 = \ln(A) \quad \text{and} \quad u_i = \left(\sum_{j=1}^{w} \beta_j\right) \ln(\theta_i) \tag{46}$$

Even if the transformation function makes sense mathematically, estimating the coefficients econometrically is not feasible. This is because the right-hand side of the equation involves endogenous variables. Remember that input-orientation implies that the objective is to minimize the set of inputs, while holding outputs constant. As soon as one of the right-hand side inputs (outputs) increase, one or several outputs (inputs), which likewise act as independent variables, will increase too. Conveniently, the fact that this analysis requires an input-oriented efficiency measure facilitates solving the problem. The assumption of exogenous outputs is not only meaningful but also suitable. As Kumbhakar et al. (2015) show, with input-orientation and under constant returns to scale, the transformation function (46) can also be written as a distance function (Shepard 1953). For this purpose, a random input serves as reference – and as dependent variable – while all the inputs on the right-hand side are normalized by the reference input. This eliminates the endogeneity of the inputs (Kumbhakar et al. 2015). After adding stochastic noise, the input distance function with the Cobb-Douglas specification is [60]

$$\ln(x_{1i}) = \alpha_0 + \sum_{r}^{q} \alpha_r \ln(y_{ri}) + \sum_{j=2}^{w} \beta_j \ln\left(\frac{x_{ji}}{x_{1i}}\right) + u_i + v_i$$
$$\begin{aligned} \text{where} \quad & \alpha_0 = \ln(A) \\ & u_i = -\ln(\theta_i) > 0 \quad \text{since} \quad 0 < \theta_i < 1 \\ & \mathbb{E}[v_{it}] = 0 \end{aligned} \tag{47}$$

Taken together, the stochastic noise and the inefficiency constitute the error term of the estimation equation (46), i.e. $\varepsilon_i = u_i + v_i$. In order to identify both, the noise term v_i and the inefficiency term u_i properly, they need to be either identically and independently distributed by assumption, or they can be independent and heteroscedastic by responding to certain environmental variables (see section 7.2). For the moment, ignore heteroscedasticity. In other words, the chance of observing a less efficient firm is equally large for any value of the random noise term v_i. While the noise term is usually assumed to follow a normal distribution with zero mean, pondering the distribution of the inefficiency term encouraged researches to come up with numerous models. For instance, the inefficiency term u_i could be half-normally distributed (Aigner et al. 1977), truncated-normally distributed (Stevenson 1980), gamma distributed (Greene 1990) or it can follow a truncated distribution with the scaling property (Wang & Schmidt 2002). The

---

[59] By definition, the transformation function with the Cobb-Douglas specification is $A \prod_r (\boldsymbol{y}_{ri})^{\alpha_r} \prod_j (\theta_i \boldsymbol{x}_{ji})^{\beta_j} = 1$. This function contains too many parameters to be identified, which justifies some normalizations. Therefore, set $\beta_1 = -1$ resulting in the transformation function $x_{1i} = A \prod_r \boldsymbol{y}_{ri}{}^{\alpha_r} \prod_{j=2} \boldsymbol{x}_{ji}{}^{\beta_j} \theta_i^{\sum_j \beta_j}$. Taking logarithms on both sides yields equation (46).

[60] To obtain the distance function, normalize $\sum_{j=1}^{w} \beta_j = -1$. The constant returns to scale assumption is simply another term for what Kumbhakar et al. (2015, p.29) called the "homogeneity restriction", which is necessary to separate the inputs from the outputs.

different approaches vary particularly in the number of parameters to be estimated. Naturally, more parameters mean more flexibility but also add complexity to the estimation.

As in other econometric issues, the exploitation of panel data opens new possibilities to identify further parameters or to identify the parameters of interest more appropriately. Concretely, having several observations of the same individual enables excluding fixed effects. The true random- and true fixed effect model splits the error term into three parts, i.e. $\varepsilon_i = u_{it} + \alpha_i + v_{it}$ (Greene 2005). The term u_{it} thereby entails only the time-variant inefficiency. On the one hand, taking all persistency out of u_{it} eliminates the risk of confounding some unexplainable individual fixed effects (as a production shock, for instance) with time-persistent inefficiency. On the other hand, a latent time-persistent inefficiency remains unidentified too (Kumbhakar et al. 2014). Even so, the aim here is to explain efficiency differences based exclusively on time-varying environmental variables, which renders time-persistent efficiency estimates redundant. The input distance function estimated with the true-random or true-fixed effect model is

$$\begin{aligned} &\text{CD: } \ln(x_{1it}) = \alpha_0 + \alpha_i + \textstyle\sum_r^q \alpha_r \ln(y_{rit}) + \sum_{j=2}^w \beta_j \ln\left(\frac{x_{jit}}{x_{1it}}\right) + u_{it} + v_{it} \\ &\quad \text{where} \quad \mathbb{E}\left[\alpha_i | x_{jit}, y_{rit}\right] = 0 \text{ [random effect] or } \mathbb{E}\left[\alpha_i | x_{jit}, y_{rit}\right] \neq 0 \text{ [fixed effect]} \\ &\quad u_{it} \overset{iid}{\backsim} N^+(0, \sigma_u^2) \\ &\quad v_{it} \overset{iid}{\backsim} N(0, \sigma_v^2) \end{aligned} \tag{48}$$

Note that Greene (2005) assumes the input-oriented technical inefficiency u_{it} to follow a half normal distribution instead of a more flexible truncated one. He thereby circumvents the usual incidental parameter problem of other non-linear fixed effect models. The inefficiency term u_{it} indicates the percentage overuse of inputs owed to inefficiency. After the estimation of equation (48), the technical efficiency can be obtained through $TE_{it} = \mathbb{E}[e^{-u_{it}} | \varepsilon_{it}]$ (Battese & Coelli 1988). The rest of the error term consists of the firm fixed effect α_i, and the stochastic noise v_{it}. As in the usual random- and fixed effects models, the decision of which suits the data better depends on the correlation of the fixed effect with the dependent variable. In case the individual fixed effects are endogenous, only the true fixed effect model is consistent. If the individual fixed effects are exogenous, the true random-effects model is consistent and, in addition, more efficient because it exploits more orthogonality conditions (Greene 2005). The Hausman-test supports the decision of which model to use.

The adequacy of the model in equation (48) becomes apparent with the estimated coefficients. Unless $\beta_j < 0 \; \forall \, j$ and $\alpha_j > 0 \; \forall \, j$ there is a problem in the specification. Such problems may arise because of the simplicity of the Cobb-Douglas function. While the function's plainness facilitates the interpretation of the coefficients, it is not very flexible and therefore might not fit the data very well. The translog specification entails more flexibility in terms of aligning to the data, while it becomes rather complex if the number of inputs or outputs is high. With the translog specification estimation equation with the input distance function reads as

TL: $\ln(x_{1it}) = \alpha_0 + \alpha_i + \sum_r \alpha_r \ln(y_{rit}) + \frac{1}{2}\sum_r \sum_k \alpha_r \ln(y_{rit}) \ln(y_{kit})$

$+\sum_{j=2} \beta_j \ln\left(\frac{x_{jit}}{x_{1it}}\right) + \frac{1}{2}\sum_{j=2}\sum_{l=2} \beta_j \ln\left(\frac{x_{jit}}{x_{1it}}\right) \ln\left(\frac{x_{jit}}{x_{1it}}\right)$

$$+\sum_{j=2}\sum_{k=2} \gamma_j \ln\left(\frac{x_{jit}}{x_{1it}}\right) \ln(y_{kit}) + u_{it} + v_{it} \tag{49}$$

where $[\alpha_i | x_{jit}, y_{rit}] = 0$ [random effect] or $\mathbb{E}[\alpha_i | x_{jit}, y_{rit}] \neq 0$ [fixed effect]

$u_{it} \overset{iid}{\backsim} N^+(0, \sigma_u^2)$

$v_{it} \overset{iid}{\backsim} N(0, \sigma_v^2)$

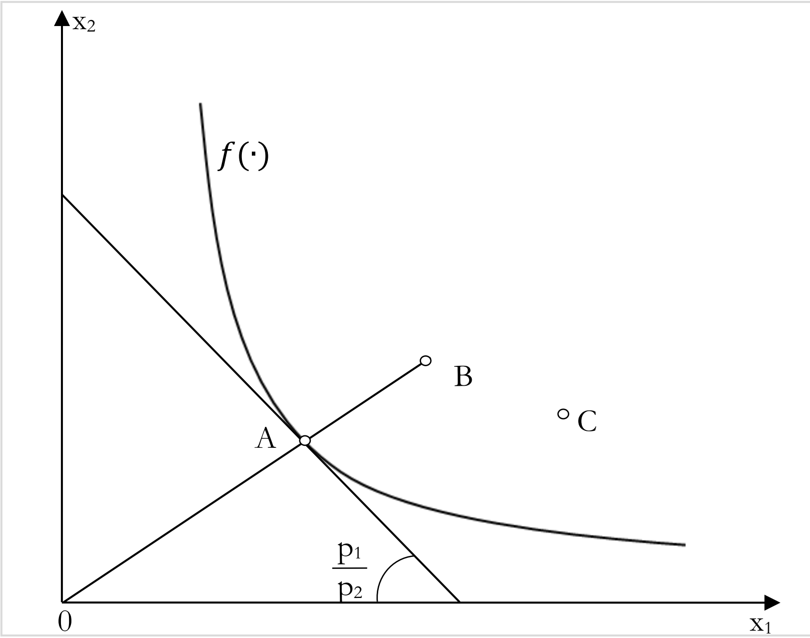
The estimation equation (49) makes apparent why the translog (TL) specification rapidly becomes much more complex than the Cobb-Douglas (CD) counterpart in equation (48). While in the Cobb-Douglas case, only the logarithmized outputs act as independent variables, the translog specification adds cross products of each output pair.

Like the DEA case, the question is raised as to what happens if only input costs but no individual amounts of inputs are available. It can be shown, that, as long as the input prices p_{jit} do not vary among cantons and years, both estimation equations (48) and (49) yield the same results when replacing $\boldsymbol{x}_{it}$ by $\boldsymbol{px}_{it}$. That is, as in the DEA case, input distance functions under the SFA framework make it possible to use the costs of each input instead of individual inputs and input prices to obtain the technical efficiency. While from a cross-sectional point of view equal prices might reasonably come from competitive markets, from which cantons purchase their resources (see section 6.4), it is less evident why prices should stay stable over time. Particularly if the time horizon is long, a certain inflation is expected. Consequently, all the input costs have to be deflated. In sum, equations (48) and (49) provide the intended estimates of the technical efficiency, when using $\ln(px_{1it})$ as a dependent variable and $\ln\left(\frac{px_{jit}}{px_{1it}}\right)$ as an independent variables.

To estimate the cost efficiency, the cost function is the starting point. Consider a firm that faces a production function $f(\boldsymbol{x})$, which is specified later. Assume further that the firm wants to minimize its input costs subject to the production function given a certain amount of outputs. To achieve its goal, standard microeconomics suggests the firm to combine the inputs so that the slope of the isoquant equals the ratio of input prices (Frank 2006).[61] The isoquant traces all input combinations with which the firm can produce the same level of outputs. For illustrative purposes, take the case where the firm utilizes only two inputs to fabricate one output. In this simplified case, the graphical illustration corresponds to the stylized example of the previous section (Figure 14). A comparison of Figure 15 with Figure 14 illustrates the difference between the parametric and the non-parametric approaches. While the isoquant in Figure 14 results from enveloping the observations, its counterpart in Figure 15 needs an assumption regarding the functional form of $f(\boldsymbol{x})$. Assuming the specific functional form of Figure 15 and the ratio of input prices, at point A the firm is both allocatively and technically efficient, while at B it could increase its technical efficiency. At C it is neither allocatively nor technically efficient.

[61] To see why, consider the firms' optimization problem: $\min \boldsymbol{p}'\boldsymbol{x}$ s.t. $y = f(\boldsymbol{x})$. The first order conditions are then $\frac{f_j}{f_1} = \frac{p_j}{p_1}, j = 2, \dots w$.

Figure 15 – Cost frontier example with SFA



Source: Adapted from Kumbhakar et al. (2015)

The inefficiency shall now also be implemented analytically. For this purpose, multiply the inputs within the production function again with an efficiency factor $\theta = e^{-u}$ where u is the input-oriented *cost* inefficiency. By definition, the inefficiency $u \geq 1$ and therefore the efficiency $1 \geq \theta > 0$ as before. Hence, multiplying the $\boldsymbol{x}$ with θ amounts to adjusting the observed inputs by the efficiency. As a result, the cost function

$$C^*(\boldsymbol{p}, \boldsymbol{y}) = \sum_j p_j x_j e^{-u} \quad (50)$$

returns the hypothetical minimum cost, which is not observable. In Figure 15, A depicts the input allocation of the hypothetical minimum cost. The researcher only observes B, which results from the actual cost function[62]

$$C^a(\boldsymbol{p}, \boldsymbol{y}) = \sum_j p_j x_j = C^*(\boldsymbol{p}, \boldsymbol{y}) e^u \quad (51)$$

As in the non-parametric case, the efficiency is the ratio of the two distances from the origin in Figure 15, i.e. the cost efficiency $CE = \frac{\overline{0A}}{\overline{0B}} = \frac{C^*(\boldsymbol{p},\boldsymbol{y})}{C^a(\boldsymbol{p},\boldsymbol{y})} = e^{-u}$. In case there is no allocative inefficiency, the ratio can also be interpreted as technical efficiency. In order to make the function stochastic and linear, logarithmize the actual cost function and add a noise term v_i. In addition, some potential individual fixed effect might interfere, which justifies adding α_i. The fixed effects, the noise term, and the inefficiency together constitute the error term of the estimation equation, i.e. $\varepsilon_i = \alpha_i + u_{it} + v_{it}$.

$$\ln C_{it}^a = \ln C^*(\boldsymbol{p}_{it}, \boldsymbol{y}_{it}) + \alpha_i + u_{it} + v_{it} \qquad \text{where} \quad u_{it} > 0, \quad \mathbb{E}[v_{it}] = 0 \quad (52)$$

[62] Applying Shephard's lemma, i.e. $\frac{\partial C^*}{\partial p_j} = x_j e^{-u}$, justifies defining the cost share equations as $S_j \equiv \frac{\partial \ln C^*}{\partial \ln p_j} = \frac{1}{p_j^{-1}} \frac{1}{C^*} \frac{\partial C^*}{\partial p_j} = \frac{p_j x_j e^{-u}}{C^*} = \frac{p_j x_j e^{-u}}{\sum_j p_j x_j e^{-u}} = \frac{p_j x_j}{\boldsymbol{p}'\boldsymbol{x}}$. Rearranging yields $p_j x_j = S_j C^* e^u$ and finally, summing up all inputs results in the actual cost function $C^a(\boldsymbol{p}, y) = \frac{\sum_j p_j x_j}{\boldsymbol{p}'\boldsymbol{x}} C^* e^u = C^* e^u$.

Again, to obtain an estimable function, the form of the minimum cost function $C^*(\boldsymbol{p}_{it}, \boldsymbol{y}_{it})$ is assumed to be either Cobb-Douglas or translog. Naturally, the aforementioned advantages and drawbacks of each specification apply here, too. Also, in either specification, the prices $\boldsymbol{p}_{it}$ appearing on the right-hand side of the estimation equation potentially imply a problem with regards to data availability. Yet, equally to the DEA models and the input distance functions, the prices drop out of the eventual estimation equation if they remain constant. The reason for their insignificance is that all coefficients related to the prices are not identified without variation and merge into the constant (Filippini & Wetzel 2014; Kumbhakar et al. 2015). Again, in order to guarantee price stability, the total cost on the left-hand side must enter in real instead of nominal terms. The same adjustment on the right-hand side ensures that prices do not vary over time and drop out. The estimation equations of the actual cost functions then read as

CD: $\ln C_{it}^{real} = \beta_0 + \boldsymbol{\beta}' \ln \boldsymbol{y}_{it} + \alpha_i + u_{it} + v_{it}$

TL: $\ln C_{it}^{real} = \beta_0 + \boldsymbol{\beta}' \ln \boldsymbol{y}_{it} + \frac{1}{2}\sum_{r=1}^{q}\sum_{s=1}^{q}\gamma_r \ln y_{itr} \ln y_{its} + \alpha_i + u_{it} + v_{it}$

where $\mathbb{E}[\alpha_i | \boldsymbol{y}_{it}] = 0$ [random effect] or $\mathbb{E}[\alpha_i | \boldsymbol{y}_{it}] \neq 0$[fixed effect] (53)

$u_{it} \overset{iid}{\sim} N^+(0, \sigma_u^2)$

$v_{it} \overset{iid}{\sim} N(0, \sigma_v^2)$

The assumptions about the distributions, the efficiency, the noise, and the fixed effect are adopted from Greene's (2005) true random-effect or true fixed effect model where the inefficiency term u_{it} follows a half-normal distribution with a constant variance σ_u^2. The constant variance assumption will be relaxed later. To sum up, given the data at hand, the estimation equations (53) represent the most appropriate to approach to estimate cost efficiency using the SFA, while the same is true for equations (48) and (49) to estimate the technical efficiency.

6.4. Inputs and Outputs

Identifying the inputs and outputs of a production process imposes challenges comparable to the operationalization of the revenue characteristics. On the one hand, one wants to represent the inputs and outputs of a production process as closely as possible. On the other hand, finding specific quantitative measures along with the usual limited availability of data restricts the possible candidates. Following many former papers, the input side considers cost data instead of individual inputs and their prices (Seifert & Nieswand 2014; Widmer & Zweifel 2012; Afonso & Fernandes 2008; Kellermann 2007; Balaguer-Coll et al. 2007). The total government expenditures proxy the cost of production.

Yearly data on total cantonal expenditures (i.e. the sum of expenses of the current account and capital expenditures) is available in a harmonized form. The Swiss federal finance administration adjusts the annual financial statements in the sense that it applies the same consolidation scope to all cantons so that the data become comparable among cantons and years (Financial Statistics Section 2011). In addition, the functional chart of accounts of the HAM2 presented in section 4.2 splits the public expenditures into different domains. While Table 3 shows the total expenditures of all cantons in one year, Table 13 presents detailed descriptive statistics of the expenditures within the different government functions.

Table 13 – Descriptive statistics of cantonal expenditures by government function 2000 to 2014

Government function	Mean	Standard deviation	Minimum	Maximum	Sum squared between (B)	Sum squared within (W)	B/W-ratio
General administration	691'482	973'748	34'291	12'336'754	1.75E+14	1.94E+14	0.90
Public order and security	984'572	1'114'694	29'423	5'490'743	4.69E+14	1.40E+13	33.40
Education	3'357'461	3'824'386	84'959	19'924'524	5.40E+15	2.93E+14	18.44
Culture	257'741	308'305	5'514	1'337'992	3.62E+13	7.66E+11	47.26
Health	2'004'178	2'138'203	57'787	8'720'812	1.69E+15	8.49E+13	19.94
Social Security	2'664'654	2'875'202	84'839	11'783'220	3.10E+15	1.12E+14	27.69
Transportation	1'047'560	1'012'224	22'692	5'483'575	3.75E+14	2.36E+13	15.87
Environment and spatial planning	222'031	189'396	6'810	1'052'621	1.25E+13	1.42E+12	8.81
National economy	784'209	751'659	66'407	4'546'419	2.06E+14	1.36E+13	15.11
Financing and taxes	919'228	1'246'362	7'079	7'210'696	5.06E+14	9.81E+13	5.16

Note: values in 1'000 Swiss francs; N = 390
Source: Federal Finance Administration (2016)

Regarding the means, Table 13 replicates the insights of Table 3 with education, social security, and health being the most expensive government functions. It is also in these domains, where the standard deviation is the highest. The extreme values in column three and four underline the rather large variation. The last three columns analyze where the variation comes from. They show the sums of the squared deviations from the means. 'Within' indicates the deviations from the group (i.e. cantonal) means and 'between' are the deviations of the group means from the grand mean; the B/W-ratio divides the between by the within variation, which simplifies the interpretation. Except the general administration domain, all government functions reveal ratios greater than one. A figure above (below) one means that the variation within cantons over the years is less (greater) than the variation between the cantons. This is not surprising given the enormous difference in scale between the cantons. The extent to which the 'between' variation exceeds the 'within' variation strongly depends on the government function though. The general administration has the most balanced variation followed by financing and taxes. Culture yields the most unbalanced variation, meaning that the amount the different cantons spend in that domain varies strongly while the individual cantons hold their expenditures constant over the years.

Following the explanations of sections 6.2 and 6.3, using only the total expenditures does not suffice to estimate the technical efficiency. In order to approximate the individual input costs, Table 26 in the appendix disaggregates the total expenditures into their types according to the chart of accounts of the HAM2 and yields the corresponding means and standard deviations. Table 14 is a more accessible summary and presents the expenditure share by expenditure type for each government function. It is noteworthy that the personnel, the operating, and the transfer expenditures together with the capital expenditures in tangible fixed assets constitute a large part of the total expenditures. The four expenditure types represent together between 68 and 90 percent of the total expenditures of six public domains. Yet, as far as regards the general administration, the protection of the environment and spatial planning, and the national economy domain, the four types are less representative and amount to 30 to 50 percent of the total expenditures. In these three domains, the extraordinary expenditures contribute largely to the total expenditures.

Table 14 – Cantonal expenditures by expenditure type and government function 2000 to 2014

Government function	General administration	Public order and security	Education	**Culture**	Health	Social Security	**Transportation**	Environment and spatial planning	National economy	Financing and taxes
Personnel expenditure [P]	15.37%	56.59%	46.25%	17.06%	25.03%	7.14%	8.09%	13.37%	5.11%	0.88%
General, administrative and operating expenditure [O]	10.90%	19.51%	7.59%	10.25%	9.96%	2.43%	10.62%	9.92%	3.03%	10.86%
Defense expenditure	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Financial expenditure [F]	0.10%	0.04%	0.00%	0.02%	0.00%	0.01%	0.25%	0.00%	0.00%	31.96%
Transfer expenditure [F]	1.65%	3.01%	18.09%	55.90%	46.08%	79.84%	21.45%	9.32%	45.94%	56.31%
Extraordinary expenditure [F]	64.25%	15.14%	24.11%	0.00%	0.58%	3.13%	0.00%	34.96%	28.73%	0.00%
Tangible fixed assets [C]	3.16%	3.47%	2.86%	6.04%	3.14%	0.33%	36.47%	7.25%	0.47%	0.00%
Capital expenditures on behalf of third parties [C]	0.01%	0.17%	0.00%	0.00%	0.00%	0.14%	1.02%	0.29%	0.00%	0.00%
Capital expenditures, intangible fixed assets [C]	1.66%	0.81%	0.18%	0.13%	0.08%	0.38%	0.24%	2.09%	0.22%	0.00%
Loans and financial interests [C]	0.53%	0.32%	0.16%	2.02%	1.23%	1.19%	7.12%	5.93%	5.11%	0.00%
Loans [C] [a]	1.08%	0.24%	0.12%	2.97%	0.28%	0.41%	8.60%	1.69%	1.33%	0.00%
Financial interests and share capital [C]	0.65%	0.07%	0.09%	0.47%	9.44%	4.27%	0.62%	0.00%	4.35%	0.00%

[a] The separation of the loans into two positions stems from the harmonization process of the federal finance administration and is not relevant in the present context (Financial Statistics Section 2011).
Notes: The letters in brackets indicate which category each expenditure type belongs to.
Source: Own calculation based on data from the Federal Finance Administration (2016)

In order to keep the number of inputs on a reasonable level, Table 14 implies the following combinations of expenditure types:

- [P] Personnel expenditure (personnel expenditure)
- [O] Operating expenditure (general, administrative and operating expenditure)
- [F] Financial expenditure (financial expenditure; transfer expenditure; extraordinary expenditure)
- [C] Capital expenditures (tangible fixed assets; capital expenditures on behalf of third parties; capital expenditures, intangible fixed assets; loans and financial interests; financial interests and share capital)

For the following estimations of the technical efficiency, these four types serve as inputs while assuming identical factor prices among the cantons.[63] The assumption of identical factor prices has provoked already clarifications in previous studies. Seifert & Nieswand (2014) make the same assumption for the French departments and argue that on the one hand the personnel costs should not vary widely because the central government largely regulates the wages. On the other hand, capital related expenditures should also face equal prices due to the equivalent access to the capital market. While the second argument seems reasonable for the Swiss cantons, too, the first is questionable. Since the cantons autonomously decide upon salary and wages regulation, the labor cost is not necessarily equivalent among cantons. The Swiss federal statistical office publishes the

[63] Similar partitions of the total expenditures also exist for studies outside of Switzerland, see e.g. Bischoff et al., 2013, Bönisch et al., 2011 or, for an overview, Narbón-Perpiñá & De Witte (2017).

respective data to assess the question of equal labor costs. The data is not available by canton but only by greater region. Table 15 summarizes the mean monthly gross salaries in the public administration for the years 2010, 2012, and 2014.

Table 15 – Monthly gross salaries in the public administration across greater regions

Year	Switzerland	Geneva region	Central plateau	North-western Switzerland	Zurich	Eastern Switzerland	Central Switzerland	Ticino
2010 [a]	7'943	7'996	8'369	7'467	8'080	7'635	7'930	7'452
2012 [a]	7'916	7'902	8'181	7'411	8'325	7'887	7'585	7'284
2014 [a]	7'864	8'197	8'009	7'369	7'815	7'841	7'717	7'199
2010 [b]	100	101	105	94	102	96	100	94
2012 [b]	100	99	103	93	105	99	95	92
2014 [b]	99	103	101	93	98	99	97	91

[a] in Swiss francs; [b] index-based where Switzerland in 2010 equals 100

Source: Federal Statistical Office (2016)

The index-based values in the lower part of Table 15 reveal a salary range of fifteen percent among the greater regions. Admittedly, looking at greater regions instead of individual cantons crops eventual extreme values. However, the fact that the canton of Zurich and the canton of Ticino possibly represent such extreme cases refute the argument. In sum, compared to the relatively large variation of the personnel expenses (c.f. Table 28 in the appendix), the variation of the wages appears reasonably small thus retaining the assumption of equal factor prices. Consequently, the variation of the personnel expenses is mainly driven by the input (i.e. the personnel count) and not the price (i.e. the wages).

Identifying and operationalizing outputs is less simple. On the one side, the true output of a public entity is already difficult to determine. On the other side, the limited data availability severely restricts the possible factors to be taken into account. These two challenges already became apparent in existing studies that have estimated the efficiency of public administrations. The selection of the outputs in the different public domains therefore mainly follows the previous literature. One important aspect concerns the separation of the outputs into quantitative and qualitative factors. While the broad empirical literature neglects qualitative measures, be it because of data unavailability or due to negligence, they play an important role in reality. There is no doubt that an agency that, for instance, operates at a higher error rate than another is less efficient, even if all other inputs and outputs are equal. Particularly under cost pressure, the quality is likely to decrease before the quantity. Thus, in order to capture the entire scope of output, the integration of qualitative measures is indispensable if the data availability allows it (Balaguer-Coll & Prior 2009). Accordingly, the itemization of the used outputs in Table 16 contains quantitative as qualitative measures and discloses the data source.

Table 16 – Descriptive statistics of output measures

Government function	Output	Mean	Standard deviation	Minimum	Maximum	N	Data Source
General administration	population (no. of inhabitants)	294’469.2	305’002.4	14’977	1’446’354	390	federal finance administration
	firms (no.)	19’166.6	18’667.2	1’004	100’002	390	federal statistical office, central business names index
	interest rate on debt (index [a])	58.9	17.8	1	100	390	federal statistical office
Public order and security	convicts (no.)	3802.7	3689.5	47	16’141	357	federal statistical office
	delinquencies registered at police (no.)	85.6	17.4	1	100	356	federal statistical office, federal office of police
	prisoner on remand (no.)	75.9	113.9	1	506	357	federal statistical office
Education	students in higher education institutions (no.)	3’577.1	3’696.2	128	16’589	390	federal statistical office
	university diplomas (no.)	643.6	691.8	15	3’565	390	federal statistical office
	successful high school graduation (no.)	1’225.2	1’204.5	36	6’171	390	federal statistical office
Culture	participation in courses and camps of Youth and Sport (no.)	20'670.1	21'887.4	703	124'804	361	federal office of sport
	museum tickets sold (no.)	558'391.4	846'930.8	122	4'224'172	385	association of Swiss museums
	museums (no.)	27.6	31.3	1	153	385	association of Swiss museums
	forest area (ha)	42'289.0	44'248.5	471	159'794	385	federal statistical office
Health	doctors (no.)	601.6	722.9	16	3’669	390	federal statistical office
	patient cases in hospitals (no.)	53’009.6	60’051.6	1’250	259’249	390	federal statistical office
Social Security	unemployed (no.)	6'958.2	7'725.7	36	42'771	389	federal statistical office
	reintegrated people (no.)	550.2	647.2	1	3'169	285	federal statistical office
	people over 65 (no.)	47'018.8	48'982.7	2'298	229'196	389	federal statistical office
Transpor-tation	road length (in km)	694.7	649.6	48	2'155	390	federal statistical office
	vehicles (no.)	199'160.8	199'069.3	9'520	906'010	390	federal statistical office
	accidents due to low road quality (index [a])	95.7	10.9	1	100	390	federal roads office
Environment and spatial planning	vacant accommodation (no.)	1'515.7	1'524.2	41	7'761	390	federal statistical office
	population (no. of inhabitants)	294'469.2	305'002.4	14'977	1'446'354	390	federal statistical office
National economy	farms (no.)	2'377.9	2'600.8	11	14'150	390	federal statistical office
	new founded firms (no.)	530.3	725.3	16	8'212	364	federal statistical office
	beds in hotels (no.)	10'385.2	12'257.5	1'135	48'163	364	federal statistical office

[a] Index as described in section 6.2
Note: N < 390 due to deletion of observations with minimum value of zero (outputs need to be strictly positive)

While Table 16 yields an overview of the outputs, Table 17 below provides information regarding the distance functions combining the inputs and outputs. The production function is specified as Cobb-Douglas technology. The regularity conditions of a well-defined distance-function require positive output coefficients and negative input coefficients (Kumbhakar et al. 2015). Also, the coefficients should be within zero and one or minus one, respectively, in order to respect the concavity condition. A violation of the two conditions means that the production function is wrongly specified or that there is a measurement error. Additionally, since the DEA efficiency scores systematically increase with a higher number of inputs and outputs ceteris paribus, only the relevant factors should enter the analysis.[64] In order to identify the appropriate outputs in terms of the sign and the magnitude of their coefficients, some simple random- and fixed effect models are estimated. The estimations are simple in the sense that they disregard possible

[64] Put differently, the convergence rate of the efficiency estimated to its true values decreases with an increasing number of inputs and outputs. The literature addresses this topic under the label ‘curse of dimensionality’ (Simar & Wilson 2011a).

inefficiencies. For culture and transportation, the results of this regression, displayed in Table 17, will be discussed subsequently together with the practical justification and the descriptive statistics of each output listed in Table 16.

The government function culture covers supportive tasks such as contributions to theaters, libraries, or sport events but also the maintenance of parks and forest trails (cf. Table 3). Accordingly, the four outputs of 'participation in courses and camps of Youth and Sport', museum tickets sold, museums, and the forest area approximate the cantonal activities in this field (Benito et al. 2010; Bischoff et al. 2017). On average, a canton registers more than 20'000 participations in Youth & Sport courses and camps per year with, naturally, huge variations due to cantonal size. Because of a change in data recording, the year 2002 is missing, thus reducing the number of observations. Regarding the number of museums, the canton of Glarus, Ob- and Nidwalden as well as Appenzell Innerrhoden closed their few institutions in some years and the respective observations drop out of the sample as DEA and SFA struggle with zero-output observations. Excluding them yields an average of 28 museums per canton and year with around 560'000 tickets sold. The forest area is averaged at around 420 kilometers squared which resembles the surface of the canton of Obwalden.

The distance function for the culture domain in Table 17 sets the inputs and outputs in relation. All of the normalized inputs have the expected sign and two of them are significant. Among the outputs, the only insignificant coefficient comes from the forest area, which raises the question of its meaningfulness. Nevertheless, the coefficient is positive and within the appropriate range of zero and one. As the number of outputs is still rather small, there is no need to drop this variable and its control function can still be valuable. A more serious issue is the Sargan-Hansen test statistic, which is highly significant. Accordingly, the fixed effects estimator would be appropriate. However, in several instances the respective estimates contradict the economic theory with opposite signs and coefficients above one. In consequence, the random-effect estimator is applied after all. The thus obtained stochastic errors accumulate to the left of the mean as the negative $\mu_3(\epsilon_{it})$ is substantiated; the statistic is not significant though. In general, the left-skewed distribution predicts no or little inefficiency (Kumbhakar et al. 2015). Note also that the distance function only explains around 60 percent of the personnel expenditures' variation, which is, compared to the other domains, rather low. This aspect influences the interpretation of the results below.

The emphasis of transportation falls on building and maintaining cantonal roads. The road length is therefore the primary output measure (Widmer & Zweifel 2012; Da Cruz & Marques 2014; Narbón-Perpiñá & De Witte 2017). In order to allow for different necessities of attrition, the number of cars reflects the frequency and intensity of use (Widmer & Zweifel 2012). Even if the cantons periodically evaluate their road quality, the results do not coalesce in a central register (Koch & Forster 2010). Because of the lack of the data, the number of accidents due to bad road quality serves as proxy (Kalb 2009). Since accidents are an undesirable variable, it enters here in the transformed form as index (see section 6.2). With an average of 95.74 and a variance of 10.87 it is strongly left-tailed; that is in 150 observations the road quality did not cause an accident. The two remaining variables differ considerably among cantons. Particularly the variation of the cantonal street length partly has its roots in the divergent coverage of the national streets. This aspect plays a role when it comes to the explanation of the efficiency differences (see section 7.3).

Looking at the estimated distance function in the seventh column of Table 17, each normalized input holds a significantly negative coefficient. On the output side, the road length and the number of vehicles are significant with the expected positive coefficients. The quality index stemming from the number of accidents due to low road quality carries a non-significant but positive coefficient. The relatively high standard error of the coefficient is unsurprising given the low variation of the variable. In consideration of the positive sign, the output still stays in the model. Like the culture domain, the significant Hansen test is problematic. The more appropriate FE-model results in economically senseless parameter estimates, making the RE-model more appropriate nevertheless. The resulting errors are significantly left-skewed as the significantly negative $\mu_3(\epsilon_{it})$ value indicates. Left-skewness puts the presence of inefficiency into question in the first instance.

Table 17 – Fixed or random effects regression by government function

General administration		Public order and security		Education		**Culture**		Health		Social Security		**Transportation**		Environment, spatial planning		National economy	
Input: O	-0.344***	Input: O	-0.073*	Input: O	-0.042**	Input: O	-0.274**	Input: O	-0.010	Input: O	-0.004	Input: O	-0.062**	Input: O	-0.156**	Input: O	-0.101**
	(0.031)		(0.032)		(0.016)		(0.040)		(0.037)		(0.016)		(0.020)		(0.031)		(0.024)
Input: F	-0.011	Input: F	-0.008	Input: F	-0.442**	Input: F	-0.673**	Input: F	-0.634**	Input: F	-0.930**	Input: F	-0.305**	Input: F	-0.097**	Input: F	-0.294**
	(0.007)		(0.009)		(0.014)		(0.033)		(0.014)		(0.020)		(0.022)		(0.017)		(0.029)
Input: C	-0.026***	Input: C	-0.004	Input: C	-0.016**	Input: C	-0.016	Input: C	-0.078**	Input: C	-0.022**	Input: C	-0.048**	Input: C	0.002	Input: C	-0.006
	(0.007)		(0.006)		(0.004)		(0.009)		(0.013)		(0.005)		(0.013)		(0.018)		(0.009)
Interest	0.026	Convicts	0.282**	Students	0.859**	Youth and	0.125**	Doctors	0.899**	Unem-	0.215**	Road	0.280*	Vacant	0.004	Farms	0.415**
rates a	(0.019)		(0.029)		(0.046)	sport	(0.026)		(0.104)	ployed	(0.035)	length	(0.130)	accomm'	(0.034)		(0.061)
Firms	0.156***	Delin-	0.226**	Univ'	0.071**	Museum	0.132**	Patient	0.055	Reint'	0.022	Vehicles	0.581**	Popu-	0.980**	Founded	0.276**
	(0.035)	quencies	(0.025)	diplomas	(0.027)	tickets	(0.024)	cases	(0.101)	people	(0.011)		(0.122)	lation	(0.081)	firms	(0.059)
Popu-	0.732***	Prisoner	0.039**	H'school	0.092**	Forest area	0.010			People	0.786**	Acci-	0.011			Beds in	0.012
lation	(0.063)	on rem'	(0.012)	graduat'	(0.032)		(0.021)			over 65	(0.058)	dents[a]	(0.023)			hotels	(0.018)
Time	0.014***	Time	0.016**	Time	0.004*	Time	0.011**	Time	-0.007	Time	0.012**	Time	0.003	Time	0.000	Time	0.020**
	(0.002)		(0.001)		(0.002)		(0.004)		(0.005)		(0.002)		(0.003)		(0.002)		(0.002)
Intercept	-17.54***	Intercept	-20.433**	Intercept	4.344	Intercept	-13.671	Intercept	23.920*	Intercept	-11.721*	Intercept	-4.826	Intercept	9.303*	Intercept	-29.974**
	(4.003)		(3.003)		(3.715)		(8.068)		(10.808)		(4.769)		(4.721)		(4.461)		(4.702)
R^2	0.902	R^2	0.963	R^2	0.986	R^2	0.625	R^2	0.969	R^2	0.967	R^2	0.860	R^2	0.854	R^2	0.791
H $\chi^2_{(7)}$	2.341	SH $\chi^2_{(7)}$	14'615***	H $\chi^2_{(7)}$	24.80***	SH $\chi^2_{(7)}$	567.20***	SH $\chi^2_{(6)}$	17.910***	H $\chi^2_{(7)}$	11.874	H $\chi^2_{(7)}$	31.08***	SH $\chi^2_{(6)}$	14.141**	H $\chi^2_{(7)}$	5.91
$\mu_3(\epsilon_{it})$	-0.284**	$\mu_3(\epsilon_{it})$	0.055	$\mu_3(\epsilon_{it})$	-0.009	$\mu_3(\epsilon_{it})$	-0.212	$\mu_3(\epsilon_{it})$	-0.828***	$\mu_3(\epsilon_{it})$	0.144	$\mu_3(\epsilon_{it})$	-0.289**	$\mu_3(\epsilon_{it})$	-0.480***	$\mu_3(\epsilon_{it})$	-0.262**
Model	RE	Model	RE	Model	FE	Model	RE	Model	RE	Model	RE	Model	RE	Model	FE	Model	RE
N	374	N	339	N	389	N	288	N	369	N	223	N	379	N	383	N	332

[a] Index as described in section 6.2

Notes: coefficients with standard errors in parentheses; all variables are in logs due to the Cobb-Douglas specification; * p=90%, ** p=95%, *** p=99%; the inputs are normalized by the dependent variable personnel expenditures, which leaves operating (O), financial (F) and capital (C) expenditures; H $\chi^2_{(df)}$ states Hausman's test statistic; SH $\chi^2_{(df)}$ is the Sargan-Hansen statistic; $\mu_3(\epsilon_{it})$ is the skewness of the residuals tested against symmetry (only significance reported). N < 390 result from expenditure values that are equal or smaller than zero (due to the harmonization of the financial data).

6.5. Efficiency estimates

The inputs and outputs assessed in section 6.4 make it possible to approach the efficiency scores using the two techniques discussed previously, i.e. DEA in section 6.2 and SFA in section 6.3. In case the two efficiency estimates differ, the literature provides little help in deciding which to trust. Badunenko et al. (2012) suggest consulting the signal-to-noise ratio $\lambda_{it} = \frac{\sigma_{u,it}}{\sigma_v}$, placing the standard error estimates of the inefficiency term ($\sigma_{u,it}$) and the stochastic noise (σ_v) into relation. The lower part of Table 18 reports the respective estimates. The authors make their preference for one or the other method dependent on three different ranges of λ. First, if the ratio is smaller than one, both approaches perform poorly, and the efficiency estimates are potentially far from their true values. Second, a ratio above one means that estimates from both approaches do well. Last, if the ratio is about one, DEA is preferable for highly efficient DMUs, whereas SFA produces better estimates for less efficient observations. However, as Andor & Parmeter (2017) detect, the accuracy of the efficiency estimates using SFA, and hence the second moments of signal ($\sigma_{u,it}$) and noise (σ_v), depend on the assumed distributions. Thus, if the two parameters are misspecified in the first place, the estimate of the signal-to-noise ratio λ_{it} is also unreliable and it cannot serve as a basis for decision-making. For pragmatic reasons, the focus lies on the usually more distinct efficiency estimates of the DEA.

In any case, the plain efficiency scores must be relativized as they leave aside the different contextual conditions under which the cantons operate. Chapter 6.6 explains the efficiency differences by means of contextual factors.

Before turning to the results, it is essential to define the necessary parameters for both estimation techniques. In the DEA case, these are the orientation, the returns to scale, and the type of bootstrap for the bias correction. For orientation, this analysis follows the argumentation of previous studies. In the public sector, the law usually leaves little leeway to alter the output, which justifies an input-orientation. The management has fairly fixed levels of public service provision and is encouraged to provide the very service with the least possible resources (Seifert & Nieswand 2014). For returns to scale, Simar & Wilson's (2002) $\hat{S}_{2n}^{crs}$ statistic tests whether the technology is subject to constant or variable returns to scale (see section 6.2). Third, to determine whether the combination of inputs is independent of the efficiency measure, Wilson's (2003) $\hat{T}_{4n}$-test approaches the question statistically.[65] Of course, while the input-orientation is valid for all the public domains, the latter two aspects have to be tested for each government function individually.

In the SFA case, constant returns to scale are a prerequisite to justify the estimation of a distance function (see footnote 60). The validity of this assumption becomes apparent by summing the output coefficients. If they sum up to one, the returns to scale are constant (Kumbhakar et al. 2015). Furthermore, it is the functional form that imposes the major concern in SFA. A well-

[65] The integrated square difference $\hat{T}_{4n} = \sum_{i=1}^{n}\left[\hat{F}_n(\delta_i, \boldsymbol{w}_i) - \hat{F}_n(\hat{\delta}_i)\hat{F}_n(\boldsymbol{w}_i)\right]^2$, where $\boldsymbol{w}_i = (\boldsymbol{y}_i, \boldsymbol{\eta}_i)$ is a vector of outputs and the decomposed efficiency score (see section 7.1) does not exceed some critical value if the combination of inputs is orthogonal to the efficiency scores (the H_0 hypothesis) (Wilson 2003). The distribution of the statistic is approximated by the bootstrapping technique.

specified production frontier meets at least the economic foundations, in the sense that $\beta_j < 0 \; \forall \, j$ and $\alpha_j > 0 \; \forall \, j$. Closely related to this concern is the assumption about the error term $\varepsilon_i = u_i + v_i$. Unless the underlying error term behaves as assumed, the efficiency estimates are biased too. It would therefore be incoherent to assume homoscedasticity up to this point and drop this assumption when it comes to the inspection of the environmental variables. Hence, in anticipation of the heteroscedasticity of u_i, the efficiency estimates presented in Table 18 rely on the estimation equation (58) and stem from the results in Table 32.

Two different Pearson correlation coefficients provide some additional information. The relationship between the estimated cost and technical efficiency $\rho_{ce,te}$ roughly identifies any misallocation of inputs; remember that the cost efficiency equals the product of the technical and the allocative efficiency (see section 6.2).[66] The same analysis is done with the efficiencies resulting from the SFA as reported in Table 18. The second coefficient $\rho_{miss,te}$ assesses a potential link between missing values of the environmental variables and the estimated efficiency. This correlation cannot be assessed with the SFA results because the environmental variables already interfere when estimating the efficiencies. Finally, Spearman's rank correlation coefficient $\rho^*_{DEA,SFA}$ gives information about the joint rank variation of the DEA and SFA efficiency estimates.

