Government
Institute for
Economic Research

Working Papers 19

Coalition formation and political decision making: Evidence from Finnish municipal mergers

Tuukka Saarimaa

Janne Tukiainen

Working Papers 19 December 2010

VATT WORKING PAPERS

19

Coalition formation and political decision making: Evidence from Finnish municipal mergers

Tuukka Saarimaa
Janne Tukiainen

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2010

We thank Essi Eerola, Elias Einiö, Ronny Freier, Ari Hyytinen, Aki Kangasharju, Kaisa Kotakorpi, Jani-Petri Laamanen, Teemu Lyytikäinen, Antti Moisio, Rohini Pande, Jaakko Pehkonen, Matti Sarvimäki, Otto Toivanen and Eric Weese for helpful comments. We also thank seminar participants at EPCS 2010, FEAAM-Tampere, HECER, Jyväskylä, STICERD at LSE and VATT for useful discussions. Oskari Harjunen and Niklas Jahnsson provided excellent research assistance. Janne Tukiainen is grateful to Emil Aaltonen Foundation, Jenny and Antti Wihuri Foundation and Yrjö Jahnsson Foundation for financial support and STICERD at the London School of Economics and Political Science for hospitality. We also thank Sitra, the Finnish Innovation Fund for funding.

Tuukka Saarimaa, Government Institute for Economic Research (VATT),
P.O. Box 1279, FI-00101 Helsinki, FINLAND, E-mail: tuukka.saarimaa@vatt.fi

Janne Tukiainen, Helsinki Center of Economic Research (HECER) and VATT,
P.O. Box 1279, FI-00101 Helsinki, FINLAND, E-mail: janne.tukiainen@vatt.fi

ISBN 978-951-561-953-2 (nid.)

ISBN 978-951-561-954-9 (PDF)

ISSN 1798-0283 (nid.)

ISSN 1798-0291 (PDF)

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Arkadiankatu 7, 00100 Helsinki, Finland

Oy Nord Print Ab

Helsinki, December 2010
Graphic design: Niilas Nordenswan

Coalition formation and political decision making: Evidence from Finnish municipal mergers

Government Institute for Economic Research
VATT Working Papers 19/2010

Tuukka Saarimaa – Janne Tukiainen

Abstract

This paper analyzes empirically the coalition formation of local governments. We introduce a novel econometric strategy involving choice based sampling from a spatial network to allow for multi-partner mergers in our empirical analysis. We apply our method to recent municipality mergers in Finland. The mergers were decided voluntarily by municipal councils but the central government promoted mergers through a subsidy scheme. Our main interest lies on the association of local politics with the merger decisions. Moreover, we are able to estimate the causal effect of the subsidy scheme on the merger decisions using a regression discontinuity design. We find that the local political environment is relevant for the merger decision making. The results are consistent with politicians' strategic behavior concerning private incentives on municipal employment possibilities and re-election. Furthermore, more concentrated political power at the local level seems to promote merging and different political parties hold different views concerning merging. The central government merger subsidy scheme has an effect on the type of mergers that took place. Overall, the results imply that local politics may hinder optimal coalition formation and that the central government may act as a corrective force by using an appropriate subsidy scheme.

Keywords: Coalition formation, Local politics, Merger subsidy, Choice based sampling, Regression discontinuity design

JEL classification numbers: H71, H72, H77, C35

Tiivistelmä

Tässä tutkimuksessa tarkastellaan empiirisesti kuntaliitosten syntyyn vaikuttavia tekijöitä. Tutkimuksessa esitellään uusi spatiaalisesta verkostosta tapahtuva ositettu valintaperusteinen otantamenetelmä, jota hyödynnetään Suomen kuntaliitosten analysoinnissa. Otantamenetelmä mahdollistaa monikuntaliitosten analysoinnin. Tutkimuksen keskeisenä mielenkiinnon kohteena on kunnallispolitiikan yhteys kuntaliitospäätöksiin. Lisäksi tarkastellaan valtion myöntämän yhdistymisavustuksen kausaalivaikutusta regressioepäjatkuvuusmenetelmän avulla. Tulosten mukaan kunnallispolitiikalla on merkittävä rooli kuntaliitosten taustalla. Kunnallispolitiikka ja kunnanvaltuutettujen henkilökohtaiset poliittiseen uraan ja siviilityöuraan liittyvät vaikuttimet näyttävät vaikeuttavan kunnan asukkaiden kannalta optimaalisten kuntaliitosten muodostumista. Toinen merkittävä tulos tutkimuksesta on se, että valtion myöntämällä yhdistymisavustuksella on ollut positiivinen vaikutus kuntaliitoksiin. Näin ollen oikein räätälöidyllä yhdistymistuella on periaatteessa mahdollista korjata paikallispolitiikan aiheuttamat ongelmat liitostilanteissa.

Asiasanat: Kuntaliitokset, kunnallispolitiikka, yhdistymistuki, valintaperusteinen otanta, regressioepäjatkuvuusmenetelmä

JEL-luokittelu: H71, H72, H77, C35

Contents

1 Introduction

The formation and dissolution of nations and local jurisdictions like school districts, municipalities or counties, agreeing on political coalitions and deciding memberships in international cooperative organizations like the UN and WTO are all examples of coalition formation. Both history and recent news are full of examples that highlight the practical importance of understanding the mechanisms that, for example, draw the borders between nations. Both the formation and dissolution of Soviet Union is an example of large scale coalition formation and in Belgium the question whether it should be one or two nations is always bubbling under. Due to its significant economic and political importance, coalition formation has attracted interest from both policy makers and theoretical and empirical researchers. The main questions in this literature are how many and what type of coalitions should be optimally formed and how is the optimal distribution of coalitions achieved.

Economic theorists have been active for long in this field. The vast literature on fiscal federalism is concentrated on the question of assignment of functions and taxes to different levels of government[1]. A central result summarized in Oates' (1972) decentralization theorem is that in the absence of cost-savings from the centralized provision of public goods and of interjurisdictional externalities, decision making should be decentralized. A related result due to Tiebout (1956) offers decentralization as a tool for efficient pricing of local public goods in a world where mobile households shop for a suitable tax and public good combination offered by different local jurisdictions. However, these strands of literature leave open the issue of where and how local jurisdictions arise in the first place.

Later work on endogenous mergers and coalition formation by e.g. Miceli (1993), Alesina and Spolaore (1997, 2003) and Bolton and Roland (1997) emphasize the fundamental trade-off between economies of scale in the production of public goods (which favors merging) and regional heterogeneity in preferences (which favors separation). This trade-off determines the optimal size of a given coalition. Bradford and Oates (1974) argue that internalizing inter-jurisdictional spillovers may also justify mergers of local jurisdictions. Alesina and Spolaore (1997), Bolton and Roland (1997) and Ellingsen (1998) discuss how politics may cause departures from welfare maximizing coalition formation. Alesina and Spolaore (1997) show that the equilibrium under within jurisdiction majority voting is suboptimal to the one achieved under a global social planner. Empirical contributions to this field are more recent. Brasington (1999, 2003a, 2003b), Gordon and Knight (2009) and Weese (2009) use different methods but are all fundamentally concerned with optimal coalition formation and the

[1] See Oates (1999) for an overview of this literature.

characteristics of optimal coalitions in terms of economies scale and preference heterogeneity. However, they abstract away from political considerations.

Our main contribution is to provide the first empirical analysis of departures from optimality in coalition formation by asking whether local politics hinder the optimal formation of local governments. To our knowledge, there is no previous empirical work that looks at the political decision making process behind coalition formation. The institutional background to our analysis is a recent surge of municipality mergers in Finland. Empirical analysis of local government coalition formation, in particular, is very useful in order for us to understand the driving forces behind coalition formation, because it allows for many observations of merger decisions that take place within the same legal, economic and cultural institutions while plausibly maintaining external validity with respect to other interesting coalition formation games. Empirical analysis using, for example, nations as the unit observation would be very challenging in comparison.

Besides providing important information on coalition formation mechanisms in general, our application of municipal mergers should be of practical interest to countries with a multilayer system of government. There have been a number of recent examples of municipality mergers, such as in Germany in the late 1990's and 2000's, in Japan in the late 1990's and early 2000's and in Denmark in 2007. In all these cases, the merger decisions were mostly made at the local level. This raises the important question of how the local political environment affects the merger decisions. For example, do local politicians really care about economies of scale in local public good or service provision or are they more interested in their own re-election chances if a particular merger goes through? Or more generally, who, the central or the local governments, should decide the number and characteristics of the coalitions?

The Finnish municipal merger data is of special interest because we observe some interesting characteristics of the municipal councils that allow us to study whether council members' private incentives are associated with their decisions concerning mergers. The mergers in our data were decided voluntarily by the municipal councils and we are interested in whether the local political environment, as measured by different municipal council characteristics, hinders otherwise sensible mergers or promotes suboptimal mergers.

Another attractive feature of our data set is that simultaneously to analyzing the potential departures from optimality caused by the local politics, we are able evaluate one potential solution to this problem. Namely, we evaluate the effectiveness of a central government merger subsidy program. The Finnish central government grants merger subsidies to merging municipalities according to a complex step function of the number of disappearing municipalities and populations of the municipalities involved in the merger. This exogenously set step function allows us to use regression discontinuity design to analyze the causal effect of the subsidy program on merger activity. That is, we are able to study whether local politics hinder optimal coalition formation and whether the central government is able to influence the decisions of local governments. In accordance with existing empirical literature, we also study how municipal population characteristics, characteristics related to costs of providing statutory services and municipal tax bases are associated with merger decisions. We also analyze within coalition heterogeneity of these characteristics.

Econometric modeling of spatial merger decisions is complicated for a number of reasons. In our analysis, the main econometric problem that we face is the high number of multi-partner mergers or one-to-many matching. We solve this problem by combining Wernicke's (2005) network detection algorithm with stratified choice-based sampling. This sampling design is our main methodological contribution. It has many attractive properties for future coalition formation studies. Namely, that we can apply the results from voluminous choice-based sampling literature directly, we are able to use computationally simple reduced form estimation techniques that, unlike the existing methods, allows for a large set of covariates and we can conduct statistical inference with a relatively small amount of actual mergers. Moreover, the sampling design allows us to incorporate existing tools of causal inference into the analysis. Using our stratified choice based sampled data we analyze coalition formation by using a conditional logit analysis where potential coalitions are used as the unit of observation and also by estimating Poirier's (1980) bivariate probit model with partial observability which can accommodate for two-sided decision making of individual municipalities. By sorting the data by population, as suggested by Brasington (2003a), the Poirier model allows us to analyze whether some of the associations are heterogeneous with respect to the size of potential merger partners. We incorporate our regression discontinuity design to both of the modeling approaches.

Our main result is that the local political environment has a very important role in the merger decision making. Councilors seem to act strategically indicating concerns over both re-election and future employment prospects of their day job. Furthermore, political fragmentation seems to hinder coalition formation. Different political parties also have different preferences concerning merging. This means that both the political environment and the private incentives of local politicians may hinder optimal coalition formation. Based on our regression discontinuity analysis, the government merger subsidy scheme has been a successful policy, at least in a sense that it has had an effect on the likelihood and type of mergers that took place. This indicates that the central government may be able act as counterforce to local politics and possibly pave way to more optimal mergers. However, analyzing the welfare effects of this policy is beyond the scope of this paper. Overall, our empirical results call for richer theoretical analyses that should model the individual decision making of each policy maker within each jurisdiction instead of abstracting to the jurisdiction level decision making or majority voting decision making as the current literature does.