[66] With bootstrap-corrected efficiency estimates, it can happen that the allocative efficiency exceeds one. Since efficiency estimates above one are nonsensical, a respective correlation coefficient proves more useful than simply reporting a mean allocative efficiency estimate.

Table 18 – Cantonal DEA and SFA technical efficiency by government function 2000 to 2014

Canton	DEA: Culture	SFA: Culture	DEA: Transportation	SFA: Transportation
AG	0.706 (0.214)	0.999 (0.000)	0.816 (0.033)	0.032 (0.003)
AI	0.544 (0.229)	0.988 (0.014)	0.678 (0.009)	0.011 (0.000)
AR	n.a.		0.646 (0.069)	0.006 (0.000)
BE	0.671 (0.196)	0.996 (0.000)	0.707 (0.109)	0.348 (0.071)
BL	0.397 (0.172)	0.992 (0.001)	0.525 (0.054)	0.539 (0.059)
BS	0.117 (0.052)	0.995 (0.000)	0.181 (0.055)	0.096 (0.001)
FR	0.578 (0.222)		0.681 (0.073)	0.005 (0.000)
GE	0.350 (0.153)	0.998 (0.002)	0.725 (0.148)	0.029 (0.001)
GL	0.701 (0.095)		0.702 (0.023)	
GR	0.684 (0.072)		0.469 (0.085)	
JU	0.487 (0.187)	0.987 (0.003)	0.726 (0.025)	0.132 (0.017)
LU	0.719 (0.205)		0.837 (0.036)	
NE	0.604 (0.208)	1.000 (0.000)	0.485 (0.086)	0.361 (0.091)
NW	0.614 (0.302)	0.999 (0.000)	0.710 (0.062)	0.032 (0.001)
OW	0.697 (0.062)	1.000 (0.000)	0.694 (0.033)	0.029 (0.001)
SG	0.725 (0.094)	1.000 (0.000)	0.716 (0.067)	0.103 (0.011)
SH	0.716 (0.111)	1.000 (0.000)	0.691 (0.073)	0.453 (0.052)
SO	0.689 (0.078)	0.993 (0.002)	0.710 (0.079)	0.527 (0.059)
SZ	0.625 (0.207)	1.000 (0.000)	0.665 (0.078)	0.596 (0.076)
TG	0.689 (0.204)		0.764 (0.014)	
TI	0.654 (0.211)	1.000 (0.000)	0.701 (0.070)	0.006 (0.000)
UR	0.630 (0.191)	0.694 (0.267)	0.737 (0.010)	0.027 (0.001)
VD	0.522 (0.196)		0.770 (0.053)	
VS	0.459 (0.156)	1.000 (0.000)	0.761 (0.028)	0.005 (0.000)
ZG	0.546 (0.206)	0.990 (0.007)	0.469 (0.051)	0.046 (0.003)
ZH	0.529 (0.179)		0.719 (0.092)	0.901 (0.005)
CH	0.572 (0.228)	0.977 (0.100)	0.671 (0.140)	0.208 (0.089)
$\hat{S}_{2n}^{crs}$ (p)	0.855 (0.98)		0.881 (0.97)	
$\hat{T}_{4n}$ (p)	0.010 (0.00)		0.023 (0.18)	
$\rho_{ce,te}$	0.692	-0.033	0.402	-0.245
$\rho_{miss,te}$	-0.044		-0.060	
$\hat{\mathbb{E}}[\sigma_{u,it}]$		0.041***		0.137***
$\hat{\sigma}_v$		0.109***		0.093***
$\hat{\mathbb{E}}[\lambda_{it}]$		0.375		1.475
N	288	214	382	298
$\rho^{*}_{DEA,SFA}$	0.189	0.189	-0.159	-0.159

Notes: * p<10%, ** p<5%, *** p<1%

DEA estimates show the mean of the year-by-year bootstrap-corrected cost efficiencies (Simar & Wilson 2011a) with standard deviation in parentheses. CH is the mean of all cantons. The $\hat{S}_{2n}^{crs}$-statistic refers to Simar & Wilson's (2002) returns to scale test; the $\hat{T}_{4n}$-statistic refers to Wilson's (2003) independence test; always the year with the smallest p-value is reported; both tests globally reject the H_0 if any year rejects it. If the p-value of the $\hat{S}_{2n}^{crs}$-test exceeds (is equal or smaller than) 0.05 then the efficiencies are estimated based on constant (variable) returns to scale; if the p-value of the $\hat{T}_{4n}$-test exceeds (is equal or smaller than) 0.05 then the heterogenous (subsampling) version of the bootstrap correction is applied. The Pearson's correlation coefficient $\rho_{ce,te}$ describes the relationship of the cost and technical efficiency, while $\rho_{miss,te}$ shows the correlation between the missing dummy and the technical efficiency. The Spearman's rank correlation coefficients $\rho^{*}_{DEA,SFA}$ describes the relationship of the DEA and SFA technical efficiency measures. N < 390 due to outlier exclusion (see footnote 57) or negative expenditures. $\hat{\mathbb{E}}[\sigma_{u,it}]$ is tested against H_0: $\hat{\mathbb{E}}[\sigma_{u,it}] = 0$; $\hat{\sigma}_v$ is tested against H_0: $\sigma_v = 0$; $\hat{\mathbb{E}}[\lambda_{it}]$ is tested against H_0: $\lambda_{it} \leq 1$; for further regression outputs see Table 20.

Like the previous sections, the focus lies on the two domains of culture and transportation (the remaining efficiency estimates are reported in Table 29 and Table 30 in the appendix). Before reacting to the efficiency estimates of the culture domain, it is worth checking the lambda. With a value of 0.375, the efficiency variation relative to the stochastic error is too small to reliably estimate the efficiencies with either method. Just as Badunenko et al. (2012) observed, the ensuing efficiency estimates from the SFA are all together close to one, whereas the estimates from the DEA show some variation. In conclusion, the authors trust neither the one nor the other estimates. As pointed out by Andor & Parmeter (2017), the surprising SFA estimates might result from a poorly specified efficiency distribution. In order to limit the risk of commenting on meaningless results, the following paragraph mainly confines itself to the brief discussion of the efficiency estimates obtained through the DEA.

The mean technical efficiency comes to 57.2 percent. In other words, on average the cantons could reduce their personnel, operational, financial, and capital expenditures concurrently by 42.8 percent if they were operating under the same environmental conditions. Given the insignificant $\hat{S}_{2n}^{crs}$ and the significant $\hat{T}_{4n}$ statistics, the estimations are based on a constant return to scale assumption and a heterogenous bootstrap to correct the bias; also note that the sum of the beta- (i.e. output-) coefficients is far from one (see Table 32), which is again an indicator that some outputs remain unobserved. This deficiency enters into the interpretation of the main results below (section 7.5). Across all cantons and years, the efficiencies range from 7 percent (Basel Stadt in 2007) to 89 percent (Fribourg in 2012). Likewise, it is the canton of Basel Stadt that, on average, provides the least efficient cultural services, while the most efficient canton is St. Gallen over these fifteen years. Estimates for the canton of Appenzell Ausserrhoden are missing as it never had any capital expenditures, and thus, with one input always zero, the observations drop out of the sample. The fact that the correlation coefficient $\rho_{ce,te}$ is comparatively high (0.7) ascribes only a small part of the cost inefficiency to a misallocation of inputs.

Within the transportation domain, the point estimate of lambda is 1.5; however, the H_0 that lambda significantly exceeds one cannot be rejected. Accordingly, for highly efficient DMUs the DEA estimates apply, whereas for the others the SFA estimates are more trustworthy. As to the DEA estimates, the returns to scale and the orthogonality of the inputs and the efficiency must be determined. For the returns to scale, the $\hat{S}_{2n}^{crs}$-statistic of 0.88 leaves the null hypothesis of constant returns to scale intact. Likewise, the sum of the beta coefficients (see Table 32) equals exactly one, which also indicates constant returns to scale (Kumbhakar et al. 2015). Therefore, in the domain of transportation, the cantons cannot benefit from scale efficiencies and the constant return to scale assumption is valid for the further analysis. The $\hat{T}_{4n}$-statistic rejects the H_0 of independence between the efficiency scores and the combination of inputs in several years. The bootstrapping procedure must accordingly account for dependency when estimating the bias-corrected efficiency scores (Wilson 2003).[67]

Column 8 of Table 29 reports the year-by-year bias-corrected DEA technical efficiency scores under the assumption of constant returns to scale. The mean efficiency is 67 percent, which means

[67] Due to the small number of observations, the heterogeneous bootstrap was not always possible to implement. In these cases, the subsampling version replaced it where the parameter $\kappa = 0.5$. Monte Carlo simulations in Kneip et al. (2008) recommend a kappa close to a half with a sample size of below 50 observations.

that the cantons could hypothetically reduce their expenditures by 23 percentage points overall. According to Badunenko et al. (2012), DEA provides more reliable results particularly for highly efficient DMUs. In the case of transportation, these are Geneva in 2007 (86.6%), Lucerne in 2008 (86.5%), and Geneva in 2006 (86.2%). Unfortunately, Lucerne is one of the cantons that did not provide data on special financings and therefore drops out of the estimation under the SFA approach. Clearly, highly efficient observations form the frontier, whereas their exclusion might strongly affect the estimation. The extremely low correlation coefficient $\rho_{miss,te}$ of -0.06 relativizes this argument, though. The sobering comparison to the SFA efficiencies in column 8 in Table 30 result in the rank correlation of -0.16. Not only do the SFA results differ strongly from the DEA scores, the two efficiency estimates are even contradictory. A possible explanation for this inconsistency comes from the right-skewed distribution of the error (see Table 17), which poses problems particularly for the SFA due to its assumptions about the distribution of the error term. The ambiguity of the efficiency estimates shall be kept in mind when it comes to the explanation of efficiency.

6.6. Interim conclusion

After an extensive review on the performance measurement methods and the operationalization of inputs and outputs, this chapter listed mean DEA and SFA estimates across the Swiss cantons from 2000 to 2014. In anticipation of when this analysis will narrow down to focus on the two domains of culture and transportation, the comments are confined to their results. As for the culture, the DEA reports a mean efficiency of about 57 percent, while the SFA's mean climbs to 98 percent. Their rank correlation amounts to only 19 percent. Accordingly, both results should be cautiously interpreted. Regarding the transportation domain, the mean efficiencies come to 67 (DEA) and 21 (SFA) percent with a rank correlation of -16 percent, respectively. Additional tests revealed that the DEA results are more trustworthy in the further analysis.

7. Explaining efficiency

In order to finally test the hypotheses, this chapter combines the previously obtained efficiency estimates with the environmental variables, namely, the earmarked revenues. As mentioned before, the combination either happens in two steps, using the DEA approach, or in one step when applying the SFA. The methodological procedure of the two approaches is discussed in two distinct sections. While the second section briefly summarizes existing SFA models involving heterogeneity, the first section revisits established two-stage DEA models and proposes a procedure to account for panel data. To this end, the section classifies the different ways which have emerged from the empirical literature that apply the two-stage DEA analyzing panel data. In order to determine which way is the most appropriate in case fixed effects are present, the section then adapts the algorithm proposed by Simar & Wilson (2007) which is designed for cross-sectional data. For ease of reference, the resulting adjustment of the Simar-Wilson approach we call here the 'quasi fixed effect algorithm'. This means that the quasi fixed effect algorithm is basically an application of the original Simar-Wilson approach with a specification for panel data.

Often, studies exploiting panel data run a DEA on the entire sample and thereby leave aside a potentially changing production frontier. In the second step, these papers usually implement year and individual dummies to account for eventual fixed effects (Selim & Bursalıoğlu 2015; Fleishman et al. 2009; Chen et al. 2005). Instead, the quasi-fixed effect algorithm first performs the DEA on each year individually and only introduces individual dummy variables in the second stage. The Monte Carlo simulations provide evidence that the quasi fixed effect algorithm performs a lot better in terms of accuracy and efficiency than the conventional approach.

The third section addresses the concerns related to reverse causality, that is, whether the models actually estimate causal effects of the environmental variables on efficiency or whether the results are simple correlations. In terms of the earmarked revenues, endogeneity should be less problematic since special financings themselves are fairly stable and the variations within existing ones are largely driven by external factors that are outside the manipulable scope of the cantonal government. Nevertheless, the fourth section introduces additional control variables to make the ceteris paribus assumption more consistent. After all, if there is some endogeneity present, the applied estimators cannot solve that issue.

Based on the methodological foundations and the data, the estimations in section five provide partly unexpected results. The cultural domain mainly supports the debate prevention theory. In particular, well-funded special financings dedicated to cultural public goods enforce the negative effect earmarking has on efficiency. At the same time, the empirical evidence suggests that the effect of earmarks can also turn positive if a canton has a restrictive debt brake. Including the findings of the transportation domain show an even more differentiated picture. While indebted special financings also allow some debate prevention (i.e. earmarking indeed negatively affects efficiency), well-funded special financings have the opposite effect. The most plausible explanation for this phenomenon are the inherent differences between the two domains. First, in the transportation domain the tax-payers and beneficiaries largely cohere (mainly road utilizers), but these two groups differ significantly in the culture domain. Second, the correlation between the taxed and the benefiting good is only strong in the transportation domain and not in culture. Third, the transportation domain has the potential to exclude those who do not pay but this is less the case for culture. These differences correspond exactly to those identified in the literature review. The public choice veterans predicted a positive effect of earmarking on efficiency if these conditions are fulfilled. The debate prevention now also explains why earmarking has a particularly negative effect on efficiency when those conditions are absent.

As it happens, the effect of earmarked revenue on efficiency can depend on other factors, which is why the sixth section outlines some conditional effects. The robustness checks and an interim conclusion round out the chapter.

7.1. Quasi fixed effects algorithm

There is a considerable amount of papers in which the authors regressed the efficiency measures from the first stage based on some external or environmental variables in the second stage.[68] Recent

[68] For an overview see Simar & Wilson (2007).

studies applying such a two-stage approach specifically in the public sector are Afonso & Fernandes (2008), Boetti et al. (2012) or Bönisch et al. (2011); Narbón-Perpiñá & De Witte (2017) provide a comprehensive review. Most of these papers account in some way for a certain specificity of the DEA measure. Because it is bound between zero and one in the input-oriented case or between zero and infinity in the output-oriented case, the measure and thus the corresponding error term inherently exclude a normal distribution. When the underlying data-generating process is not specifically discussed, there are two basic approaches authors often use to consider the boundedness of the error term. First, some authors argue that the underlying efficiency measure as a dependent variable is indeed normally distributed, but only values within the bounds exist and are observable (e.g. Kalb, 2010). In this case, statistical theory proposes truncated models. Second, other researchers believe in the censoring of the true error term in the sense that values above or below a certain threshold take the threshold value (e.g. Kirjavainen & Loikkanent, 1998). These researchers apply a censored estimator as, for instance, the Tobit model (Honoré 1992).

As Simar & Wilson (2007) and Simar & Wilson (2011b) point out however, all of these studies share two common shortcomings. First, none of these authors describe a coherent data-generating process that would explain the non-normal distribution of the error term. Consequently, the use of censoring and truncation methods as well as whatever transformations of the efficiency score are ad hoc makeshift solutions rather than theoretically grounded ones. Second, neither censored nor simple truncation models yield estimates of standard errors that would allow for statistical inference. The reason is that these methods, by default, do not take into account the correlation between the efficiency scores. Correlation is most likely because in finite samples every efficient observation expands the frontier and thereby alters the efficiency estimates of all or most of the other observations. Unfortunately, because very little is known about the theoretical distribution of efficiency scores obtained through DEA, the correlation cannot be simply integrated out of the data-generating process.

Acknowledging the correlation as well as the non-normally distributed error term, Balaguer-Coll et al. (2007) chose to use smoothing techniques. Although such techniques facilitate graphical illustrations of the effects observed in the sample, the authors did not report any generalization in terms of statistical inference. A second approach to overcome the issues of the error term comes from Banker & Natarajan (2008). They impose fairly strict assumptions on the data-generating process to enable simple OLS estimation. However, the narrow framework in which these assumptions hold, almost prevents the use of the model in reality (Simar & Wilson 2011b).[69] Yet another idea is to use bootstrap estimation to conduct inference (Simar & Wilson 2007; Xue & Harker 1999; Hirschberg & Lloyd 2002).

The idea of bootstrapping, here in particular the parametric bootstrap with resampling residuals, entails the following steps (Greene 2003). The first is to estimate the wanted coefficients out of the sample at hand together with an estimate of the variance of the error term. Then, some pseudo errors are drawn based on the assumed distribution of the error term and its estimated variance. Combining the pseudo errors with the coefficient estimates and the observed values of the independent variable leads to a pseudo sample of the dependent variable. Next, the pseudo sample

[69] Admittedly, Johnson & Kuosmanen (2012) later found that the OLS estimator is still consistent under less restrictive assumptions.

of the dependent and the original values of the independent variables yield a bootstrap estimate of the wanted coefficients. Repeating that procedure multiple times eventually reveals the sought distribution of the coefficients and enables inference. Drawing pseudo samples from the estimated coefficients mimics the underlying data-generating process (Simar & Wilson 2000). Accordingly, "in order to simulate the disturbances, we need either to know (or assume) the data-generating process that produces [the error term]. […] The obvious disadvantage of the parametric bootstrap is that one cannot learn of the influence of an unknown DGP for [the error term] by assuming it is known. For example, if the bootstrap is being used to accommodate unknown heteroscedasticity in the model, a parametric bootstrap that assumes homoscedasticity would defeat the purpose" (Greene 2003, p.652).

The description of the data-generating process given here intends to provide a rough overview, whereas Simar & Wilson (2007) and Simar & Wilson (2011b) treat it in more detail. With reference to section 6.2, still let $\boldsymbol{x}_i \in \mathbb{R}_+^p$ be the ith row of the input matrix $\boldsymbol{X}$ with p inputs, and let $\boldsymbol{y}_i \in \mathbb{R}_+^q$ denote the ith row of the output matrix $\boldsymbol{Y}$ with q outputs. Now introduce a matrix $\boldsymbol{Z}$ whose ith row $\boldsymbol{z}_i \in \mathbb{R}^m$contains m environmental variables. Assume that these sample observations ($\boldsymbol{x}_i$, $\boldsymbol{y}_i$, $\boldsymbol{z}_i$) are realizations of identically, independently-distributed random variables ($\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$) with probability density function $f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Under the condition that the production frontier is fixed, $\boldsymbol{x}$ can be written in terms of polar coordinates by $(\boldsymbol{\eta},\theta)$, where $\boldsymbol{\eta} = [\eta_1, \dots, \eta_{q-1}]$ is a vector of angles and θ is the input efficiency measure.[70] Using polar coordinates, now write the probability density function as $f(\boldsymbol{\eta},\theta,\boldsymbol{y},\boldsymbol{z})$ which is a series of conditional densities that describe the data-generating process in three steps (Simar & Wilson 2007):

$$f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = f(\boldsymbol{\eta},\theta,\boldsymbol{y},\boldsymbol{z}) \;= f(\boldsymbol{z})f(\theta|\,\boldsymbol{z})f(\boldsymbol{y},\boldsymbol{\eta}|\theta,\boldsymbol{z}) \tag{54}$$

The interpretation is straightforward. Starting from the left of equation (54), a decision-making unit operates in the environment $\boldsymbol{z}$ that is distributed as $f(\boldsymbol{z})$. Concretely z_i could be, for instance, the normally distributed topographic conditions of a canton i. Second, conditional on this environment $\boldsymbol{z}$ the decision-making unit draws an input efficiency level θ from the conditional probability density $f(\theta|\,\boldsymbol{z})$. Again, to be concrete, a canton i might have a higher or lower probability to be efficient depending on the topographic conditions. Third, the decision-making unit draws $\boldsymbol{y},\boldsymbol{\eta}$ from $f(\boldsymbol{y},\boldsymbol{\eta}|\theta,\boldsymbol{z})$. To conclude the example, imagine that the canton can achieve a certain input-output combination based on θ and $\boldsymbol{z}$. Note however, that the third density is hard to interpret due to the partitioned output. The take-home message here is that the unadjusted inputs $\boldsymbol{\eta}$ and the outputs $\boldsymbol{y}$ are simultaneously achieved from the same distribution and both do not directly depend on the environment variables $\boldsymbol{z}$. Put differently, the environmental variable does not affect the efficient frontier itself, but the probability of a certain decision-making unit to be closer or further away from the frontier (i.e. to be more or less efficient).[71] This condition is the reason why the inputs $\boldsymbol{x}$ were rewritten in its polar coordinates $(\boldsymbol{\eta},\theta)$ in the first place. The second

[70] The original model bases on output-oriented efficiency measures δ that are equal or bigger than one. As demonstrated in section 6.2, in the CCR model the input-oriented efficiency measure θ is simply the inverse of the output oriented one. In the BCC model this simple transformation does not apply.
[71] Daraio et al. (2010) propose how to test this so-called separability condition.

conditional density $f(\theta|\,\boldsymbol{z})$ is the crucial step because it operates by assumption through a specific mechanism:

$$0 \leq \theta = \psi(\boldsymbol{z}, \boldsymbol{\beta}) + \varepsilon \leq 1 \qquad \text{where } \varepsilon \overset{iid}{\backsim} N(0, \sigma_\varepsilon^2) \text{ which is truncated left at } -\psi(\boldsymbol{z}, \boldsymbol{\beta}) \text{ and right at } 1 - \psi(\boldsymbol{z}, \boldsymbol{\beta}) \tag{55}$$

ψ is a smooth, continuous function, $\boldsymbol{\beta}$ is a vector of parameters, and ε is the error term. Additionally, henceforth $\psi(\boldsymbol{z}, \boldsymbol{\beta}) = \boldsymbol{z}'\boldsymbol{\beta}$ by assumption. Remember the previous distinction between censored and truncated models. On the first sight, it seemed puzzling to assume that in fact only values within certain bounds of efficiency exist as it is the case for truncation models. Based on equation (55) it becomes clear why Simar & Wilson (2007) advocate for truncated regression in the second stage. They suppose that the environmental variable influences efficiency, which then in turn creates values of inputs and outputs. Since the inefficiency itself has a truncated distribution by construction, the environmental variables $\boldsymbol{z}$ cannot load into the efficiency measure over their entire range.

Assuming the data-generating process in (54) together with its specification in (55) the adequate bootstrap procedure runs through the subsequent steps (Simar & Wilson 2007):

[1] Compute the estimated bias-corrected efficiency scores $\hat{\hat{\theta}}_i$ using (43).

[2] Use the maximum likelihood method to obtain the estimates $\widehat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\varepsilon^2$ from the truncated regression of $\boldsymbol{z}_i$ on $\hat{\hat{\theta}}_i$. The truncation boundaries are 0 and 1.

[3] Repeating the following steps [3.1 to 3.3] L times, generates L bootstrap samples ($\hat{\theta}_i^*$) and accordingly L bootstrap estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$:

[3.1] Draw n different ε_i^* from a truncated normal distribution with zero mean and the variance $\hat{\sigma}_\varepsilon^2$ estimated in [2]. The truncation on the left is $-\boldsymbol{z}_i'\widehat{\boldsymbol{\beta}}$ and on the right $1 - \boldsymbol{z}_i'\widehat{\boldsymbol{\beta}}$.

[3.2] Compute the bootstrap estimates of the efficiency scores by $\hat{\hat{\theta}}_i^* = \boldsymbol{z}_i'\widehat{\boldsymbol{\beta}} + \varepsilon_i^*$

[3.3] Use the maximum likelihood method to obtain the estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$ from the truncated regression of $\hat{\hat{\theta}}_i^*$ on $\boldsymbol{z}_i$.

[4] The L resulting bootstrap estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$ provide the necessary information to construct estimates of confidence intervals for every element in $\boldsymbol{\beta}$.

Having introduced the Simar-Wilson algorithm, an interim conclusion is appropriate. The estimator this section ultimately looks for needs to be ready for three data characteristics: First, the DEA score is truncated. Second, the error term is correlated in an unknown way. Third, the data comes in a panel format containing most likely unobserved individual- and time-invariant effects. The Simar-Wilson algorithm is the current standard used to accomplish the two-step approach with cross sectional data as it simultaneously accounts for the truncation and the correlation. However, the literature has so far left aside the aspect of potential fixed effects in this regard. Therefore, it is worth checking how the empirical literature treats longitudinal data characteristics in the spirit of two steps, without necessarily considering the other two data characteristics.

The common practice to implement panel data in the two-step approach is to run first a DEA over the entire observation set including all individuals and all time periods. In section 6.2 such a reference set was called intertemporal. The second step then either simply ignores panel effects when explaining the efficiency scores, as in Dimas et al. (2012), or uses some types of fixed and random effect models. Examples advocating for the latter estimation strategies are Selim & Bursalıoğlu (2015), Fleishman et al. (2009) or Chen et al. (2005). Probably the least sophisticated papers in this family come from Kittelsen & Rehnberg (2008) and Poveda (2011) who simply used OLS, linear fixed, and random effects models. They thereby leave aside not only the correlation between, but also the truncation of the efficiency scores.

Still based on intertemporal efficiency scores in the first stage, several authors used slightly more elaborate models in the second stage. For instance, Staub et al. (2010) and Souza & Gomes (2015) used moment conditions to estimate some dynamic panel data models which allowed for dynamic responses of the efficiency scores and autoregressive errors. Alternatively, if a policy change as a one-time event figures as a subject of main interest, the difference-in-difference as a quasi-experimental design has lately become very popular in the most diverse fields of social sciences. Consequently this design has also found its way into the explanation of efficiency scores (Tiemann & Schreyögg 2012).[72] Similarly, Puenpatom & Rosenman (2008) include interactions with the control and dummy variables for those years when some policy program was implemented.

Running DEA over the entire time span for all individuals compares production processes of different years. The implicit assumption is that there is no technology change or innovation and thus the efficient frontier remains equal over time (see section 6.2). Very few papers challenge this strict assumption by using contemporaneous efficiency scores (Tiemann & Schreyögg 2012; Wanke 2012) or the results from a window analysis (Zhang et al. 2008).

With the seminal paper by Simar & Wilson (2007), the empirical two-stage DEA literature increasingly followed them by applying truncated regression and bootstrapping also with longitudinal data. However, now that these two characteristics came into focus, some authors leave other specificities of longitudinal data aside. Barros et al. (2010) applied the cross sectional truncated regression with bootstrapped standard errors and thereby pooled the data. In their subsequent studies (Barros & Dieke 2008a; Barros & Dieke 2008b), they introduced a trend variable without any further consideration of a possible correlation due to the panel structure. The authors did not mention their implicit and rather strict assumption of a constant trend in efficiency. In contrast, Kalb (2010), Chortareas et al. (2013), and Agasisti & Wolszczak-Derlacz (2014) insert dummies while running the Simar-Wilson algorithm in order to account for fixed effects. Note that all three base their regressions on intertemporal efficiency scores.

The studies referenced above illustrate various possibilities for accounting for panel data in the two-step procedure. Up to now very few papers have simultaneously incorporated all three characteristics. Among them are those of Kalb (2010), Chortareas et al. (2013), and Agasisti & Wolszczak-Derlacz (2014) who all estimate DEA scores using the intertemporal reference set in the first step. They thereby implicitly assume that the production frontier is constant (Tulkens & Eeckenhaut 2006). As already mentioned, this assumption is stronger the longer the time period

[72] Note that Tiemann & Schreyögg (2012) actually use contemporaneous efficiency scores as indicated below.

analyzed. Assuming a constant frontier suits the second step, where all three studies run the Simar-Wilson bootstrap procedure to consider the unknown correlation between the efficiency scores as well as the truncated error. Following the previously explained data-generating process in equation (54), the second step needs the fix frontier assumption, because of the separability condition. In order to write the inputs $\boldsymbol{x}$ by its polar coordinates $(\boldsymbol{\eta}, \theta)$, the production frontier has to remain fixed (Simar & Wilson 2007). By including time dummies among the regressors in the second step, one inevitably questions the assumption of a fixed frontier, which actually also prevents writing the inputs by its polar coordinates. Hence, even if the idea of combining fixed effects with the Simar-Wilson algorithm is striking, the implementation suggested so far does not satisfy.[73]

A further problem that appears when using a fixed effects model in the second stage lies in the truncation. In linear fixed effects models, it is well known that consistent coefficient estimates are the result of inserting individual dummies. This least squares dummy variables (LSDV) estimator controls for individual fixed effects. Indeed, the number of dummies, and therefore the coefficients to be estimated, increases with the number of individuals – an issue known as the incidental parameter problem. Consequently, the coefficients are not necessarily consistent (i.e. possibly asymptotically biased) if the number of years is kept fixed. When applying the OLS framework, the estimates resulting from the LSDV approach equal those of the regression with de-meaned or first difference transformed variables. The OLS estimates from this transformed data can be shown to be consistent and, thus, so are those of the LSDV estimator (Greene 2003). Now, remember that truncation leads to non-normally distributed error terms. Due to the non-linearity, estimating a fixed effect model by subtracting individual means or by taking first differences is not possible anymore. Put differently, since the efficiency scores underlie a truncation process, deviations of group means or first differences are meaningless (Greene 2008). Kalb (2010), Chortareas et al. (2013) and Agasisti & Wolszczak-Derlacz (2014) unintentionally avoid this issue by using dummies instead while possibly suffering from the incidental parameter problem.

The challenge therefore is to find the middle course between relaxing the assumption of a fixed frontier and accounting for individual and time fixed effects. The trick is to combine the data transformation and the dummy variable strategy, where the consideration of the appropriate reference set is the quasi-transformation and removes the year fixed effect. The dummies tackle the individual fixed effect. For further reference, this is called the quasi fixed effect algorithm.

Concretely, in order to compare only decision-making units that could possibly operate on the same efficient frontier, the DEA incorporates only observations of one year in the first step. In terms of section 6.2, a year-per-year DEA uses the contemporaneous reference set, which automatically excludes unobservable shocks (i.e. fixed effects) that shift the production frontier in one year for all DMUs.

Implementing the dummy variables strategy in the second step controls for individual fixed effects. However, as explained above, the dummy approach provokes the incidental parameter issue. Greene (2008, p.117 et seqq.) points out, though, that the bias caused by the incidental parameters differs from one model to another: "Until recently, analysis of this sort was limited to

[73] Note that Agasisti & Wolszczak-Derlacz (2014) also estimated a model where they allowed varying frontiers between but not within individuals, i.e. they run a DEA over one individual and several years. However, as they included year dummies, their approach does not make more sense than Kalb's (2010).

binary choice models, but it was, by and large, taken as a given […] that similar results apply to other models. In fact, this appears not to be the case. […] The end result would seem to be that estimating fixed effects models with censoring and truncation presents no practical obstacle. The incidental parameters problem is to be reckoned with, but if the Monte Carlo results given here have any generality, then the IP problem in this setting is far less severe than in the binary choice case." As a consequence, the dummy variables strategy to control for individual fixed effects can be used in the present case without too much doubt. Therefore, the second stage of the truncated quasi fixed effects estimator with bootstrap inference is simply the Simar-Wilson algorithm with individual dummies.

Analytically the underlying assumed data-generating process enriched by time and individual fixed effects is

$$\begin{aligned} f(\boldsymbol{x}_{it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it}, \alpha_i, \tau_t) &= f(\boldsymbol{\eta}_{it}, \theta_{it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it}, \alpha_i, \tau_t) \\ &= f(\boldsymbol{z}_{it}) f(\alpha_i | \, \boldsymbol{z}_{it}) f(\tau_t | \, \boldsymbol{z}_{it}) f(\theta_{it} | \, \boldsymbol{z}_{it}, \alpha_i) f(\boldsymbol{y}_{it}, \boldsymbol{\eta}_{it} | \theta_{it}, \boldsymbol{z}_{it}, \alpha_i, \tau_t) \end{aligned} \tag{56}$$

where α_i are individual fixed effects and τ_t are time fixed effects. Note that fixed effects estimators even allow a correlation between the regressors $\boldsymbol{z}_{it}$ and the fixed effects while still providing unbiased coefficients. Again, the interesting conditional density function is $f(\theta | \, \boldsymbol{z}_{it}, \alpha_i)$ because it operates by assumption through a specific mechanism:

$$0 \leq \theta_{it} = \boldsymbol{z}_{it}{}'\boldsymbol{\beta} + \varepsilon_{it} + \alpha_i \leq 1 \qquad \begin{array}{ll} \text{where } \varepsilon_{it} \overset{iid}{\backsim} & N(0, \sigma_\varepsilon^2) \text{ is truncated} \\ \text{left at} & -\boldsymbol{z}_{it}{}'\boldsymbol{\beta} - \alpha_i \\ \text{and right at} & 1 - \boldsymbol{z}_{it}{}'\boldsymbol{\beta} - \alpha_i \end{array} \tag{57}$$

Accordingly, the quasi fixed effect algorithm is

[1] Compute the bias-corrected efficiency scores $\hat{\hat{\theta}}_{it}$ applying (43) a contemporaneous reference set.

[2] Using the maximum likelihood method yields the estimates $\widehat{\boldsymbol{\beta}}$, $\hat{\sigma}_\varepsilon^2$ and α_i from the truncated regression of $\hat{\hat{\theta}}_{it}$ on $\boldsymbol{z}_{it}$ and individual dummies. The truncation boundaries are 0 and 1.

[3] Repeating the following steps [3.1 to 3.3] L times, generates L bootstrap samples ($\hat{\theta}_{it}^*$) and accordingly L bootstrap estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$:

[3.1] Draw n different ε_{it}^* from a truncated normal distribution with zero mean and the variance $\hat{\sigma}_\varepsilon^2$ estimated in step [3]. The truncation bound on the left is at $-\boldsymbol{z}_{it}{}'\widehat{\boldsymbol{\beta}} - \alpha_i$ and on the right at $1 - \boldsymbol{z}_{it}{}'\widehat{\boldsymbol{\beta}} - \alpha_i$.

[3.2] Compute the bootstrap estimates of the efficiency scores by $\hat{\hat{\theta}}_{it}^* = \boldsymbol{z}_{it}{}'\widehat{\boldsymbol{\beta}} + \alpha_i + \varepsilon_{it}^*$

[3.3] Use the maximum likelihood method to obtain the estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$ from the truncated regression of $\boldsymbol{z}_{it}$ on $\hat{\hat{\theta}}_{it}^*$.

[4] The distribution of the L resulting bootstrap estimates $\widehat{\boldsymbol{\beta}}^*$ and $\hat{\sigma}_\varepsilon^{2^*}$ provides the necessary information to construct confidence intervals for every coefficient in $\boldsymbol{\beta}$ (see Simar & Wilson, 2000).

The assumption about the data-generating process (57) together with the quasi fixed effects algorithm covers all three data characteristics requested in the beginning of this section.

7.2. Heteroscedasticity in Stochastic Frontier Analysis

Unlike the Simar-Wilson algorithm, which estimates the coefficients of the environmental variable in two steps, the stochastic frontier approach enables a one-step estimation. If done directly, the impact of the environmental variables on efficiency integrates more naturally into the model. Recall the two estimation equations (53) where the inefficiency u_{it} follows a half-normal distribution with constant variance σ_u^2. Introducing some environmental variables $\boldsymbol{z}_{it}$ challenges the constant variance (i.e. homogeneity) assumption and conditions the variance on contextual factors (Kumbhakar et al. 2015). In other words, for most econometric models heteroscedasticity leads to issues that must be resolved, but when estimating frontier models it achieves an economic meaning.

Recall Greene's (2005) true fixed effects model in section 6.3, which now incorporates the heteroscedastic inefficiency and heteroscedastic error. The thus obtained results compare to the ones from the two-step approach of the previous section, which also only accounts for time-variant efficiency. Formally, the heteroscedastic true fixed- and random-effect model estimating the actual cost functions read as[74]

$$\begin{aligned}
&\text{CD: } \ln C_{it}^{real} = \alpha_i + \boldsymbol{\beta}' \ln \boldsymbol{y}_{it} + u_{it} + v_{it} \\
&\text{TL: } \ln C_{it}^{real} = \alpha_i + \boldsymbol{\beta}' \ln \boldsymbol{y}_{it} + \tfrac{1}{2}\textstyle\sum_{r=1}^{q}\sum_{s=1}^{q} \gamma_r \ln y_{itr} \ln y_{its} + u_{it} + v_{it} \\
&\quad \text{where } \mathbb{E}[\alpha_i | u_{it}, v_{it}] = 0 \text{ [random effect] or } \mathbb{E}[\alpha_i | u_{it}, v_{it}] \neq 0 \text{ [fixed effect]} \qquad (58)\\
&\qquad u_{it} \backsim N^+(0, \sigma_{u,it}^2) \text{ and } \sigma_{u,it}^2 = e^{\boldsymbol{\delta}' \boldsymbol{z}_{it}} \\
&\qquad v_{it} \backsim N(0, \sigma_v^2) \text{ and } \sigma_v^2 = e^{\omega_0}
\end{aligned}$$

Again, CD is the actual cost function with Cobb-Douglas specification, whereas the term TL stands for the translog specification. In contrast to the homoscedastic version, the time-variant inefficiency u_{it} is not identically distributed anymore by the very definition of the term. Yet, it is still independently distributed conditional on $\boldsymbol{z}_{it}$ meaning that one materialization of u_{it} given $\boldsymbol{z}_{it}$ is independent from all other values. The interpretation of the coefficients of the environmental variables ($\boldsymbol{\delta}$) is not intuitive. As they appear in a non-linear fashion in the estimation equation, a direct interpretation would mislead. Instead, Wang (2002) provides the necessary modifications to obtain their marginal effects on the inefficiency. He also shows that the effects are monotonic in the model (58). In other words, even if the coefficient per se might not be informative without modifications, its sign is. Naturally, the effect of interacted variables is non-monotonic. Finally, it is crucial to recognize that the coefficients link the environmental variables with the *in*efficiency u_{it}. A coefficient above zero therefore means that the respective variable negatively affects efficiency.

7.3. Reverse causality

Irrespective of the method used to estimate efficiency, endogeneity issues might affect the estimates. That is, in order to actually claim that the environmental variables *cause* changes in

[74] The adaptations translate one-to-one into the estimation equation of the distance functions.

efficiency – rather than they are simply correlated – it is essential that the revenue characteristics are exogenous. Namely, potential endogeneity issues of the revenue characteristics may arise from reverse causality and unobserved factors that simultaneously affect efficiency and the revenue characteristics. At this point, the estimators accounting for this type of endogeneity are still in a rather early stage of development and, particularly in the empirical literature, they are not yet broadly established (Amsler et al. 2016; Karakaplan & Kutlu 2017). Because of this, the endogeneity issues cannot be solved here. Discussing their possible influence on the results is therefore vital.

Several mechanisms may manifest under the label of reverse causality. In case of earmarked revenues, either the executive or the parliament could endogenously modify the part of earmarked revenues. The first case is ruled out by the legal prerequisite of special financings. As both the HAM1 and the HAM2 stipulate, the earmarking of revenues requires a statutory basis (Conference of cantonal finance directors 1981; Conference of cantonal finance directors 2008). As such, special financings need to pass the parliament, be it at their implementation, their modification, or their repeal. This means that the executive has scarcely any direct and immediate influence on the part of earmarked revenues.[75]

The second case where the parliament endogenously manipulates the earmarked share builds on the Leviathan argument (Brennan & Buchanan 1978). Imagine that in reality earmarks cause a higher efficiency. The parliament, being aware of this causal relationship, introduces new earmarked revenues if it detects an inefficient service provision in order to discipline the executive. Exploiting data from this world would lead to a negative coefficient. Yet, the negative correlation would merely indicate the legislator's reaction to the executive's behavior rather than the causal relationship between the earmarked revenues and efficiency. As a counterargument, note that from the parliament's point of view, the introduction of a new earmarking restriction only makes sense if it essentially increases the part of the thereby assigned expenditures. Accordingly, one would observe an active use of earmarking as a disciplining instrument. The data sets this type of endogeneity in the naught. The mean variation coefficient of the earmarked revenue share (about 40%) largely exceeds that of the special financings count (about 14%). Taking the low correlation coefficient (about 20%) of the two variables into account indicates that the parliament not only seldom adjusts the number of special financings, but also that these alterations hardly find expression in higher or lower earmarked revenue shares. As it happens, the variation of the earmarked revenue shares seems to come from external (and exogenous) sources such as a higher fuel demand (which oftentimes contributes to the street fund) or a hunting boom (whose earnings from the licenses usually accumulate in the hunting fund). There is no obvious reason why these demand shocks should be correlated with efficiency.

An endogenous manipulation of earmarks by the parliament might also result in a positive estimated coefficient between the earmark share and efficiency. If the parliament introduces a new earmark in reaction to sparse financial resources, then one possibly observes a positive correlation even if no such causal relationship exists. A domain struggling with too few resources arguably operates very productively (i.e. a high input-to-output ratio). Furnishing this domain with a special financing would seemingly show a positive causal relationship between earmarks and efficiency,

[75] In this regard, also remember the discussion about the externality requirement in section 5.1 to consider an account as being earmarked.

even if it is not there. Here, again, such a misestimation cannot be fully excluded, while the above argument also reduces the problem somewhat. Since the parliament seldom introduces or repeals a special financing, the major part of the variation of the earmarked revenue share is due to external incidents.

The tax system likewise potentially imposes a misdirection in the form of endogeneity. Oates (1985) warns, for instance, that the government might spread its tax sources in order to expand its expenditures. That is, in more general terms, the government actively exploits the fiscal illusion in order to increase the expenditures while holding the outputs on the same level. By now it is clear that this kind of behavior leads to inefficiency. Starting with the tax structure as a factor of fiscal illusion, the legal prerequisites from above also apply here. Any alterations of the tax base (referring to tax complexity and diversification) as well as of the tax rate (referring to the progressivity) must pass the parliament. These checks and balances principally prevent the executive from taking advantage of the tax configuration as a form of fiscal illusion. The question remains however, as to whether the parliament could instrumentalize the tax law to fight inefficient behavior of the executive. If so, one would observe a negative correlation between the number of tax sources, the tax complexity, and the progressivity on the one side, and the efficiency on the other side, even if the true underlying causal relationship is positive. The empirical literature so far lacks evidence for such an instrumentalizing parliamentary strategy. Indeed, the results by Feld & Reulier (2009) suggest that the parliament does set the tax structure strategically, namely in reaction to the surrounding cantons. Because of the relatively distinct tax competition among the Swiss cantons, the parliament has to account for other jurisdictions' tax rates in order to attract tax payers. Admittedly, this competition still leaves some room to bring the tax law into line with other strategies, but it severely limits the scope.

Transfers can likewise trigger the endogeneity problem. The government has a strong incentive to seize every centime of grants as a funding source for discretionary slacks (see section 3.3). In doing so, higher transfers are automatically associated with an inefficient government even if the underlying causal mechanism does not exist (Bischoff et al. 2017). When the federal government became aware of the possible misconduct, it reformed the fiscal equalization scheme between the central state and the cantons. The new equalization scheme bows out of an output-compensating system and builds on the idea of adjusting financial potentials as well as compensating burdens (Swiss Federal Council 2001). It is in the very nature of the new equalization scheme to prevent the cantonal governments from being able to directly manipulate the incoming grants. At least since 2008, the new system ensures a certain externality of the total transfers, to which those of the equalization scheme (item 462 HAM2) contributed in 2014 around 13 percent. Another 17 percent stemmed from the unswayable shares of the federal tax, of other revenues from the central state and of public owned companies (item 460 HAM2). Operating contributions of other entities (item 463 HAM2) represent the largest part of the transfers (around 46% in 2014) (Federal Finance Administration 2016), which the cantonal governments potentially influence to some part. Therefore, the exogeneity of the transfers is mixed and the engendered bias is positive (i.e. an overestimation is possible).

Finally, endogeneity can arise from the debt as a last element of fiscal illusion. The government can be tempted to increase the debt in order to increase its discretionary slack (see section 3.4).

This behavior would indicate a seemingly negative relationship between the debt and efficiency even if there is truly no such causality. However, the debt brake most of the cantons have implemented limits the government's ability to arbitrarily increase the debt. These fiscal rules sometimes allow for additional debt in case of economic downturns. At the same time, most of them require the amortization of outstanding debt in the medium or long term (Yerly 2013). Together, the two principles generate a certain exogeneity of the debt variation.