The rest of the paper is organized as follows. Section 2 presents a more detailed description of Finnish municipalities and the institutional background for municipality mergers. Section 3 looks through previous empirical studies on local government mergers and coalition formation. Section 4 describes the econometric methods and data. Section 5 presents the empirical results and section 6 concludes.

2 Institutional background

Local governments play an important role in the economies of many countries. This is especially true in the Nordic countries like Finland where there has been a long tradition of decentralized political decision making. In Finland, public goods and services are provided by two tiers of government where municipalities constitute the local level. Because of the variety of tasks assigned to them, municipalities are of considerable importance to the whole economy. The GDP share of municipality spending is roughly 18 percent and they employ around 20 percent of the total workforce. Unlike in the mainstream economic models of local public finance, where local governments provide local public goods, the bulk of Finnish municipalities' expenditures come from producing welfare services with a strong redistributive character. Municipalities are responsible for providing the most of social and health care services, primary education and culture services. In addition, municipalities provide the basic environment and technical infrastructure services along with public transportation. Municipalities also have a zoning monopoly within their borders.

Municipalities fund their spending mostly through their own revenue sources. The most important revenue source is the flat municipal income tax which constitutes 40 percent of all municipality revenues. Municipalities are also entitled to a share of the corporate income tax paid by corporations in the municipality, but corporate taxes make up only about 4 percent of revenues. The role of the property tax is even smaller at only 2.5 percent of revenues. Due to the fact that most municipal services in Finland are statutory and large disparities in municipal population structure and density, a central government grant system is used to equalize local cost and revenue disparities. The grant system is based on calculatory costs and compensates municipalities with unfavorable cost conditions, expensive age structure and social problems. The system also includes a separate net equalization system for tax bases. The grant system covers 20 percent of total municipal spending. Growing regional disparities have led to unsustainable deficits in many small and remote municipalities and for them the grant system has become the most important source of revenue. Currently, the grant system covers more than 50 percent of all revenues for every fourth municipality. The rest of municipal revenues consist of user-fees and sales incomes.

Another important feature of the Finnish system is the large number of municipalities relative to population with a large variation in municipal population size. The largest municipality is the country's capital, Helsinki, which has over half a million inhabitants, whereas the smallest mainland municipality, Suomenniemi, has roughly 800 inhabitants. The median municipal population size is less than 6,000. Finland is also sparsely populated and population density varies a great deal between municipalities. The number of municipalities has diminished considerable since the 1940's as can be seen from Figure 1. There have been two major merger waves. The reduction in the number of municipalities in the 1940's was due to losing land areas to the Soviet Union after the WWII. In the 1960's almost all and in the 2000's all of the mergers have been voluntary.

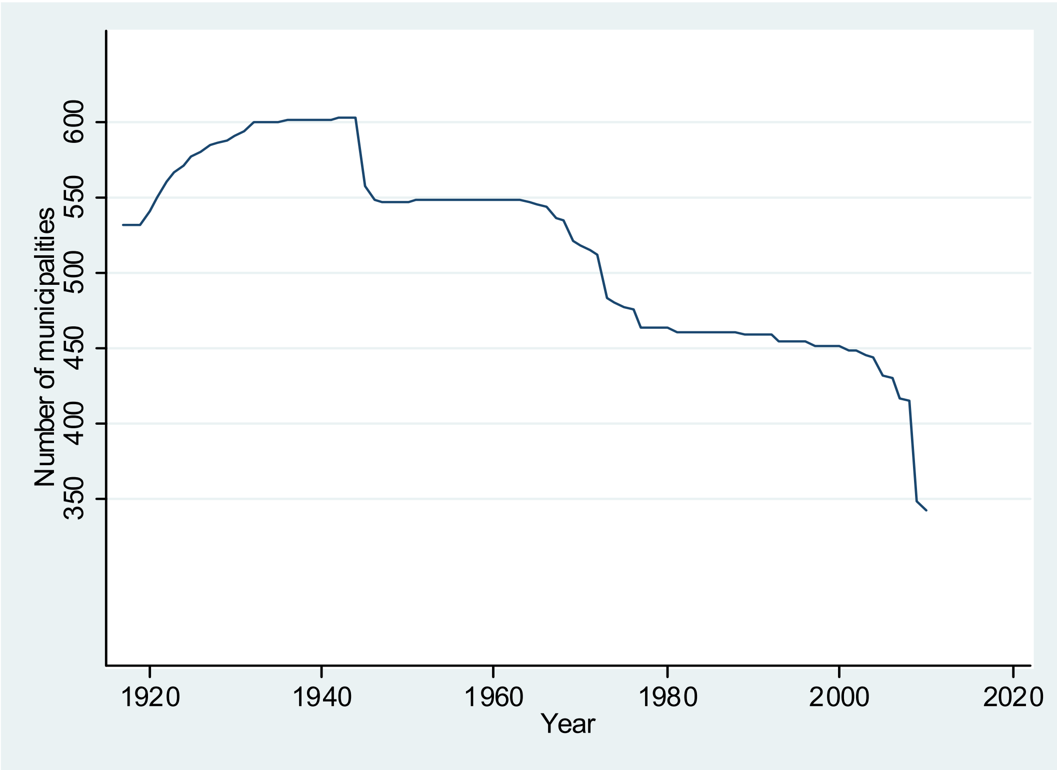

Figure 1. Number of municipalities in Finland, 1917–2010.

Figure 1 illustrates that municipality structure has been quite stable through time but every now and then we observe a wave of municipality mergers. This raises the question of why do we observe municipality mergers at a given time. Ellingsen (1998) argues that economies of scale are a questionable rationale for merging since they can be just as easily exploited through contracting or cooperation. In fact, cooperation in service production through joint authorities is quite common among Finnish municipalities.[2] As Miceli (1993) argues, cooperation in producing some services but not others is a way to deal with the fact that different services have different economies of scale. One must therefore

[2] Currently, there are 226 joint authorities set up by two or more municipalities. Joint authorities are independent legal entities governed by municipal legislation. They are financed by selling their services to municipalities. The most important joint authorities are hospital districts, basic health care, care for disabled and vocational education. Membership is voluntary, except that every municipality has to belong to a hospital district.

ask what the additional benefits or costs are that merging provides compared to cooperation.

One apparent difference is that merging depends on and alters political power structures in a way that cooperation does not. There are many aspects to this argument. First, it is obvious that the costs and benefits of mergers are not equally divided among all members. Clearly there are winners and losers. In a democratic system with majority voting, mergers may not occur even when merging would be efficient. When all parties must accept the merger, a majority against the merger in just one municipality is enough to block it. Second, efficient mergers may be blocked or inefficient mergers may take place because of local politicians and bureaucrats self-interest. For example, municipal employees who are also council members may have private incentives either to vote for or against merging regardless of the preferences of their constituents. This is especially true when merging is motivated mainly on cost efficiency grounds. Third, council members of a particular party may find a merger undesirable because in the new municipality their party's share of council seats would be lower or that they would face stiffer competition from members of their own party. Fourth, large municipalities have a stronger leverage in negotiations between the central government and the municipal sector.[3]

On the other hand, if households have sorted themselves to municipalities in the spirit of Tiebout (1956), it is not clear whether one should promote or even expect merger activity. Some recent developments in Finland have, however, changed the operating environment of municipalities. A major recession in the early 1990s hit different regions with a different force due to differences in industrial composition. The subsequent recovery was regionally uneven and regional disparities started to grow. In particular, during the last fifteen years or so, regional disparities have grown rapidly because the workforce has concentrated into few growing centers with better employment possibilities. This means that both the municipalities that are losing population and the ones gaining have experienced major changes in their operating environment. Furthermore, rapid population aging continues to challenge municipal finance in the future because a bulk of municipal expenditures is related to health and elderly care. This development may be enough to encourage municipalities to search for new ways of cooperation in service production or even to merge.

Due to aging-related expected increase in municipal spending and disparities in revenue base the central government initiated a plan (so-called PARAS initiative) in 2005 that aimed at reforming municipal revenue structure and more importantly making the production of statutory municipal services more efficient.

[3] Finnish municipalities are self-governing bodies by constitution which means that central government cannot assign new responsibilities to municipalities without passing new legislation. However, the central government may, for example, change municipal tax bases and assign limits to property tax rates quite easily.

In 2007 these goals were made concrete by a provisional law. The main tool for strengthening the operating environment of municipalities set forth in the new law was municipality mergers. The law clearly states that municipalities should have strong enough revenue and labor force bases to cope with the production of statutory public services. Municipalities are allowed to decide for the mergers themselves. However, the government initiated a new subsidy scheme to encourage merger activity. We discuss the subsidy scheme in detail later, when we show how to use regression discontinuity analysis to study its effectiveness.

Finland has a multi-party system. Currently there are eight parties in the Finnish parliament and these parties also dominate municipal politics, but some local single-issue groups exist as well. The municipal council is responsible for strategic and financial outlines and main objectives of municipal activities. The council also chooses the municipal board. The composition of the board is based on party shares in the council, i.e. each party in the council get seats in the municipal board according to their share of council seats. Thus, there is no real opposition in local politics. Municipal boards have a strong role because the most important issues prepared for the council are agreed in advance by the board.

In this paper, we focus on mergers that took place in 2008 and 2009. The councils in our data were elected in October 2004 and these councils voted for all the actual mergers in our analysis. Thus, we can treat the mergers as a result of a one period coalition formation game. A given merger takes place if it gains the majority vote in all the individual municipalities contemplating the merger. New municipal councils are elected every four years.[4] The councils are elected using list elections that apply the D'Hondt method. Each municipality has only one electoral district. This goes also for the merged coalitions. Council size is a step function of municipality's population and is determined by law. As can be seen from Table A1 in Appendix A, the council size varies from 13, for municipalities with population 2,000 or less, up to 85 for municipalities with population 400,000 or more. Therefore, in case of a merger of a very large municipality and a very small municipality, it is likely that all the council members in the new coalition will be residents of the former larger municipality.

[4] See Moisio et al. (2010) for more details on municipal politics.

3 Previous empirical studies

Political coalition formation has received considerable attention from both theoretical and empirical economists. In this section we concentrate on empirical contributions so that these results can be compared to ours. Alesina et al. (2004) examine the number of political jurisdiction, such as school districts and municipalities, within counties in the U.S. In particular, they test whether a trade-off between economies of scale and population heterogeneity can explain the number and size of jurisdictions. Using data from the period of 1960–1990, Alesina et al. (2004) find evidence of a trade-off between economies of scale and heterogeneity in race and income. However, they do not analyze actual merging decisions of these entities.

Brasington (1999) uses Poirier's partial observability bivariate probit which allows for two-sided decision making to study school district consolidations in Ohio U.S. He finds that racial composition and the difference in racial composition have no independent effect on consolidation, nor do income levels or school quality. He argues that consolidation happens mainly with cost savings in mind. In particular, he finds a general tendency of small and large school districts to merge, whereas middle-sized districts merge less often. Brasington (2003a) uses the same data as Brasington (1999) but sorts the data according to population in order to see whether small and large communities differ in their preferences to merge. Brasington (2003a) finds clear differences in the way large and small communities react to differences in population size. The larger the population gap between the large and small community, the more the large community wants to merge. The opposite is true for small communities. He speculates that a large community gains scale economies while probably maintaining political control over schooling decisions. This result is in line with predictions from Ellingsen's (1998) model. Brasington (2003b) uses the same sorting idea as Brasington (2003a) but now sorting is done according to income and race (percentage of white). That is, richer communities are compared to poorer and whiter to darker in each potential consolidation pair. Brasington (2003b) finds that the higher property values in a community are the less likely it is to merge. This is true for richer and poorer potential partners. When data are arranged according to percentage of white, the effect of property value is still negative but it is statistically significant only for less white communities.