The arguments discussed above also somewhat address potential endogeneity stemming from unobserved external variables. For instance, if it is convincing that the earmarked revenue share is mainly externally driven, then it is immaterial whether the endogeneity would originate from reverse causality or unobserved variables. In either case endogeneity is abolished. Surely, it would be more convenient to have a persuasive instrument at hand which captures the endogenous part of the variables at stake. But to date, no method has been developed that could deal with endogeneity of environmental variables by incorporating instruments.

7.4. Environmental variables

Although chapter 3 pointed to the importance of certain revenue characteristics to explain the efficiency of the public service provision, other variables might be equally important. Holding these variables fixed is necessary in order to interpret the coefficients of the revenue characteristics as causal effects rather than sheer correlations (for a broader discussion of endogeneity see section 7.3). This section thus aims to identify the relevant control variables, to review previous findings in the literature with respect to the effect of these variables, and to give an overview of their manifestation in the Swiss cantons. Table 19 lists variables with the descriptive statistics and the source. As the bottom of the table states, the panel is unbalanced, which obviously impacts the variables' moments as well as their minimum and maximum values.

Table 19 – Descriptive statistics of the environmental variables

Variable	Mean	Standard deviation	Minimum	Maximum	Data source [a]
Nom. GDP per capita (in 1'000 CHF)	62.311	20.615	36.196	163.457	BAK Basel Economics
Population density (1'000 inh. per km²)	0.326	0.509	0.032	5.107	federal statistical office
Fragmentation (1'000 inh. per municipality)	4.139	5.440	0.829	63.319	federal statistical office
Foreigners (in percent of total inhabitants)	18.685	6.920	7.832	40.984	federal statistical office
Referendum (mark from 1 [hard] to 6 [easy])	3.870	0.960	1.000	6.000	mainly federal statistical office
Initiative (mark from 1 [hard] to 6 [easy])	4.692	1.138	2.333	6.000	mainly federal statistical office
Voter turnout (in percent)	46.188	7.056	27.431	69.262	federal statistical office
Bourgeois parties in executive (in percent)	42.572	22.027	0.000	100.000	federal statistical office
National street length (in meters per km²)	54.500	30.337	0.000	256.272	federal statistical office

[a] All data are subject to own calculations based on the data source indicated in the last column

Notes: Without missing values there should be 390 observations (26 cantons and 15 years); the true N = 301; missing cantons (and years) are BS (if year < 2013), VD (all years), TG (all years), LU (all years), GR (all years), GL (all years), GE (2014); the N can be slightly smaller in some regressions in section 7.5 due to excluded outliers.

The theoretical argument connecting the GDP per capita to the explaining variables relates to the progressivity of the tax system (Oates 1985). As argued in section 3.4, according to the fiscal illusion theory people err in perceiving the true marginal tax bill if the increase stems from a larger income rather than a rise of the tax rate. Accordingly, only an increase in income paired with a progressive tax system negatively affects efficiency. To approach the respective effect, the estimation equation has to include an interaction term of the progressivity variable and income. The parameter of interest is the marginal effect of the income. Presumably due to the deficiency of an adequate elasticity measure, the empirical research disregarded the interaction term so far and only focused on the income alone (Dollery & Worthington 1996).[76] Like other studies, this analysis lacks comprehensive panel data on income and has to draw on the GDP instead. When introducing the interaction term, the GDP usefully has to be in nominal terms, because the cantons compensate the cold progression through an adjustment of the tax rate (Eidgenössische Steuerverwaltung 2015). Recent studies repeatedly found a positive income or GDP elasticity with respect to government expenditures (Mitias & Turnbull 2001; Deller & Maher 2005). Papers on efficiency often use the GDP as an output rather than as an explaining variable (e.g. Adam et al., 2008). Indeed, by operationalizing the concept with the income variable, De Borger & Kerstens (1996), Widmer & Zweifel (2012) and Seifert & Nieswand (2014) report a negative effect on efficiency.

Across all observations, the nominal GDP averages 62'311 Swiss Francs per capita. The smallest value in the sample (35'196 CHF per capita) comes from the canton of Appenzell Innerrhoden in 2000, whereas the canton of Basel Stadt tops the list (163'457 CHF per capita) in 2014. Given that the values are in nominal terms, a within comparison has little meaning. Based on the full sample (i.e. N=390), the Canton of Geneva has the highest mean value over the years (32'139 CHF per capita) followed by Basel Stadt (26'024 CHF per capita) and Vaud (10'641 CHF per capita). On the lower end are the canton of Appenzell Innerrhoden (1'890 CHF per capita), Schwyz (2'178 CHF per capita) and Appenzell Ausserrhoden (2'681 CHF per capita). Ergo, the between-variation is considerable.

The population density approximates the extent of the urbanity or rurality of a canton. Cities potentially benefit from agglomeration economies such that a higher density leads to higher efficiency. At the same time, cities might bear greater burdens with respect to social welfare, culture, et cetera which hampers an efficient provision of public services. Accordingly, the theory has no clear forecast on the influence of population density on efficiency. Empirical findings support the former hypothesis in German municipalities (Geys et al. 2010; Bischoff et al. 2017; Kalb 2010b), as well as Portuguese (Afonso & Fernandes 2008), and Belgian municipalities (De Borger & Kerstens 1996). There is also a contradictory and more recent study about Portuguese municipalities, however(Da Cruz & Marques 2014). Finally, no evidence is found for the OECD countries (Adam et al. 2011).

Statistics describe Switzerland as a sparsely densely populated country. The cantonal population densities vary from 32 (canton of Uri in 2007) to 5'107 (canton of Basel Stadt in 2014) with an average of 326 inhabitants per kilometer squared. Assorting the variable by its values reveals that only about half of the cantons change their rank from one year to another. This comes from a poor within-variation compared to the between-variation. The cantons of Nidwalden, Schwyz, Fribourg, and Bern are clustered round the median of about 200 inhabitants per kilometer squared.

[76] Under careful consideration of the theory, regressing the log-expenditures on the log-income only tests Wagner's Law, that is whether the income elasticity is larger than unity (Rodden 2003). Concluding the existence of a fiscal illusion from a high elasticity is premature.

The institutional fragmentation divides the cantonal inhabitants by the number of municipalities within the canton. It reveals the possible situation that a highly fragmented canton diversifies the central burden among different geographical areas, thus granting relief to the canton from costly interventions (see Grossman et al. (1999) for a similar argument for U.S. cities). This argument associates a higher fragmentation with an efficient canton. On the contrary, including many different municipalities also means a rag rug of different rules, systems et cetera, which complicates cantonal tasks (Kellermann 2007). The estimates the literature provides are not directly adaptable to the present case. Grossman et al. (1999) reports a positive effect of an equally distributed population among cities on the efficiency. Note though, that even a highly fragmented canton can still be focused on one center. Kirchgässner et al. (2005) find no influence of the fragmentation on the cantonal economic growth. Here again, the result does not automatically translate into an effect on efficiency.

A view on cantonal data shows a highly fragmented country with a decreasing trend; decreasing means a declining fragmentation and an ascending value of the variable. Unsurprisingly, the top cantons are the small ones with big cities; this is most importantly Basel Stadt with a very low fragmentation with 63'319 inhabitants per municipality in 2013 (Basel Stadt's data before 2013 are not part of the sample). On the other end the canton of Jura has a pronounced fragmentation with only 828 inhabitants per municipality in 2000. The canton of St. Gallen represents about the mean with approximately 6000 inhabitants per municipality.

The proportion of foreigners in a canton might become important if this group of people has a different perception of the price of public services (see section 3.4). Predicting the direction of the effect is difficult, as the resulting demand for public services can be larger or smaller compared to the average population. Some studies mention the possibility of including this factor as a further demographic control variable (e.g. Bönisch et al., 2011) without discussing the influence in detail and without actually estimating their effect.

The cantonal differences of the percentage of foreigners are marked and range from 7.8 percent in Uri (2001) to 41.0 percent in Geneva (2014). The average value is 18.68 percent with an increasing trend. Naturally, the geography of the area has a strong influence, which finds expression in higher percentages of foreigners in border cantons; after Geneva, these are notably Basel Stadt with 35.02% and Ticino with 27.91%, each in 2014. In general, francophone-dominated and urban cantons rank among the more popular destinations and they have accordingly higher proportions.

If there are low obstacles to participation in the democratic process, citizens face a stronger incentive to be better informed and to exercise their control function (Widmer & Zweifel 2012; Geys et al. 2010). Borge et al. (2008) argue that citizens who participate frequently in polls monitor politicians more intensively which brings about a higher level of efficiency. These obstacles are often operationalized using the index for financial initiatives and referenda as proposed by Frey & Stutzer (2000). Analyzing the Swiss cantons from 2000 to 2004, Widmer & Zweifel (2012) detect a negative effect on efficiency, which contradicts the theory.

By construction, the two indices can take values between six (low barrier) and one (high barrier) whereas only the referendum-index yields observations over its entire range (1: Geneva in 2013; 6: Schwyz and Appenzell Ausserrhoden, both over the entire time span). To launch an initiative, the index assigns the highest obstacles to the canton of Bern (2000-2007) and the best mark to the canton of Appenzell Ausserrhoden in 2001. The two measures correlate by 55 percent, whereas both have a similar standard deviation.

While the possibility of an initiative or a referendum incentivizes contributing to the political process, the voter turnout can also be measured directly, using data from national ballots. The expected effect is the same. Indeed, Borge et al. (2008), Geys et al. (2010), as well as Adam et al. (2011) find some evidence for a positive effect of voter turnout on efficiency.

At 41.83 percent, the distance from the lowest (Valais in 2003, 27.43%) to the highest (Schaffhausen, 2014, 69.26%) voter turnout reaches almost half of a cantonal electorate. The within (standard deviation 3.89) and the between variation (standard deviation 4.54) balance almost, which is a rare feature amongst the variables discussed here. The varying nationwide attractiveness of federal ballots is one reason why the within variation nearly approaches the between variation. In consequence, there is no trend from or towards a higher voter turnout.

Although the idea is nearly only based on intuition and not on a detailed theory, several studies assess the effect of political party affiliation on efficiency. The typical argument is that left-wing parties are less sensitive with respect to public expenditures. More precisely, the socialist ideology tends towards supporting wage claims of worker unions, which increases the costs (Brunner & Squires 2013). Also, their preference for public services tends to impede advocating for hard budget constraints (Borge et al. 2008). Kalb (2010b) objects to the latter point by rectifying that higher public expenditures do not automatically lead to a lower efficiency. Nevertheless, his results support the purported lower efficiency of left-wing parties. Also Borge et al. (2008) and Adam et al. (2011) find a – however not very robust – relationship between socialist parties and inefficiency. On the contrary, results published by Boetti et al. (2012) tie inefficient governing to bourgeois parties. The operationalization used here leans on the share of bourgeois parties in the executive.[77]

The statistics show that both extremes exist in the cantons. While two cantons (Appenzell Innerrhoden and Jura in multiple years) have no bourgeois representative among their ministers, the canton of Appenzell Innerrhoden's executive consisted exclusively of bourgeois party members in three consecutive years (2003 – 2005). The standard deviation of 22.03 is mainly driven by between-canton variation; the between standard deviation reaches 19.52 in contrast to the within standard deviation of only 1.47. The composition of the governing council is thus stable without any trend.

When it comes to specific cantonal tasks, it is useful to account for unique contextual factors. The construction and maintenance of roads is a task that is distributed among all state layers. Each layer, i.e. the municipal, the cantonal and the federal level, assumes a partial responsibility over the road network. While municipal roads cover only a limited area, cantonal and national roads connect larger regions. That is, national and cantonal roads substitute each other to some extent (Koch & Forster 2010). Therefore, if there is a long national road system within a canton, the cantonal road system has to absorb less of the heavy traffic and the cantonal expenditures per road kilometer shrink. The length of the national roads is measured in meters per kilometer squared of cantonal area.

[77] The following parties are considered bourgeois: *FDP.Die Liberalen*, *Liberale Partei Schweiz* (until their amalgamation with the *FDP* in 2008), *Schweizer Demokraten*, *Bürgerlich-Demokratische Partei Schweiz*, *Schweizerische Volkspartei*, *Lega dei Ticinesi*.

A surprising fact, when looking at the statistics of the national road length per canton is the minimum (0 meter per km^2). The two semi-cantons Appenzell Innerrhoden and Ausserrhoden manage their road network only with their municipalities; no federal road crosses through the two cantons. It is again the canton of Basel Stadt that figures on the top of the list with 256 road meters per square kilometer of cantonal area. Expectedly, the variable hardly varies within cantons (standard deviation within is 4.00 vs. 30.18 between). However, there is an increasing trend of building up the national road network.

7.5. Results

This section contrasts the regression coefficients of the two frameworks. Naturally, the question arises as to which results to trust more when they contradict each other. The discussion in section 6.5 on the efficiency estimates provides a clue. Particularly, if a low lambda value (i.e. a low signal-to-noise ratio) puts the DEA results into question, the doubt translates automatically into the second stage, i.e. the regression. The lambda's impact is less clear in the case of the SFA approach. The primary goal is to estimate the frontier and the estimation of the environmental coefficients relies on the stochastic properties of the estimated frontier (i.e. on the estimated variances of u_{it} and v_{it}). The efficiency itself accrues after the frontier estimation, that is, after the environmental variable took effect on the frontier. Consequently, if the coefficients' standard errors are small enough to reject the null hypothesis, the results might still be valid even with a low variation of the efficiency. Clearly, if both frameworks, i.e. the one-step and the two-step approach, produce coefficients with the same directions, this can be considered strong evidence that the respective relationship is indeed present in reality.

In order to give the regression on the DEA scores a meaning, Simar & Wilson (2007) assume that the support of inputs is independent of the environmental variables (see section 7.1). This separability condition impacts the interpretation of the coefficients associated with the environmental variables. It means that the only channel through which the environment affects the production process is the distribution of the efficiency scores (Bădin et al. 2014). Such an interpretation is utterly comparable to the environmental factors in the SFA framework, where the inefficiency follows by assumption an independent distribution. Thus, SFA also sees the only possibility in how the environmental variables cause the production process through the distribution of the inefficiency. In this aspect, the coefficients of the two frameworks measure the same effect.

The comparability of the SFA and DEA results becomes trickier when it comes to the very coefficients. Coefficients stemming from the quasi fixed effects algorithm can be interpreted as such without any more ado; of course, the interaction terms need a special treatment too. On the contrary and as briefly mentioned in section 7.2, the environmental variables in the SFA framework link to the inefficiency in a highly non-linear fashion. In fact, the way the variables enter the model produces a specific marginal effect for every observation depending on the values of all the other environmental variables (Wang 2002). In other words, the otherwise familiar *ceteris paribus* interpretation does not hold in the SFA. Even at the sample mean, marginal effects only have a limited validity. Graphical illustrations provide remedy here. A plot of the marginal effects against the relevant counter-variable proves much more insightful than simple considering coefficient estimates (Kumbhakar et al. 2015).

The different roles of efficiency and inefficiency is the final nuisance in comparing the coefficients of the two methods. As alluded repeatedly before, the coefficients stemming from the SFA approach measure the effect that environmental variables have on inefficiency. On the contrary, the coefficients of the quasi fixed effect algorithm indicate the environmental variables' effect on efficiency. Qualitatively, the coefficients can be interpreted simply vice versa. Quantitatively, a comparison of the coefficients is difficult. To facilitate the graphical interpretation, the illustrations provide marginal effects on efficiency for both, the one-step and the two-step approach.[78] Unfortunately, the formula to compute the marginal effects on efficiency can lead to missing values, something which particularly appears with the estimates of the transportation domain. Therefore, the respective graphs show marginal effects on inefficiency nevertheless. For the sake of coherence, the tables always show marginal effects on inefficiency as proposed by Wang (2002).

[78] Wang (2002) only provides the formula to estimate the marginal effects of the environmental variables on *in*efficiency. To make the results of the SFA approach more comparable to the two-step approach, appendix A. 3 develops the marginal effect on efficiency.

Table 20 – Influence of the revenue characteristics on efficiency

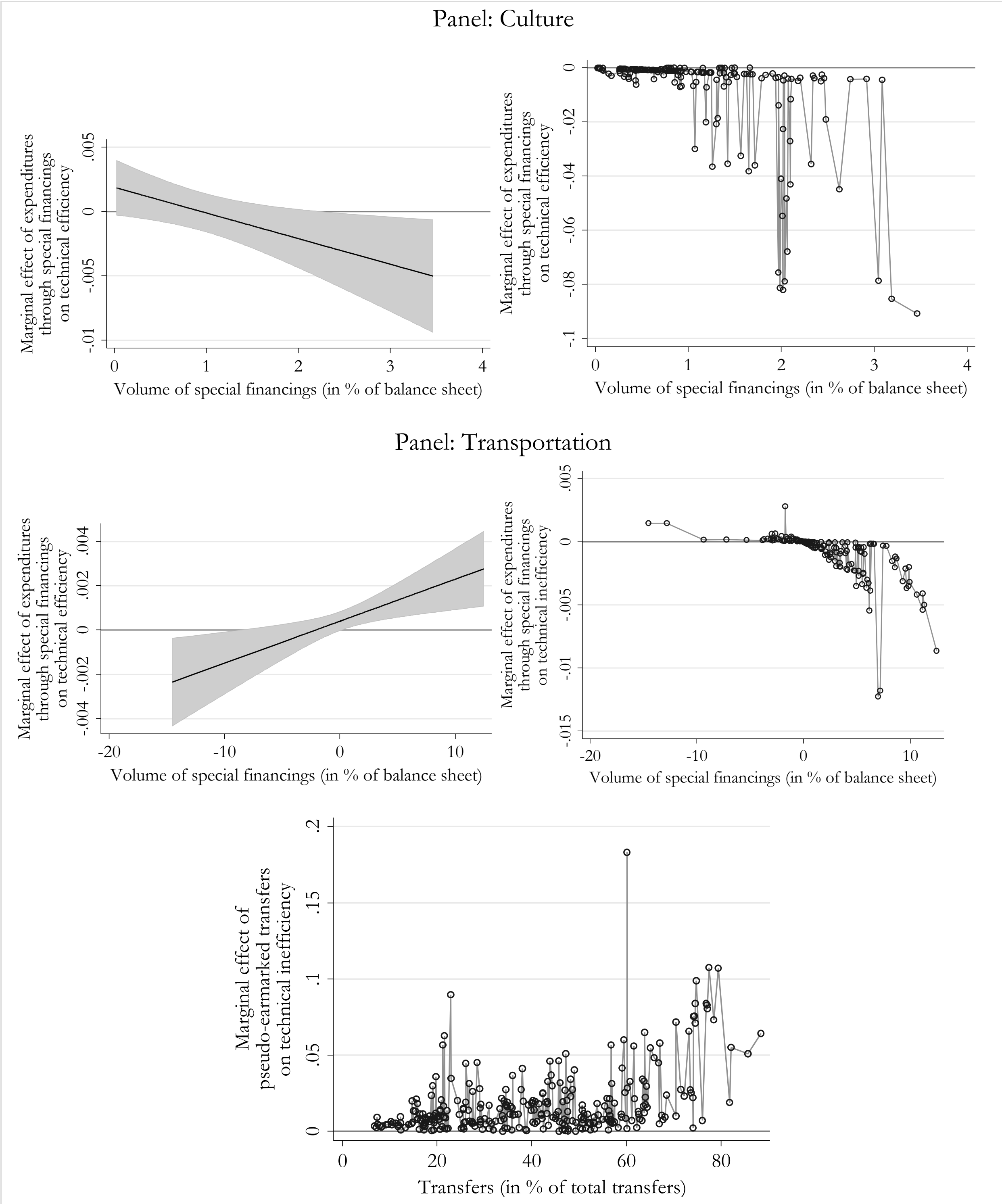
	Culture	Transportation	Culture	Transportation	Transportation
Delta coefficients of $\sigma_{u,it}^2 = e^{\boldsymbol{\delta}' \boldsymbol{z}_{it}}$			n.a.	n.a.	
[A] Expenditures through special financings (in % of total expenditures)	0.017 (0.068)	0.020 (0.018)	0.002* (0.001)	4.00E-04* (0.000)	
[B] Volume of special financings (in % of balance sheet's total)	0.515 (0.890)	0.349* (0.190)	0.034 (0.039)	0.001 (0.003)	
Interaction [A]x[B]	0.004 (0.032)	-0.003 (0.005)	-0.002** (0.001)	1.89E-04*** (0.000)	
[C] Tax progressivity (curvature parameter)	4.392*** (1.058)	0.856 (1.166)	0.004 (0.038)	-0.013 (0.021)	-1.195 (1.836)
[D] GDP per capita (in 1'000 CHF)	0.502*** (0.179)	0.029 (0.080)	-0.002 (0.004)	-0.004** (0.002)	-0.047 (0.132)
Interaction [C]x[D]	-0.057*** (0.02)	-0.025 (0.032)	0.001 (0.001)	0.001*** (0.000)	0.022 (0.032)
Tax complexity (normalized HHI)	0.043 (0.055)	-0.021 (0.077)	-0.002 (0.002)	2.57E-04 (0.001)	-0.143** (0.063)
Tax diversification (no. of tax sources)	-2.530** (0.997)	0.815** (0.345)	0.062** (0.022)	-0.013 (0.012)	-0.102 (1.058)
Pseudo-earmarked transfers (in % of total transfers)				-4.71E-04 (0.000)	-0.091*** (0.03)
Transfers (in % of total revenues)	-0.031** (0.015)	0.012 (0.027)	-0.001 (0.001)	0.008** (0.003)	-0.629*** (0.18)
Gross debt per capita (in 1'000 CHF)	0.502* (0.276)	0.264*** (0.086)	0.002 (0.005)	-0.599** (0.294)	0.006*** (0.002)
Population density (1'000 inh. per km²)	20.401* (12.098)	1.327 (7.142)	0.254 (0.462)	-0.013 (0.014)	0.148 (3.185)
Referendum (1 [hard] to 6 [easy])	0.022 (1.149)	-1.311 (0.927)	-0.048* (0.028)	0.033 (0.037)	-0.889 (0.732)
Initiative (1 [hard] to 6 [easy])	1.300* (0.785)	0.621 (0.565)	-0.139** (0.057)	0.001** (0.000)	2.143 (2.889)
Bourgeois parties in executive (in %)	-0.051 (0.054)	0.029 (0.020)	-0.001 (0.001)	0.003 (0.022)	0.022 (0.018)
Fragmentation (1'000 inh. per municipality)	-1.658* (0.967)	-0.005 (0.389)	0.008 (0.039)	-0.002** (0.001)	-0.141 (0 394)
Voter turnout (in %)	0.035 (0.023)	0.047 (0.074)	0.001 (0.001)	0.003 (0.005)	0.029 (0.057)
Foreigners (in % of total inhabitants)	-1.194** (0.511)	-0.066 (0.098)	0.028*** (0.009)	-0.001 (0.001)	-0.366 (0 485)
National street length (in km)		0.009 (0.018)		4.00E-04* (0.000)	0.019 (0.013)
Intercept (δ_0)	-15.61*** (5.95)	-13.22** (6.106)			7.056 (33.589)
Marginal effects on (in)efficiency					
Expenditures special financings	0.000 (n.a.)	0.000 (n.a.)	-0.000 (0.001)	0.001*** (0.000)	
GDP per capita	-0.012 (n.a.)	-0.015 (n.a.)	0.000 (0.002)	0.001*** (0.001)	0.010 (n.a.)
Tax complexity	0.001 (n.a.)	-0.002 (n.a.)	n.a.	n.a.	-0.012 (n.a.)
Tax diversification	-0.084 (n.a.)	0.089 (n.a.)	n.a.	n.a.	-0.009 (n.a.)
Pseudo-earmarked transfer share					0.019 (n.a.)
Transfers	-0.001 (n.a.)	0.001 (n.a.)	n.a.	n.a.	0.051 (n.a.)
Gross debt per capita	0.017 (n.a.)	0.029 (n.a.)	n.a.	n.a.	0.033 (n.a.)
N	214	298	214	298	298
Model	TRE	TRE	Quasi-FE algorithm	Quasi-FE algorithm	TRE

Notes: * p<10%, ** p<5%, *** p<1%

TRE models: Delta coefficients with robust standard errors clustered at cantonal level; the dependent variable is the logged personnel expenditures (full output reported in Table 32 in the appendix); a positive delta coefficient indicates a positive (negative) effect on inefficiency (efficiency); marginal effects reported as means without significance tests

Quasi FE models: Coefficients with bootstrap-based standard errors in parentheses (Simar & Wilson 2007); the dependent variable is the DEA technical efficiency score with a contemporaneous reference set (i.e. year-by-year DEA); individual dummies included but not reported; marginal effects at mean of interaction variable

Figure 16 – Marginal effects of expenditures through special financings



Notes: Negative values on the abscissa indicate indebted special financings; illustrations only cover range of sample
Source: Own illustration

As explained earlier (see section 5.4), the main text focusses the results of the two domains culture and transportation.[79] The upper part of Table 20 lists the pure coefficients of the environmental variables. The lower part shows the marginal effects of the revenue characteristics among the environmental variables (see Table 32 in the appendix for coefficients of distance

[79] The remaining results follow in Table 31 and Table 32 in the appendix.

function). Note that no standard errors and therefore no significance tests are estimated for the SFA marginal effects. Thus, the absence of stars is no indicator of an insignificant effect. Column-wise, the first, second, and fifth estimates rest upon the SFA approach, while column three and four show the results of the quasi-FE algorithm.

Looking at the expenditures through and the volume of special financings, the four models appear initially sobering. The two variables as well as their interaction term are always insignificant in the SFA case.[80] The quasi-FE algorithm only provides significant coefficients for the interactions. Focusing on the marginal effect, though, proves that it is worth digging deeper. According to the **culture** panel in Figure 16, the expenditures through special financings affect efficiency differently depending on the volume of special financings. The better funded the special financing, the more intense the effect. The DEA approach suggests a significant and negative effect for special financings that constitute more than 2% of the balance sheet. The illustration of the marginal effect stemming from the SFA approach provides further evidence for this finding. It seems that debate prevention effectively plays a role in this domain, particularly if the special financing is well-funded.[81]

On the contrary, the results of the **transportation** domain point in the opposite direction. In fact, the marginal effect even changes its sign over the range of the volumes (in both the one-step and the two-step approach). For poorly-funded special financings the marginal effect is negative and increases constantly until it becomes significantly positive for special financings that make up a material part of the balance sheet. The quasi-FE results work well for a quantitative interpretation: When a canton stores 10% of its balance sheet in special financings for the transportation, it provides its service with 0.2 percentage points less inputs (personnel costs, etc.) if it increases the expenditure share through special financings by one percentage point.[82] Again, the SFA support these findings (note that the one-step graph of the transportation panel shows the marginal effect on the *in*efficiency).

The positive effect of the expenditures through special financings on efficiency in the transportation domain contradicts the theoretical expectation. Finding possible explanations requires some background knowledge about the concrete special financings in these domains. In transportation, most cantons uphold special financings to construct and maintain roads. This domain attracts considerably high political attention because there is a clear link between the taxed good (i.e. the vehicle, for instance) and the benefitting good (i.e. the roads) (Brennan & Buchanan 1978). Given this link and the relatively precise demarcation of the public service (and therefore those who pay and those who do not), the lobbying groups can easily organize themselves and monitor the government, whereas the focus lies on the effective assignment of the earmarked

[80] When including both—the expenditures through and the volume of special financings—the question of their correlation arises. If they are strongly correlated, they impose a multicollinearity problem. Given the correlation coefficient in the transportation domain (0.081) and in the cultural domain (0.338) this issue seems negligible.

[81] Likewise, the domain of social security results in a negative effect, which is only significant for poorly-funded special financings though (see Figure 33 in the appendix).

[82] In 2010, the canton of Zug provides a suitable example. The canton stored 166 million Swiss francs in special financings earmarked for transportation services, amounting to 9.25% of the balance sheet total. The government funded 16.87% (19 million Swiss francs) of its expenditures for transportation through special financings. Increasing this share by 1% (1 million CHF) would lead to a 0.2 percentage points reduction of personnel (20'000 CHF), operating (31'000 CHF), financial (64'000 CHF), and capital expenditures (110'000 CHF) according to the estimates.

revenues for the intended purpose (Lee & Tollison 1991).[83] The better funded the special financing, the higher the risk of widening the intended purpose and increased political attention. Attention in turn might force the government to provide the service efficiently (Davis & Hayes 1993; Dhillon & Perroni 2001). If the special financing is poorly funded, there is less risk that the earmarked revenue is misused, and this reduces the attention. With the reduced attention, the debate prevention theory comes into play again. For the domain of culture, sport and leisure, the dominant special financing comes from lottery money. Since there is no organized lobbying group that monitors the usage of these funds, the debate prevention theory takes effect immediately. The better funded the special financing is, the less debate takes place and the more heedlessly the government dispenses the money. Naturally, this negligence translates into a lower efficiency.

A further explanation for the specific pattern of the results are the characteristics of the benefiting goods. For culture, the special financings fund several small transient projects while the earmarked revenue for transportation purposes finances the roads. In the latter case, special financings guarantee a constant maintenance of tangible assets which prevents their complete deterioration. According to infra suisse (2017), the accumulated expenditures are higher if the value retention of roads is postponed than if the roads are constantly maintained. Constant maintenance is less guaranteed for poorly-funded special financings, which possibly explains the different effects earmarking has depending on the financial endowment.

Given these findings, it is interesting to test the effect earmarked transfers have on efficiency. The last model of Table 20 replaces the expenditures through special financings by the pseudo-earmarked transfers' share of total transfers and thereby also drops the interaction term with the volume of special financings. The respective graph in Figure 16 reveals marginal effects that follow a mounting trend, suggesting an increasing positive (negative) effect of earmarked transfers on inefficiency (efficiency). Note that the model controls for the transfers' fraction of the total revenue. Thus, for earmarked transfers, the refined fiscal illusion theory developed by Wyckoff (1990) does not apply to the transportation domain, which, initially, contradicts the findings of earmarked revenue in general. However, reconsider the argument of the positive effect of earmarked revenue on efficiency: Cantonal interest groups form due to the strong link between the tax their members pay (for their vehicles for instance) and the good they consume (roads). These interest groups then monitor the government. Now, earmarked transfers come mainly from the *national* government where the payer circle is broader. In turn, the cantonal interest group has less incentive to inform themselves about the transfers as they pay only a small part of it and because the free-riding problem becomes more severe (Dhillon & Perroni 2001). In other words, they cherish the fiscal illusion and they struggle to monitor the *cantonal* government properly. The lack of proper monitoring then allows the latter to branch off some rents.[84]

[83] On June 5, 2016, the voters decided upon an initiative that called for an even narrower usage of the earmarked revenues. The initiative failed, but the episode shows that this domain is kept under critical observation.

[84] Anecdotal evidence comes from the heavy vehicle charges (*Leistungsabhängige Schwerverkehrsabgabe*, LSVA), which primarily flows to the central state and is then partly transferred to the cantons. In turn, the cantons ought to use their share exclusively for expenses related to road traffic (art. 19(3) SVAG). However, while most cantons established a special financing dedicated to roads where they allocate – by law – the revenues of the cantonal vehicle tax, only some proceed the same way with the national heavy vehicle charges (Federal Office for Spatial Development 2009). As a concrete example take § 35 GSW, in which the canton of Zug lists all revenues for the special financing for roads without mentioning the heavy vehicle charges.

To summarize the findings of the variable of main interest, the empirical insights partly support the debate prevention theory. At the same time, they put the completeness of the theory into question, as debate prevention currently does not explicitly predict qualitatively different effects depending on the volume of the special financings. To increase the credibility of the results, the after next section presents some robustness checks.

For the effect of the GDP per capita interacted with the tax progressivity, the results are patchy. Remember that the theory is directly geared towards to the interaction coefficient, predicting a downward sloping curve with an always negative marginal effect. The one-step approach makes it difficult to bring the numerical results in line with the theory. The illustrations in Figure 31 and Figure 34 show, however, a slightly upward sloping trend with an always positive marginal effect of the GDP. For **transportation**, the effect becomes even significant for highly progressive tax systems. The results of the two- and one-step approach are very similar.[85] In sum, the data does not support the fiscal illusion stemming from progressive taxes. It rather seems that an increasing GDP positively affects efficiency under tax regimes with a strong progressivity. One possible explanation for this surprising relationship is that expenditures in transportation do not react immediately to changes in demand (i.e. through a greater number of vehicles), whereas the latter comes hand in hand with a higher GDP. Therefore, a higher GDP is associated with a higher efficiency to start with. Why this effect accentuates under a progressive tax system is difficult to explain.

Focusing on tax diversification and complexity, the results provide no evidence for the fiscal illusion due to a complex tax system. For both domains, the one-step and the two-step approach yield insignificant coefficients. On the contrary and in contradiction to Carroll's (2009) idea, tax diversification indeed affects efficiency. While more tax sources lead to more efficient service provision for culture, a diversified tax system decreases efficiency in terms of transportation services. These results are significant in both approaches. For **transportation** services, the effect is unsurprising given the previous interpretation of the effect the expenditures have on efficiency through special financings. The well-organized interest groups who monitor the transportation domain struggle to keep an overview if the number of tax sources increases, which allows the administration to be less efficient. For **culture**, where no such organized interest group exists per se, the additional tax sources might actually attract the attention of an important entity who could then more closely monitor that domain.[86]

The variable related to the flypaper effect does not confirm the fiscal illusion. The transfers show a significant coefficient only in one case (culture domain estimated with the one-step approach); otherwise the estimates are always close to zero. The effects are possibly confounded

[85] The remaining state functions provide a rather vague picture. For instance, for the general administration, there also seems to be an upward-sloping trend in both approaches. On the contrary but in line with the theory, the domains of health, transportation, and national economy show a downward sloping trend, whereas the former two are significant in the two-step approach.

[86] Another anecdotal example comes from the entertainment tax of the city of St. Gallen. Until 2008, the city collected a tax on event tickets to cover the additional costs of security or waste collection. The organizer acted as an interest group pressuring the city to provide its service efficiently. In 2008, the city reformed the tax system and abolished the entertainment tax whereby the organized interest group lost its interest in monitoring the public service provision (Klingenberg 2007).

because of the 2008 regime change regarding transfer payments between the federal state and the cantons.

The fiscal illusion stemming from the debt per capita receives just as little support. Albeit the SFA provides significant coefficients with the expected negative effect on efficiency, the two-step approach provides the opposite results.

Turning to the control variables, the effects on efficiency are mostly somewhat more stable and intuitive. First and foremost, the population density influences efficiency rather negatively in both domains. In the **transportation** domain, the effect is significant in the two-step approach and has the same direction, however without significance, in the one-step case. While the coefficient for the **culture** domain has the same sign and is also significant using the SFA approach, the quasi-FE algorithm provides an insignificant positive sign. Intuitively, the negative effect of a dense population on efficiency receives clearer support for transportation services than for culture. Here it is worth expanding the focus to the remaining government functions (see Table 31 and Table 32 in the appendix). Taking the regressions where the two models' coefficients are compliant and with partly significant results, the population density's negative effect repeats. It is particularly striking that this is a matter in the domains of health and social security. These government functions can be challenging for urban cantons who seem to struggle with an efficient service provision.

When looking at the institution of Switzerland's direct democracy, the intuition holds that initiatives tend to induce inefficient public service provision in the **culture** domain. It is imaginable that a low threshold for initiatives encourages culture enthusiasts to demand further expenditures while the output either simply does not increase enough or is simply difficult to measure; the rather low value of the distance function's R^2 speaks in favor of the latter argument (see Table 17). The inefficiency effect of the initiatives does not repeat for the remaining government functions. The effect of the referendum is ambiguous. In any case it seems reasonable to disentangle the direct democracy variable in two separate measures in order to prevent premature conclusions made in earlier studies (Widmer & Zweifel 2012).

The voter turnout has a less clear but similar influence on efficiency as the initiative. The effect is significant and negative in case of the **transportation** domain estimated by the two-step approach (with the same sign using the one-step approach). Moreover, the domains of national economy and education provide further evidence for a negative effect. Particularly in these three domains, people seem to participate more actively in polls favoring more public service even if they are provided less efficiently. Counterintuitively, politicians are not urged to manage their domain more efficiently because of a higher voter turnout and its stronger monitoring effect (Borge et al. 2008; Geys et al. 2010).

The variable measuring the bourgeois party share in the government council provides further interesting insights, even if the results for the two main domains are vague. A closer look at the remaining functions singles out social security, where the influence is positive and significant in both approaches. The fairly high R^2 of the distance function in Table 17 refutes the potential argument that left-wing governments would approach social security in a broader manner, resulting in a lower efficiency measure. If this was the case, the considered inputs would explain the outputs insufficiently ending up in a low R^2.

While fragmentation has a significant and positive effect on efficiency in the **transportation** domain when estimated with the two-step approach, the one-step approach contradicts this finding. Social security is the only function with mutually confirming results: highly significant positive effects in both models. Here, the results support the small-is-good hypothesis.

Finally, the share of foreigners relative to the entire population has a positive influence on efficiency in both domains, whereas the results of the **culture** domain are highly significant in both approaches. On the contrary, in the two domains public order and security on one side and social security on the other side, the foreign share affects efficiency rather negatively. Intuitively, one could argue that, on the one hand, foreigners demand less extravagant cultural goods (e.g. museum exhibitions) keeping expenditures on a reasonable level. On the other hand, they induce higher costs per output in the domains of public order and security as well as social security.

7.6. Conditional effects of earmarking on efficiency

As the previous section revealed, concentrating on the plain coefficient of the expenditures through special financings on the efficiency possibly overlooks some potential conditional effects. This section therefore interacts the respective variable with different factors that influence the marginal effect. To simplify the interpretation, all the estimations are based on the two-step approach which yields linear marginal effects by construction. Furthermore, only the estimates with significant marginal effects are shown.[87] Table 21 lists the estimated coefficients together with the marginal effects, while Figure 17 illustrates the latter. Regarding the illustrations, at a first glance the always upward-sloping trend stands out. However, this is more by accident as some of the unreported models resulted in decreasing (though non-significant) marginal effects.

The first row of Table 21 indicates the domain followed by the variable with which the expenditures through special financings is interacted. The first specification interacts the expenditures through special financings with the debt brake. A restrictive debt brake might incentivize ministers to generally monitor more accurately and thereby restrain debate prevention. If this argumentation were true, the marginal effect of the expenditures through special financings on efficiency would, while always remaining negative, approach zero with a more restrictive debt brake. Indeed, applied to culture, the marginal effect is (insignificantly) negative for small volumes of the special financing and increases along the debt brake axis. However, at a certain point, the marginal effect becomes positive and even turns significant for very restrictive debt brakes. In combination with the results above, where the effect of expenditures through special financings was always negative, this specification provides a further consideration. Even for culture, special financings can positively affect efficiency, if the debt brake of a canton is strong enough.

The remaining models are based on the data of the transportation domain, with the fraction of tax revenues as the first interaction variable.[88] The idea is that ministers generally monitor their

[87] In addition to the reported interaction variables, the debt per capita, the transfers, the initiative index, and the lagged budget balance were tested. None of these variables provided a significant marginal effect over their entire scope for either domain.

[88] Note that the ceteris paribus interpretation still holds, since the fraction of tax revenue is measured on the cantonal level, while the share of expenditures through special financings ties on expenditures at the domain level.

colleagues more carefully when the public services are overall funded with taxes that the politicians have to sell hard to their voters (compared to transfers, for instance). This means that the expectation is that a higher fraction of tax revenues bans debate prevention, i.e. a marginal effect that approaches zero from below. As Table 21 reveals, the marginal effect is positive and significant at the mean value of the tax revenue fraction. Moreover, the larger the fraction of tax revenues becomes, the stronger the marginal effect becomes (see Figure 17). If the tax revenues amount to more than 70 percent of the total revenues, an additional percentage point of expenditures through special financings decreases the input requirements by more than 0.2 percentage points. Hence, tax revenues do indeed restrain debate prevention, but it turns its effect even in the positive scope.

The next conditioning variable aims for the fractionization within the governing council. If the governing council is composed of very few parties, it is likely that there is less monitoring going on and thus fractionization stimulates the debate prevention. Again, this idea ought to translate into a negative but increasing marginal effect. Indeed, the marginal effect is (insignificantly) negative for unity government councils.[89] However, the effect turns eventually and becomes significant for fractionization values above 60 percent. Note that this does not mean that highly fractionalized government councils lead to an efficient public service provision.

Finally, it is again the political context that motivates the last model. With a similar logic as above, the concordance possibly fosters debate prevention, as the governing council feels less pressure from the parliament. That is, ministers supposedly monitor less if their party is equally represented in the legislative government branch. Given the similarity of the arguments, it comes as little surprise that this specification yields comparable results with significant marginal effects for concordance values above 40 percent.

In sum, the three models analyzing the transportation domain provide little further insights. It seems that earmarking generally has a positive effect on efficiency for this government function – at least as long as the special financing is not in debt. On the contrary, the previously found negative effect in the culture domain can be rescued if the canton operates under a restrictive debt brake.

[89] The government council of the canton of Appenzell Innerrhoden, composed of seven members of the Christian democratic people's party from 2008 to 2007.