More recently, Gordon and Knight (2009) and Weese (2009) have proposed structural econometric methods to analyze spatial mergers. Gordon and Knight (2009) apply their spatial simulated method of moments estimator to school district mergers in Iowa in the 1990's. They find that state subsidies were an important factor influencing merger decisions of school districts. They also find that economies of scale matter. Furthermore, heterogeneity in education level of district population (percentage with a college degree), heterogeneity in spending and distance between potential merger partners makes merging less likely. Weese (2009) introduces a novel method of estimating the structural parameters of a coalition utility function in a coalition formation game and applies the method to municipality mergers in Japan. He finds that high amount of government services and low population are preferred.

4 Sampling and econometric modeling

Econometric modeling of merger decisions is complicated for a number of reasons. First, merger decision making is two-sided in the sense that we observe a merger only if all merger parties agree to merge. Second, each municipality faces multiple potential merging partners but can merge only once during a given period. It may be challenging both to account for all the potential merger partners and to restrict the sum of the probabilities of all the potential mergers to one. Third, municipal merger choices are spatially interdependent so that a merger between two or more municipalities changes the choice set of adjacent municipalities. Finally, there is the possibility of one-to-many matching, meaning that coalitions that are larger than two are possible. This means that the number of potential partners and coalitions may be very large. In this case, it is not clear what the underlying population from which to sample is and it may be very costly to collect all the relevant data.

There are three estimators in the literature that account for some or all of these features. First, Poirier's (1980) bivariate probit model with partial observability can accommodate for two-sided decision making but not the other features. This approach has been used by Brasington (1999, 2003a and 2003b). Second, the matching estimator proposed by Gordon and Knight (2009) takes into account two-sided decision making, spatial interdependence and multiple potential partners but does not allow for one-to-many matching. Mergers including multiple municipalities are a prominent feature of our data which makes the Gordon-Knight estimator infeasible in our case. Third, Weese (2009) introduces a method of estimating the structural parameters of a political coalition formation game. Weese's estimator allows for all the above features including one-to-many matching, which is accomplished with a clever application of Wernicke's (2005) network algorithm. However, both Gordon and Knight's and Weese's analyses are more restricted than Brasington's because their method abstracts away from individual heterogeneity in the decision making of potential merger partners by looking only at coalition level data.

Besides theoretical limitations, Brasington's (1999, 2003a, 2003b) applications of the Poirier model are somewhat problematic. Perhaps the main problem in many of his models is that the correlation coefficient of the error terms of the bivariate probit equations is exactly or close to -1 or 1 indicating a corner solution to the maximization problem. According to Butler (1996), error correlation that approaches -1 or 1 violates the regularity conditions of maximum likelihood, and causes the Hessian to be noninvertible or implausible *t*-values. This problem is generally related to the poor convergence properties of the Poirier (1980) model. A further problem with Brasington's applications is that he uses post-merger district characteristics to explain mergers. If merged school districts tend to become more similar after merging, using post-merger data from

1990 to explain mergers in the 1960's as he does is bound to produce errors in analyzing the effects of heterogeneity. One of the contributions of our empirical strategy is that we can account for both of these practical problems in our analysis. Moreover, our set up allows for extending Brasington's approach to a multi-partner setting, thus relaxing the assumptions under which individual heterogeneity in merger decision making can be studied.

Due to computational reasons, a number of important factors contributing to merger decisions are left out in both Gordon and Knight's and Weese's empirical applications. Moreover, the underlying theoretical models in these studies do not allow for political economy type questions, since they are based on maximizing utility functions of coalitions or individual jurisdictions (or populations of coalitions). For these two reasons, the characteristics of all potentially merging municipalities' councils cannot be easily included in their analysis.

Since one of our main contributions is a rich set of interesting explanatory variables, we will not estimate a computationally cumbersome structural model. Nevertheless, we can account for most of the four econometric problems raised earlier. We solve the first potential problem by using coalition level data, which abstracts away from individual municipality choices, and by applying the Poirier model. We do not address the second problem in this paper. In principle, it should be possible to adjust the likelihood functions to limit the probability sum of all the potential mergers involving a given partner to be below one. Moreover, it is possible to check after estimation whether this condition is violated in a particular data set. We address the third potential problem by using data only from a single period of the coalition formation game. Assuming the decision making is simultaneous, the choice set remains stable over the period. Most importantly, we introduce a novel choice based sampling procedure that addresses the fourth potential problem.

In summary, our method deals with most of the associated econometric problems while allowing for much richer analysis than previous studies in terms of the number of covariates that can be analyzed and in terms of the heterogeneity of individual decision making. Our sampling method also easily accommodates typical methods of causal inference such as the regression discontinuity design. Existing structural tools are very problematic in terms of potential omitted variable bias since they provide no methodological tool to deal with the issue of unobserved heterogeneity and these models can incorporate only a limited number of observed variables.

4.1 Choice based sampling from a network

We introduce a novel stratified choice based sampling procedure to allow for multi-partner mergers in a spatial context. In choice based sampling, the idea is to keep all the rare events in the data as a treatment group, in our case we keep all the mergers, and then randomly sample a control group out of the common events, in our case the potential coalitions that did not merge. Choice based sampling can be made more efficient by conditioning it on the observed characteristics of the treatment group. Sampling conditioned on explanatory variables is typically referred to as exogenous stratified sampling while choice based sampling is an example of endogenous sampling. Combination of these two is often referred to as stratified choice based sampling. This sampling method has been widely applied especially in epidemiology with rare diseases but it is also common in econometrics.[5] Our sampling procedure is an extension to the existing literature on empirical spatial merger analysis and any matching or coalition formation games in general. This contribution may have its uses in many other applications because the statistical theory on choice based sampling is well developed and estimators that handle data from this type of sampling are readily available.[6]

We follow Weese (2009) and utilize the network algorithm by Wernicke (2005) but in a different way. This algorithm allows us to identify the set of all the potential mergers from which we can sample the control coalitions. The sampling from all the potential coalitions of a given size can be done using the FANMOD software by Wernicke and Rasche (2006). The FANMOD software takes into account geographic borders so that municipalities can form a coalition only if all of them share a geographic border. Naturally, some municipalities may merge even if they are not neighbors. This happens when they are a part of a coalition including multiple municipalities that jointly form a geographically coherent municipality.

Our sampling procedure is as follows. First, we select all the actual mergers into our sample. This endogenous sampling in necessary because a random sample would consist only of zeros with a very high probability (see Table 1). For each actual merger in the data of size n (n = number of merging municipalities), we sample 4 potential coalitions of equal size that did not actually take place as controls.[7] We repeat this for all n that actually took place (in our case 2, 3, 4, 5, 6 and 10). The limitation of the FANMOD software is that it can perform the calculations only up to 8 merging units. Therefore, we form the four 10 member

[5] See e.g. Breslow (1996) and Manski (1995) for a review.

[6] See e.g. Manski and Lerman (1977), Imbens (1992) and Breslow (1996).

[7] The number of controls should be decreasing in the costs of collecting the variables of interest. In the logit model, as large a number as possible is preferred, since rare events correction is possible (see King and Zeng 2001). In the Poirier model, some balance could be preferred over many observations due to the convergence problems.

potential mergers in our data by sampling from 8 partner potential mergers and including two neighbors that make the new coalition most circular. We stratify the sampling on the number of merging municipalities in this way because otherwise the control group would consist almost only of mergers with many partners thus making the estimation very inefficient. This sort of stratification is standard practice in choice based sampling literature. One way to interpret this set up is that the population of interest is the set consisting of all the potential coalitions.

When using stratified choice based sampled data such as ours, Cram et al. (2010) show that it is important to account for two potential errors one of which threatens internal validity and the other external validity of the results. First, we need to condition the analysis on strata fixed effects. In our case the strata groups are defined by the different number of merger partners. Failure to do so is akin to introducing an omitted variable bias and will therefore result in inconsistent estimates regardless of the estimator that is used and downward biased standard errors when the typical logit estimation is applied. Second, we need to weight the observations based on the sampling rates, for example by using the Manski and Lerman (1977) weighted exogenous sampling maximum likelihood estimator (WESML). These weights can be calculated using the numbers in Table 1. It shows the number of actual mergers and the number of potential mergers sampled to our data set. Table 1 also presents the total number of potential mergers that could take place calculated using FANMOD. WESML weighting is not necessary if we are not interested in the intercepts because all the slope parameters are consistent when logit estimation is used, given that the model is fully saturated with respect to stratification. However, if we are interested in the intercepts or are concerned that stratification may be correlated with some of our discrete explanatory variables, only weighting guarantees consistent estimates. WESML weighting can be applied within any maximum likelihood estimation method.[8]

[8] In STATA, this can be done either using the pweight option or the svy survey package. Imbens (1992) provides a more general semiparametric estimator based on the method of moments estimation.

Table 1. Number of coalitions in the choice based sampling procedure.

Coalition size	Actual	Control	Potential
2	17	68	787
3	8	32	2 171
4	4	16	6 551
5	1	4	20 647
6	2	8	66 534
7	0	0	Not calc.
8	0	0	698 481
9	0	0	NA
10	1	4	NA

Note: The table presents the number of mergers that actually took place, the sampled control coalitions and the total number of potential coalitions.

Consistent estimation does not make this sampling approach robust to any endogeneity issues such as the omitted variable bias. However, one of the main advantages of this sampling set up is that usual methods of causal inference such as regression discontinuity design and instrumental variables can be used without any further complications. Consider the case of regression discontinuity which we apply in this study. Obviously, if stratification is not correlated with the selection into the treatment, regression discontinuity treatment effect is unbiased. However, if stratification is correlated with the selection into the treatment the issue is not as clear. Let us consider the worst case scenario of perfect correlation with an abstract example. Assume that we stratify based on individuals being under 40-years-old or over 40-years-old and the discontinuity of interest is at 40 years. Due to a balanced stratification scheme, this would result in having exactly the same relation of ones and zeros in the outcome variable in either side of the discontinuity threshold and on average this identical relation would persist near the discontinuity. If we estimated the model without weighting, the treatment effect would always be zero. Similar scenarios can be contemplated where this sort of correlation would lead to either upward or downward biased treatment effects. Fortunately, by using the WESML weights, the population level group effects are estimated consistently. Thus, even with perfect correlation between the selection into the treatment and the stratification, weighting allows for unbiased estimation of the treatment effect in the regression discontinuity design.