Table 21 – Truncated regressions with different interactions

Domain	Culture	Transportation	Transportation	Transportation
Conditional variable	Debt brake[1]	Tax revenue share[2]	Fractionization in governing council[3]	Concordance[4]
[A] Expenditures through special financings (in % of total expenditures)	-0.002* (0.001)	-0.002* (0.001)	0.000 (0.001)	0.000 (0.000)
[B] Condition variable	-0.073*** (0.024)	-0.002 (0.002)	-0.117 (0.108)	-0.175*** (0.061)
Interaction [A]x[B]	0.002** (0.001)	-0.002 (0.002)	0.002 (0.001)	0.002*** (0.001)
Volume of the special financing (in % of balance sheet's total)	-0.041 (0.025)	0.006** (0.003)	0.006** (0.003)	0.007*** (0.002)
[C] Tax progressivity (curvature parameter)	-0.020 (0.038)	-0.060** (0.026)	-0.050* (0.027)	-0.043 (0.026)
[D] GDP per capita (in 1'000 CHF)	-0.005 (0.004)	-0.004 (0.002)	-0.004 (0.002)	-0.003 (0.002)
Interaction [C]x[D]	0.001* (0.001)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Tax complexity (normalized HHI)	-0.002 (0.002)	0.001 (0.001)	0.002 (0.001)	0.002 (0.001)
Tax diversification (no. of tax sources)	0.079*** (0.022)	-0.002 (0.014)	-0.011 (0.014)	-0.013 (0.014)
Transfers (in % of total revenues)	-0.001* (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Gross debt per capita (in 1'000 CHF)	-0.003 (0.005)	0.006* (0.004)	0.006 (0.004)	0.007* (0.004)
Population density (1'000 inh. per km^2)	0.419 (0.460)	-0.594* (0.343)	-0.424 (0.340)	-0.419 (0.335)
Referendum (1 [hard] to 6 [easy])	-0.020 (0.026)	-0.014 (0.016)	-0.017 (0.016)	-0.024 (0.016)
Initiative (1 [hard] to 6 [easy])	-0.144*** (0.057)	0.007 (0.040)	0.006 (0.040)	-0.013 (0.041)
Bourgeois parties in executive (in percent)	-0.002** (0.001)	0.002** (0.001)	0.001** (0.001)	0.001** (0.001)
Fragmentation (1'000 inh. per municipality)	0.012 (0.039)	0.014 (0.027)	0.014 (0.028)	0.014 (0.027)
Voter turnout (in percent)	0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Foreigners (in % of total inhabitants)	0.029*** (0.009)	-0.006 (0.006)	-0.004 (0.006)	-0.008 (0.006)
Marginal effect on inefficiency				
Expenditures special financings	0.000 (0.001)	0.001*** (0.000)	0.001** (0.000)	0.001*** (0.000)
N	214	298	298	298

Notes: * p<10%, ** p<5%, *** p<1%; coefficients with bootstrap-based standard errors in parentheses (Simar & Wilson 2007); the dependent variable is the DEA technical efficiency score with a contemporaneous reference set (i.e. year-by-year DEA); individual dummies included but not reported; marginal effects at mean of interaction variable

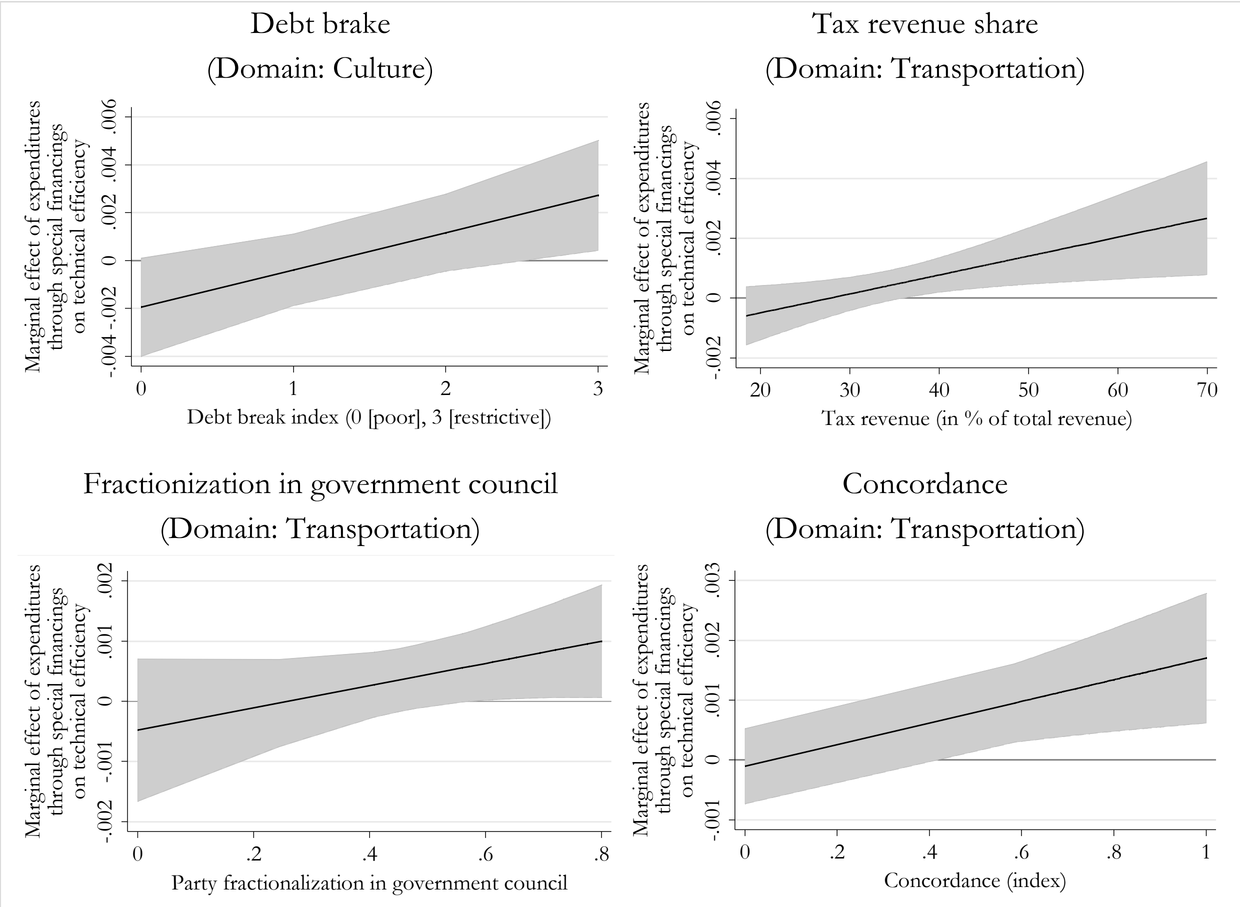
[1] Index proposed by Feld & Kirchgässner (2008): Each of the following elements adds one point to the index: (1) a strong connection of budget planning with actual budget execution; (2) strong numerical constraints; (3) highly effective sanctions in the form of automatic tax adjustments. Thus, a high index indicates a restrictive debt brake.

[2] In percent of total revenue

[3] Index proposed by Rae (1967): $Fractionization = 1 - \sum_{p=1}^{n} f_p^2$ where n is the number of parties in the governing council and f_p the fraction of seats each party occupies in the council. A large number indicates a high fractionization.

[4] Index proposed by Martin (2008): $Concordance = \sum_{p=1}^{n} f_p^2$ where p includes all parties represented in the governing council and f_p is their fraction of seats in the parliament. A large number indicates a high congruence between the parties in the governing council and in the parliament and hence a high concordance.

Figure 17 – Marginal effects of expenditures through special financings with interactions



Source: Own illustration

7.7. Robustness checks

The relatively assertive statements above regarding the appropriateness of special financings call for some robustness tests. For the sake of limiting the number of reported results, the different tests do not recur for both government functions. The focus lies on the transportation domain which faces eleven different specifications or estimators to substantiate the earlier claimed effects. At the same time, the robustness checks for the culture domain involve two tests. The two regressions give an idea of the additional tests, which are not reported. Both, the two-step (models (I) through (VII)) and the one-step approach (models (VIII) to (XII)) are taken up again, although, with partly different estimators. Since the goal of this section is to test the robustness of the effect special financings have on efficiency, the graphs in Figure 18 only illustrate the respective marginal effect and leave the other variables aside.

The first model (I) questions the quasi-fixed effect algorithm and re-estimates the original model using the <u>conventional estimator</u> proposed by Simar & Wilson (2007) including individual- and time fixed effects in the second step. The DEA efficiency scores of the first step come from an intertemporal reference set, which implicitly assumes a fix frontier along the time axis (see discussion in section 7.1). The respective graph in Figure 18 basically replicates the marginal effect of the base model and supports the earlier estimates.

The original estimates withstand the further tests using the two-step approach. Model (II) re-estimates the base model while dropping the fixed effects in both steps, whereas model (III) drops the interaction with the volume of special financings; note that the calculation of marginal effects becomes unnecessary without interactions in the two-step approach. In the latter case, the variable of interest remains positive with a 90 percent confidence interval above zero. Still without the interaction term, model (IV) introduces a first and second lag of the variable of interest in order to test for long-term effects. All respective coefficients are positive (on a very low scale though) but no longer significant. In order to test the influence of the hilliness, the specification (V) again drops the fixed effects and introduces the altitude of the cantonal capital city as a further control variable and its interaction with the street length (not reported). Both coefficients are not significantly different from zero and the marginal effect of the expenditures through special financings remains unchanged. Preventing a misestimation due to an alternating effect with a time lag, model (VI) re-estimates the baseline model using a five-year moving average of all variables. Still the qualitative and quantitative results hardly differ.[90] The last two-step model (VII) applies the specification of the first robustness test of Model (I) using data from the culture domain. Here also, the marginal effect of the variable of interest is the same if the year fixed effects only come into play in the second step. However, the estimates of the standard errors are less conservative which yields positive marginal effects for poorly-funded (but not indebted) special financings.

The remaining robustness test is based on the one-step approach, that is, the models follow the SFA logic. Model (VIII) re-estimates the base model using the estimator proposed by Battese & Coelli (1995), which uses a time trend in the estimation of the frontier instead of fixed effects. In addition, the heteroscedasticity draws on the mean of the inefficiency distribution instead of the standard error (i.e. $u_{it} \backsim N(\mu_{u,it}, \sigma^2)$ and $\mu_{u,it} = e^{\varphi' z_{it}}$), and the reported estimates represent the φ-coefficients. As the respective plot in Figure 18 indicates, the marginal effects of the variable of interest follow the previously found trend (note that the graph 'Transportation (VIII)' shows the marginal effect on *in*efficiency). Applying the same estimator to the culture domain replicates the findings of model (VI) where the marginal effect is positive for poorly-funded special financings. Using the Battese & Coelli (1995) model, lifts the trend even more, while keeping its negative slope.

Staying within the transportation domain, model (X) expresses the expenditures through special financings in per capita terms. Interestingly, this modification renders the other variables highly significant, while the implication on the marginal effect turns out to be jejune. While some single observations show pronounced marginal effects, most of them are very close to zero. From a theoretical point of view, such a fruitless result is satisfying as debate prevention builds on the earmarked revenues' share on the total expenditures rather than on absolute values. Hence, earmarking has – in absolute terms – little influence on efficiency.

To account for the restrictiveness of the Cobb-Douglas specification, model (XI) allows for further flexibility of the production function using the translog form. As this specification uses up many more estimations, the degrees of freedom shrink while the standard errors typically rise.

[90] Even if the results withstand this robustness test, the problem might still be present, since the five-year average might be too short. Due to the lack of data, a longer time range becomes increasingly inept.

Nevertheless, the marginal effects are by and large unchanged as illustrated in Figure 18. The same is true when the base model is re-estimated using the debt brake as an additional variable as in model (XII).

The last model provides little additional insights since the graph illustrating the marginal effects is dominated by a few extreme values. The base for these two plots is model (XIII), which replaces the expenditures through special financings with their own normalized HHI. The idea behind this last estimation is that it might make a difference whether there is one dominant special financing or several small ones. As mentioned above, the marginal effect mostly moves around zero.

In sum, the estimates for the transportation domain are very robust against different specifications and methods. As far as the culture is concerned, the base estimate represents the upper bound in terms of the negative effect earmarking has on efficiency. Nevertheless, it seems reasonable to conclude that the larger the special financing, the more pronounced the negative effect of earmarking on efficiency.

Table 22 – Robustness checks

Government function	Transportation	Transportation	Transportation	Transportation	Transportation	Transportation	Culture	Transportation
Estimation	(I)	(II)	(III)	(IV)	(V)[1]	(VI) [2]	(VII)	(VIII)
[A] Expenditures through special financings (in % of total expenditures)	-2.61E-04 (0.000)	1.99E-04 (0.000)	3.83E-04* (0.000)	3.83E-04 (0.000)	2.18E-04 (0.000)	0.001*** (0.000)	0.004*** (0.001)	3.30E-04 (0.002)
Lag (1 year)				1.29E-04 (0.000)	-0.009** (0.004)			
Lag (2 years)				9.08E-05 (0.000)				
[B] Volume of special financings (in % of balance sheet's total)	-0.001 (0.003)	-0.014*** (0.004)	0.008*** (0.002)	0.010*** (0.002)		-0.001 (0.003)	0.153*** (0.038)	0.091*** (0.031)
Interaction [A]x[B]	2.01E-04*** (0.00)	3.18E-04*** (0.00)			2.74E-04*** (0.00)	2.31E-04*** (0.00)	-0.003*** (0.001)	-0.001* (0.001)
[C] Tax progressivity (curvature parameter)	0.089*** (0.026)	0.048*** (0.015)	0.022 (0.014)	0.015 (0.019)	0.010 (0.018)	-0.036* (0.021)	-0.001 (0.049)	0.019 (0.115)
[D] GDP per capita (in 1'000 CHF)	0.002 (0.002)	0.002** (0.001)	3.36E-04 (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.007*** (0.002)	-0.004 (0.004)	-0.009 (0.007)
Interaction [C]x[D]	-0.001* (0.000)	-0.001*** (0.000)			-0.001*** (0.000)	0.001*** (0.000)	0.001 (0.001)	0.002 (0.002)
Tax complexity (normalized HHI)	0.030 (0.299)	0.002 (0.001)	-4.52E-04 (0.001)	0.001 (0.001)	0.002** (0.001)	0.001 (0.001)	-0.005** (0.002)	-0.008 (0.008)
Tax diversification (no. of tax sources)	-0.030** (0.011)	0.007 (0.005)	-0.013 (0.012)	-0.020 (0.014)	0.014*** (0.005)	0.025* (0.013)	0.107*** (0.023)	-0.037 (0.048)
Transfers (in % of total revenues)	0.001** (0.000)	-0.001*** (0.000)	-1.33E-04 (0.000)	-3.65E-04 (0.000)	-0.001** (0.000)	-0.002*** (0.000)	-0.001 (0.001)	0.002 (0.003)
Gross debt per capita (in 1'000 CHF)	-0.013*** (0.003)	-0.003* (0.002)	0.008*** (0.003)	0.008** (0.004)	0.003 (0.002)	-0.012** (0.005)	0.013** (0.006)	0.059*** (0.022)
Marginal effects on (in)efficiency								
Expenditures special financings	-0.000	0.001***	n.a.	n.a.	0.001**	0.002***	0.001	-0.001
GDP per capita	-0.002	-0.004***	n.a.	n.a.	-0.005***	0.000	-0.001	0.002
Tax complexity	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-0.005
Tax diversification	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-0.022
Transfers	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.001
Gross debt per capita	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.035
$\hat{\mathbb{E}}[\sigma_{u,it}]$	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.202***
$\hat{\sigma}_v$	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.051***
$\hat{\mathbb{E}}[\lambda_{it}]$	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	3.953***
Control variables	yes	yes	yes	yes	yes	yes	yes	yes
Year trend	no	no	no	no	no	no	no	yes
Year fixed effects	yes (2nd step)	no	yes (1st step)	yes (1st step)	no	yes (1st step)	yes (2nd step)	no
Individual fixed effects	yes (2nd step)	no	yes (2nd step)	yes (2nd step)	no	yes (2nd step)	yes (2nd step)	no
N	298	298	298	253	298	195	214	298
Model	Simar & Wilson (2007)	Simar & Wilson (2007)	Quasi FE algorithm	Quasi FE algorithm	Simar & Wilson (2007)	Simar & Wilson (2007)	Simar & Wilson (2007)	Battese & Coelli (1995)

Notes: * p<10%, ** p<5%, *** p<1%**;** **c**oefficients with robust standard errors clustered at cantonal level; the dependent variable is the logged personnel expenditures (in the SFA case) or the input oriented efficiency score (in the DEA case); a positive coefficient indicates a positive (negative) effect on inefficiency (efficiency) in case of the SFA approach – vice versa signs with the DEA approach; mean marginal effects without significance tests in the SFA case; $\hat{\mathbb{E}}[\sigma_{u,it}]$ is tested against H_0: $\hat{\mathbb{E}}[\sigma_{u,it}] = 0$; $\hat{\sigma}_v$ is tested against H_0: $\sigma_v = 0$; $\hat{\mathbb{E}}[\lambda_{it}]$ is tested against H_0: $\lambda_{it} \leq 1$

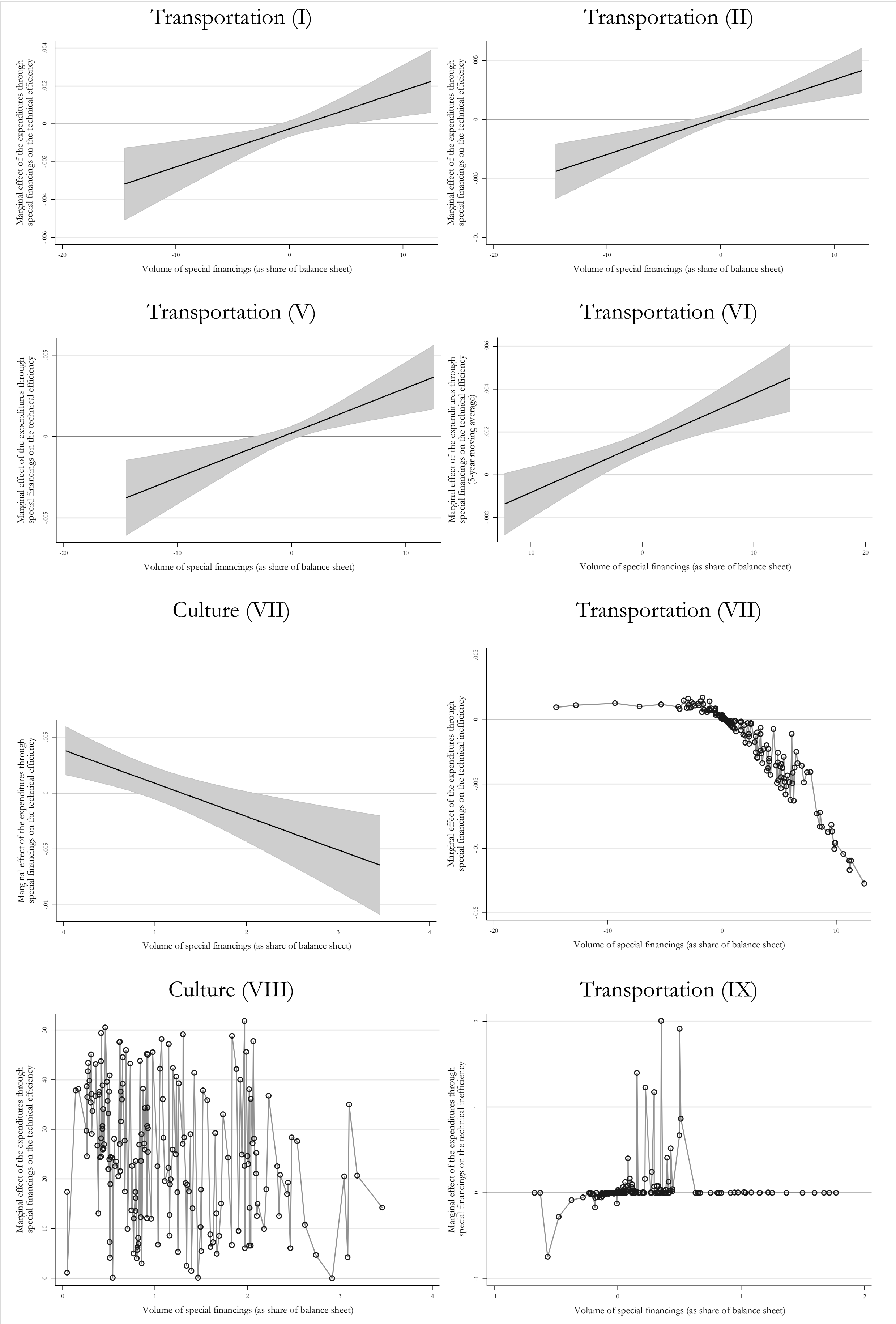
[1] Includes altitude and its interaction with national street length as control variables; [2] All variables converted to five year moving averages

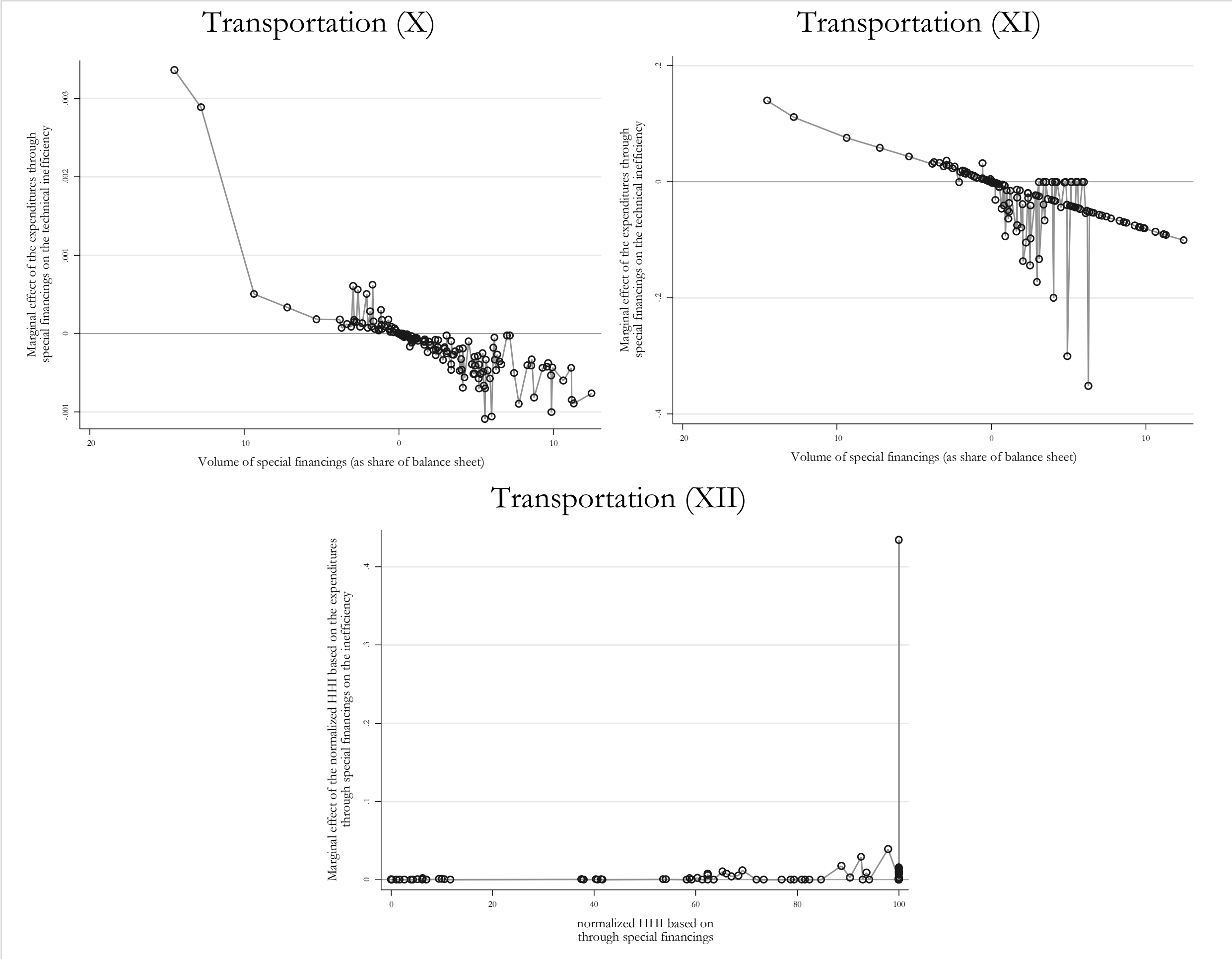
Table 22 continued – Robustness checks

Government function	Culture	Transportation	Transportation	Transportation	Transportation
Estimation	(IX)	(X) [3]	(XI)	(XII)	(XIII)
[A] Expenditures through special financings (in % of total expenditures)	0.003 (0.005)	-6.405* (3.397)	0.012 (0.009)	-0.006 (0.025)	
[B] Volume of special financings (in % of balance sheet's total)		-27.037*** (4.76)	0.080 (0.196)	0.152 (0.514)	
Interaction [A]x[B]		40.12*** (11.14)	-0.001 (0.005)	0.002 (0.006)	
[C] Tax progressivity (curvature parameter)	-0.016 (0.284)	5.121*** (1.677)	0.450 (0.676)	-2.101 (2.004)	2.991 (4.566)
[D] GDP per capita (in 1'000 CHF)	-0.006 (0.004)	1.004*** (0.218)	0.004 (0.048)	-0.158 (0.219)	0.378 (0.591)
Interaction [C]x[D]	0.359*** (0.126)	-0.199*** (0.039)	-0.012 (0.011)	0.035 (0.035)	-0.100 (0.101)
Tax complexity (normalized HHI)	0.027*** (0.009)	0.127*** (0.046)	-0.148*** (0.051)	-0.238** (0.102)	-0.297 (0.348)
Tax diversification (no. of tax sources)	-0.005* (0.003)	-2.102*** (0.268)	-0.813** (0.374)	0.290 (0.536)	-1.315* (0.716)
Transfers (in % of total revenues)	-0.002 (0.008)	-0.173*** (0.031)	-0.026** (0.012)	0.015 (0.019)	-0.009 (0.065)
Gross debt per capita (in 1'000 CHF)	-0.046 (0.068)	-0.235 (0.239)	0.366*** (0.105)	0.642 (0.537)	0.684 (0.590)
Debt brake ([0] weak, [3] restrictive)	0.003** (0.001)			0.298 (0.666)	
Normalized HHI of expenditures through special financings	-0.007 (0.016)				0.086*** (0.022)
Marginal effects on (in)efficiency					
Expenditures special financings	-0.003	0.036	0.000	0.000	
Normalized HHI of expenditures through special financings					0.005
GDP per capita	0.000	-0.024	-0.004	0.011	-0.033
Tax complexity	-0.002	0.003	-0.009	-0.013	-0.016
Tax diversification	-0.041	-0.047	-0.050	0.016	-0.072
Transfers	0.003	-0.004	-0.002	0.001	-0.001
Gross debt per capita	-0.006	-0.005	0.023	0.036	0.037
$\hat{\mathbb{E}}[\sigma_{u,it}]$	0.256***	0.028***	0.078***	0.069***	0.070***
$\hat{\sigma}_v$	0.010	0.066***	0.039***	0.064***	0.059***
$\hat{\mathbb{E}}[\lambda_{it}]$	2.573	0.429	1.980***	1.083	1.150***
Control variables	yes	yes	yes	yes	yes
Year trend	yes	no	no	no	no
Year fixed effects	no	no	no	no	no
Individual fixed effects	no	yes	yes	yes	yes
N	214	298	298	298	170
Model	Battese & Coelli (1995)	TRE	TRE	TRE	TRE
Specification of distance function	Cobb-Douglas	Cobb-Douglas	translog	Cobb-Douglas	Cobb-Douglas

Notes: [3] Expenditures through special financings and volume of special financing measured in per-capita-units

Figure 18 – Robustness checks of marginal effects of expenditures through special financings





Notes: Negative values on the abscissa indicate indebted special financings; illustrations only cover range of sample
Source: Own illustration

7.8. Interim conclusion

This chapter started with the proposition of an enhanced estimator when explaining DEA efficiency scores in a panel data environment. Often, studies with such a data structure run the DEA on the entire sample and in this way leave aside a potentially changing production frontier. In the second step, these papers usually implement year and individual dummies to account for eventual fixed effects (Selim & Bursalıoğlu 2015; Fleishman et al. 2009; Chen et al. 2005). Instead, the quasi-fixed effect algorithm first performs the DEA on each year individually and only introduces individual dummy variables in the second stage. The Monte Carlo simulations provide evidence, that the quasi fixed effect algorithm performs a lot better in terms of accuracy and efficiency than the conventional approach.

Still concerned with methodological issues, the second section introduced heteroscedasticity in SFA in order to make the inefficiency dependent on some environmental variables. The third section rounded up the methodological part by discussing endogeneity problems. In terms of the earmarked revenues, endogeneity should be less problematic since special financings themselves are fairly stable and the variations within existing ones are largely driven by external factors that are outside the manipulable scope of the cantonal government. Nevertheless, the fourth section introduced additional control variables to make the ceteris paribus assumption more consistent.

Based on the methodological foundations and the data, the estimations provide partly unexpected results. The cultural domain mainly supports the debate prevention theory. In particular, well-funded special financings dedicated to cultural public goods enforce the negative effect earmarking has on efficiency. At the same time, the results show that the effect of earmarks can also turn positive if a canton has a restrictive debt brake. Including the findings of the transportation domain suggests an even more differentiated picture. While indebted special financings also allow some debate prevention (i.e. earmarking indeed negatively affects efficiency), well-funded special financings have the opposite effect. The most plausible explanation for this phenomenon are the inherent differences between the two domains. First, in the transportation domain the tax-payers and beneficiaries largely cohere (mainly road utilizers), but these two groups differ significantly in the culture domain. Second, the correlation between the taxed and the benefiting good is only strong in the transportation domain and not in culture. Third, the transportation domain has the potential to exclude those who do not pay but this is less the case for culture. These differences correspond exactly to those identified in the literature review. The public choice veterans predicted a positive effect of earmarking on efficiency if these conditions are fulfilled. The debate prevention now also explains why earmarking has a particularly negative effect on efficiency when those conditions are absent.

8. Qualitative Analysis

After the previous chapters quantitatively tested the debate prevention theory, a deepening qualitative approach proves useful to directly examine the hypothesized causal chains. For this purpose, this chapter assesses the budgeting process of the canton of Solothurn in 2016 for the domains of culture and transportation. Besides adding some empirical evidence, this kind of case analysis also identifies further aspects which the theory so far ignores. In order to equip the case study with an appropriate methodological foundation, the process-tracing method provides the structural frame. Note that applying process-tracing as a main method would be enough for a separate thesis. This means that this chapter cannot actually present enough evidence to claim the validity of the debate prevention theory. However, looking deeply into a specific case helps set aside the quantitative lenses and make better founded claims.

Probably one of the most theory-challenging insights concerns the point where the actual individual budget cuts (and consequently the budget allocation) take place. While the debate prevention theory regards the governing council as the dominant allocation authority, the case study proposes a different venue. In fact, the governing council actually sets the overall budgetary objective, but it does not debate the effective budget allocation. This happens instead in the bilateral meetings with the finance department (i.e. the minister and the head of the finance office) and the individual spending departments (i.e. the minister and the department secretary), whereas the former usually demands specific budget cuts while the latter justifies his financial needs. Still, the debate prevention theory holds as far as the finance department asks for less budgetary cuts from services funded through special financings, as they have no influence on the budget balance. From a theoretical point of view, this means that the spending ministers still have an incentive to increase the quality of their dossier in order to protect their budget share. However, it is the finance minister who puts some effort into monitoring the spending ministers in order to achieve the parliamentary and the governing council's objective to balance the budget.

Apart from that, the process-tracing provides some evidence that the parliament aims to cut the transportation budget instead of the cultural budget. Even if both domains are funded with special financings, the transportation domain is monitored by an organized interest group that promotes an efficient use of resources. At the same time, the culture domain was hardly debated during the budgeting procedure.

8.1. Process-tracing method

The fundamental difference between the previously applied quantitative approaches and the process-tracing method as a member of the qualitative family is the ontological understanding of causal relationships. That is, while the quantitative large-n approaches is usually based on the regularity assumption (i.e. a regular association between X and Y whose underlying link is unobservable), the qualitative small-n process-tracing believes in mechanisms that are open to scrutiny (Chalmers 2013). Such a mechanism is not necessarily regular because it can be infrequent. At the same time, even if the link is found empirically, "[n]o claims can be made […] about whether the mechanism was the only cause of the outcome" (Beach & Pedersen 2013, p.3). Yet, there must necessarily exist an X that follows a path linked with Y. As to the second big methodological characterization, the process-tracing method resembles the DEA in the sense that both are deterministic. As opposed to this, the SFA has a probabilistic ontology and allows a certain causal fuzziness by including a stochastic error.

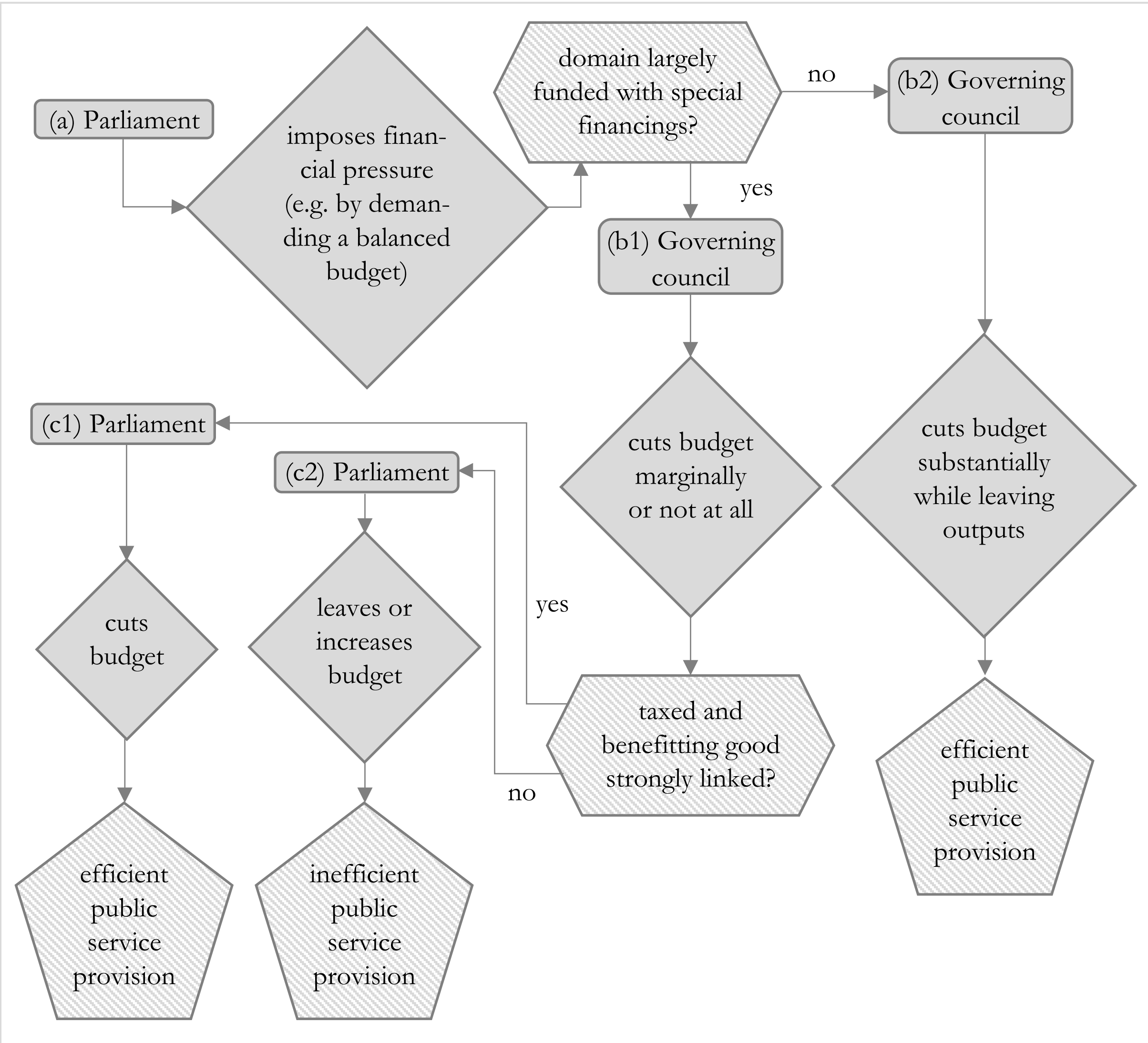
Process-tracing goes further in imagining a mechanism than simply assuming its existence. In fact, the aim is to open the black box as much as possible and to go beyond the mere stringing together of variables (Beach & Pedersen 2013). Instead, a mechanism involves entities that engage in activities and thereby they actually transmit the causal force from a cause to an outcome. The logic behind the mechanism derives from the theory. If the hypothesized mechanism truly exists, it should leave observable fingerprints that can be studied empirically. Detecting these fingerprints therefore produces some evidence for the causal force between X and Y and consequently for the theory.

Inevitably, the design of the mechanism automatically brings up the question of the level of analysis. Some authors see the micro-foundation as the only place where the true mechanisms actually take place (George & Bennett 2005). In their view, the analysis should focus on individuals in order to truly understand the causal forces. In terms of the debate prevention theory, this claim is hardly rejectable since the idea ties on individual utility functions of the ministers. Accordingly, the best understanding of the true mechanism surely would result from a micro-level analysis. From a more pragmatic viewpoint though, it makes more sense to choose the level of analysis at which, at the same time, the mechanism is still observable and the necessary information is available (Beach & Pedersen 2013). In the present case this is the actor level in the institutional sense, i.e. the government council and the parliament with its commissions.

8.2. Empirical fingerprints

After the previous section briefly introduced process-tracing as a method, the debate prevention theory and its conditions identified through the quantitative analysis now translate into an expected mechanism with entities (actors) and activities (events) as illustrated in Figure 19. The mechanism closely follows the budgeting process introduced earlier (section 4.1). The soft-edge rectangles signify actors with their activities in the diamond boxes. The contextual factors appear in hexagons and the pentagons denote the outcome of the process.

Figure 19 – Expected mechanism how earmarking leads to inefficiency



Source: Own illustration

Before going through the mechanism demonstrating the empirical fingerprints, the question of the case selection remains. In light of the fact that the method is information-demanding, there must be enough documentation about the selected case. The canton of Solothurn appears to be an optimal choice because of its unique public nature of the governing council meetings. To see behind the curtain of the executive branch of the government is essential since the hypothesized debate prevention largely takes place there.[91] Besides the sessions of the government council, the events before and after have their impact on the outcome too and need to be taken into account. Accordingly, the entire documentation stems from several meetings, sessions, and interviews as listed in Table 23. Besides the event itself, where the statements of the actors are part of the information, the related documents also serve as a source.

Table 23 – Sources for process tracing

Date	Event	Actors	Documents
2016/04/27	FICO meeting (budgetary target)	FICO members, finance minister, head of financial office, head of internal audit	financial plan, minutes, letter to government council
2016/05/09	government council meeting (response to budgetary target)	ministers, state secretary, head of financial office	letter to parliament, internal budget directive
2016/08/29	interview (budgeting process)	head of financial office, controller of finance department	minutes
2016/09/06	government council meeting (budget passage)	ministers, state secretary, head of financial office	financial plan, draft budget
2016/09/14	meeting of FICO committee for education and culture (budget)	committee members, finance minister, state secretary, head of financial office, controller of finance department	minutes, draft budget
2016/09/18	meeting of FICO committee for building and justice (budget)	committee members, finance minister, state secretary, head of financial office, controller of finance department	minutes, draft budget
2016/10/25 2016/10/26 2016/11/23	FICO meeting (budget)	FICO members, ministers, office heads, state secretary, department controllers, head of internal audit	draft budget, budget supplement, minutes of committees, financial semester report
2016/11/22	government council meeting (comments on amendments)	ministers, state secretary, head of financial office	amended draft budget
2016/12/06 2016/12/07 2016/12/14	parliament (budget)	parliament members, ministers, office heads, state secretary, department controllers, head of internal audit	enacted budget

Notes: FICO is the Cantonal Finance Commission; the named actors involve the most relevant ones excluding further subordinates

The first part of the mechanism in Figure 19 refers to step [1.2] of the budgeting process (Table 1), where the finance commission decides upon the budgetary target based on the financial plan.[92] Strictly speaking, this first step happens before the hypothesized causal chain starts. The earmarked revenue (special financings) only comes into play in the second part (step [1.6]) where the process separates into two distinct paths. Remember though, that debate prevention assumes a balanced

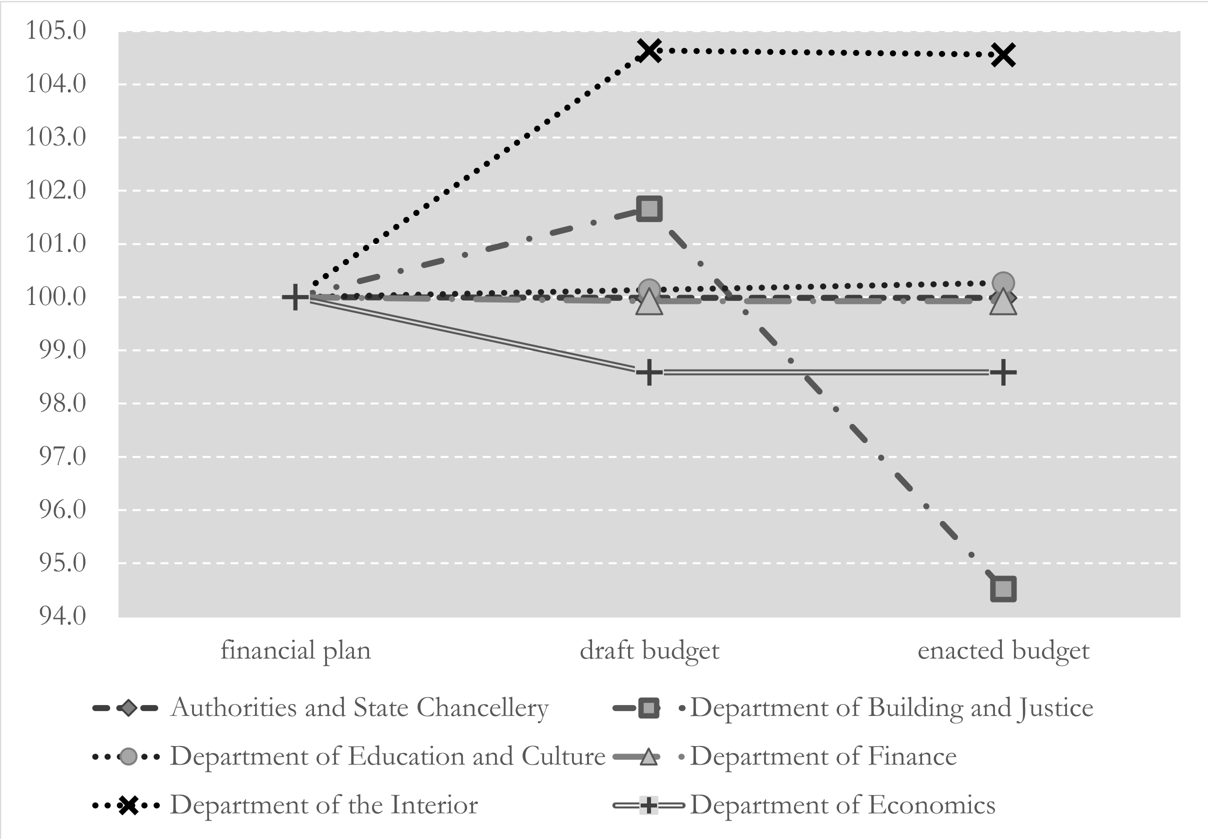
[91] The choice of the budget year 2017 (planned in 2016) is for solely practical reasons. It was at that time when enough data of the quantitative part was available to reasonably classify the information and to ask the right questions when meeting with the relevant actors.
[92] In the budgeting process in Table 1, the financial plan corresponds to the 4-year financial blueprint (IAFP, *Integrierter Aufgaben- und Finanzplan)*.

budget, which the financial plan rarely satisfies. With the balance requirement, the parliamentary commission underlines the practical importance of that assumption. In case of the budget 2017, the financial plan 2017-2020 foresees for 2017 a loss (deficit) of the financial performance of 47.1 million Swiss Francs.[93] Proceeding from this base, the finance commission decided during their meeting on April 27th, 2016 to impose a balance requirement and hence to improve the financial result. In their subsequent meeting on May 9th, 2016, the government council started the administrative budgeting procedure and at the same time it responded to the parliaments' budgetary target. In their letter, the government council expressed their concern about the strict budget requirement and held out for the prospect of a slightly broader budget.

The administrative budgeting procedure is worth looking at in more detail. According to the theoretical mechanism, the government council's reaction of increased financial pressure differs by domain (corresponds to step [1.6] of Table 1). A government function (here the department) which benefits mostly from earmarked revenues presumably prevents the debate by hiding behind their special financings. In contrast, the other departments depend on the now reduced general fund and allocate the remaining resources according to their dossier quality and the monitoring effort of the ministers. In line with the descriptive statistics of all its counterparts, the canton of Solothurn's domains with the highest expenditures through special financings in terms of total expenditures are culture (52% on average between 2000 and 2014) followed by transportation (51%). It is therefore in these two domains where the theory predicts no budget cuts (i.e. path b1). At the same time, the other domains supposedly face stronger budgetary restrictions. The result of the budgetary debate in the government council is manifested in Figure 20.[94]

[93] The financial performance (*Saldo der Erfolgsrechnung*) comprises all revenues and expenses of the income statement (*Erfolgsrechnung*) but it excludes the transactions of the statement of capital expenditures (*Investitionsrechnung*).
[94] Figure 37 in the appendix shows the same graph including the capital expenditures.

Figure 20 – Budgetary amendments (expenditures) of the Solothurn cantonal budget 2017



Source: Own illustration

The different lines show the planned expenditures of each department according to the stage of the budgetary process. The focus lies on the expenditures instead of the balance because the theory stipulates the ministers' objective to maximize the former. In order to account for the different sizes of the departments, the planned expenditures of the financial plan serve as reference. [95]

The draft budget column shows the government council's reaction to the financial pressure (i.e. reducing the deficit by 47.1 million Swiss Francs) imposed by the finance commission. Apparently, the council decided not to apply the broad-brush approach and reduce anyone's expenditures by a certain percentage point. The Department of Economics was the least powerful in the budget debate and lost 1.6 percent of its initial budget. Surprisingly, the Department of the Interior was the most successful and increased its budget share by 4.6 percent.[96] As expected, both, the Department of Building and Justice (+1.7%) – managing the transportation domain – and the Department of Education and Culture (+0.3%) slightly increased their share.[97] In sum, the

[95] The expenditure shares after the financial plan divide as follows: Authorities and State Chancellery (0.9%), Department of Building and Justice (12.0%), Department of Education and Culture (21.8%), Department of Finance (7.5%), Department of the Interior (45.4%), Department of Economics (11.3%). The courts account for the rest, but they follow a different logic and are therefore not part of this analysis. These shares can be interpreted as $\bar{\lambda}$ of the theory section.

[96] During the interview, the head of the finance office names the Department of the Interior as an example with a lot of cantonal autonomy, which is a precondition to reduce the budget in a fiscally stressful situation (Bühlmann 2016).

[97] For the transportation domain, the head of the finance office sees a relationship between the existence of a special financing and higher expenditures. In his view, this effect was much stronger under the previous accounting system (HAM1), when the special financing for roads actually had no influence on the fiscal balance. For the cultural domain,

governing council managed to reduce the loss (deficit) of the financial performance to 8.2 million Swiss francs and thereby failed to balance it entirely.

A deeper empirical assessment of the actual process shows where the numbers come from. The official (but non-public) decision of the government council as a result of its meeting on May 9, 2016 is particularly insightful. The decision takes the form of budgetary directives which state that "the departments must develop measures to reduce the budget on the expense side […] within their scope" (Government Council of the Canton of Solothurn 2016).[98] In other words, the government council only decides formally on the budget allocation, whereas the budget cuts effectively evolve during dialogues between the finance department and the spending departments. Concretely, the finance minister, backed up by the head of the finance office and one of his controllers, meet the spending minister who in turn brings his department secretary along. The finance delegation then asks the spending minister to reduce the expenditures by a certain amount derived from the financial plan, previous budgets, and the most recent financial statement. The spending delegation then either complies or it needs to explain why the expected cuts are not possible (Bühlmann 2016); according to the head of the finance office, the focus lies on expenditure cuts that improve the fiscal balance, which implicitly excludes special financings. These activities are found in the debate prevention theory. While the questioning of the financial delegation comes rather close to the idea of the monitoring effort (i.e. e), the rationalization of the spending delegation corresponds to the dossier quality (i.e. q). Since these talks are poorly documented and not public, it turns out to be challenging to find even more explicit empirical fingerprints to show the influence of earmarking. Therefore, Figure 19 does not depict these activities in detail.

Once the government council submits the draft budget to the parliament, the theory does not predict specific amendments to domains without large special financings. These domains are supposed to provide their services efficiently since they were already subject to monitoring by the other ministers. For the domains of culture and transportation, the further procedure depends on the intensity of the link between the taxed and the benefitting good. When the link is strong, the parliament supposedly advocates for a well-organized interest group that asks for less inputs (path c1 in Figure 19). Since it is the same group that pays and benefits, their aim is to make the specific service more efficient. When the link is weak, no such pressure exists as the payers differ from the beneficiaries (path c2). In this case, it might even appeal to some parliamentary groups to serve their clientele to increase the inputs and thus reduce efficiency.

In reality, the draft budget first passes by the parliamentary commissions before it arrives at the assembly (steps [3.1] – [3.4] of Table 1). Particularly, the finance commission organizes in committees following the departmental division to debate the various domains. Accordingly, the committee for building and justice treats the transportation domain, while the committee for education and culture obviously covers the cultural domain. In the latter, the committee accepted the government council's budgetary proposal for the cultural domain without any discussion in their meeting on October 18, 2016. Likewise, in their meeting on September 14, 2016 the

he refuses to give his opinion regarding the efficiency of the Museum *Altes Zeughaus* (a recent and concrete cultural project partly funded with a special financing) which he classifies as a subjective question. Applying the efficiency concept there is difficult according to him.

[98] Translated from German by the author.

committee for building and justice waved through the budgetary proposals for the transportation domain without discussion; upon request, several commission members confirmed their limited interest in discussing domains where most of the revenues stem from either earmarked revenues, transfers, or charges. This means that the first stage of the parliamentary phase does not entirely echo the predictions of the theory. Indeed, they muddled through without encountering critical questions, but likewise they did not benefit from additional allowances.

Subsequently, the different committees carried their pre-discussed parts of the budget in the meeting of the plenary finance commission on October 25 and 26, 2016. Suddenly, a previously ignored domain now moved into more focus. Several attendees, typically related to the road lobby, accused the chief of the road office of gilding the roads. Consequently, they demanded a budget reduction (for capital expenditures) of 10 percent, but this did not gain a necessary majority in the vote. Still, the culture domain did not spark a debate. This incident shows the increasing pressure on the transportation budget in comparison to the funds for the cultural domain. At the end of the commission's debate, the budget foresaw a financial performance surplus of 1.9 million Swiss Francs that was mainly caused by higher expected tax revenues.

Having passed the commissions, the budget reached the assembly (step [3.7] of Table 1), which debated the budget in three sessions (December 6, 7 and 14, 2016). Unsurprisingly, the initiated discussion about the transportation domain continued heatedly. The transportation minister identified the reason for such high emotions during the debate: "There are discussions because we have many interest groups as cyclists, the motorized individual transport, the public transport, the pedestrian, and people with disabilities." Likewise, he brought up an objection to the accusation that his service is excessive: "The expansions are often based on norms, on discussions with the municipalities, and on concessions to the different interest groups" (Parliament's secretariat of the canton of Solothurn 2016, p.920).[99] Despite the dedicated debate, the renewed proposal to reduce the budget by 10 percent failed. The significantly lower expenditures in the enacted budget of the Department of Building and Justice were the result of lower appropriations due to delayed capital expenditures for building constructions. The budget reductions were simply an accounting artefact and had nothing to do with the predicted cuts in the transportation domain. Apparently, the organized interest group acts as a disciplinary factor only if it can achieve its power in the parliament. Indeed, Figure 38 (in the appendix) supports this hypothesis as the positive effect of the expenditures through special financings on efficiency increases with the share of civic parties in the parliament.[100]

Note that in the entire process, neither the governing council, nor the parliament or its commissions decided on output reductions. In consequence, all reductions of the budget automatically translate into efficiency gains. Surely, it is another question as to whether these gains are achievable in the execution of the budget. In any case, it puts the administration under pressure to reduce the bureaucratic slack if there is any. In addition, the maintenance of the output level supports the earlier claim, that the input-oriented efficiency measures are adequate.

In sum, this search for empirical fingerprints confirming the theorized mechanism does support some of the predictions. First of all, the parliament did exert some financial pressure on the

[99] Both quotations are translated from German by the author.
[100] The respective estimates are not reported in any table but available upon request.

governing council which then triggered the latter to cut the budget. Yet, the mutual appearance of these two elements does not mean that the governing council would not have cut the budget to the same extent even if the parliament had given the governing council a long leash financially. Second, the budget in the domains of transportation and culture increases in a first stage while most of the other departments have to accept cuts. There is some evidence, based on a statement made by the head of the financial office, that the special financing has an influence on the demands for budget cuts. However, ministers with large special financings do not seem to prevent the debate because their colleagues are aware of the funds' fixed allocation, but more because an expenditure reduction within special financings is irrelevant for the budget balance.[101] Third, the strong congruence of the two groups paying in and benefitting from special financings does seem to have a disciplining effect on the expenditures. The respective effect only has an influence, however, if the group has enough power to prevail in the parliament.

8.3. Interim conclusion

Process-tracing as part of the qualitative method family has a completely different ontological understanding of the world than the previously applied quantitative approaches. Because of its rigorous demand for a mechanism linking X and Y instead of a simple correlation, process-tracing can identify necessary clarifications and even gaps in the theory. As an analytical grid, the debate prevention theory is translated into a mechanism that should leave empirical fingerprints if it really exists. This grid then tests the validity of the theory using the budgeting procedure of the canton of Solothurn in 2016.

Probably one of the most theory-challenging insights concerns the point where the actual individual budget cuts (and hence the budget allocation) take place. While the debate prevention theory regards the governing council as the dominant allocation authority, the case study proposes a different venue. In fact, the governing council actually sets the overall budgetary objective, but it does not debate the effective budget allocation. This happens instead in the bilateral meetings with the finance department (i.e. the minister and the head of the finance office) and the individual spending departments (i.e. the minister and the department secretary), whereas the former usually demands specific budget cuts while the latter justifies his financial needs. Still, the debate prevention theory holds as far as the finance department asks for less budgetary cuts from services funded through special financings, as they have no influence on the budget balance. From a theoretical point of view, this means that the spending ministers still have an incentive to increase the quality of their dossier in order to protect their budget share. However, it is the finance minister who puts some effort into monitoring the spending ministers in order to achieve the parliamentary and the governing council's objective to balance the budget.

---

[101] Nevertheless, in the previous year's budgetary debate (i.e. 2015), the finance minister requests "not to appropriate the entire revenue of the heavy vehicle fee to the special financing for roads but only half of it. The special financing has the effect that it ties available equity as soon as the fund becomes positive. In the worst case, we need to trigger the deficit brake even if the canton closes with a positive financial result because the special financing ties so many resources" (Parliament's secretariat of the canton of Solothurn 2015, p.861) (Translated from German by the author). This statement shows his discontent about the reduced scope of domains that cannot benefit from that special financing.

Apart from that, process-tracing provides some evidence that the parliament aims to cut the transportation budget instead of the cultural budget. Even if both domains are funded with special financings, the transportation domain is monitored by an organized interest group that promotes an efficient use of resources. At the same time, the culture domain was hardly debated during the budgeting procedure.

9. General conclusion

Public finance scholars and practitioners might already have had the feeling that earmarking revenues is a widespread practice in the Swiss cantons. But because earmarked funds as a subject have successfully snuck past any thorough scientific study, that there is not enough temporal and spatial comparable data to make substantiated statements possible. To fill this gap in the knowledge, the first step was to set about creating a database that to work with. Establishing a database of the Swiss cantons from 2000 to 2014 initially involved a careful definition of the term 'earmarked fund'. A fleeting glimpse over the cantonal financial statements is enough to observe the heterogeneity among cantons in terms of the interpretation and the utilization of special financings and comparable instruments (such as special funds, prefinancings etc.). Accordingly, a simple definition based on plain accounts would have missed the target and an identification of the relevant characteristics more appropriate. An earmarked fund is considered as such, if the financial statement reports it as a stock account that is exclusively funded by externally set earmarked revenues and that must be exclusively used for a specific and narrowly predefined purpose.

The database resulting from this definition revealed that the shares of earmarked revenues differed vastly across government functions. On average over all cantons from 2000 to 2015, the traffic (24 percent) and culture (35 percent) domain reported the highest shares of expenditures through special financings. These two government functions amounted to seven (traffic) and two (culture) percent of the total cantonal expenditures. For the accumulated earmarked revenues (i.e. special financings), they represented about one percent of the balance sheet on average, but they also increased to nearly four percent in some cantons. In a temporal comparison, the expenditures funded with special financings remained roughly stable over time. This stagnation is surprising given the reform of the fiscal equalization between the central state and the cantons in 2008 which was meant to drastically cut the earmarked part of transfers (Swiss Federal Council 2001). Conversely, the volume of special financings related to transportation expanded significantly in some cantons after 2008; there was no comparable pattern for special financings containing earmarked revenues for cultural expenditures. It seems that the earmarked revenues in some cantons exceeded the expenditures in the traffic domain such that the respective special financings accumulate more and more appropriated funds. This phenomenon leaves open the question as to whether earmarking enhances efficiency – and thus spares financial resources that accumulate in special financings – or if earmarking fosters inefficiency.

A literature review provides little help in predicting whether earmarking has a positive or negative effect in terms of efficiency. The overall conclusion that can be drawn from the literature is mainly based on the presumed motive of politicians. On the one side, public administration scholars disapprove of earmarking as it prevents the benevolent minister from optimally allocating the budget (Schönenberger 2013). On the other side, public choice proponents focus on the controlling function that earmarking can have to restrain selfish politicians (Buchanan 1963). At the least, public choice literature works out some conditions under which their arguments in favor of earmarking should hold. First, a strong link between the taxed good and the benefiting public service or good; second, a significant differentiation from other public goods or services; third, a strong homogeneity among payers of the taxed good and the beneficiaries of the good financed by the earmarked revenue along with an excludability of those who do not pay (Brazer 1984; Lee & Wagner 1991; McCleary 1991). Nonetheless, supporters of earmarking have so far failed to explain why it might actually have an opposite effect if these conditions are not fulfilled. This is where the debate prevention theory can fill in some of the gap.

Within the public choice framework, the debate prevention theory is grounded on established median voter models with uncertainty and ideological preferences (Persson & Tabellini 2002). These models show that even under competition, politicians can extract rents depending on transaction cost. Now, debate prevention – referring to the Swiss cantonal government system with several equally empowered ministers – claims that the transaction cost is a function of the monitoring activities of the other ministers. The ministers in turn have less incentive to monitor their colleagues if the latter benefit from earmarked revenues. Since these funds are appropriated for a specific domain, the ministers cannot shift the earmarked part of the budget to their department even when engaging in extensive monitoring. This means, then, that higher shares of earmarked revenues are predicted to diminish efficiency. This mechanism might particularly intervene in the Swiss cantonal budgeting process with its rather specific organization of the executive branch of government. Yet, also ministers working for a president fight for their part of the total budget and the debate prevention potentially also has its effects in such a governing system.

To test the prediction, two approaches proved successful in the literature. Albeit both use a production frontier as a reference to estimate the efficiency of each canton, their underlying assumptions are fundamentally different. The Stochastic Frontier Analysis (SFA) specifies a functional form which makes it possible to separate efficiency from stochastic noise. Also, the method already developed in the spheres of panel data and exploits them accordingly. In contrast, the Data Envelopment Analysis (DEA) forms the production frontier using the observed data points and thereby follows a deterministic logic. Even if today's DEA estimators acknowledge a certain noise in the data, their application on panel data leaves room for improvement. The study remedies this shortcoming by proposing an algorithm to estimate the effect of environmental variables on efficiency with fixed effects. The respective Monte Carlo simulations reveal the dominance of the proposed algorithm in terms of accuracy in comparison to conventional estimators.

The empirical test of the debate prevention theory indeed yields significant results – both in a statistical and a practical sense. Apparently, the effect that earmarking has on efficiency depends not only on the domain under investigation but also on how well the special financings are funded, on the restrictiveness of the debt brake, on the share of tax revenues, on the fractionization within the government council, and on the concordance between the executive and the legislative branch of government. In general, earmarking negatively affects efficiency in the cultural domain whereas the opposite is true for the transportation domain. A possible explanation can be found by returning to the public choice literature. With reference to the three criteria for a successful implementation of earmarking, the transportation domain meets all of them considerably well. When looking at the cultural domain however, the correlation between the lottery (the taxed good) and the benefitting service (culture) appears rather weak. Furthermore, many different goods and services can be gathered into the category of culture, which makes it difficult to clearly differentiate from other goods. Finally, the non-paying beneficiaries of culture are difficult to clearly exclude.

Based on the estimations and on the results from the qualitative analysis, the following policy recommendations impose themselves. Even when politically alluring, the government should carefully contemplate before installing a new special financing funded with earmarked revenues. It might well be that linking specific taxes to a policy program can enhance their acceptance among voters (Baranzini & Carattini 2017). Also, it is undoubtedly tempting to bind future governments through earmarked funds (Brett & Keen 2000; Jackson 2013). However, if the designated government function fails to meet the above-mentioned criteria, it is very likely that the special financing will lead to a waste of taxpayers' money. It is therefore welcome, that the Swiss harmonized accounting standards ask for a legal foundation for every special financing (Conference of cantonal finance directors 2008). Since there is considerable autonomy in terms of how cantons may implement the HAM2 standards, some cantons consider executive ordinances or parliamentary decrees as a sufficient legal foundation. This juridical dodge circumvents the necessity of a popular vote, which would set a rather high hurdle for any new special financing.

In case the three above-mentioned criteria are met, a special financing can initiate a positive effect on efficiency. In order to make the mechanism work, the payers must correspond to the beneficiaries of the tax and they need a powerful representation in the parliament. Optimally, there is a high degree of concordance. That is, the party representation in the parliament resembles the one in the executive. Likewise, when the budget is generally more dependent on taxes (instead of other revenue sources like third-party contributions or transfers), then earmarking has a more positive effect on efficiency. When, however, the criteria do not apply, a strong debt brake can potentially offset the mechanism. Hence, besides other positive consequences (Luechinger & Schaltegger 2013; Feld et al. 2017), the debt brake favorably affects the debate prevention mechanism.

In addition, these insights on earmarked transfers shed new light on the fiscal equalization reform from 2008. Admittedly, the non-earmarked transfers decreased conspicuously in the year of the reform, but their share has since gradually increased. Taking into account their significant negative influence on efficiency in the transportation domain justifies a reconsideration of the system, in which, currently, the federal state collects the taxes (notably the petroleum tax on fuel and the heavy vehicle charges) and redistributes them to the cantons. During the recurring debate about road pricing, the Swiss Federal Council already considered in 2004 the possibility of repealing these taxes and replacing them with differentiated user charges (Swiss Federal Council 2004). Based on the results of the previous analysis, a direct cantonal collection of these charges appears reasonable.

Besides the policy implications, this study also provides material for future research based on gaps revealed within the existing literature or inherent unsolved shortcomings of this study.

Theory-wise, the study limited its scope relatively narrowly to the assessment of technical efficiency, while earmarking undoubtedly also affects the budget allocation. Accordingly, the results make no statement whatsoever regarding the contingent increase or loss in social welfare due to earmarked funds. Moreover, the study proposes one possible mechanism through which earmarking might affect efficiency. At the same time, it cannot exclude other channels that could potentially lead to the same outcome; one being, for instance, the efficiency gained by focusing on the actual provision of public services instead of putting lots of effort into political discussions. Future research projects could therefore try to disentangle the hypothesized mechanism and integrate a measure that approximates the monitoring activities within the executive government branch. Equally, the explanation for the opposing effects in the two government domains could be substantiated further. While the three conditions derived from the previous literature are reasonable, other reasons seem plausible too. Namely, the two government functions vary widely in terms of the dominant type of expenditures. While the cultural domain spends its budget share for current expenditures, the transportation function typically invests in roads. Constant maintenance of roads is typically cheaper than repairing them once they are in bad shape (Koch & Forster 2010). Earmarked funds make the resources available immediately which helps to smooth expenses. This might also explain the positive effect found in the transportation domain. Given the almost identical effect of earmarks on efficiency in the short-run (Model I) and the long-run (Model VI) model, this latter argument appears less pertinent. The same robustness check refutes a further skepticism referring to the delay in the production process. When the transportation office spends an earmarked Swiss franc in one year, increasing thereby the earmarked share, one can hardly measure a longer road in the very same year. Ignoring the other outputs, one might expect a lower efficiency together with a higher share of expenditures through special financings. This is not the case though and, again, there is apparently no different long- and short-run effect.

One criticism almost every efficiency study faces concerns the lack of sufficient output measures. Particularly in the transportation domain, the unobserved variables might trigger a serious bias. Imagine most of the cantons fund their roads mainly through earmarked revenues, while the public transportation provision – which is also part of this government function – relies on the general fund. Note furthermore that all estimations entirely exclude any measure approximating the amount of public transport. If a canton puts more resources into public transport, it decreases its share of earmarked funding, while the *observed* output remains stable. Consequently, the efficiency decreases. On the contrary, if a canton spends more for roads funded by earmarked revenues, the share of expenditures through special financings rises together with the output. If the output increases more than the input, then efficiency improves. Altogether, in such a world, earmarking would be positively correlated with efficiency only due to the unobserved outputs related to public transportation. The crucial question is, whether cantons fund private and public transport differently. In principle, a comparison of the road fund laws with the expenses covered by the transportation function indicates that this should not systematically be the case. In addition, one would observe exactly the opposite in the cultural domain, which does not seem reasonable either. Indeed, there are also certainly unobserved outputs in the cultural domain. But no obvious argument suggests that the unobserved rather than the observed outputs are funded with earmarked revenues.

Digging deeper into the effective mechanisms in the government councils might also involve investigating the drivers of special financings, i.e. explaining the varying popularity of earmarks among cantons. The newly designed database on special financings could be moved from the left- to the right-hand side of the estimation equation providing insights regarding the motivation to introduce, adjust, and remove earmarks. A relatively broad theoretical literature already holds available a number of hypotheses waiting to be tested.

While the further revenue features associated with the fiscal illusion theory were not particularly responsive in the estimated models, their operationalizations might still prove useful for further research. The progressivity of the tax system is worth mentioning explicitly. By fitting a logarithmic function into the income-tax-rate-plot the herein used approach copes without micro data and still covers a very broad income range. With the discernible variation between cantons and years, the variable might likewise be used as dependent variable in a future research project.

In terms of the interest of the research community, the methodological strand also seems promising. In their review of papers related to efficiency analysis, Lampe & Hilgers (2015) highlight the phenomenal increase of papers applying DEA and SFA. Given this impressive development, the research on how to implement panel data in the two-stage DEA approach definitely needs to be pushed forward in order to receive more accurate results. The herein proposed quasi fixed effects algorithm presents one possibility.

Bibliography

Adam, A., Delis, M. & Kammas, P., 2011. Public sector efficiency: leveling the playing field between OECD countries. *Public Choice*, 146(1), pp.163–183.

Adam, A., Delis, M.D. & Kammas, P., 2008. Fiscal decentralization and public sector efficiency: evidence from OECD countries.

Afonso, A. & Fernandes, S., 2008. Assessing and explaining the relative efficiency of local government. *The Journal of Socio-Economics*, 37(5), pp.1946–1979.

Afonso, W.B., 2015. Leviathan or flypaper: Examining the fungibility of earmarked local sales taxes for transportation. *Public Budgeting & Finance*, 35(3), pp.1–23.

Afriat, S.N., 1972. Efficiency estimation of production functions. *International Economic Review*, pp.568–598.

Agasisti, T. & Wolszczak-Derlacz, J., 2014. Exploring universities' efficiency differentials between countries in a multi-year perspective: an application of bootstrap DEA and Malmquist index to Italy and Poland, 2001-2011.

Aigner, D., Lovell, C.A.K. & Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. *journal of Econometrics*, 6(1), pp.21–37.

Amsler, C., Prokhorov, A. & Schmidt, P., 2016. Endogeneity in stochastic frontier models. *Journal of Econometrics*, 190(2), pp.280–288.

Andor, M. & Parmeter, C., 2017. Pseudolikelihood estimation of the stochastic frontier model. *Applied Economics*, 49(55), pp.5651–5661.

Anesi, V., 2006. Earmarked taxation and political competition. *Journal of Public Economics*, 90(4), pp.679–701.

Aragon, Y., Daouia, A. & Thomas-Agnan, C., 2005. Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21(02), pp.358–389.

Ashworth, J. et al., 2014. Competition in the political arena and local government performance. *Applied Economics*, 46(19), pp.2264–2276.

Bădin, L., Daraio, C. & Simar, L., 2014. Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, 214(1), pp.5–30.

Badunenko, O., Henderson, D.J. & Kumbhakar, S.C., 2012. When, where and how to perform efficiency estimation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4), pp.863–892.

Badunenko, O. & Mozharovskyi, P., 2016. Nonparametric frontier analysis using Stata. *Stata Journal*, 16(3), pp.550–589.

Bailey, S.J. & Connolly, S., 1998. The flypaper effect: Identifying areas for further research. *Public choice*, 95(3-4), pp.335–361.

Balaguer-Coll, M.T. & Prior, D., 2009. Short-and long-term evaluation of efficiency and quality. An application to Spanish municipalities. *Applied Economics*, 41(23), pp.2991–3002.

Balaguer-Coll, M.T., Prior, D. & Tortosa-Ausina, E., 2007. On the determinants of local government performance: A two-stage nonparametric approach. *European Economic Review*, 51(2), pp.425–451.

Banker, R.D., Charnes, A. & Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), pp.1078–1092.

Banker, R.D., Janakiraman, S. & Natarajan, R., 2004. Analysis of trends in technical and allocative efficiency: An application to Texas public school districts. *European Journal of Operational Research*, 154(2), pp.477–491.

Banker, R.D. & Natarajan, R., 2008. Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations research*, 56(1), pp.48–58.

Baranzini, A. & Carattini, S., 2017. Effectiveness, earmarking and labeling: testing the acceptability of carbon taxes with survey data. *Environmental Economics and Policy Studies*, 19(1), pp.197–227.

Barro, R.J., 1988. *The Ricardian approach to budget deficits*, National Bureau of Economic Research Cambridge, Mass., USA.

Barros, C.P., Assaf, A. & Sá-Earp, F., 2010. Brazilian football league technical efficiency: a Simar and Wilson approach. *Journal of Sports Economics.*

Barros, C.P. & Dieke, P.U.C., 2008a. Measuring the economic efficiency of airports: A Simar–Wilson methodology analysis. *Transportation Research Part E: Logistics and Transportation Review*, 44(6), pp.1039–1051.

Barros, C.P. & Dieke, P.U.C., 2008b. Technical efficiency of African hotels. *International Journal of Hospitality Management*, 27(3), pp.438–447.

Battese, G.E. & Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, 20(2), pp.325–332.

Battese, G.E. & Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of econometrics*, 38(3), pp.387–399.

Battese, G.E. & Corra, G.S., 1977. Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. *Australian journal of agricultural economics*, 21(3), pp.169–179.

Beach, D. & Pedersen, R.B., 2013. *Process-tracing methods: Foundations and guidelines*, University of Michigan Press.

Benito, B., Bastida, F. & García, J.A., 2010. Explaining differences in efficiency: an application to Spanish municipalities. *Applied economics*, 42(4), pp.515–528.

Bird, R.M., 1982. Budgeting and Expenditure Control in Columbia. *Public Budgeting & Finance*, 2(3), pp.87–99.

Bird, R.M. & Jun, J., 2005. Earmarking in theory and Korean practice. *ITP Paper*, 513.

Bischoff, I. et al., 2017. Vertical grants and local public efficiency: The inference-disturbing effect of fiscal equalization. *Public Finance Review*, p.1091142117725393.

Blankart, C.B., 2012. *Öffentliche Finanzen in der Demokratie: eine Einführung in die Finanzwissenschaft*, Vahlen.

Boetti, L., Piacenza, M. & Turati, G., 2012. Decentralization and local governments' performance: how does fiscal autonomy affect spending efficiency? *FinanzArchiv: Public Finance Analysis*, 68(3), pp.269–302.

Bogetoft, P., 2013. *Performance benchmarking: Measuring and managing performance*, Springer Science & Business Media.

Bönisch, P. et al., 2011. *Municipality size and efficiency of local public services: Does size matter?*, Halle Institute for Economic Research.

Borge, L.-E., Falch, T. & Tovmo, P., 2008. Public sector efficiency: the roles of political and budgetary institutions, fiscal capacity, and democratic participation. *Public Choice*, 136(3-4), pp.475–495.

De Borger, B. & Kerstens, K., 1996. Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, 26(2), pp.145–170.

Bös, D., 2000. Earmarked taxation: Welfare versus political support. *Journal of Public Economics*, 75(3), pp.439–462.

Bradford, D.F. & Oates, W.E., 1971. The analysis of revenue sharing in a new approach to collective fiscal decisions. *The Quarterly Journal of Economics*, pp.416–439.

Brazer, H.E., 1984. Some observations on earmarking. *Proceedings of the annual conference on taxation held under the auspices of the national tax association-tax institute of America*, 77, pp.269–273.

Brennan, G., 2012. The political economy of public debt. *Constitutional Political Economy*, 23(3), pp.182–198.

Brennan, G. & Buchanan, J.M., 1978. Tax instruments as constraints on disposition of public revenues. *Journal of Public Economics*, 9(3), pp.301–318.

Brett, C. & Keen, M., 2000. Political uncertainty and the earmarking of environmental taxes. *Journal of Public Economics*, 75(3), pp.315–340.

Brunner, E.J. & Squires, T., 2013. The bargaining power of teachers' unions and the allocation of school resources. *Journal of Urban Economics*, 76, pp.15–27.

Buchanan, J.M., 1967. *Public finance in democratic process: Fiscal institutions and individual choice*, UNC

Press Books.

Buchanan, J.M., 1963. The economics of earmarked taxes. *The Journal of Political Economy*, pp.457–469.

Bühlmann, A., 2016. Interview on 08-29-2016.

Carroll, D.A., 2009. Diversifying municipal government revenue structures: Fiscal illusion or instability? *Public Budgeting & Finance*, 29(1), pp.27–48.

Cazals, C., Florens, J.-P. & Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *Journal of econometrics*, 106(1), pp.1–25.

Chalmers, A.F., 2013. *What is this thing called science?*, Hackett Publishing.

Charnes, A., Cooper, W.W. & Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp.429–444.

Chen, A., Hwang, Y. & Shao, B., 2005. Measurement and sources of overall and input inefficiencies: Evidences and implications in hospital services. *European Journal of Operational Research*, 161(2), pp.447–468. Available at: http://www.sciencedirect.com/science/article/pii/S0377221703007021.

Chernick, H., 2005. On the determinants of subnational tax progressivity in the US. *National Tax Journal*, pp.93–112.

Chortareas, G.E., Girardone, C. & Ventouri, A., 2013. Financial freedom and bank efficiency: Evidence from the European Union. *Journal of Banking & Finance*, 37(4), pp.1223–1231.

Conference of cantonal finance directors, 1981. *Handbook of Accounting of the Public Authorities*, Berne: Verlag Paul Haupt Bern.

Conference of cantonal finance directors, 2008. *Harmonized accounting model for the Swiss cantons and municipalities* Version 5., Berne: Conference of cantonal finance directors. Available at: http://www.srs-cspcp.ch/en/recommendations-conference-cantonal-finance-directors-n17985.

Cooper, W.W., Seiford, L.M. & Tone, K., 2007. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software* Second Edi., Springer US. Available at: https://books.google.ch/books?id=XW0sswC0RzsC.

Cordero, J.M. et al., 2017. Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach. *Journal of Productivity Analysis*, 48(1), pp.1–24.

Courant, P.N., Gramlich, E.M. & Rubinfeld, D.L., 1979. The stimulative effects of intergovernmental grants: Or why money sticks where it hits. In P. Mieszkowski & W. H. Oakland, eds. *Fiscal Federalism and Grants-in-Aid*. Washington DC: The Urban Institute, pp. 127–143.

Crowley, G. & Hoffer, A., 2012. *Dedicating tax revenue: Constraining government or masking its growth*, Mercatus center working paper 12-17. George Mason University.

Da Cruz, N.F. & Marques, R.C., 2014. Revisiting the determinants of local government performance. *Omega*, 44, pp.91–103.

Daraio, C., Simar, L. & Wilson, P.W., 2010. Testing whether two-stage estimation is meaningful in non-parametric models of production. *Institut de Statistique, UCL, Belgium.*

Davis, M.L. & Hayes, K., 1993. The demand for good government. *The review of economics and statistics*, pp.148–152.

Deller, S.C. & Maher, C.S., 2005. Categorical municipal expenditures with a focus on the flypaper effect. *Public Budgeting & Finance*, 25(3), pp.73–90.

Deprins, D., Simar, L. & Tulkens, H., 2006. Measuring labor-efficiency in post offices. In *Public goods, environmental externalities and fiscal competition*. Springer, pp. 285–309.

Deran, E., 1965. Earmarking and expenditures: A survey and a new test. *National Tax Journal*, 18(4), pp.354–361.

Dhillon, A. & Perroni, C., 2001. Tax earmarking and grass-roots accountability. *Economics Letters*, 72(1), pp.99–106.

Dimas, G., Goula, A. & Soulis, S., 2012. Productive performance and its components in Greek public hospitals. *Operational Research*, 12(1), pp.15–27.

Dollery, B.E. & Worthington, A.C., 1996. The empirical analysis of fiscal illusion. *Journal of Economic Surveys*, 10(3), pp.261–297.

Dye, R.F. & McGuire, T.J., 1992. The effect of earmarked revenues on the level and composition of expenditures. *Public Finance Review*, 20(4), pp.543–556.

Eidgenössische Steuerverwaltung, 2015. *Die kalte Progression*, Berne.

Evans, W.N. & Zhang, P., 2007. The impact of earmarked lottery revenue on K–12 educational expenditures. *Education*, 2(1), pp.40–73.

Farrell, M.J., 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), pp.253–290.

Federal Finance Administration, 2016. *Financial statistic - Reporting*, Berne. Available at: https://www.efv.admin.ch/efv/en/home/themen/finanzstatistik/berichterstattung.html.

Federal Office for Spatial Development, 2009. *Verwendung des Ertragsanteils der Kantone an der Leistungsabhängigen Schwerverkehrsabgabe (LSVA)*, Berne. Available at: https://www.are.admin.ch/are/de/home/medien-und-publikationen/publikationen/verkehr/verwendung-des-ertragsanteils-der-kantone-an-der-leistungsabhang.html.

Federal Statistical Office, 2016. *Monatlicher Bruttolohn nach Wirtschaftszweigen und Grossregionen - Privater und öffentlicher Sektor zusammen*, Available at: https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/wages-income-employment-labour-costs/wage-levels-major-regions.assetdetail.327882.html.

Feld, L.P. et al., 2017. Sovereign bond market reactions to no-bailout clauses and fiscal rules–The Swiss experience. *Journal of International Money and Finance*, 70, pp.319–343.

Feld, L.P. & Kirchgässner, G., 2008. On the effectiveness of debt brakes: the Swiss experience. *Sustainability of public debt*, pp.223–255.

Feld, L.P. & Reulier, E., 2009. Strategic tax competition in Switzerland: Evidence from a panel of the Swiss cantons. *German Economic Review*, 10(1), pp.91–114.

Filippini, M. & Wetzel, H., 2014. The impact of ownership unbundling on cost efficiency: Empirical evidence from the New Zealand electricity distribution sector. *Energy Economics*, 45, pp.412–418.

Financial Statistics Section, 2011. *Financial Statistics Methods and Concepts in Switzerland*, Berne. Available at: https://www.efv.admin.ch/dam/efv/en/dokumente/finanzstatistik/methoden/FS-

Methodenband_e.pdf.download.pdf/FS-Methodenband_e.pdf.

Finnigan, T., 2007. *All About Pork: The Abuse of Earmarks and the Needed Reforms*, Washington DC.

Fleishman, R. et al., 2009. Does regulation stimulate productivity? The effect of air quality policies on the efficiency of US power plants. *Energy Policy*, 37(11), pp.4574–4582. Available at: http://www.sciencedirect.com/science/article/pii/S0301421509004273.

Frank, R.H., 2006. *Microeconomics and behavior*, McGraw-Hill/Irwin New York.

Frey, B.S. & Stutzer, A., 2000. Happiness, economy and institutions. *The Economic Journal*, 110(466), pp.918–938.

George, A.L. & Bennett, A., 2005. *Case studies and theory development in the social sciences*, mit Press.

Geys, B., Heinemann, F. & Kalb, A., 2010. Voter involvement, fiscal autonomy and public sector efficiency: Evidence from German municipalities. *European Journal of Political Economy*, 26(2), pp.265–278.

Geys, B. & Moesen, W., 2009. Exploring sources of local government technical inefficiency: evidence from Flemish municipalities. *Public Finance and Management*, 9(1), p.1.

Gijbels, I. et al., 1999. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, 94(445), pp.220–228.

Glazer, A. & Proost, S., 2007. *Earmarking: Bundling to signal quality*,

Goel, R.K. & Nelson, M.A., 1999. The political economy of motor-fuel taxation. *The Energy Journal*, pp.43–59.

Goel, R.K. & Nelson, M.A., 2003. Use or abuse of highway tax revenues? An economic analysis of highway spending. *Applied Economics Letters*, 10(13), pp.813–819.

Goetz, C.J., 1968. Earmarked taxes and majority rule budgetary processes. *The American Economic Review*, 58(1), pp.128–136.

Governing Council of the Canton of Obwalden, 2005. *Botschaft des Regierungsrats zur Teilrevision des Steuergesetzes auf den 1. Januar 2006*, Sarnen. Available at: http://www.ow.ch/dl.php/de/4656f2506856c/Steuer_Botschaft.pdf.

Governing Council of the Canton of Uri, 2008. *Bericht und Antrag des Regierungsrats an den Landrat zur Teilrevision des Gesetzes über die direkten Steuern (StG)*, Altdorf. Available at: http://www.ur.ch/dl.php/de/50066e896f65f/25.pdf.

Government Council of the Canton of Solothurn, 2016. *Weisungen und Terminplan zur Erstellung des Voranschlages 2017 sowie Einleitung erster Korrekturmassnahmen (Budgetvorentscheide I)*, Solothurn: Cantonal Council Decree 2016/871.

Gramlich, E.M. et al., 1973. State and local fiscal behavior and federal grant policy. *Brookings Papers on Economic Activity*, 1973(1), pp.15–65.

Greene, W.H., 1990. A gamma-distributed stochastic frontier model. *Journal of econometrics*, 46(1-2), pp.141–163.

Greene, W.H., 2008. Censored Data and Truncated Distributions. *SSRN Working Paper Series.*

Greene, W.H., 2003. *Econometric analysis*, Pearson Education India.

Greene, W.H., 2005. Fixed and random effects in stochastic frontier models. *Journal of productivity analysis*, 23(1), pp.7–32.

Greene, W.H., 1980. Maximum likelihood estimation of econometric frontier functions. *Journal of econometrics*, 13(1), pp.27–56.

Grossman, P.J., Mavros, P. & Wassmer, R.W., 1999. Public sector technical inefficiency in large US cities. *Journal of Urban Economics*, 46(2), pp.278–299.

Gwilliam, K. & Shalizi, Z., 1999. Road funds, user charges, and taxes. *The World Bank Research Observer*, 14(2), pp.159–186.

Hirschberg, J.G. & Lloyd, P.J., 2002. Does the technology of foreign-invested enterprises spill over to other enterprises in China? An application of post-DEA bootstrap regression analysis. *Modelling the Chinese economy, Edward Elgar Press, London.*

Honoré, B.E., 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica: journal of the Econometric Society*, pp.533–565.

Jackson, J., 2013. Tax earmarking, party politics and gubernatorial veto: Theory and evidence from US states. *Public Choice*, pp.1–18.

Jankowski, J.E., 1984. The practice and prevelance of earmarking. *Proceedings of the Annual Conference on Taxation Held under the Auspices of the National Tax Association-Tax Institute of America*, 77, pp.258–267.

Jann, W. & Wegrich, K., 2007. Theories of the policy cycle. *Handbook of public policy analysis: Theory, politics, and methods*, 125, pp.43–62.

Johnson, A.L. & Kuosmanen, T., 2012. One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research*, 220(2), pp.559–570.

Jung, C., 2002. The effect of local earmarking on capital spending in Georgia counties. *State & Local Government Review*, pp.29–37.

Kalb, A., 2010a. *Public Sector Efficiency: Applications to Local Governments in Germany*, Springer Science & Business Media.

Kalb, A., 2010b. The impact of intergovernmental grants on cost efficiency: theory and evidence from German municipalities. *Economic Analysis and Policy*, 40(1), pp.23–48.

Kalb, A., 2009. *What determines local governments' technical efficiency? The case of road maintenance*,

Karakaplan, M.U. & Kutlu, L., 2017. Endogeneity in panel stochastic frontier models: an application to the Japanese cotton spinning industry. *Applied Economics*, 49(59), pp.5935–5939.

Kellermann, K., 2007. *Die öffentlichen Ausgaben der Kantone und ihrer Gemeinden im Quervergleich*,

Kirchgässner, G. et al., 2005. Fiskalischer Föderalismus und wirtschaftliche Entwicklung: Evidenz für die Schweizer Kantone. *Jahrbuch für Regionalwissenschaft*, 25, pp.3–25.

Kirjavainen, T. & Loikkanent, H.A., 1998. Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. *Economics of Education Review*, 17(4), pp.377–394.

Klingenberg, D., 2007. Ein begrenztes Vergnügen. *St. Galler Tagblatt*. Available at: http://www.tagblatt.ch/altdaten/tagblatt-alt/tagblattheute/sg/stgallen/tb-st/art742,136355.

Klopp, G.A., 1985. *The Analysis of the Efficiency of Productive Systems with Multiple Inputs and Outputs*. University of Illinois.

Kneip, A., Simar, L. & Wilson, P.W., 2008. Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24(06), pp.1663–1697.

Kneip, A., Simar, L. & Wilson, P.W., 2016. Testing hypotheses in nonparametric models of production. *Journal of Business & Economic Statistics*, 34(3), pp.435–456.

Kneip, A., Simar, L. & Wilson, P.W., 2015. When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory*, 31(02), pp.394–422.

Koch, B. & Forster, M., 2010. *Zustandsanalyse und Werterhaltung bei den Kantonsstrassen in der Schweiz*, Infra Suisse.

Kumbhakar, S.C., Lien, G. & Hardaker, J.B., 2014. Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis*, 41(2), pp.321–337.

Kumbhakar, S.C., Wang, H. & Horncastle, A.P., 2015. *A practitioner's guide to stochastic frontier analysis using Stata*, Cambridge University Press.

Lampe, H.W. & Hilgers, D., 2015. Trajectories of efficiency measurement: a bibliometric analysis of DEA and SFA. *European Journal of Operational Research*, 240(1), pp.1–21.

Lee, D.R. & Tollison, R.D., 1991. Rent seeking and tax earmarking. In *Charging for Government (Routledge Revivals): User Charges and Earmarked Taxes in Principle and Practice*. Routledge, p. 125.

Lee, D.R. & Wagner, R.E., 1991. The political economy of tax earmarking. In *Charging for Government (Routledge Revivals): User Charges and Earmarked Taxes in Principle and Practice*. Routledge, p. 110.

Liang, L., Li, Y. & Li, S., 2009. Increasing the discriminatory power of DEA in the presence of the undesirable outputs and large dimensionality of data sets with PCA. *Expert Systems with Applications*, 36(3), pp.5895–5899.

Lindbeck, A. & Weibull, J.W., 1987. Balanced-budget redistribution as the outcome of political competition. *Public Choice*, 52(3), pp.273–297.

Loikkanen, H.A. & Susiluoto, I., 2005. Cost efficiency of Finnish municipalities in basic service provision 1994-2002.

Luechinger, S. & Schaltegger, C.A., 2013. Fiscal rules, budget deficits and budget projections. *International Tax and Public Finance*, 20(5), pp.785–807.

Malmquist, S., 1953. Index numbers and indifference surfaces. *Trabajos de Estadistica y de Investigacion Operativa*, 4(2), pp.209–242.

Martin, M.-J., 2008. *Soldes financiers des collectivités publiques: Explications théoriques et modélisation simultanée des recettes et des dépenses des cantons suisses*, Peter Lang.

Martínez, L.R., 2014. Sources of revenue and government performance: Theory and evidence from colombia. *Working Paper*.

Le Maux, B., 2009. Governmental behavior in representative democracy: a synthesis of the theoretical literature. *Public Choice*, 141(3-4), pp.447–465.

McCleary, W., 1991. The earmarking of government revenue: A review of some World Bank experience. *The World Bank Research Observer*, 6(1), pp.81–104.

McMahon, W.W. & Sprenkle, C.M., 1970. A theory of earmarking. *National Tax Journal*, 23(3), pp.255–261. Available at: http://www.jstor.org/stable/41791721.

Meeusen, W. & Van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, pp.435–444.

Meyers, R.T., 1996. *Strategic budgeting*, University of Michigan Press.

Migué, J.-L., Belanger, G. & Niskanen, W.A., 1974. Toward a general theory of managerial discretion. *Public Choice*, 17(1), pp.27–47.

Mitias, P.M. & Turnbull, G.K., 2001. Grant illusion, tax illusion, and local government spending. *Public Finance Review*, 29(5), pp.347–368.

Musgrave, R.A. & Musgrave, P.B., 1989. *Public finance in theory and practice* Internatio., McGraw-Hill.

Nagin, D.S. et al., 2002. Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *The American Economic Review*, 92(4), pp.850–873.