One practical problem with this sampling procedure in our application to municipality data is that large coalitions may take many different geographic shapes and some of the sampled coalitions may be quite unrealistic, and thus, would not serve as very good controls.[9] For example, some coalitions of size 6

[9] Weese (2009) faced a similar problem because the largest merger that took place during his analysis period in Japan included 15 municipalities. This creates problems for Weese's estimator because the number of potential coalitions with a size of fifteen or less is huge. Weese circumvents the problem by introducing an upper bound on the number of bordering municipalities that any potential coalition can

may be string-like and result in high transportation costs. In order to circumvent this problem, we restrict the potential coalitions so that they cannot cross county borders. The numbers in Table 1 include this restriction. Figure A1 in Appendix A illustrates municipality and county borders. From the Figure it's clear that coalitions that stay within county borders (bold line) are much less likely to be problematic in this sense. This choice is also supported by the fact that no mergers took place across county borders. There are at least two reasons for this. First, counties constitute a middle-level in regional policy-making in Finland even though counties have very limited political power. For example, all regional administrative authorities by the central government should follow county division. This makes counties a natural cooperation environment for municipalities even in the absence of formal mergers and is probably the main reason, along with cultural identity and such, why merging across a county border does not happen. Second, county division is almost identical to hospital district division and every municipality has to belong to a hospital district.[10]

4.2 Estimation strategies

We use two different estimation strategies that provide answers to different questions. Both analyses can be conducted using the WESML-estimator to account for choice based sampling. The estimation proceeds as follows. First, using the binary logit model we study which types of coalitions are likely to form.[11] This approach can be seen as estimating a coalition formation game in reduced form. Let y_k denote the merger decision of coalition k, i.e. y_k equals one if the coalition forms an actual municipality and zero otherwise. The logit model can be written as

$$(1) \qquad y_k = 1\left(\mathbf{x}'_{1k}\boldsymbol{\beta}_1 + \mathbf{x}'_{2k}\boldsymbol{\beta}_2 + \mathbf{z}'_k\boldsymbol{\delta} + \varepsilon_k > 0\right),$$

where 1(.) is an indicator function that equals one if the statement in the parentheses is true and zero otherwise. The model includes three sets of covariates along with stratification group fixed effects. The vector $\mathbf{x}_1$ includes coalition characteristics, $\mathbf{x}_2$ is a vector of variables describing coalition

have. This upper bound basically restricts the geographic shape of the coalitions to be reasonably circular. In our case restricting the mergers to within county mergers only takes care of this problem.

[10] We also extract the counties of Kainuu and Lapland from our analysis. Kainuu County is experimenting with a county level council, and thus, the municipalities will not merge while the experiment is ongoing. Lapland is an outlier in the data, with large land area and low population municipalities. Therefore, drawing control coalitions from Lapland would not be a good idea. Moreover, there were no mergers in Lapland during our analysis period. The map in Appendix A highlights these counties.

[11] We use the clogit command in STATA as Cram et al. (2010) suggest, instead of the logit command with group dummies. In our analysis, there is not much qualitative difference between using logit or clogit or whether the WESML weights are used or not.

heterogeneity and **z** a vector of municipal council characteristics at the coalition level.

In addition to coalition formation, we are also interested in how individual municipalities' own and potential merger partners' characteristics and council characteristics are associated with mergers. In answering these questions, we follow the approach by Brasington (2003a) and use the Poirier model. Although this model is not perfect in our setup, it accounts for the most important feature of our data, which is two-sided decision making where each potential merger partner has a veto power on the merger. Brasington's (2003a and 2003b) insight in using the Poirier model was to sort the data in a way that gives meaningful interpretations of the coefficients in the bivariate probit equations. We follow Brasington (2003a) and sort the data according to municipal population. This allows us to focus on the political power structure which changes differently for smaller and larger merging partners due to a merger.

Let y_{1i} and y_{2j} denote the merger decisions of larger municipality i (type 1) and its smaller potential merger partner j (type 2). Poirier's (1980) bivariate probit model can be written as

$$(2) \qquad y_{1i} = 1\left(\mathbf{x}'_{1i}\boldsymbol{\beta}_{11} + \mathbf{x}'_{2i}\boldsymbol{\beta}_{12} + \mathbf{z}'_{i}\boldsymbol{\delta}_{1} + u_{1i} > 0\right),$$

$$(3) \qquad y_{2j} = 1\left(\mathbf{x}'_{1j}\boldsymbol{\beta}_{21} + \mathbf{x}'_{2j}\boldsymbol{\beta}_{22} + \mathbf{z}'_{j}\boldsymbol{\delta}_{2} + u_{2j} > 0\right),$$

where 1(.) is again an indicator function and the vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and **z** correspond to municipal level variables introduced in equation (1). The error terms (u_1 and u_2) follow a standardized bivariate normal distribution with correlation ρ. Thus, the model allows for same unobservables affecting the merger decisions of both parties. Poirier's model differs from the standard bivariate probit in that instead of observing both y_1 and y_2 we only observe their product ($y_1 * y_2$).

Table 2 clarifies the data setup in the Poirier model. Panel A illustrates potential mergers with only two partners. The first column in Table 2 is simply the identification number. The first municipality in each potential match pair is the larger one corresponding to equation (2). This sorting of the data allows one to distinguish the effects of explanatory variables separately for "large" and "small" municipalities. One should note, though, that a given municipality may be either "large" or "small" depending on the potential match pair, as is the case for municipality A in Table 2.

Table 2. Example of data setup in the Poirier model.

Panel A:								
Potential match pair	**Mun. 1**	**Mun. 2**	**Merged**	**Population 1**	**Population 2**	**Difference in population 1**	**Difference in population 2**	**Gov. Subsidy**
1	A	B	0	20 000	9 000	11 000	-11 000	400
2	D	A	1	67 000	20 000	47 000	-47 000	200
3	C	B	0	21 000	9 000	12 000	-12 000	1500
4	G	H	0	14 500	5 000	9 500	-9 500	320

Panel B:								
Potential match pair	**Mun. 1**	**Mun. 2**	**Merged**	**Population 1**	**Population 2**	**Difference in population 1**	**Difference in population 2**	**Gov. Subsidy**
..	..	..	..	..	..			..
101	I	J,K,L	1	60 000	42 000	18 000	-18 000	900
102	I,K,L	J	1	81 000	21 000	60 000	-60 000	900
103	I,J,L	K	1	90 000	12 000	78 000	-78 000	900
104	I,J,K	L	1	93 000	9 000	84 000	-84 000	900
..	..	..	..	..	..			..