Narbón-Perpiñá, I. & De Witte, K., 2017. Local governments' efficiency: A systematic literature review—part I. *International Transactions in Operational Research*.

Narbón-Perpiñà, I. & De Witte, K., Local governments' efficiency: A systematic literature review-part II. *International Transactions in Operational Research*.

Nash, J.F., 1950. The bargaining problem. *Econometrica: Journal of the Econometric Society*, 18(2), pp.155–162.

Navarro, N., 2002. *Does Earmarking matter? The case of state lottery profits and education spending*,

Nesbit, T.M. & Kreft, S.F., 2009. Federal grants, earmarked revenues, and budget crowd-out: State highway funding. *Public Budgeting & Finance*, 29(2), pp.94–110.

Niskanen, W.A., 1994. *Bureaucracy and Public Economics*, E. Elgar. Available at: https://books.google.ch/books?id=QDB-QgAACAAJ.

Niskanen, W.A., 1975. Bureaucrats and politicians. *The Journal of Law & Economics*, 18(3), pp.617–643.

Oakland, W.H., 1984. Earmarking and decentralization. *Proceedings of the Annual Conference on Taxation Held under the Auspices of the National Tax Association-Tax Institute of America*, 77, pp.274–277. Available at: http://www.jstor.org/stable/42911907.

Oates, W.E., 1975. Automatic increases in tax revenues: The effect on the size of the public budget. *Financing the New Federation. Revenue Sharing, Conditional Grants and Taxation, Johns Hopkins University Press, Baltimore*.

Oates, W.E., 1985. On the nature and measurement of fiscal illusion: A survey. In G. Brennan et al., ed. *Taxation and Fiscal Federalism: Essays in Honour of Russell Mathews*. Sydney: Australian National University Press, pp. 65–82.

OECD Publishing, 2001. *Measuring productivity-OECD manual: Measurement of aggregate and industry-level productivity growth*, Paris: Organisation for Economic Co-operation and Development. Available at: http://www.oecd.org/std/productivity-stats/2352458.pdf.

Olsen, K.R. & Rehnberg, C., 2008. Hospital productivity and the Norwegian ownership reform.

Ostrom, V., 2008. *The intellectual crisis in American public administration*, University of Alabama Press.

Parliament's secretariat of the canton of Solothurn, 2015. Verhandlungen des Kantonsrats - VII. Session – 8. Dezember 2015. In M. Ackermann, ed. *16. Sitzung*. Solothurn, pp. 825–888.

Parliament's secretariat of the canton of Solothurn, 2016. Verhandlungen des Kantonsrats - VII. Session 2016 – 6./7./14. Dezember. In M. Ackermann & B. Steinbrunner, eds. *16. Sitzung*. Solothurn, pp. 897–1051.

Pérez-López, G., Prior, D. & Zafra-Gómez, J.L., 2015. Rethinking new public management delivery forms and efficiency: long-term effects in Spanish local government. *Journal of Public Administration Research and Theory*, 25(4), pp.1157–1183.

Persson, T. & Tabellini, G.E., 2002. *Political economics: Explaining economic policy*, MIT press.

Pfäffli, S., 2011. *Budgetierung im öffentlichen Sektor: ein Handbuch für Studium und Praxis aus finanzwissenschaftlicher Sicht*, Haupt Verlag AG.

Polo, M., 1998. Electoral competition and political rents. *IGIER Working Paper*, 144. Available at: https://ssrn.com/abstract=143851.

Pommerehne, W.W. & Schneider, F., 1978. Fiscal illusion, political institutions, and local public spending. *Kyklos*, 31(3), pp.381–408.

Porcelli, F., 2014. Electoral accountability and local government efficiency: quasi-experimental evidence from the Italian health care sector reforms. *Economics of Governance*, 15(3), pp.221–251.

Porter, R. & Walsh, S., 2008. Earmarks in the federal budget process. In R. Allen & R. Brown, eds. *Federal Budget Policy Seminar, Briefing Paper*. Harvard Law School.

Poveda, A.C., 2011. Economic development and growth in Colombia: An empirical analysis with super-efficiency DEA and panel data models. *Socio-Economic Planning Sciences*, 45(4), pp.154–164. Available at: http://www.sciencedirect.com/science/article/pii/S0038012111000358.

Puenpatom, R.A. & Rosenman, R., 2008. Efficiency of Thai provincial public hospitals during the introduction of universal health coverage using capitation. *Health care management science*, 11(4), pp.319–338.

Puviani, A., 1903. *Teoria della illusione finanziaria* R. Sandron, ed., Palermo: Tip. Fratelli Vena.

Rae, D.W., 1967. *The Political Consequences of Electoral Laws*, Yale University Press. Available at: https://books.google.ch/books?id=ItwWAAAAIAAJ.

Ramli, N.A. & Munisamy, S., 2013. Modeling undesirable factors in efficiency measurement using data envelopment analysis: a review. *Journal of Sustainability Science and Management*, 8(1), pp.126–135.

Ray, S.C., 2004. *Data envelopment analysis: theory and techniques for economics and operations research*, Cambridge university press.

Richmond, J., 1974. Estimating the efficiency of production. *International economic review*, pp.515–521.

Roberts, A., 2002. Lockbox government: Segregated funding strategies and the decline of governmental flexibility. *Governance*, 15(2), pp.241–270.

Rodden, J., 2003. Reviving Leviathan: fiscal federalism and the growth of government. *International Organization*, 57(04), pp.695–729.

Rubin, I.S., 1990. *The politics of public budgeting: Getting and spending, borrowing and balancing*, CQ Press.

Schmidheiny, K., 2006. Income segregation and local progressive taxation: Empirical evidence from Switzerland. *Journal of Public Economics*, 90(3), pp.429–458.

Schönenberger, A., 2013. Öffentliche Finanzen der Schweiz. In A. Ladner et al., eds. *Handbuch der öffentlichen Verwaltung in der Schweiz*. Zurich: Verlag Neue Zürcher Zeitung, pp. 565–585.

Seifert, S. & Nieswand, M., 2014. What drives intermediate local governments' spending efficiency: The case of French Départements. *Local Government Studies*, 40(5), pp.766–790.

Seiford, L.M. & Zhu, J., 2002. Modeling undesirable factors in efficiency evaluation. *European journal of operational research*, 142(1), pp.16–20.

Selim, S. & Bursalıoğlu, S.A., 2015. Efficiency of Higher Education in Turkey: A Bootstrapped Two-Stage DEA Approach 1. *International Journal of Statistics and Applications*, 5(2), pp.56–67.

Shepard, R.W., 1953. *Cost and production functions*, Princepton University Press.

Simar, L. & Wilson, P.W., 1999. Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*, 115(3), pp.459–471.

Simar, L. & Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, 136(1), pp.31–64.

Simar, L. & Wilson, P.W., 2002. Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1), pp.115–132.

Simar, L. & Wilson, P.W., 2011a. Performance of the bootstrap for DEA estimators and iterating the principle. In *Handbook on data envelopment analysis*. Springer, pp. 241–271.

Simar, L. & Wilson, P.W., 2000. Statistical inference in nonparametric frontier models: The state of the art. *Journal of productivity analysis*, 13(1), pp.49–78.

Simar, L. & Wilson, P.W., 2011b. Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, 36(2), pp.205–218.

Social Democratic Party, 2014. NAF: Priorität auf Agglomerationen legen. Available at: https://www.sp-ps.ch/de/publikationen/espress/naf-prioritaet-auf-agglomerationen-legen [Accessed February 15, 2018].

Souza, G. da S. & Gomes, E.G., 2015. Management of agricultural research centers in Brazil: A DEA application using a dynamic GMM approach. *European Journal of Operational Research*, 240(3), pp.819–824.

Staub, R.B., e Souza, G. da S. & Tabak, B.M., 2010. Evolution of bank efficiency in Brazil: A DEA approach. *European journal of operational research*, 202(1), pp.204–213.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13(1), pp.57–66.

Stratmann, T., 2013. The effects of earmarks on the likelihood of reelection. *European Journal of Political Economy*, 32, pp.341–355.

Suits, D.B., 1977. Measurement of tax progressivity. *The American Economic Review*, 67(4), pp.747–752.

Swiss Federal Council, 2001. *Botschaft zur Neugestaltung des Finanzausgleichs und der Aufgaben zwischen Bund und Kantonen (NFA)*, Berne.

Swiss Federal Council, 2015. *Botschaft zur Schaffung eines Nationalstrassen- und Agglomerationsverkehrs-Fonds, zur Schliessung der Finanzierungslücke und zum Strategischen Entwicklungsprogramm Nationalstrassen (NAF-Botschaft)*, Berne: Federal Council.

Swiss Federal Council, 2012. *Botschaft zur Volksinitiative «Für den öffentlichen Verkehr» und zum* *direkten Gegenentwurf*, Berne: Federal Council.

Swiss Federal Council, 2013. *Mehr Transparenz bei Spezialfonds und Spezialfinanzierungen*, Berne. Available at: https://www.parlament.ch/centers/eparl/curia/2013/20134214/Bericht BR D.pdf.

Swiss Federal Council, 2004. *Stellungsnahme des Bundesrates vom 24.11.2004 zur Interpellation Walker - Verkehr. Finanzpolitische Standortbestimmung*, Bern: Executive.

Swiss Federal Council, 2014. *Wirksamkeitsbericht 2012–2015 des Finanzausgleichs zwischen Bund und Kantonen*, Berne.

Swiss People's Party, 2014. Vernehmlassung Vorlage zur Schaffung eines Nationalstrassen- und Agglomerations- verkehrs-Fonds (NAF); zur Schliessung der Finanzierungslücke; zum Strategischen Entwicklungsprogramm Nationalstrasse (STEP). Available at: https://www.svp.ch/partei/positionen/vernehmlassungen/vorlage-zur-schaffung-eines-nationalstrassen-und-agglomerations-verkehrs-fonds-naf-zur-schliessung-der-finanzierungsluecke-zum-strategischen-entwicklungsprogramm-nationalstrasse-step/ [Accessed February 15, 2018].

Sydsæter, K. & Hammond, P.J., 2009. *Mathematik für Wirtschaftswissenschaftler: Basiswissen mit Praxisbezug*, Pearson Deutschland. Available at: https://books.google.ch/books?id=OCjBcEwPN_EC.

Tauchmann, H., 2011. Partial frontier efficiency analysis for stata.

Teja, R.S., 1988. The case for earmarked taxes. *Staff Papers*, 35(3), pp.523–533.

Tiemann, O. & Schreyögg, J., 2012. Changes in hospital efficiency after privatization. *Health care management science*, 15(4), pp.310–326.

Tulkens, H. & Eeckenhaut, P. Vanden, 2006. Nonparametric efficiency, progress and regress measures for panel data: Methodological aspects. In P. Chander et al., eds. *Public goods, environmental externalities and fiscal competition*. Springer US, p. pp 395–429.

Tullock, G., 2008. Public choice. In S. N. Durlauf & L. E. Blume, eds. *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

Vatter, A. & Freitag, M., 2007. The contradictory effects of consensus democracy on the size of government: Evidence from the Swiss cantons. *British Journal of Political Science*, 37(02), pp.359–367.

Wagner, R.E., 1976. Revenue structure, fiscal illusion, and budgetary choice. *Public choice*, 25(1), pp.45–61.

Wang, H.-J., 2002. Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18(3), pp.241–253.

Wang, H.-J. & Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *journal of Productivity Analysis*, 18(2), pp.129–144.

Wanke, P.F., 2012. Determinants of scale efficiency in the Brazilian 3PL industry: a 10-year analysis. *International Journal of Production Research*, 50(9), pp.2423–2438.

Widmer, P. & Zweifel, P., 2012. Fiscal equalization, Tiebout competition, and incentives for efficiency in a federalist country. *Public Finance Review*, 40(1), pp.3–29.

Wilkinson, M., 1994. Paying for public spending: is there a role for earmarked taxes? *Fiscal studies*, 15(4), pp.119–135.

Wilson, P.W., 2003. Testing independence in models of productive efficiency. *Journal of Productivity Analysis*, 20(3), pp.361–390.

Winsten, C.B., 1957. Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society*, 120, pp.282–284.

Wu, J. et al., 2013. Congestion measurement for regional industries in China: A data envelopment analysis approach with undesirable outputs. *Energy Policy*, 57, pp.7–13.

Wyckoff, P.G., 1988. Bureaucracy and the "publicness" of local public goods. *Public Choice*, 56(3), pp.271–284.

Wyckoff, P.G., 1990. The simple analytics of slack-maximizing bureaucracy. *Public Choice*, 67(1), pp.35–47.

Xue, M. & Harker, P.T., 1999. Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach. *Unpublished Working Paper, Wharton Financial Institutions Center, University of Pennsylvania*.

Yerly, N., 2013. *The Political Economy of Budget Rules in the twenty-six Swiss cantons: institutional analysis, preferences and performances*. University of Fribourg Switzerland.

Zhang, N.J., Unruh, L. & Wan, T.T.H., 2008. Has the Medicare prospective payment system led to increased nursing home efficiency? *Health services research*, 43(3), pp.1043–1061.

Appendix

A. 1 Derivations of the mathematical statements related to the debate prevention

From equation (25), recall minister A's optimized objective with respect to e:

$$(1-\bar{\eta})\bar{g}\bar{\vartheta}_A\left(\frac{\partial\Psi_A(e,q)}{\partial e}\Phi(e,q)+\frac{\partial\Phi(e,q)}{\partial e}\Psi_A(e,q)\right)=\frac{\partial C(e,q)}{\partial e} \quad (59)$$

The first step is to show that equation (59) has only one solution e^* if $\bar{\eta}$ is fixed. For that to be the case, the curve on the left-hand side must intersect the curve on the right-hand side once. Equation (59)'s left-hand side term in the large brackets is the unweighted marginal benefit of the monitoring effort. For simplicity define this term as

$$z\equiv\left(\frac{\partial\Psi_A(e,q)}{\partial e}\Phi(e,q)+\frac{\partial\Phi(e,q)}{\partial e}\Psi_A(e,q)\right) \quad (60)$$

Remember the assumption from the text regarding the marginal cost, i.e. $\frac{\partial C(e=0,q)}{\partial e}=0$. Furthermore, note that, based on equation (60), $z>0$ if $e=0$. In other words, the left-hand side of (59) is positive for $e=0$ while the right-hand side is zero. Ergo, to obtain an intersection curve, the left-hand side's derivative must be smaller than the one of the right hand's side. If this is the case, the two curves necessarily intersect at one point as e becomes larger than zero. Unfortunately, the imposed assumptions made so far are not sufficient to guarantee a single intersection. Mathematically, the necessary condition for one single intersection curve is:

$$(1-\bar{\eta})\bar{g}\frac{\partial z}{\partial e}<\frac{\partial C(e,q)}{\partial e} \qquad \Leftrightarrow \qquad \frac{\partial z}{\partial e}>0>\frac{\frac{\partial C(e,q)}{\partial e}}{(\bar{\eta}-1)\bar{g}}\;\forall\, e\in[0,\infty) \text{ and } \forall\,\bar{\eta}\in[0,1] \quad (61)$$

For simplicity assume that $\frac{\partial z}{\partial e}>0$. It can be shown that a necessary condition for this assumption to hold is a sufficiently large second order derivative of the offensive bargaining power $\frac{\partial^2\Omega_D(e)}{\partial e^2}$. As stated in the text, the positive second order derivative means that marginal effect of the monitoring activity on the other minister's bargaining power decreases. The additional assumption simply requires the second derivative of the bargaining power to be sufficiently high, which translates into a rapidly shrinking bargaining power of the other ministers rapidly when minister A increases his monitoring activities.

Knowing that equation (59) has only one solution e^* if $\bar{\eta}$ is fixed, the next step is to determine the dependency of the e^* if η is variable. Put differently, what is the minister's optimal effort level depending on the earmarked share? To answer this question, solve (59) for $\bar{\eta}$ and take the first derivative with respect to e:

$$\frac{\partial\eta}{\partial e}=-\frac{\frac{\partial^2 C(e,q)}{\partial e^2}}{\bar{g}\left(\frac{\partial\Psi_A(e,q)}{\partial e}\Phi(e,q)+\frac{\partial\Phi(e,q)}{\partial e}\Psi_A(e,q)\right)}-\frac{\frac{\partial C(e,q)}{\partial e}}{\left[\bar{g}\left(\frac{\partial\Psi_A(e,q)}{\partial e}\Phi(e,q)+\frac{\partial\Phi(e,q)}{\partial e}\Psi_A(e,q)\right)\right]^2}*\frac{\partial z}{\partial e}<0 \quad (62)$$

The inequality in (62) justifies as follows. The first derivative of z with respect to e is:

$$\frac{\partial z}{\partial e}=\frac{\partial^2\Psi_A(e,q)}{\partial e^2}\Phi(e,q)+2\frac{\partial\Phi(e,q)}{\partial e}\frac{\partial\Psi_A(e,q)}{\partial e}+\frac{\partial^2\Phi(e,q)}{\partial e^2}\Psi_A(e,q) \quad (63)$$

The necessary derivatives of the relative bargaining power $\Psi_A(e,q)$ and the residual share $\Phi(e,q)$ are:

- $\frac{\partial \Phi(e,q)}{\partial e} = -\frac{\sum_D \frac{\partial \Omega_D(e)}{\partial e}}{1+\bar{n}} > 0$
- $\frac{\partial^2 \Phi(e,q)}{\partial e^2} = \frac{\partial \frac{1+\bar{n}-\sum_D \Omega_D(e)-\Omega_A(q)}{1+\bar{n}}}{\partial^2 e} = -\frac{\sum_D \frac{\partial^2 \Omega_D(e)}{\partial e^2}}{1+\bar{n}} < 0$
- $\frac{\partial \Psi_A(e,q)}{\partial e} = \frac{\partial \frac{\Omega_A(q)}{\sum_D \Omega_D(e)+\Omega_A(q)}}{\partial e} = -\frac{\Omega_A(q)}{[\sum_D \Omega_D(e)+\Omega_A(q)]^2} \sum_D \frac{\partial \Omega_D(e)}{\partial e} > 0$
- $\frac{\partial^2 \Psi_A(e,q)}{\partial e^2} = \frac{\partial \frac{\Omega_A(q)}{\sum_D \Omega_D(e)+\Omega_A(q)}}{\partial^2 e} = \Omega_A(q) \frac{\left[\sum_D \frac{\partial \Omega_D(e)}{\partial e}\right]^2}{[\sum_D \Omega_D(e)+\Omega_A(q)]^3} - \Omega_A(q) \frac{\sum_D \frac{\partial^2 \Omega_D(e)}{\partial e^2}}{[\sum_D \Omega_D(e)+\Omega_A(q)]^2} \lesseqgtr 0$

The uncertainty about the fourth derivative necessitates the additional assumption above, i.e. $\frac{\partial z}{\partial e} > 0$. To reproduce the signs of the other derivatives, remember the following assumptions:

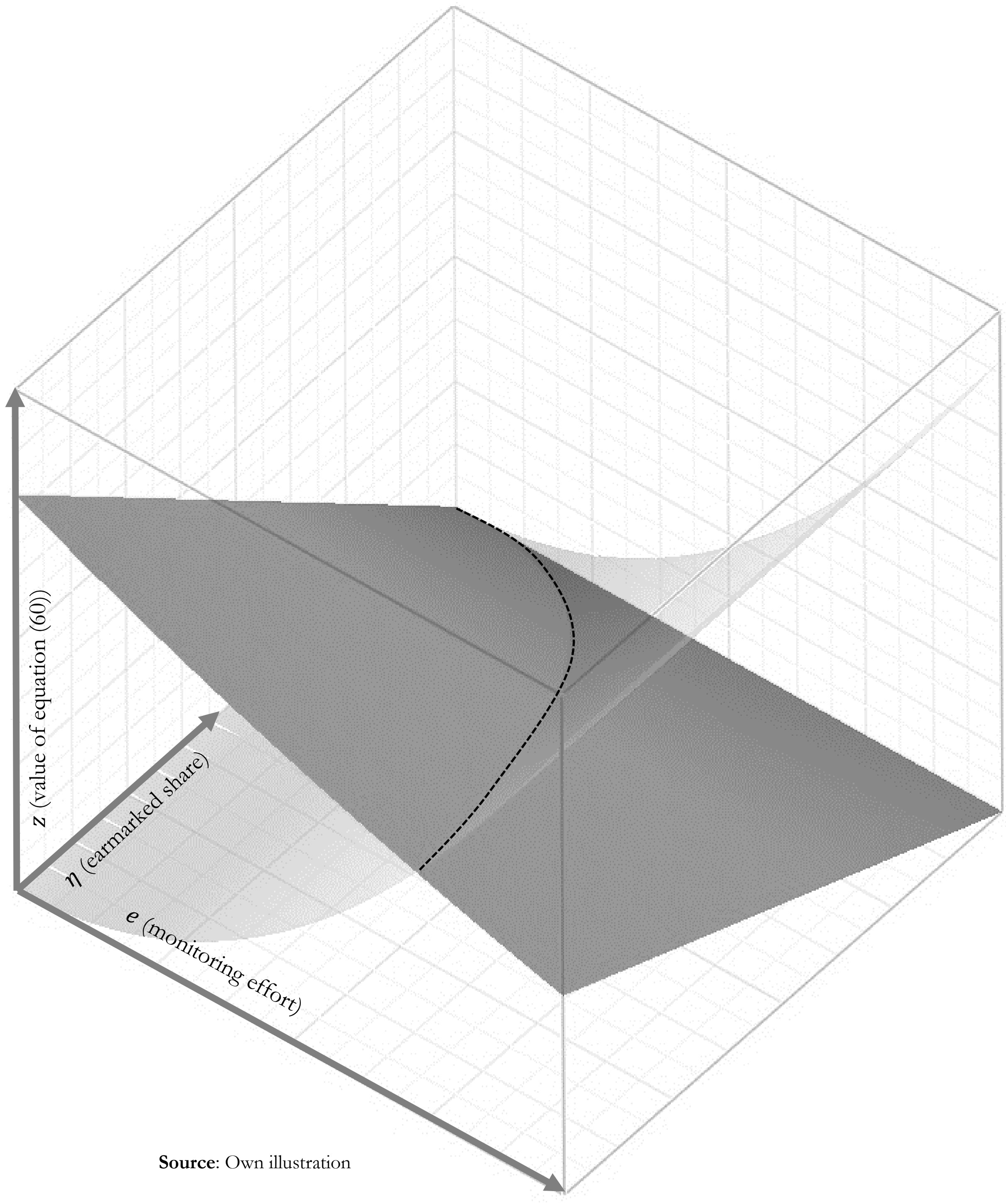
- $\Omega_A(q) \in [0,1]$ [absolute defensive bargaining power]
- $\Omega_D(e) \in [0,1]$ [absolute offensive bargaining power]
- $\frac{\partial \Omega_D(e)}{\partial e} < 0$ [by assumption; see text]
- $\frac{\partial^2 \Omega_D(e)}{\partial e^2} < 0$ [by assumption; see text]
- $\frac{\partial^2 C(e,q)}{\partial e^2} > 0$ [by assumption; see text]

Equation (62) states that a higher monitoring effort comes with a higher earmarked share. Or put differently, the higher the earmarked share, the lower the monitoring effort will be.

For illustrative purposes consider the following calibration: The total budget $\bar{g} = 1$; the offensive bargaining power function $\Omega_D(e) = \frac{1}{1+x}$; the defensive bargaining power $\Omega_A(q) = 0.5$; there are $n = 4$ ministers besides minister A in the council; the ministers' cost function $C(e) = \frac{1}{3}x^3$.

Figure 21 traces the marginal cost, i.e. the left-hand side of equation (59), in the lightly shaded transparent surface. Since the marginal cost is independent of the earmarked share η, the respective function is simply a convex curve in the e-z-plane. The darker shaded opaque surface spans the marginal benefit of the monitoring effort, which has a negative slope in the e-z-plane. Thus, the two functions chosen for $\Omega_D(e)$ and $C(e)$ meet the requirement imposed in equation (61) and the two surfaces intersect only once. As deduced analytically, the marginal benefit of the monitoring effort has a negative slope in the η-z-plane. Finally, the interesting perspective is the e-η-plane, where the intersection of the two surfaces produce a negatively sloped curve. A decreasing intersection curve means that the higher the earmarked share η, the lower minister A must set his monitoring effort e in order to maximize his objective.

Figure 21 – Minister A's reaction on an increased earmarked share



Source: Own illustration

A. 2 Monte Carlo simulation with the quasi fixed effect algorithm

In order to evaluate the performance of the proposed estimation algorithm, some Monte Carlo simulations were conducted. They reveal whether the resulting confidence intervals cover the true value. In addition, they contrast the quasi fixed effect algorithm with the usually applied simultaneous regression. Likewise, because most of the parameters are equal to those set by Simar and Wilson (2007), a comparison to their study is straightforward. Nevertheless, since the estimator's behavior depending on the number of individuals and time periods is the focus, the simulation does not include different numbers of inputs, outputs, and environmental variables. Concretely, the controlled data-generating process produces two inputs, two outputs, and one environmental variable. Table 24 presents the formula or distribution imposed to generate the respective variable. The last column entails the parameters to be estimated.

Table 24 – Distributions and parameters used for the data-generating process

Symbol	Meaning	Distribution / Formula	Parameters
z_{it}	environmental variable	$z_{it} \backsim U(0,2)$	
α_i	individual fixed effect	$\alpha_i = \bar{z}_{1,i} - 1$	$\bar{z}_{1,i} = \frac{1}{T}\sum_{t=1}^{T} z_{1,it}$
τ_t	time fixed effect	$\tau_t = 3 * \bar{z}_{1,t} - 2$	$\bar{z}_{1,t} = \frac{1}{N}\sum_{n=1}^{N} z_{1,it}$
ε_{it}	error term of efficiency	$\varepsilon_{it} \backsim N(0, \sigma_\varepsilon^2)$ truncated left at $1 - \beta_0 - z_{1,it}\beta_1 - z_{2,it}\beta_1 - \alpha_i$ right at $0 - \beta_0 - z_{1,it}\beta_1 - z_{2,it}\beta_1 - \alpha_i$	$\sigma_\varepsilon = 0.2; \beta_0 = 0.2; \beta_1 = 0.2$ $\beta_2 = 0.2$
θ_{it}	input efficiency	$\theta_{it} = \beta_0 + z_{1,it}\beta_1 + z_{2,it}\beta_1 + \alpha_i + \varepsilon_{it}$	
$\boldsymbol{y}_{it}$	outputs	$x_{(1,2)it} \backsim U(6,16)$	
u	weight of first input	$u \backsim U(0,1)$	
$x_{(1)it}$	first input	$y_{(1)it} = u\theta_{it}^{-1}\tau_t\left(x_{1it}^{3/4} + x_{2it}^{3/4}\right)$	
$x_{(2)it}$	second input	$y_{(2)it} = (1-u)\theta_{it}^{-1}\tau_t\left(x_{1it}^{3/4} + x_{2it}^{3/4}\right)$	

Notes: Simulations were conducted using Stata with 1000 replications each time

Source: Own composition inspired by Simar and Wilson (2007)

The environmental variable z_{it} follows a uniform distribution with the boundaries zero and two. It thereupon affects the value of the fixed effects. Remember that in terms of an OLS regression, independently distributed individual effects make it possible to use the more efficient random-effects estimator whereas the fixed effect estimator is still consistent. Since the main concern is consistency here, a non-orthogonality between the two fixed effects and the environmental variable as a regressor is imposed, which takes advantage of the fixed effect estimator. Concretely, both, the time- and the-individual fixed effects depend upon either the individual mean per year or the annual mean per individual, respectively. The rationale behind the particular specification, i.e. $\alpha_i = \bar{z}_{1,i} - 1$ and $\tau_t = 3 * \bar{z}_{1,t} - 2$, is to ensure a sensible output efficiency whose error term lies within the truncation boundaries. That is, on the one hand, if α_i becomes too small, it does not affect the output efficiency. On the other hand, a too high α_i dominates the output efficiency and annihilates the influence of the environmental variable z_{it}. At the same time, a too negative α_i provokes a very high left truncation boundary which in turn leads to a problem in finding a randomly distributed error term ε_{it}. The same arguments hold for the time fixed effect τ_t. The chosen distribution parameters proved to take into account accurately these concerns.

The next step entails setting the parameter value $\beta_1 = 0.2$, which is the effect the environmental variable has on efficiency, and the error term's variance $\sigma_\varepsilon = 0.2$. Note that $\beta_0 = 0.2$ is simply the constant which is irrelevant for this analysis. Having z_{it}, α_i, τ_t, β_0, β_1, and σ_ε at hand enables generating values for the efficiency's error term ε_{it}. Because the input efficiency θ_{it} must stay within zero and one by definition, the error term's truncation boundary on the left is $1 - \beta_0 - \beta_1 z_{it} - \alpha_i - \tau_t$ and on the right $0 - \beta_0 - \beta_1 z_{it} - \alpha_i - \tau_t$. As Cooper et al. (2007) outline, the input efficiency is simply the inverse of the output efficiency. The outputs are drawn twice from an independent uniform distribution between 6 and 16. Based on these values, the total inputs are the result of $\theta_{it}^{-1}\tau_t\left(y_{1it}^{3/4} + y_{2it}^{3/4}\right)$. The term in parentheses indicates the combination of the two outputs in the production process and efficiency multiplier in the front increases the input by a certain percentage if it exceeds one. In order to appropriate the two input proportions on the total, the first input's weight u is determined by a uniform distribution between zero and one.

Applying the quasi fixed effect algorithm presented in section 7.1 produces the parameter estimates listed in the left part of Table 25. As a comparison, the right part presents the estimates running the DEA in the first step with an intertemporal reference set (i.e. with pooled data). The second step is always a truncated regression with Simar & Wilson (2007) type bootstrapped confidence intervals. In the simultaneous case, year dummies control for fixed effects in addition to the individual dummies to account for individual fixed effects. As in usual Monte Carlo simulations, the algorithms were run with different numbers of individual observations (n) and years (t) in order to reveal the small sample biases.

Table 25 – Estimated coverage rates of confidence intervals

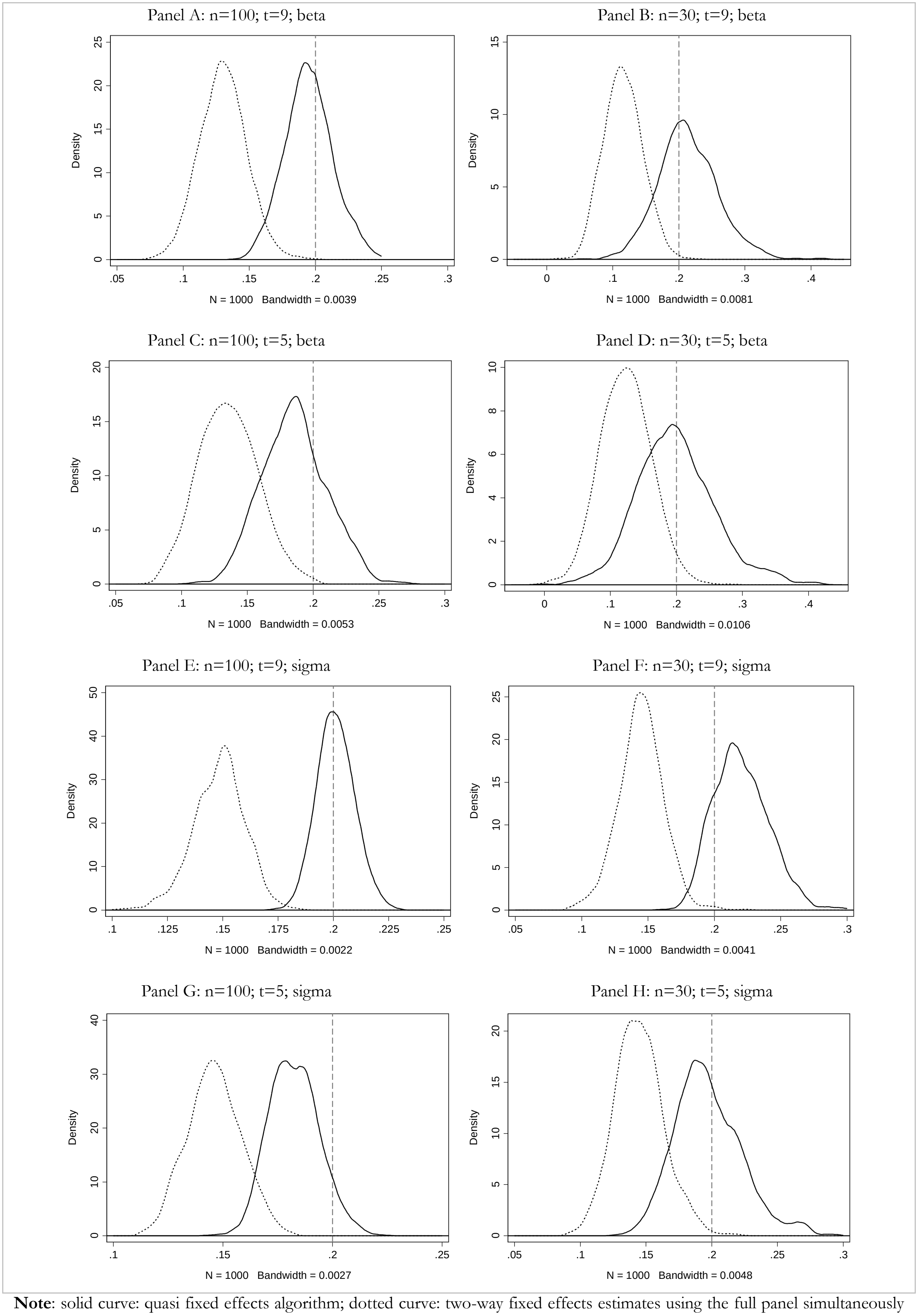
		Para-	Quasi fixed effect algorithm				Simultaneous regression			
n	t	meter	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
100	9	β_1	0.661	0.791	0.865	0.953	0.001	0.001	0.002	0.008
100	9	σ_ε	0.187	0.29	0.399	0.608	0.000	0.000	0.000	0.000
30	9	β_1	0.765	0.875	0.926	0.982	0.019	0.026	0.048	0.104
30	9	σ_ε	0.737	0.831	0.913	0.974	0.001	0.002	0.005	0.010
100	5	β_1	0.454	0.588	0.692	0.844	0.016	0.031	0.038	0.081
100	5	σ_ε	0.000	0.001	0.004	0.024	0.000	0.000	0.000	0.000
30	5	β_1	0.690	0.801	0.873	0.952	0.099	0.171	0.227	0.342
30	5	σ_ε	0.212	0.312	0.414	0.629	0.001	0.003	0.004	0.010

Notes: Simulations were conducted using Stata with 1000 replications each time

Source: Own calculations

Before reflecting the coverage rates of the confidence intervals, consider the graphical representation of the coefficient estimates in Figure 22.

Figure 22 – Densities of sample estimates



Note: solid curve: quasi fixed effects algorithm; dotted curve: two-way fixed effects estimates using the full panel simultaneously
Source: Own illustration

The most important insight the results provide is the dominance of the quasi fixed effect algorithm over the simultaneous regression. Irrespective of the number of individual observations or the number of years, the coverage rate of the quasi fixed effect algorithm always outperforms the simultaneous regression. Admittedly, the coverage rates are not stunningly high compared to the results that Simar & Wilson (2007) find in the cross-sectional case. However, if time and individual fixed effects as implemented in the data-generating process are present, the quasi fixed effect algorithm is apparently the better way estimate the parameters.

The graphical illustration in Figure 22 supports the dominance of the quasi fixed effect algorithm. It becomes particularly visible that the simultaneous regression estimates the coefficients more optimistically (i.e. with a narrower distribution of the point estimates) than the quasi fixed effect algorithm. Consequently, the confidence interval also becomes too narrow so that the coverage rate of the simultaneous regression is poor (c.f. Table 25). Also, the point estimates itself of the quasi fixed effect algorithm accumulate closer to the true value. This becomes more evident the higher the number of individual observations. While the difference between the two estimators is less evident for only 30 individual observations (panels B and D), the estimated betas of the quasi fixed effect algorithm come much closer to its true values if the number of observations increases (panels A and C).

In the case where the Swiss cantons represent the individual observations, the econometrician finds himself clearly in the panels B and D. Even if the illustrations in panel B and D suggest that the number of years is almost irrelevant if the number of individuals is so small, the coverage results in Table 25 object. Consider the 80% confidence interval of the quasi fixed effect algorithm. While the coverage rate of β_1 only reaches 38% with n=30 and t=5, it increases to 51% when the researcher includes 4 additional years. In the present case, where 15 years are observed, the coverage rate can undoubtedly be expected to lie even higher.

A. 3 Marginal effects on efficiency in the Stochastic Frontier framework

Once the estimates of equation (58) are disposable, the technical efficiency computes as (Battese & Coelli 1988)

$$TE_{it} = \mathbb{E}[\exp(-u_{it})\,|\varepsilon_{it}] = \exp(-\mu_{*it} + 0.5\sigma_*^2) * \frac{\Phi\left(\frac{\mu_{*it}}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_{*it}}{\sigma_*}\right)}$$

$$\begin{aligned} \text{where}\quad \mu_{*it} &= -\varepsilon_{it}\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2} - \frac{\mu_{it}\sigma_v^2}{\sigma_u^2+\sigma_v^2} \\ \sigma_*^2 &= \frac{\sigma_u^2\sigma_v^2}{\sigma_u^2+\sigma_v^2} \\ \sigma_u^2 &= \exp(\boldsymbol{\delta}'\boldsymbol{z}) \\ \sigma_v^2 &= \exp(\boldsymbol{\omega}'\boldsymbol{z}) \\ \mu_{it} &= \boldsymbol{\varphi}'\boldsymbol{z} \end{aligned} \qquad (64)$$

As usual, $\Phi(\cdot)$ indicates the cumulative distribution function and $\phi(\cdot)$ is the probability density function. The marginal effect of each environmental variable z_1 on efficiency is the result of the first partial derivative:

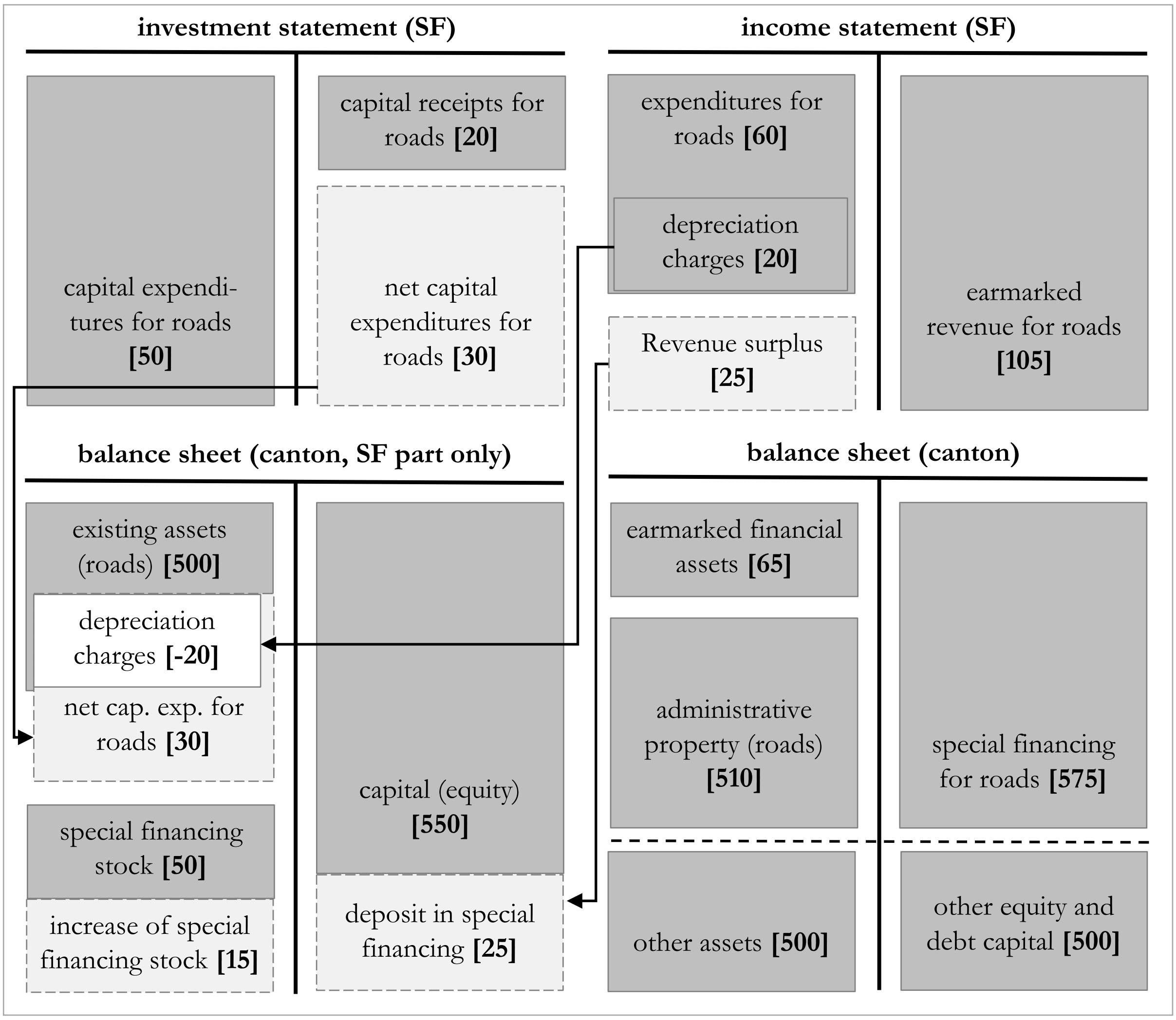
$$\frac{\partial TE_{it}}{\partial z_1} = f'(\cdot)\frac{g(\cdot)}{h(\cdot)} + f(\cdot)\frac{g'(\cdot)h(\cdot) - g(\cdot)h'(\cdot)}{h(\cdot)^2} \qquad \text{where}$$

$$\begin{aligned} f(\cdot) &= \exp(-\mu_{*it} + 0.5\sigma_*^2) \quad \text{and } f'(\cdot) = \frac{\partial f(\cdot)}{\partial z_1} \\ g(\cdot) &= \Phi\left(\frac{\mu_{*it}}{\sigma_*} - \sigma_*\right) \quad \text{and } g'(\cdot) = \frac{\partial g(\cdot)}{\partial z_1} \\ h(\cdot) &= \Phi\left(\frac{\mu_{*it}}{\sigma_*}\right) \quad \text{and } h'(\cdot) = \frac{\partial h(\cdot)}{\partial z_1} \\ f' &= \exp(-\mu_{*it} + 0.5\sigma_*^2)\frac{\sigma_*^2}{\sigma_u^2+\sigma_v^2}[0.5(\omega_1\sigma_u^2 - \delta_1\sigma_v^2) - (\sigma_u^2 - \delta_1)(\mu_{it} + \varepsilon_{it} + \varphi_1\sigma_v^2)] \\ g' &= \phi\left(\frac{\mu_{*it}}{\sigma_*} - \sigma_*\right)\frac{1}{\sigma_*}\left[\sigma_*^2\frac{\omega_1-\delta_1}{\sigma_u^2+\sigma_v^2}(\mu_{it} + \varepsilon_{it} + \varphi_1\sigma_v^2) - \frac{\omega_1\sigma_u^2+\delta_1\sigma_v^2}{2(\sigma_u^2+\sigma_v^2)}(\mu_{*it} + \sigma_*^2)\right] \\ h' &= \phi\left(\frac{\mu_{*it}}{\sigma_*}\right)\left[\sigma_*\frac{\omega_1-\delta_1}{\sigma_u^2+\sigma_v^2}(\mu_{it} + \varepsilon_{it} + \varphi_1\sigma_v^2) - \frac{\mu_{*it}}{2\sigma_*(\sigma_u^2+\sigma_v^2)}(\gamma_1\sigma_u^2 + \delta_1\sigma_v^2)\right] \end{aligned} \qquad (65)$$

This expression simplifies when the estimations stem from the true random- or true fixed effect model, since in this case $\mu_{it} = 0$. Further, the utilized models assume homogeneity of the stochastic error term which means $\sigma_v^2 = \exp(\omega_0)$.

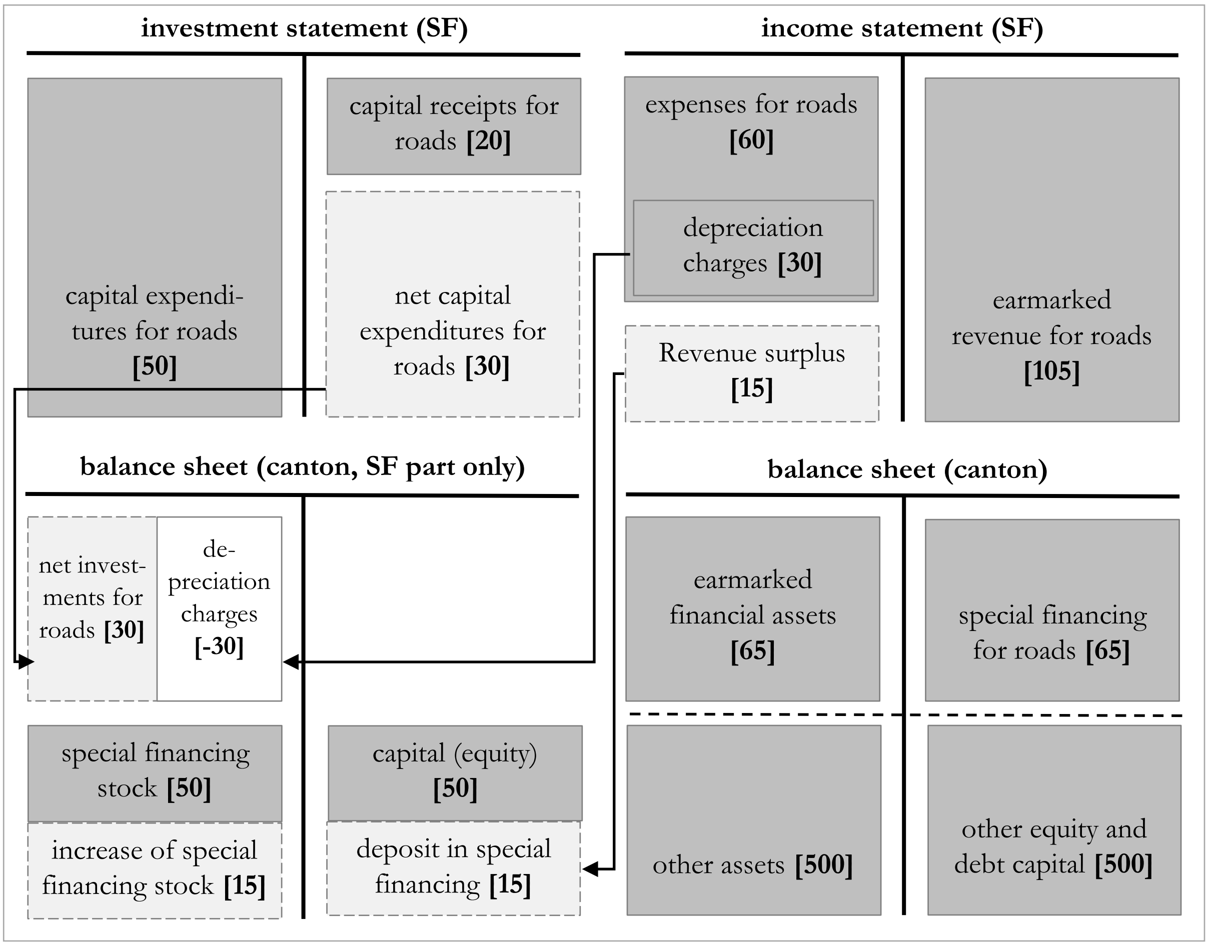
A. 4 Figures

Figure 23 – Accounting scheme for the special financing (SF) with non-immediate depreciation



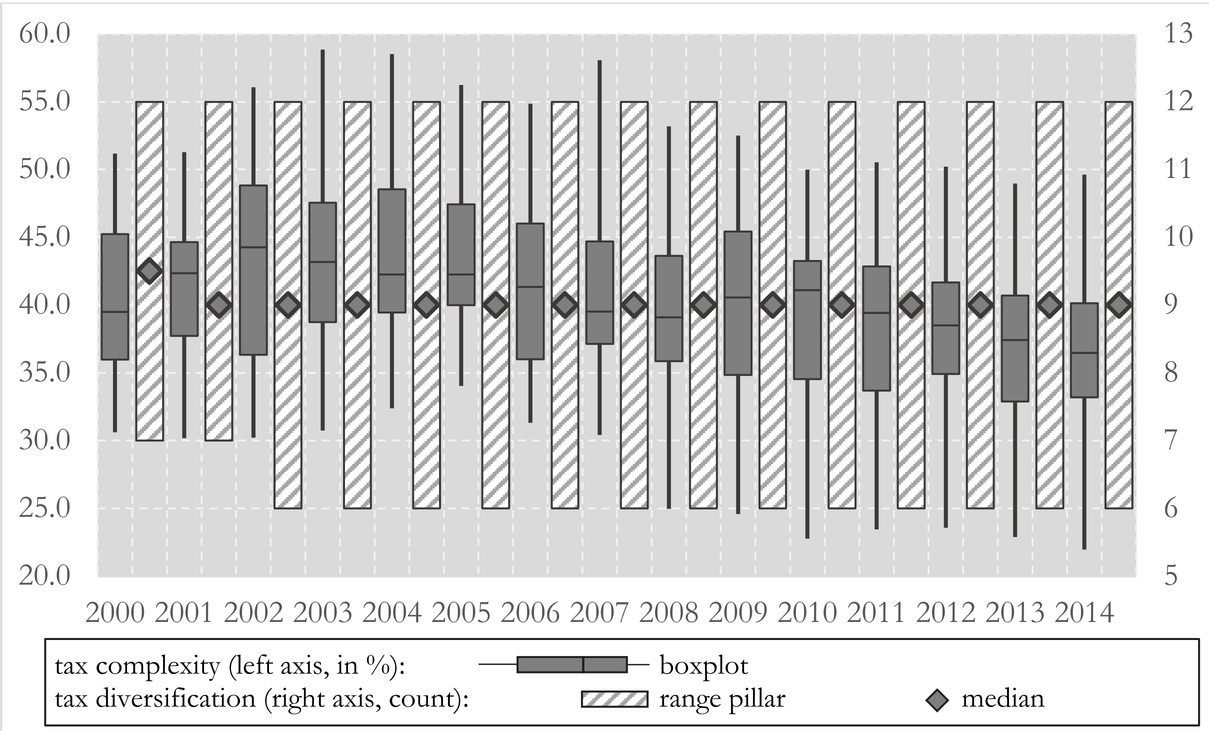
Source: Own illustration

Figure 24 – Accounting scheme for the special financing (SF) with immediate depreciation



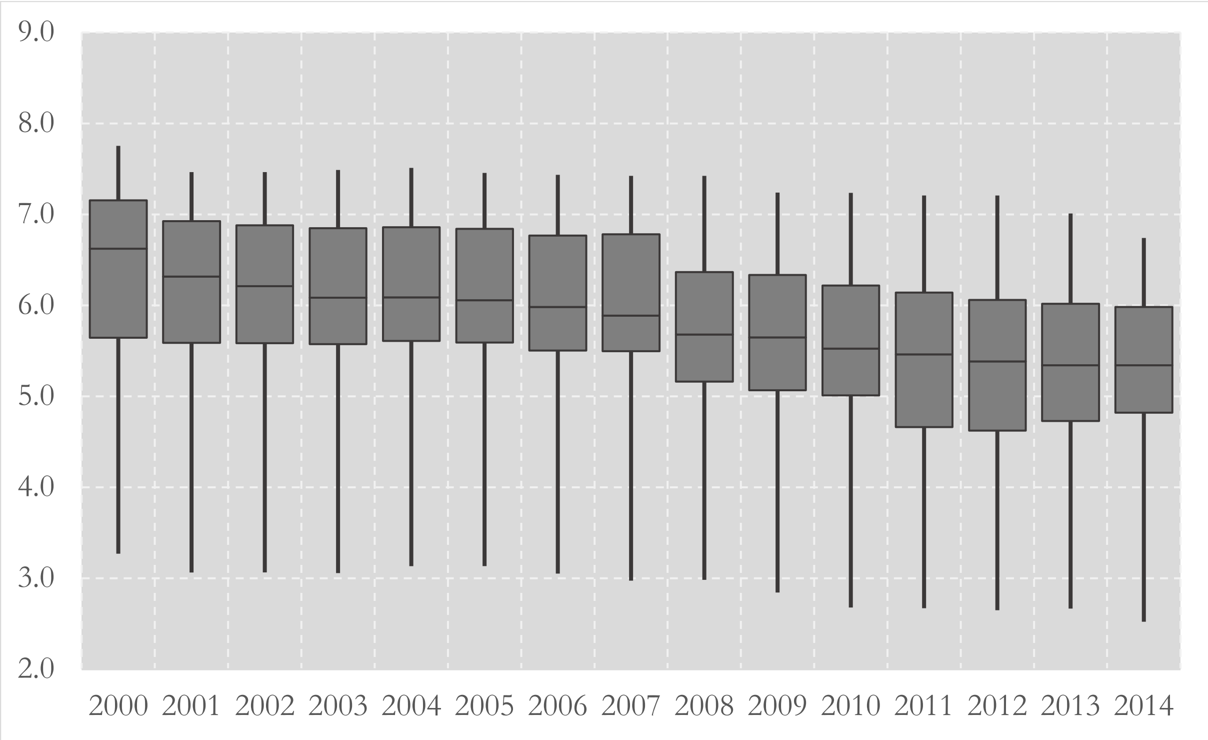
Source: Own illustration

Figure 25 – Tax complexity and diversification by year



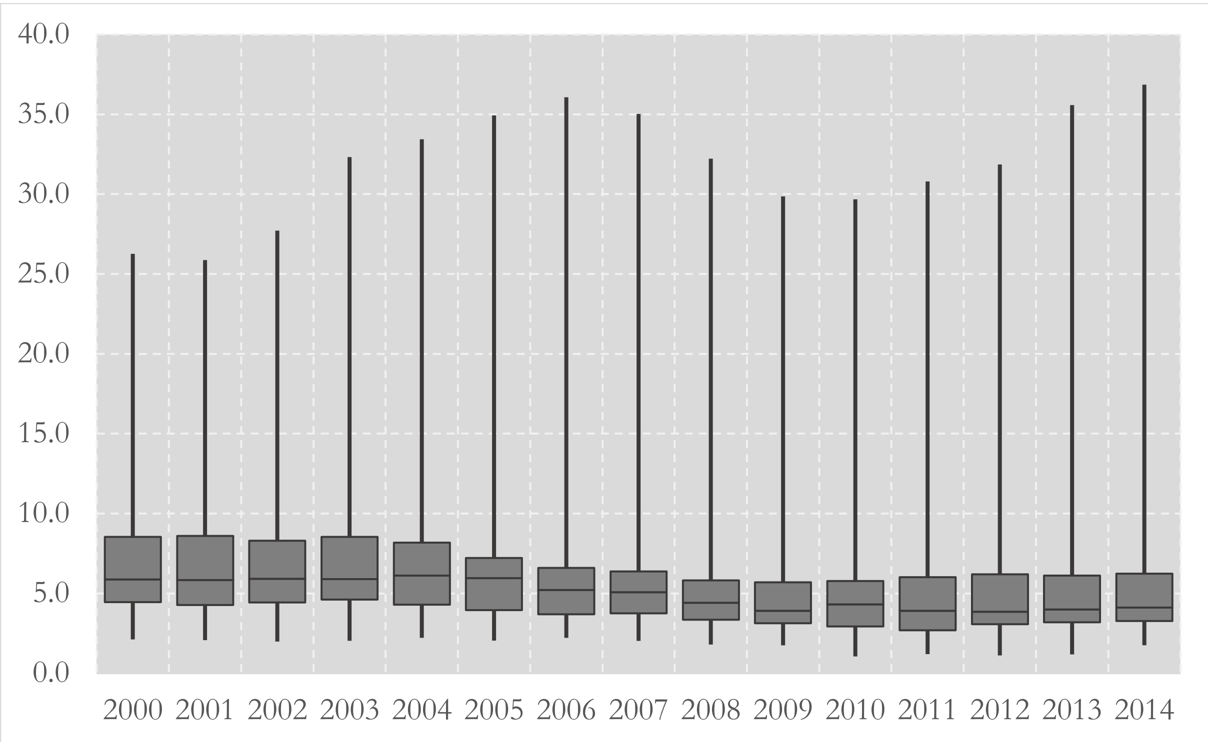
Source: Own illustration

Figure 26 – Progressivity of the tax system by year



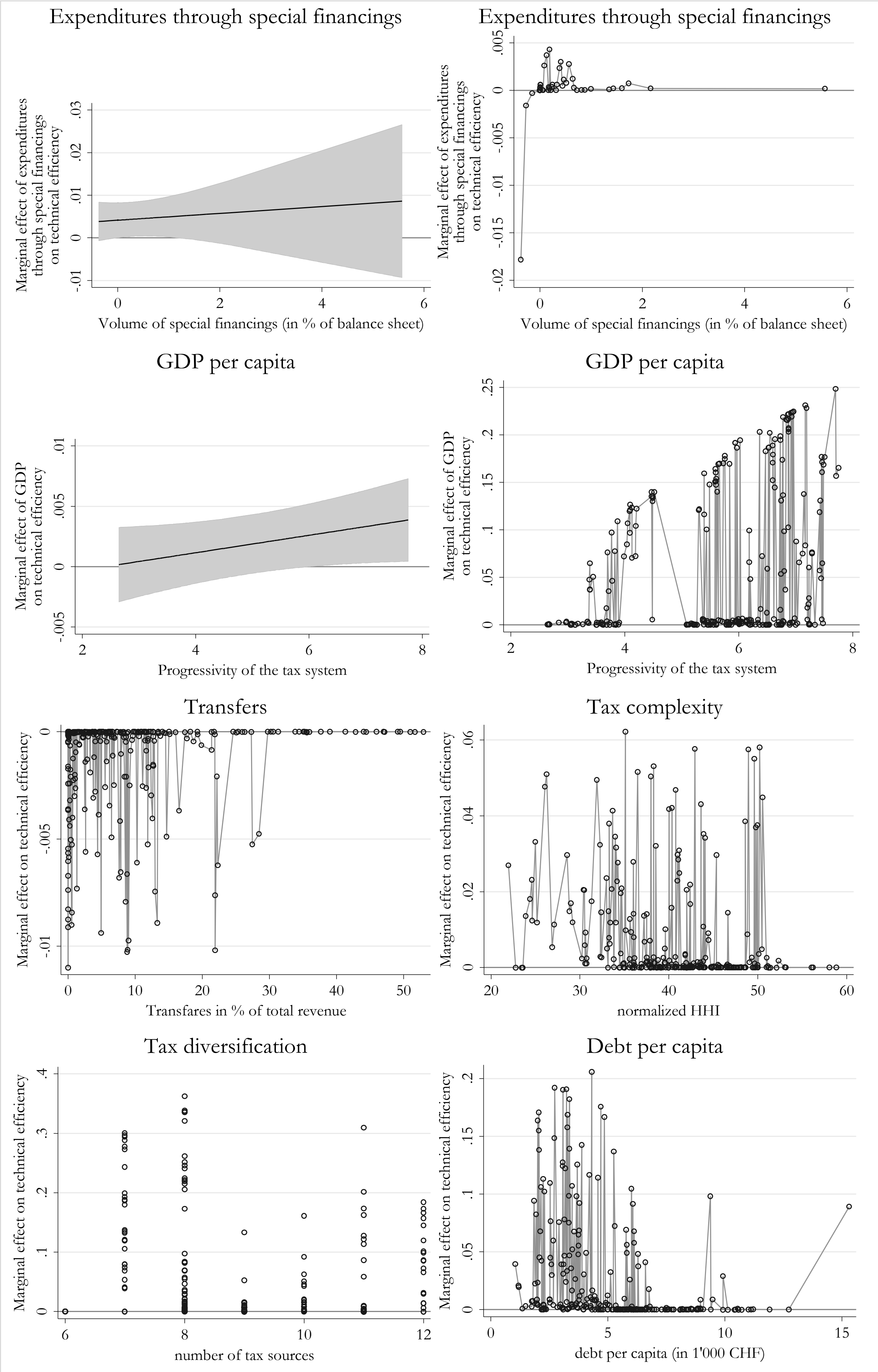
Notes: The ordinate shows the estimated a-value of the logarithmic function $t(y_i) = a * \ln(y_i) + b$ fitted into data on the tax rate (t) and income (y_i) of a household consisting of a married couple and two children.
Source: Own illustration

Figure 27 – Gross debt (in 1'000 CHF per capita) by year



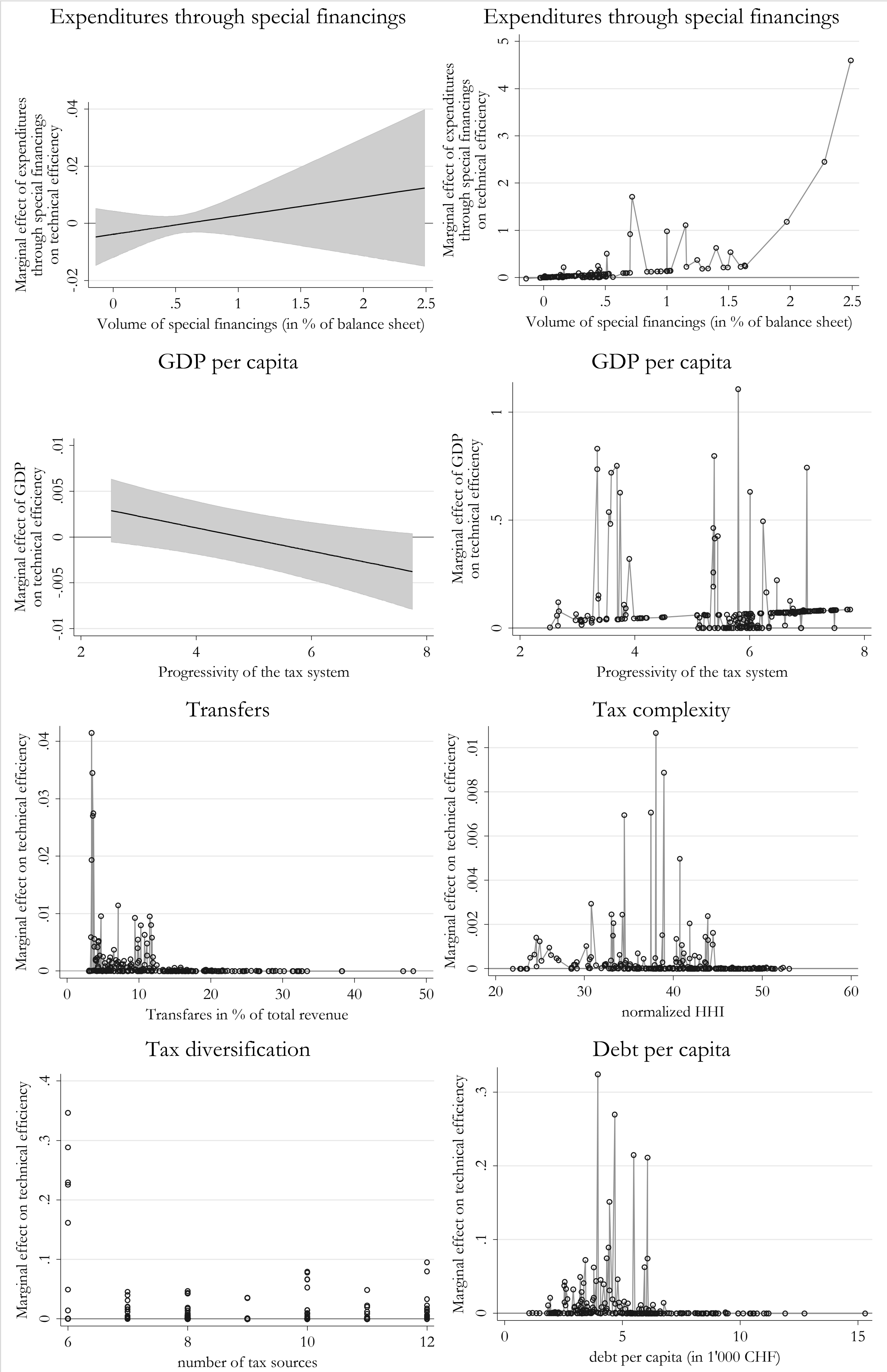
Source: Own illustration

Figure 28 – Marginal effects within the domain of general administration



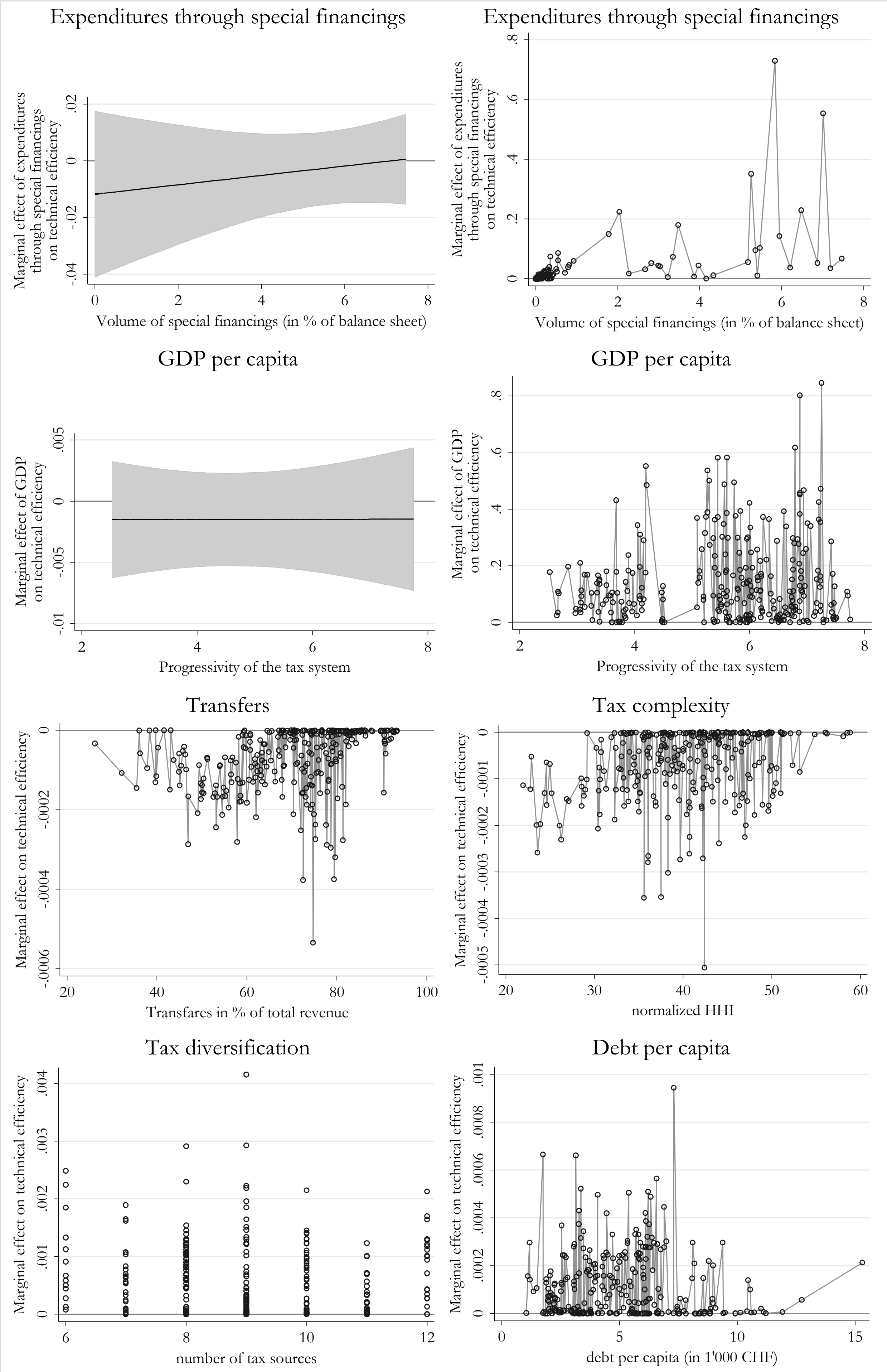
Source: Own illustration

Figure 29 – Marginal effects within the domain of public order and security



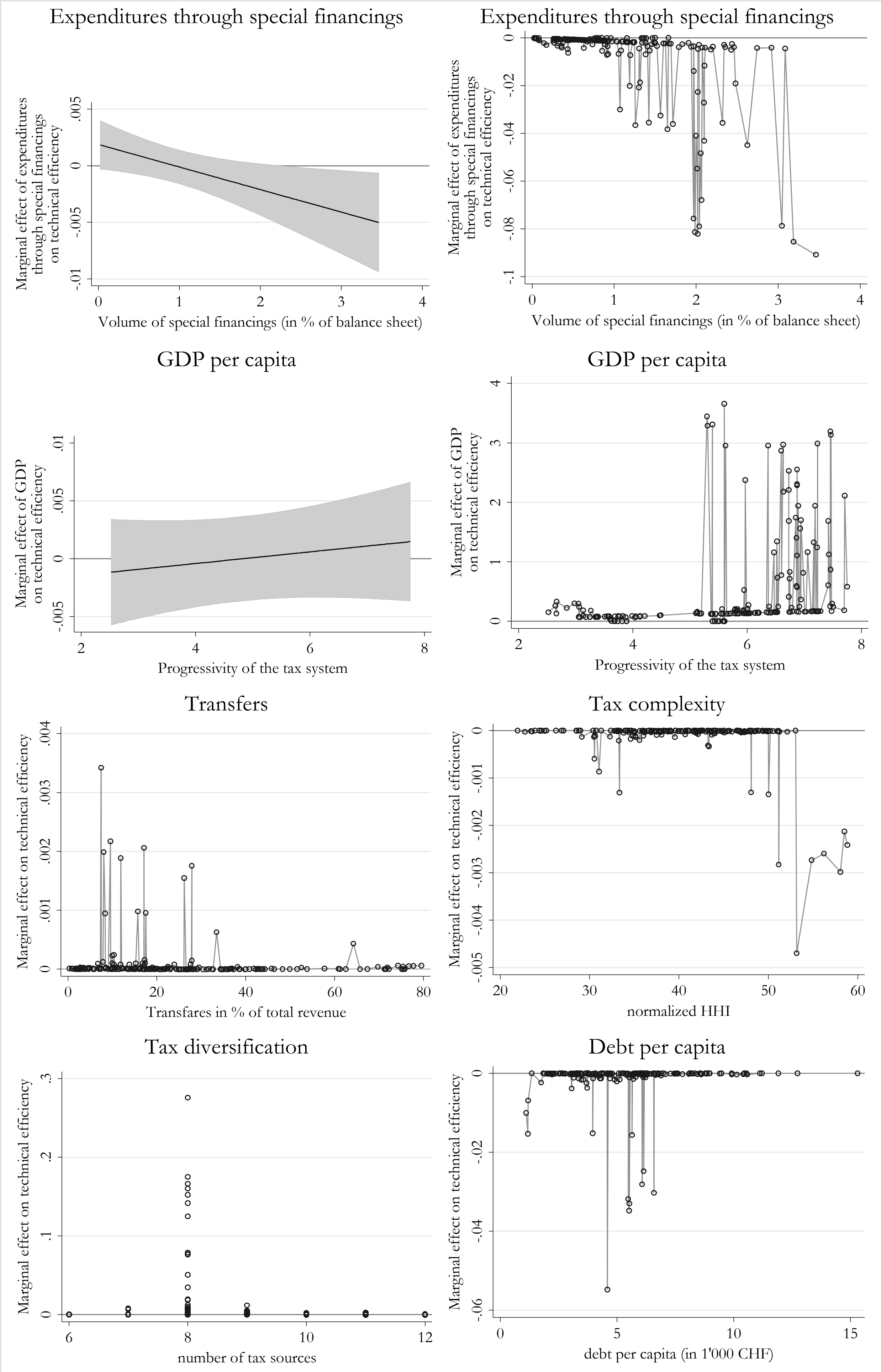
Source: Own illustration

Figure 30 – Marginal effects within the domain of education



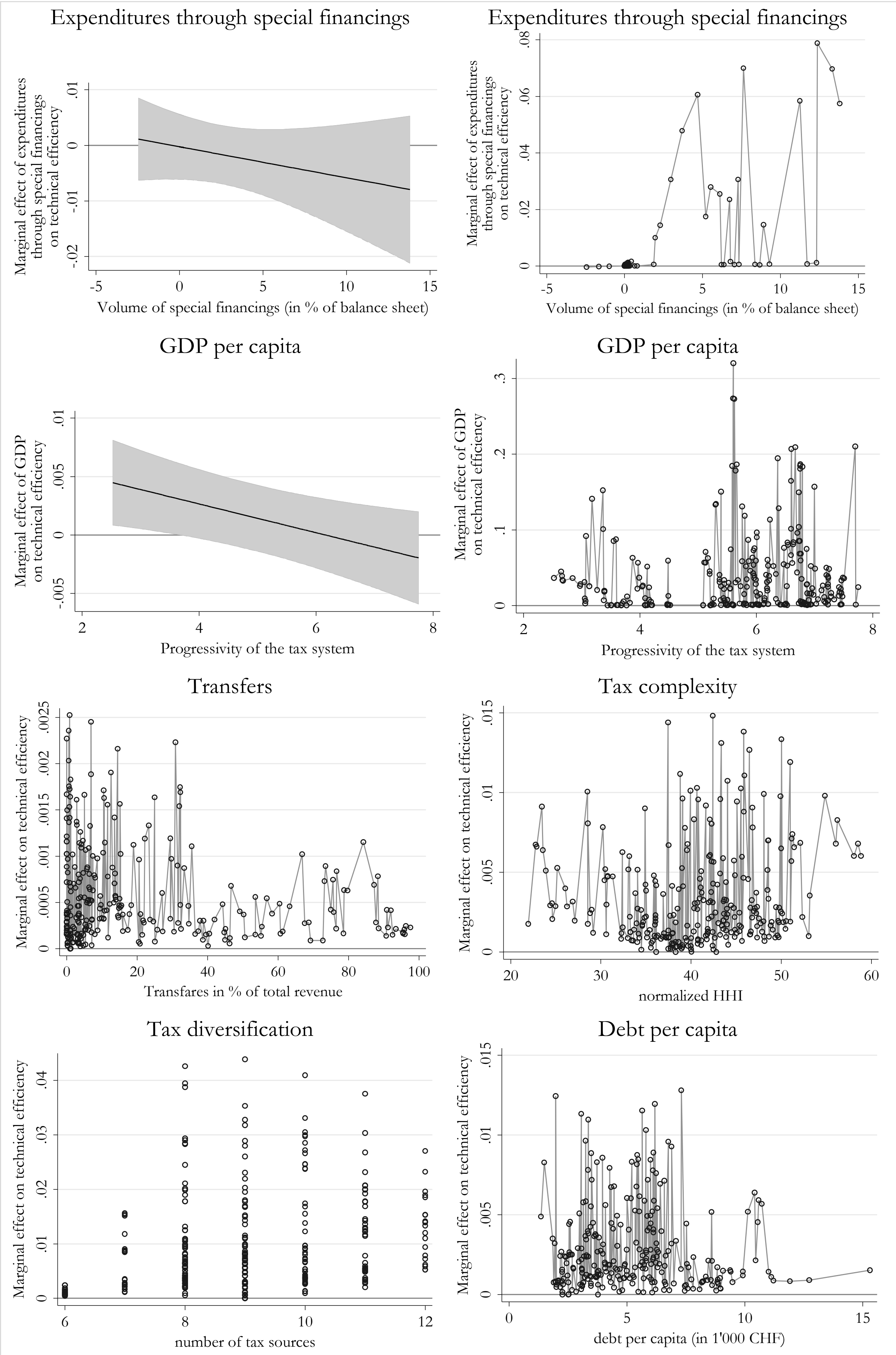
Source: Own illustration

Figure 31 – Marginal effects within the domain of culture



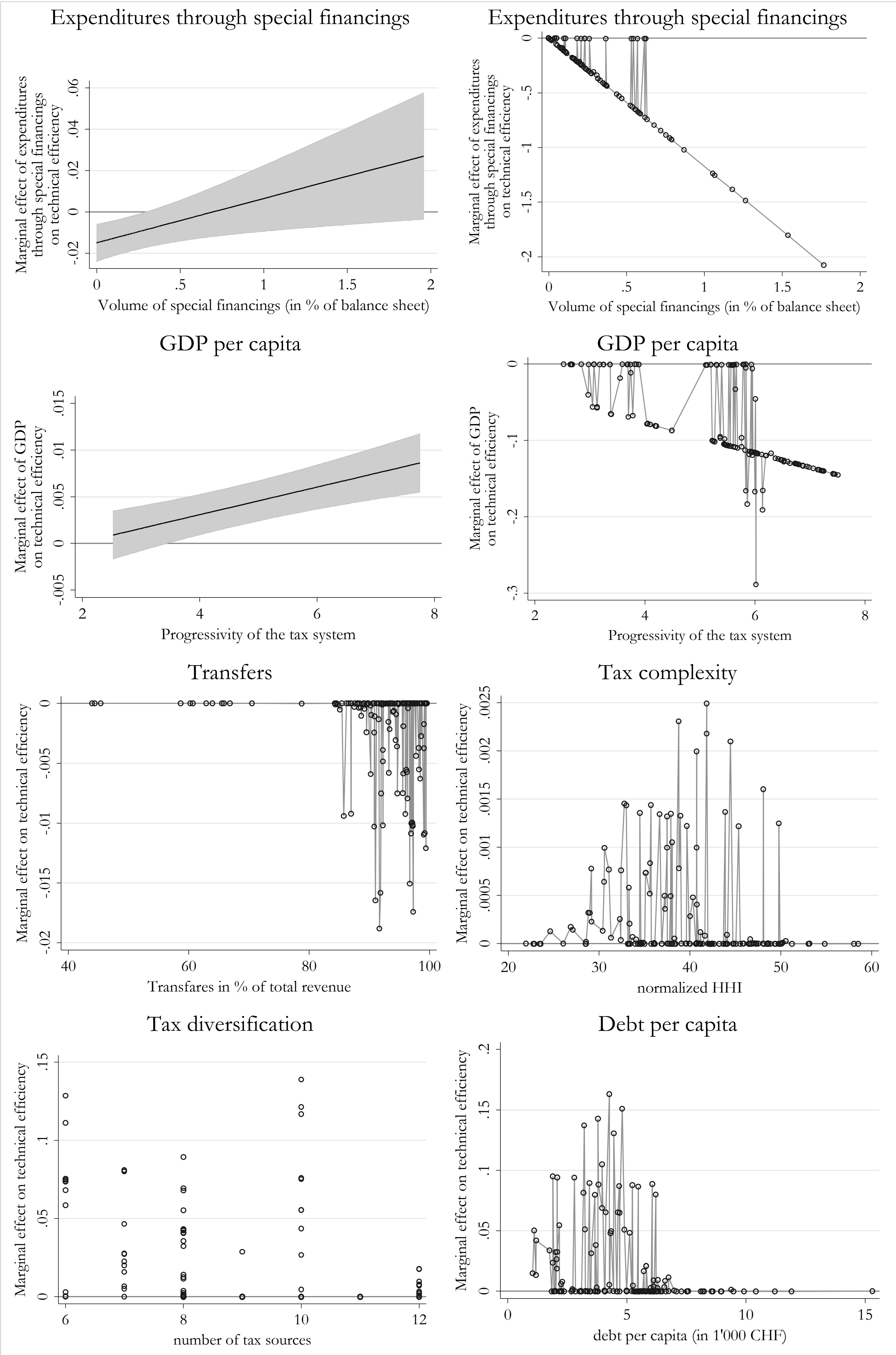
Source: Own illustration

Figure 32 – Marginal effects within the domain of health



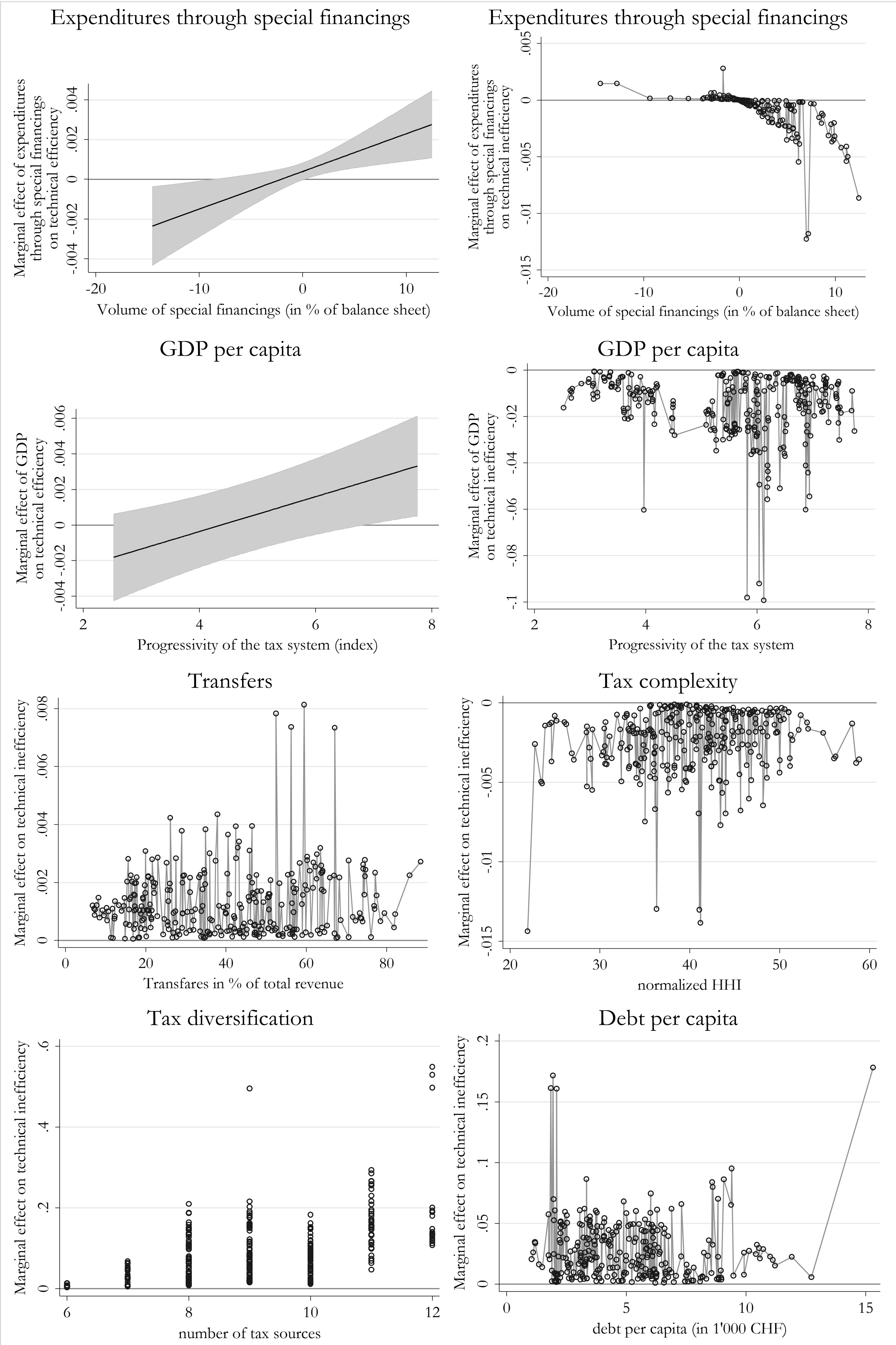
Source: Own illustration

Figure 33 – Marginal effects within the domain of social security



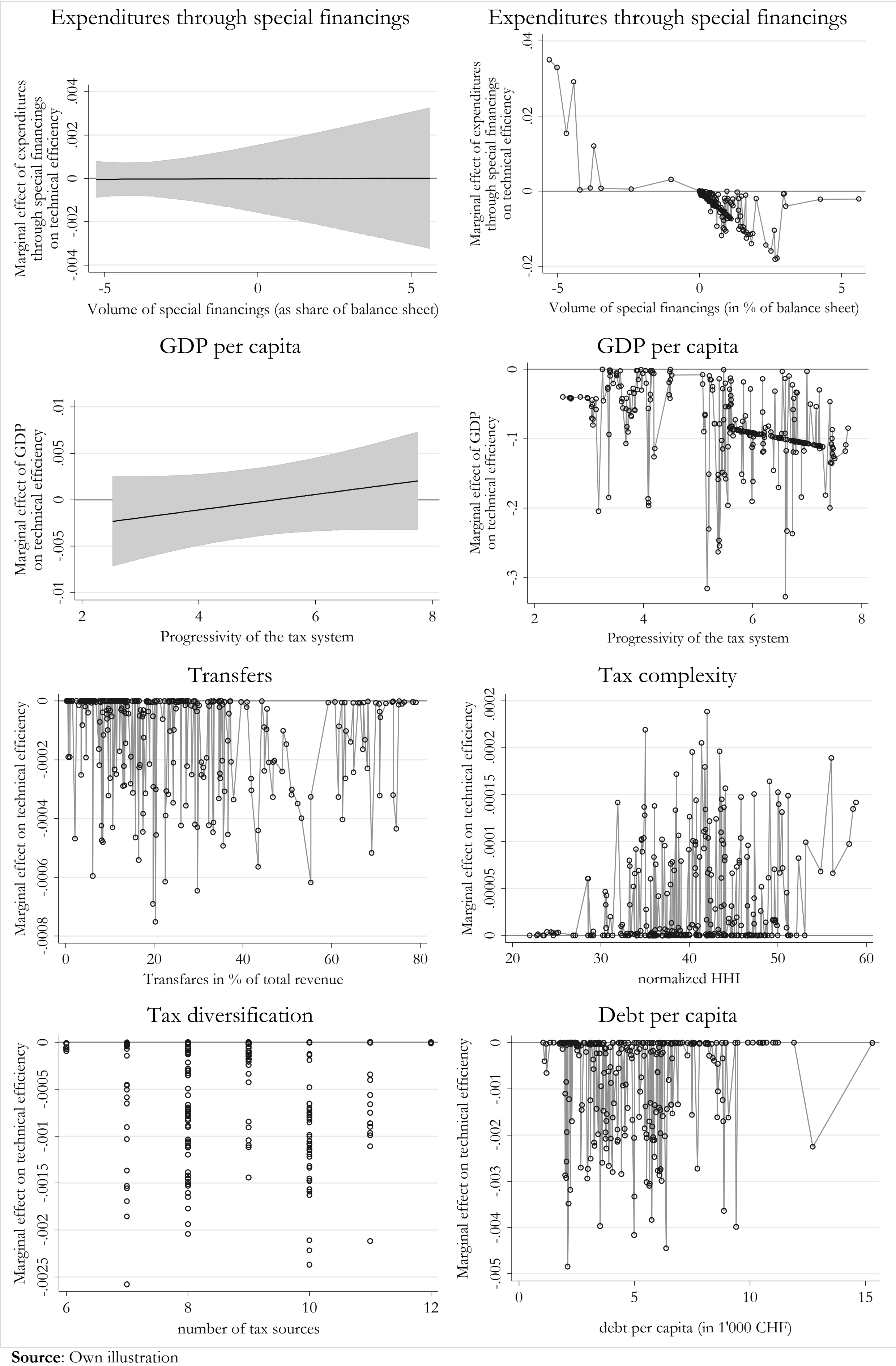
Source: Own illustration

Figure 34 – Marginal effects within the domain of transportation



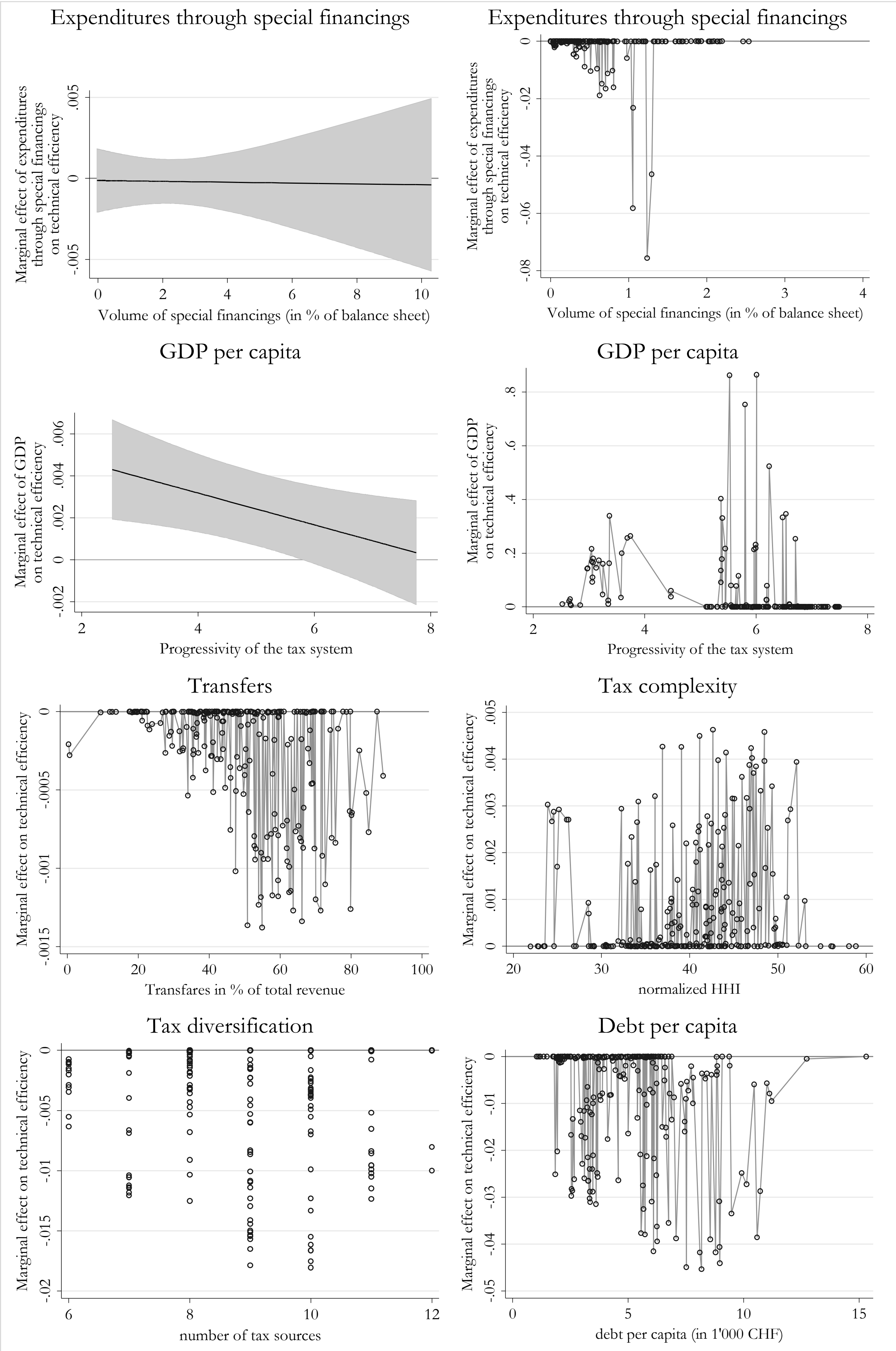
Source: Own illustration

Figure 35 – Marginal effects within the domain of environment and spatial planning



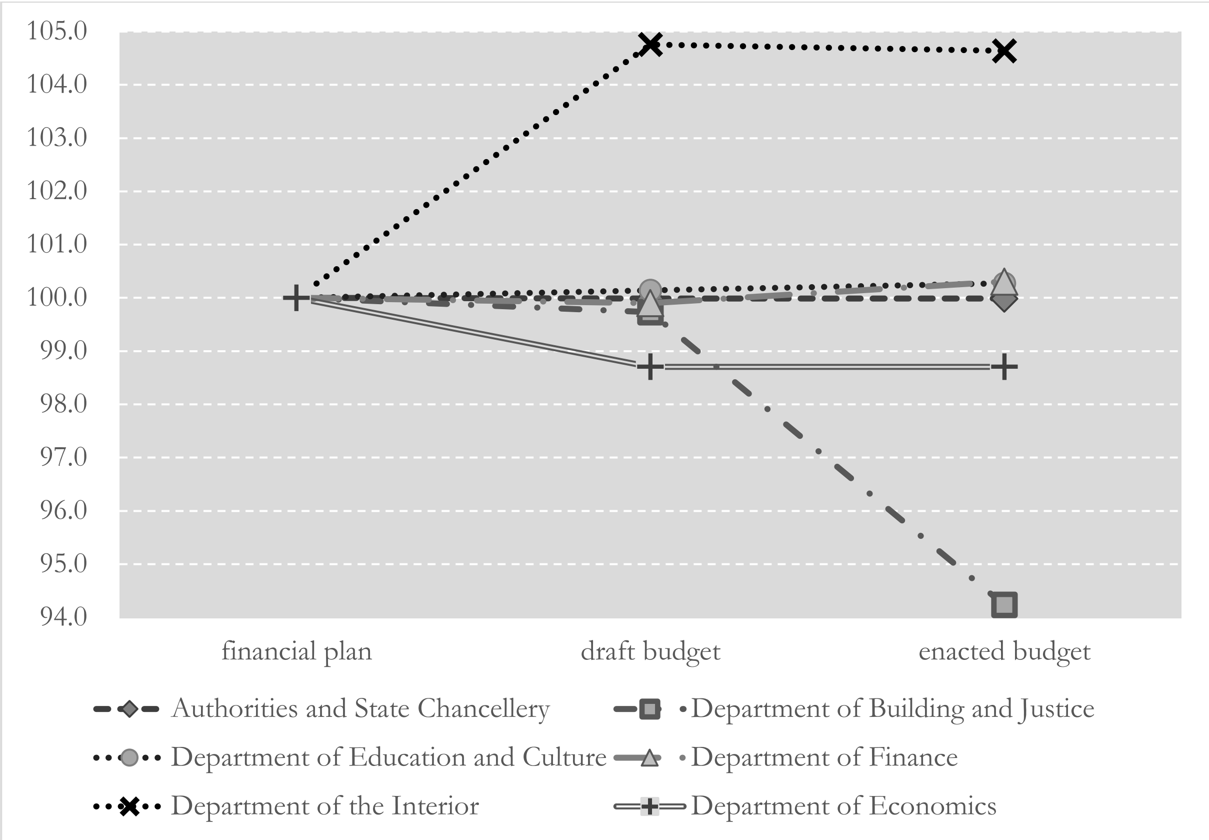
Source: Own illustration

Figure 36 – Marginal effects within the domain of national economy



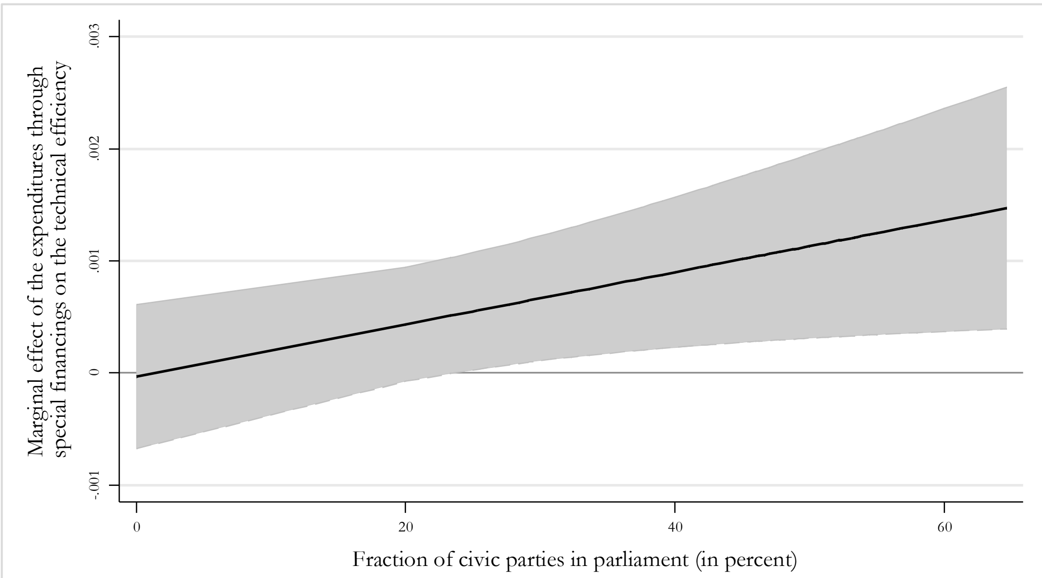
Source: Own illustration

Figure 37 – Budgetary amendments of the Solothurn cantonal budget 2017 (operating and capital expenditures)



Source: Own illustration

Figure 38 – Marginal effects of expenditures through special financings conditional to the party dominance in parliament



Source: Own illustration

A. 5 Tables

Table 26 – Mean and standard deviation of cantonal expenditures by expenditure type and government function 2000 to 2014

Government function	General administration	Public order and security	Education	Culture	Health	Social Security	Transportation	Environment and spatial planning	National economy	Financing and taxes
Personnel expenditure [P]	70'550	166'342	485'076	10'480	128'050	43'043	23'773	12'973	13'554	1'659
	(63'848)	(185'797)	(680'727)	(12'973)	(238'313)	(63'279)	(18'554)	(12'078)	(12'218)	(1'992)
General, administrative and operating expenditure [O]	50'020	57'357	79'604	6'292	50'961	14'639	31'184	9'628	8'030	20'572
	(47'334)	(72'989)	(111'832)	(11'182)	(95'621)	(17'583)	(39'944)	(13'673)	(9'210)	(34'819)
Defense expenditure	0	0	0	0	0	0	0	0	0	0
	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
Financial expenditure [F]	473	108	4	9	0	60	736	1	6	60'543
	(2'008)	(285)	(8)	(21)	(n.a.)	(88)	(1'115)	(1)	(9)	(88'984)
Transfer expenditure [F]	7'575	8'861	189'751	34'334	235'754	481'139	62'993	9'050	121'948	106'681
	(14'627)	(12'309)	(142'629)	(47'375)	(320'690)	(520'032)	(72'093)	(11'741)	(121'382)	(170'113)
Extraordinary expenditure [F]	294'856	44'513	252'820	0	2'990	18'846	0	33'933	76'253	0
	(508'374)	(62'811)	(516'365)	(n.a.)	(n.a.)	(23'354)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
Tangible fixed assets [C]	14'510	10'194	29'983	3'708	16'057	1'981	107'106	7'033	1'258	0
	(24'510)	(13'277)	(43'376)	(9'044)	(22'397)	(4'455)	(103'331)	(8'714)	(3'062)	(n.a.)
Capital expenditures on behalf of third parties [C]	33	50	0	0	0	819	2'997	285	0	0
	(15)	(706)	(n.a.)	(n.a.)	(n.a.)	(500)	(5'343)	(363)	(n.a.)	(n.a.)
Capital expenditures, intangible fixed assets [C]	7'631	2'367	1'936	79	406	2'269	700	2'026	588	0
	(12'493)	(3'421)	(2'370)	(69)	(392)	(3'007)	(1'698)	(3'527)	(1'314)	(n.a.)
Loans and financial interests [C]	2'427	927	1'714	1'243	6'286	7'194	20'906	5'758	13'571	0
	(6'346)	(675)	(2'888)	(3'230)	(18'874)	(10'092)	(59'342)	(7'515)	(37'435)	(n.a.)
Loans [C] [a]	4'961	716	1'292	1'824	1'412	2'497	25'273	1'644	3'539	0
	(n.a.)	(315)	(2'177)	(2'444)	(2'826)	(2'555)	(34'944)	(1'706)	(4'053)	(n.a.)
Financial interests and share capital [C]	2'962	197	925	290	48'289	25'747	1'809	0	11'537	0
	(6'880)	(266)	(1'248)	(474)	(106'524)	(15'509)	(1'720)	(n.a.)	(22'569)	(n.a.)

Notes: N = 390; the letters in brackets indicate to which category each expenditure type belongs (see section 6.3)

[a] The separation of the loans into two positions stems from the harmonization process of the federal finance administration and is not relevant to the present context (Financial Statistics Section 2011).

Source: Own calculation based on data from the Federal Finance Administration (2016)

Table 27 – Mean and standard deviation of variables related to revenue characteristics by government function

Government function	General administration	Public order and security	Education	Culture	Health	Social Security	Transportation	Environment and spatial planning	National economy
Expenditures through special financings [a]	0.832 (2.540)	1.841 (4.253)	0.512 (1.694)	35.431 (20.296)	1.020 (2.787)	2.509 (6.700)	23.628 (33.456)	12.076 (33.019)	3.320 (5.538)
Volume of the special financings [b]	0.125 (0.422)	0.273 (0.440)	0.512 (1.427)	1.104 (0.776)	0.710 (2.344)	0.348 (0.486)	1.131 (3.253)	0.359 (1.121)	1.014 (1.516)
Transfers [c]	9.668 (11.558)	14.410 (9.027)	70.021 (13.656)	26.783 (21.609)	21.930 (27.894)	88.605 (11.917)	3.580 (12.892)	26.888 (21.671)	48.660 (17.320)
Pseudo-earmarked transfers [d]	88.203 (30.411)	98.009 (7.500)	99.137 (1.813)	96.324 (11.946)	87.526 (30.790)	99.787 (0.647)	97.284 (11.622)	98.740 (5.836)	99.861 (0.394)

[a] In percent of total expenditures within function
[b] In percent of the balance sheet's total
[c] In percent of total revenues within function
[d] In percent of total transfers within function
Note: N = 301
Source: Own calculation based on data from the federal finance administration and the annual financial statements of the cantons

Table 28 – Cantonal expenditure shares by state function and expenditure type 2000 to 2014

Government function	General administration	Public order and security	Education	Culture	Health	Social Security	Transportation	Environment and spatial planning	National economy	Financing and taxes
Personnel expenditure [P]	15.37%	56.59%	46.25%	17.06%	25.03%	7.14%	8.09%	13.37%	5.11%	0.88%
General, administrative and operating expenditure [O]	10.90%	19.51%	7.59%	10.25%	9.96%	2.43%	10.62%	9.92%	3.03%	10.86%
Defense expenditure	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Financial expenditure [F]	0.10%	0.04%	0.00%	0.02%	0.00%	0.01%	0.25%	0.00%	0.00%	31.96%
Transfer expenditure [F]	1.65%	3.01%	18.09%	55.90%	46.08%	79.84%	21.45%	9.32%	45.94%	56.31%
Extraordinary expenditure [F]	64.25%	15.14%	24.11%	0.00%	0.58%	3.13%	0.00%	34.96%	28.73%	0.00%
Tangible fixed assets [C]	3.16%	3.47%	2.86%	6.04%	3.14%	0.33%	36.47%	7.25%	0.47%	0.00%
Capital expenditures on behalf of third parties [C]	0.01%	0.17%	0.00%	0.00%	0.00%	0.14%	1.02%	0.29%	0.00%	0.00%
Capital expenditures, intangible fixed assets [C]	1.66%	0.81%	0.18%	0.13%	0.08%	0.38%	0.24%	2.09%	0.22%	0.00%
Loans and financial interests [C]	0.53%	0.32%	0.16%	2.02%	1.23%	1.19%	7.12%	5.93%	5.11%	0.00%
Loans [C] [a]	1.08%	0.24%	0.12%	2.97%	0.28%	0.41%	8.60%	1.69%	1.33%	0.00%
Financial interests and share capital [C]	0.65%	0.07%	0.09%	0.47%	9.44%	4.27%	0.62%	0.00%	4.35%	0.00%

Notes: The letters in brackets indicate to which category each expenditure type belongs (see section 6.3)
[a] The separation of the loans into two positions stems from the harmonization process of the federal finance administration and is not relevant in the present context (Financial Statistics Section 2011).
Source: Own calculation based on data from the Federal Finance Administration (2016)

Table 29 – Mean cantonal DEA technical efficiency by government function 2000 to 2014

Canton	General administration	Public order and security	Education	**Culture**	Health	Social Security	**Transportation**	Environment, spatial planning	National economy
AG	0.732 (0.031)	0.772 (0.040)	0.778 (0.231)	0.706 (0.214)	0.760 (0.084)	0.793 (0.027)	0.816 (0.033)	0.775 (0.093)	0.746 (0.046)
AI	0.579 (0.086)	0.620 (0.176)	0.776 (0.074)	0.544 (0.229)	0.762 (0.062)	0.823 (0.031)	0.678 (0.009)	0.639 (0.130)	0.626 (0.037)
AR	0.582 (0.071)	0.479 (0.114)	0.744 (0.133)	n.a.	0.734 (0.083)	0.807 (0.036)	0.646 (0.069)	0.567 (0.159)	0.773 (0.059)
BE	0.743 (0.056)	0.677 (0.069)	0.771 (0.102)	0.671 (0.196)	0.750 (0.061)	0.648 (0.040)	0.707 (0.109)	0.779 (0.089)	0.749 (0.075)
BL	0.517 (0.085)	0.635 (0.088)	0.699 (0.213)	0.397 (0.172)	0.826 (0.078)	0.667 (0.004)	0.525 (0.054)	0.410 (0.142)	0.628 (0.102)
BS	0.312 (0.070)	0.716 (0.095)	0.600 (0.125)	0.117 (0.052)	0.505 (0.067)	0.547 (0.118)	0.181 (0.055)	0.286 (0.091)	0.550 (0.063)
FR	0.677 (0.076)	0.658 (0.053)	0.757 (0.072)	0.578 (0.222)	0.601 (0.201)	0.678 (0.041)	0.681 (0.073)	0.765 (0.076)	0.561 (0.031)
GE	0.447 (0.114)	0.732 (0.032)	0.852 (0.065)	0.350 (0.153)	0.713 (0.142)	0.590 (0.098)	0.725 (0.148)	0.606 (0.173)	0.604 (0.066)
GL	0.585 (0.075)	0.578 (0.133)	0.814 (0.080)	0.701 (0.095)	0.620 (0.162)	0.806 (0.076)	0.702 (0.023)	0.698 (0.065)	0.627 (0.105)
GR	0.700 (0.068)	0.737 (0.100)	0.815 (0.068)	0.684 (0.072)	0.797 (0.071)	0.880 (0.032)	0.469 (0.085)	0.412 (0.040)	0.770 (0.019)
JU	0.540 (0.076)	0.737 (0.068)	0.809 (0.050)	0.487 (0.187)	0.553 (0.119)	0.789 (0.040)	0.726 (0.025)	0.539 (0.130)	0.430 (0.064)
LU	0.729 (0.084)	0.831 (0.058)	0.646 (0.077)	0.719 (0.205)	0.660 (0.078)	0.678 (0.011)	0.837 (0.036)	0.833 (0.058)	0.765 (0.029)
NE	0.527 (0.069)	0.789 (0.055)	0.695 (0.076)	0.604 (0.208)	0.628 (0.084)	0.661 (0.085)	0.485 (0.086)	0.681 (0.147)	0.358 (0.039)
NW	0.676 (0.069)	0.628 (0.166)	0.761 (0.186)	0.614 (0.302)	0.581 (0.110)	0.846 (0.067)	0.710 (0.062)	0.580 (0.094)	0.772 (0.079)
OW	0.647 (0.108)	0.619 (0.064)	0.749 (0.076)	0.697 (0.062)	0.525 (0.092)	0.797 (0.101)	0.694 (0.033)	0.566 (0.103)	0.774 (0.065)
SG	0.739 (0.104)	0.746 (0.038)	0.669 (0.081)	0.725 (0.094)	0.663 (0.051)	0.799 (0.092)	0.716 (0.067)	0.687 (0.081)	0.753 (0.050)
SH	0.602 (0.092)	0.554 (0.125)	0.802 (0.126)	0.716 (0.111)	0.686 (0.076)	0.772 (0.041)	0.691 (0.073)	0.681 (0.050)	0.616 (0.045)
SO	0.622 (0.052)	0.746 (0.085)	0.730 (0.126)	0.689 (0.078)	0.770 (0.078)	0.776 (0.000)	0.710 (0.079)	0.692 (0.119)	0.659 (0.044)
SZ	0.654 (0.120)	0.600 (0.135)	0.826 (0.067)	0.625 (0.207)	0.609 (0.064)	0.714 (0.010)	0.665 (0.078)	0.846 (0.064)	0.755 (0.069)
TG	0.774 (0.038)	0.654 (0.120)	0.764 (0.067)	0.689 (0.204)	0.701 (0.096)	0.721 (0.021)	0.764 (0.014)	0.715 (0.053)	0.681 (0.061)
TI	0.537 (0.059)	0.835 (0.032)	0.788 (0.073)	0.654 (0.211)	0.743 (0.060)	0.714 (0.100)	0.701 (0.070)	0.807 (0.053)	0.616 (0.040)
UR	0.564 (0.135)	0.000 (0.000)	0.743 (0.125)	0.630 (0.191)	0.543 (0.083)	0.833 (0.053)	0.737 (0.010)	0.291 (0.161)	0.762 (0.074)
VD	0.567 (0.094)	0.733 (0.040)	0.797 (0.082)	0.522 (0.196)	0.722 (0.116)	0.723 (0.046)	0.770 (0.053)	0.838 (0.068)	0.535 (0.081)
VS	0.798 (0.065)	0.776 (0.019)	0.830 (0.055)	0.459 (0.156)	0.729 (0.106)	0.865 (0.027)	0.761 (0.028)	0.691 (0.107)	0.673 (0.054)
ZG	0.745 (0.031)	0.602 (0.111)	0.549 (0.092)	0.546 (0.206)	0.767 (0.073)	0.749 (0.047)	0.469 (0.051)	0.708 (0.116)	0.608 (0.072)
ZH	0.837 (0.052)	0.766 (0.066)	0.830 (0.075)	0.529 (0.179)	0.810 (0.035)	0.877 (0.016)	0.719 (0.092)	0.761 (0.091)	0.681 (0.038)
CH	0.637 (0.137)	0.697 (0.121)	0.754 (0.128)	0.572 (0.228)	0.684 (0.131)	0.746 (0.107)	0.671 (0.140)	0.652 (0.182)	0.658 (0.123)
$\hat{S}_{2n}^{crs}$ (p)	0.965 (0.93)	0.888 (1.00)	n.a.	0.855 (0.98)	0.849 (1.00)	n.a.	0.881 (0.97)	0.889 (1.00)	0.960 (0.98)
$\hat{T}_{4n}$ (p)	0.001 (0.00)	0131 (0.00)	0.009 (0.00)	0.010 (0.00)	0.019 (0.00)	0.012 (0.00)	0.023 (0.18)	0.010 (0.00)	0.019 (0.00)
$\rho_{ce,te}$	0.606	0.640	0.115	0.692	0.452	0.674	0.402	0.594	0.602
$\rho_{miss,te}$	-0.063	0.101	-0.028	-0.044	-0.055	-0.070	-0.060	-0.032	0.019
N	372	339	389	288	369	223	382	378	332

Notes: * p<10%, ** p<5%, *** p<1%. Mean of year-by-year bootstrap-corrected cost efficiencies (Simar & Wilson 2011a) with standard deviation in parentheses. CH is the mean of all cantons. The $\hat{S}_{2n}^{crs}$-statistic refers to Simar & Wilson's (2002) returns to scale test; the $\hat{T}_{4n}$-statistic refers to Wilson's (2003) independence test; always the year with the smallest p-value is reported; both tests globally reject the H_0 if any year rejects it. If the p-value of the $\hat{S}_{2n}^{crs}$-test exceeds (is equal or smaller than) 0.05 then the efficiencies are estimated based on constant returns (variable) to scale; if the p-value of the $\hat{T}_{4n}$-test exceeds (is equal or smaller than) 0.05 then the heterogenous (subsampling) version of the bootstrap correction is applied. The Pearson's correlation coefficient $\rho_{ce,te}$ describes the relationship of the cost and technical efficiency, while $\rho_{miss,te}$ shows the correlation between the missing dummy and the technical efficiency. N < 390 due to outlier exclusion (see footnote 57) or negative expenditures.

Table 30 – Mean cantonal SFA technical efficiency by government function 2000 to 2014

Canton	General administration	Public order and security	Education	**Culture**	Health	Social Security	**Transportation**	Environment, spatial planning	National economy
AG	0.998 (0.001)	0.993 (0.004)	0.918 (0.052)	0.999 (0.000)	0.956 (0.024)	0.851 (0.134)	0.032 (0.003)	0.938 (0.002)	0.993 (0.002)
AI	0.999 (0.001)	1.000 (0.000)	0.947 (0.032)	0.988 (0.014)	0.898 (0.033)	0.858 (0.073)	0.011 (0.000)	0.600 (0.093)	0.611 (0.018)
AR	0.999 (0.000)	1.000 (0.000)	0.963 (0.019)		0.805 (0.189)	1.000 (0.000)	0.006 (0.000)	0.668 (0.026)	0.784 (0.030)
BE	1.000 (0.000)	1.000 (0.000)	0.938 (0.057)	0.996 (0.000)	0.942 (0.015)	1.000 (0.000)	0.348 (0.071)	1.000 (0.000)	0.998 (0.000)
BL	1.000 (0.000)	1.000 (0.000)	0.849 (0.094)	0.992 (0.001)	0.890 (0.068)	1.000 (0.000)	0.539 (0.059)	0.805 (0.021)	0.989 (0.005)
BS	0.000 (0.000)	1.000 (0.000)	0.994 (0.001)	0.995 (0.000)	0.957 (0.001)	1.000 (0.000)	0.096 (0.001)	0.989 (0.000)	0.993 (0.002)
FR	0.999 (0.000)	0.998 (0.004)	0.889 (0.071)		0.972 (0.008)	1.000 (0.000)	0.005 (0.000)	1.000 (0.000)	0.998 (0.000)
GE	1.000 (0.000)	1.000 (0.000)	0.972 (0.027)	0.998 (0.002)	0.957 (0.024)	1.000 (0.000)	0.029 (0.001)	1.000 (0.000)	0.416 (0.514)
GL									
GR									
JU	1.000 (0.000)	1.000 (0.000)	0.942 (0.036)	0.987 (0.003)	0.984 (0.004)	1.000 (0.000)	0.132 (0.017)	0.991 (0.000)	0.998 (0.000)
LU									
NE	0.999 (0.000)	1.000 (0.000)	0.958 (0.031)	1.000 (0.000)	0.990 (0.006)	1.000 (0.000)	0.361 (0.091)	1.000 (0.000)	0.998 (0.000)
NW	1.000 (0.000)	1.000 (0.000)	0.989 (0.005)	0.999 (0.000)	0.942 (0.017)	1.000 (0.000)	0.032 (0.001)	0.969 (0.000)	0.667 (0.034)
OW	1.000 (0.000)	1.000 (0.000)	0.994 (0.004)	1.000 (0.000)	0.970 (0.006)	1.000 (0.000)	0.029 (0.001)	0.526 (0.028)	0.744 (0.062)
SG	0.997 (0.000)	0.990 (0.006)	0.926 (0.064)	1.000 (0.000)	0.896 (0.048)	1.000 (0.000)	0.103 (0.011)	1.000 (0.000)	0.997 (0.000)
SH	0.998 (0.000)	0.861 (0.127)	0.989 (0.006)	1.000 (0.000)	0.944 (0.033)	0.971 (0.050)	0.453 (0.052)	0.959 (0.000)	0.943 (0.013)
SO	0.999 (0.001)	0.990 (0.009)	0.872 (0.088)	0.993 (0.002)	0.941 (0.023)	1.000 (0.000)	0.527 (0.059)	0.997 (0.000)	0.991 (0.004)
SZ	0.998 (0.000)	0.841 (0.105)	0.952 (0.042)	1.000 (0.000)	0.884 (0.047)	0.859 (0.128)	0.596 (0.076)	0.907 (0.008)	0.962 (0.017)
TG									
TI	0.998 (0.000)	0.996 (0.002)	0.979 (0.011)	1.000 (0.000)	0.947 (0.014)	0.986 (0.016)	0.006 (0.000)	1.000 (0.000)	0.790 (0.040)
UR	0.999 (0.001)	0.000 (0.000)	0.929 (0.066)	0.694 (0.267)	0.968 (0.006)	0.981 (0.041)	0.027 (0.001)	0.890 (0.012)	0.729 (0.038)
VD									
VS	0.999 (0.001)	0.995 (0.005)	0.924 (0.062)	1.000 (0.000)	0.981 (0.011)	1.000 (0.000)	0.005 (0.000)	0.985 (0.000)	0.900 (0.067)
ZG	0.999 (0.000)	0.991 (0.012)	0.992 (0.003)	0.990 (0.007)	0.968 (0.014)	0.927 (0.079)	0.046 (0.003)	0.987 (0.000)	0.941 (0.032)
ZH	1.000 (0.000)	0.999 (0.001)	0.973 (0.017)		0.795 (0.078)	0.860 (0.098)	0.901 (0.005)	0.999 (0.000)	0.996 (0.001)
CH	0.999 (0.001)	0.981 (0.060)	0.945 (0.061)	0.977 (0.100)	0.934 (0.075)	0.958 (0.083)	0.208 (0.089)	0.910 (0.030)	0.876 (0.193)
$\hat{\mathbb{E}}[\sigma_{u,it}]$	0.025**	0.021**	0.076***	0.041***	0.141***	0.026**	0.137***	0.152***	0.227**
$\hat{\sigma}_v$	0.066***	0.055***	0.029***	0.109***	0.147***	0.058***	0.093***	0.085***	0.031***
$\hat{\mathbb{E}}[\lambda_{it}]$	0.381	0.377	2.611***	0.375	0.960	0.453	1.475	1.797***	7.256***
$\rho_{CE,TE}$	-0.016	0.183	0.140	-0.033	-0.0006	0.0274	-0.245	0.529	-0.113
$\rho^*_{DEA,SFA}$	-0.103	-0.013	0.023	0.189	-0.365	-0.123	-0.159	0.303	-0.312
N	285	258	300	214	282	161	298	294	260

Notes: * p<10%, ** p<5%, *** p<1%**.** The Spearman's rank correlation coefficients $\rho^*_{DEA,SFA}$ describes the relationship of the DEA and SFA technical efficiency measures, while the Pearson's correlation coefficient $\rho_{CE,TE}$ indicates how the cost and technical efficiency estimates interrelate; $\hat{\mathbb{E}}[\sigma_{u,it}]$ is tested against H_0: $\hat{\mathbb{E}}[\sigma_{u,it}] = 0$; $\hat{\sigma}_v$ is tested against H_0: $\sigma_v = 0$; $\hat{\mathbb{E}}[\lambda_{it}]$ is tested against H_0: $\lambda_{it} \leq 1$; for further regression outputs see Table 20.

Table 31 – Truncated regression of the DEA efficiency score on environmental variables by government function (two-step approach)

	General administration	Public order and security	Education	Health	Social Security	Environment, spatial planning	Economy
[A] Expenditures through special financings (in % of total expenditures)	0.004** (0.002)	-0.004 (0.004)	-0.012 (0.015)	0.000 (0.003)	-0.015*** (0.005)	0.000 (0.001)	0.000 (0.001)
[B] Volume of special financings (in % of balance sheet's total)	-0.012 (0.023)	-0.123** (0.058)	-0.016 (0.016)	0.001 (0.005)	-0.022 (0.037)	-0.016 (0.015)	0.000 (0.004)
Interaction [A]x[B]	0.001 (0.002)	0.007 (0.007)	0.002 (0.002)	-0.001 (0.001)	0.021** (0.008)	0.000 (0.000)	0.000 (0.000)
[C] Tax progressivity (curvature parameter)	0.010 (0.027)	0.091*** (0.032)	0.016 (0.041)	0.048 (0.030)	-0.006 (0.027)	0.002 (0.041)	0.011 (0.021)
[D] GDP per capita (in 1'000 CHF)	-0.002 (0.002)	0.006** (0.003)	-0.002 (0.004)	0.008*** (0.003)	-0.003 (0.002)	-0.004 (0.004)	0.006*** (0.002)
Interaction [C]x[D]	0.001* (0.000)	-0.001*** (0.000)	0.000 (0.001)	-0.001** (0.000)	0.001*** (0.000)	0.001 (0.001)	-0.001** (0.000)
Tax complexity (normalized HHI)	-0.001 (0.001)	-0.003* (0.002)	0.000 (0.002)	0.000 (0.002)	0.002* (0.001)	-0.003 (0.002)	0.002 (0.001)
Tax diversification (number of tax sources)	-0.003 (0.014)	0.026 (0.018)	-0.028 (0.023)	-0.010 (0.019)	-0.009 (0.016)	0.009 (0.022)	-0.002 (0.011)
Transfers (in % of total revenues)	0.001 (0.001)	-0.004*** (0.001)	-0.001 (0.001)	0.000 (0.000)	0.002** (0.001)	0.001** (0.001)	0.000 (0.001)
Gross debt per capita (in 1'000 CHF)	0.000 (0.004)	0.003 (0.004)	0.008 (0.006)	0.011*** (0.004)	0.005 (0.003)	-0.001 (0.006)	0.003 (0.003)
Population density (1'000 inh. per km^2)	-0.889*** (0.327)	-0.841** (0.358)	-0.439 (0.539)	-1.981*** (0.392)	-0.232 (0.311)	0.368 (0.514)	-0.015 (0.243)
Referendum (1 [hard] to 6 [easy])	0.054* (0.029)	-0.012 (0.016)	-0.008 (0.025)	-0.078*** (0.028)	0.017** (0.009)	0.013 (0.024)	-0.010 (0.010)
Initiative (1 [hard] to 6 [easy])	-0.122*** (0.041)	0.090** (0.042)	-0.078 (0.061)	0.050 (0.044)	-0.049* (0.026)	0.050 (0.062)	0.048* (0.026)
Bourgeois parties in executive (in %)	-0.001* (0.001)	0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.002*** (0.001)	-0.001 (0.001)	-0.001*** (0.000)
Fragmentation (1'000 inh. per municipality)	0.085*** (0.027)	0.012 (0.030)	0.006 (0.043)	0.020 (0.035)	0.087*** (0.021)	0.022 (0.050)	-0.046** (0.019)
Voter turnout (in %)	0.002* (0.001)	0.001 (0.001)	-0.003** (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	0.001* (0.001)
Foreigners (in % of total inhabitants)	-0.003 (0.006)	-0.003 (0.007)	0.016 (0.010)	-0.014* (0.008)	-0.005 (0.005)	0.017* (0.010)	0.012*** (0.005)
National street length (in km)							
Marginal effect: **Expenditures through special financings**	0.004*	-0.002	-0.011	-0.001	-0.009***	-0.000	-0.000
Marginal effect: **GDP per capita**	0.002*	-0.001	-0.001	0.001	0.005***	0.000	0.002**
N	285	258	300	282	161	294	260

Notes: * p<10%, ** p<5%, *** p<1%**.** Coefficients with bootstrap-based standard errors in parentheses (Simar & Wilson 2007); the dependent variable is the DEA technical efficiency score with a contemporaneous reference set (i.e. year-by-year DEA); individual dummies included but not reported; marginal effects at mean of interaction variable.

Table 32 – True fixed- and true random-effect model estimates

	General administration (RE)		Public order and security (RE)		Education (FE)		Culture (RE)		Health (RE)		Social Security (RE)		Transportation (RE)		Environment, spatial planning (FE)		National Economy (RE)	
Frontier	Input: O	-0.109***	Input: O	0.049***	Input: O	-0.035	Input: O	-0.316***	Input: O	-0.028	Input: O	-0.043	Input: O	0.048	Input: O	-0.093**	Input: O	-0.032**
		(0.029)		(0.019)		(0.052)		(0.019)		(0.046)		(0.027)		(0.047)		(0.043)		(0.013)
	Input: F	-0.011	Input: F	0.000	Input: F	-0.401***	Input: F	-0.662***	Input: F	-0.708***	Input: F	-0.926***	Input: F	-0.255***	Input: F	-0.071**	Input: F	-0.899***
		(0.009)		(0.012)		(0.114)		(0.019)		(0.022)		(0.021)		(0.062)		(0.033)		(0.020)
	Input: I	-0.008	Input: I	-0.007	Input: I	0.000	Input: I	-0.004	Input: I	-0.033***	Input: I	-0.014	Input: I	-0.007	Input: I	0.006	Input: I	-0.007
		(0.010)		(0.006)		(0.005)		(0.010)		(0.011)		(0.011)		(0.040)		(0.018)		(0.005)
	Interest rate [a]	0.045**	Convicts	0.213***	Students	0.148	Youth and sport	0.099***	Doctors	1.098***	Unemployed	0.169***	Road length	0.539***	Vacant accomm'	-0.026	Farms	0.631***
		(0.017)		(0.039)		(0.106)		(0.028)		(0.084)		(0.028)		(0.086)		(0.037)		(0.022)
	Firms	0.137***	Delinquencies [a]	0.188***	Univers' diplomas	0.033	Museum tickets	0.075***	Patient cases	-0.094	Reintegr' people	-0.005	Vehicles	0.458***	Population	0.619	Founded firms	0.076***
		(0.019)		(0.032)		(0.041)		(0.021)		(0.082)		(0.014)		(0.134)		(0.748)		(0.019)
	Population	0.831***	Prisoner on remand	-0.008	H' school grad'	0.034	Forest area	-0.011			People over 65	0.856***	Accidents [a]	0.002			Beds in hotels	-0.026***
		(0.017)		(0.007)		(0.028)		(0.008)				(0.019)		(0.005)				(0.003)
	Year	0.007**	Time	0.017***	Time	0.024***	Time	0.017***	Time	0.008	Time	0.006**	Time	0.897	Time	0.003	Time	0.023***
		(0.003)		(0.002)		(0.005)		(0.004)		(0.006)		(0.003)		(10.343)		(0.008)		(0.001)
	Intercept	-0.109***	Intercept	-22.87***			Intercept	-26.44***	Intercept	-0.028	Intercept	-0.043	Intercept	0.048			Intercept	-36.26***
		(0.029)		(3.722)				(8.080)		(0.046)		(0.027)		(0.047)				(1.228)

	General administration (RE)	Public order and security (RE)	Education (FE)	Culture (RE)	Health (RE)	Social Security (RE)	Transportation (RE)	Environment, spatial planning (FE)	National Economy (RE)
Sigma u_{it}									
Expenditures special financings	0.096 (0.327)	0.248 (0.188)	0.578 (0.465)	0.017 (0.068)	0.243 (0.156)	-12.977** (5.312)	0.020 (0.018)	-0.034** (0.016)	-0.026 (0.068)
Volume special financings	4.059 (2.749)	0.829 (2.237)	0.354 (0.277)	0.515 (0.890)	0.015 (0.157)	-2.097 (8.420)	0.349* (0.190)	-1.074* (0.610)	0.105 (0.152)
Interaction	-0.200 (0.274)	-0.346 (0.266)	-0.084 (0.068)	0.004 (0.032)	-0.008 (0.016)	2.943 (14.165)	-0.003 (0.005)	0.016** (0.008)	0.013 (0.013)
Tax progressivity	3.492 (4.463)	1.089 (1.517)	3.522*** (1.222)	4.392*** (1.058)	0.380 (1.238)	-11.258*** (3.452)	0.856 (1.166)	-3.616* (2.057)	1.265** (0.620)
GDP per capita	0.335 (0.343)	0.030 (0.055)	0.107** (0.044)	0.502*** (0.179)	-0.002 (0.104)	0.124 (0.141)	0.029 (0.080)	-0.222** (0.106)	0.073* (0.037)
Interaction	-0.081 (0.061)	-0.028 (0.022)	-0.047*** (0.013)	-0.057*** (0.021)	-0.020 (0.033)	0.048** (0.019)	-0.025 (0.032)	0.038** (0.017)	-0.030** (0.012)
Tax complexity	-0.240 (0.296)	-0.059 (0.083)	0.041 (0.050)	0.043 (0.055)	-0.162* (0.098)	-0.027 (0.075)	-0.021 (0.077)	-0.012 (0.053)	-0.045 (0.056)
Tax diversification	-1.396 (1.653)	-1.916*** (0.585)	-0.338 (0.207)	-2.530** (0.997)	-0.480 (0.351)	-1.520 (1.162)	0.815** (0.345)	0.131 (0.519)	0.173 (0.178)
Transfers	0.042 (0.246)	-0.229** (0.092)	0.043* (0.025)	-0.031** (0.015)	-0.028 (0.027)	0.206 (0.130)	0.012 (0.027)	0.038 (0.027)	0.013 (0.015)
Gross debt p. cap.	-0.793 (0.602)	-1.792*** (0.258)	-0.077 (0.177)	0.502* (0.276)	-0.140 (0.489)	-1.784*** (0.530)	0.264*** (0.086)	0.246 (0.415)	0.436*** (0.110)
Population density	-12.959 (62.591)	-5.664 (3.967)	2.985 (7.223)	20.401* (12.098)	11.934 (14.591)	28.404** (11.420)	1.327 (7.142)	-0.517 (4.056)	0.892 (2.424)
Referendum	0.558 (1.827)	0.384 (0.318)	1.193** (0.605)	0.022 (1.149)	-0.722 (0.866)	-0.271 (0.477)	-1.311 (0.927)	-1.513 (0.955)	-0.250 (0.494)
Initiative	0.018 (2.644)	1.847*** (0.622)	-0.698 (0.576)	1.300* (0.785)	-0.672 (1.431)	-3.600** (1.696)	0.621 (0.565)	6.954** (3.402)	0.433 (0.277)
Bourgeois parties	0.000 (0.101)	-0.064** (0.026)	-0.037 (0.027)	-0.051 (0.054)	0.055*** (0.019)	-0.141** (0.071)	0.029 (0.020)	0.034 (0.021)	-0.008 (0.016)
Fragmentation	-1.419 (1.537)	-0.935*** (0.344)	-0.020 (0.461)	-1.658* (0.967)	-0.606 (0.752)	-6.212*** (2.331)	-0.005 (0.389)	0.013 (0.268)	-0.129 (0.152)
Voter turnout	-0.065 (0.088)	0.012 (0.036)	-0.024 (0.026)	0.035 (0.023)	-0.005 (0.047)	0.063 (0.055)	0.047 (0.074)	-0.004 (0.025)	0.036** (0.017)
Foreigners	1.137 (1.397)	2.069*** (0.278)	0.215 (0.176)	-1.194** (0.511)	-0.208 (0.209)	0.989** (0.474)	-0.066 (0.098)	-0.315* (0.165)	0.051 (0.076)
National streets							0.009 (0.018)		
Intercept (δ_0)	-11.253 (43.124)	-25.965*** (8.507)	-21.355 (14.028)	-15.608*** (5.946)	18.365 (13.239)	45.338 (40.745)	-13.223** (6.106)	-14.772 (22.041)	-13.344*** (5.075)
Sigma v									
Intercept (ω_0)	-5.426*** (0.194)	-5.809*** (0.169)	-7.073*** (0.202)	-4.435*** (0.154)	-3.832*** (0.207)	-5.707*** (0.194)	-4.746*** (0.623)	-4.938*** (0.230)	-6.930*** (0.216)

[a] Index as described in section 6.2

Notes: * p<10%, ** p<5%, *** p<1%. Beta, delta and omega coefficients with robust standard errors clustered by canton in parentheses; the upper part (frontier) describes the frontier function and hence all variables are in logs due to the Cobb-Douglas specification; the middle (Sigma u_i) part shows the coefficients of the $\sigma_{u,it}$ parametrization, where the environmental variables coincide with those in Table 20; the lower part (Sigma v) shows the estimate of $\omega_0 = 2\ln(\sigma_v)$; (RE) and (TE) indicate the true random or true fixed effect model.