A particular challenge in using the Poirier model comes from mergers that included more than two municipalities. In these cases, we model the situation so that each municipality chooses individually whether to join the rest of the coalition. Panel B in Table 3 illustrates. Municipalities I, J, K and L have merged. In the Poirier model, we treat this observation so that in each case, one of the municipalities considers merging with the three others that form a coalition. All the potential coalitions from our FANMOD sampling procedure are treated the same way. The down-side of this procedure is that now the coalition formed by the three other municipalities is treated as if it really was a decision making entity. Clearly this is not the case in reality. Unfortunately, it is not possible to exclude these artificial entities from the data if we want to include multi-partner mergers in the Poirier estimations. We conduct sensitivity analysis to check whether these artificially generated additional data points drive the results by using according weights. For the Poirier model, we report the unweighted parameter estimates and conduct sensitivity analysis with respect to using the WESML weights. The reason for this is that we ran a number of other sensitivity tests concerning the Poirier results that were easier to implement and faster to run using the unweighted estimator. The sensitivity tests are reported in more detail after the main results.

4.3 The subsidy scheme and regression discontinuity design

In order to encourage mergers, the central government grants merger subsidies to merging municipalities. The subsidy scheme consists of two parts. The basic part is determined according to the population of the new coalition and the combined populations of all the municipalities involved in the merger with the exception of the largest municipality. An additional supplement amount is paid if the number of municipalities diminishes at least by two and increases as the number of disappearing municipalities rise. This subsidy scheme is in place from 2008 to 2013. The subsidy to a given merger is increased by 80 percent if the merger took place in 2008 or 2009 and by 40 percent if the merger takes place in 2010 or 2011.

Table 3 illustrates the subsidy scheme in 2008–2009. It is important to note that, in addition to the subsidy, the government guarantees that central government grants are not reduced for five years after the merger even if the new municipality is entitled to a smaller overall grant than the merging municipalities individually. The subsidy scheme is clearly designed to promote economies of scale from the mergers as the amount paid is higher if the new municipality has over 20,000 inhabitants. In addition, the subsidy scheme rewards savings in bureaucracy because the amount paid increases in the number of municipalities involved in the merger. Each additional municipality in a merger increases the subsidy by 1.26 million Euros. In 2008 and 2009, the total amount of merger subsidies paid was about 220 million Euros. It is important to emphasize that the discontinuities in the subsidy scheme offer us a way to distinguish the effects of the subsidy from the role of economies of scale and from any unobserved heterogeneity that would potentially be correlated both with the size of the subsidy (whether absolute or per capita or any other function of subsidy and some municipal characteristics) and the propensity to merge.

Table 3. The central government subsidy scheme in 2008–2009.

Population in thousands (first number is the new merged municipality, second number is coalition population minus largest):	Number of disappearing municipalities:									
	1	2	3	4	5	6	7	8	9	10
over 20, over 10	7.20	8.46	9.72	10.98	12.24	13.50	14.76	16.02	17.28	18.54
over 20, 5-10	6.48	7.74	9.00	10.26	11.52	12.78	14.04	15.30	16.56	17.82
over 20, less than 5	5.76	7.02	8.28	9.54	10.80	12.06	13.32	14.58	15.84	17.10
under 20, over 7	5.40	6.66	7.92	9.18	10.44	11.70	12.96	14.22	15.48	16.74
under 20, 3.5-7	4.50	5.76	7.02	8.28	9.54	10.80	12.06	13.32	14.58	15.84
under 20, under 3.5	3.60	4.86	6.12	7.38	8.64	9.90	11.16	12.42	13.68	14.94

Notes: Figures in the table are millions of Euros. The subsidy continues to increase by the same amount if there are more than 10 disappearing municipalities.

The subsidy scheme has discontinuities in three dimensions. First, two newly formed coalitions can be identical in all respects but the one that involved more disappearing municipalities receives a larger subsidy. However, it is not possible to utilize this dimension of discontinuity because this forcing variable is discrete and follows exactly the according treatment indicator. Second, a coalition with just over 20,000 inhabitants gets more subsidy than an otherwise identical coalition with just under 20,000 inhabitants. To simplify the analysis, we do not use this dimension either. Using it would give us only a few more identifying observations but would require us to control for forcing variables in two dimensions. Besides complicating the analysis, it would make it more difficult to study the optimal population size or economies of scale effects due to having many different population variables in the models. Instead, we use discontinuities in the third dimension of the subsidy scheme; the combined population of all the municipalities in the coalition with the exception of the largest one. Thus, the forcing variable is defined as "total coalition population minus the population of the largest municipality in the coalition".

Our regression discontinuity design (RDD) is illustrated in Figure 2. We construct one dummy variable that gets a value of one if the potential coalition is "just above" the threshold of getting a larger subsidy and a value zero otherwise (treatment 1 in the subsequent analysis). We construct a second dummy variable that equals one if the potential merger is "just below" the threshold (reference group in the subsequent analysis). We define "just above" and "just below" as closer than 1,000 inhabitants. Robustness of our results with respect to this definition of "just" is discussed in Appendix B along with all the other robustness and validity checks.[12] To make sure that the discontinuities in the other population dimension do not interfere, we assign a value zero to both these dummies for all the observations for which the total population of the coalition is between 19,000 and 21,000. We construct a third dummy variable that equals one when both the "just above" and "just below" variables are zero (treatment 2 in the subsequent analysis). We use the treatment 1 and treatment 2 variables in the regressions while the "just below" variable acts as the reference group. This allows us to interpret the parameter estimate of the treatment 1 variable as the effect of crossing the discontinuity, and thus, getting more subsidy.[13]

[12] Lee and Lemieux (2010) and Imbens and Lemieux (2008) provide surveys on RDD literature and also provide a guide-to-practice on how to conduct research based on the design. However, they concentrate on the standard case of one discontinuity point. Our analysis has more in kind with, for example, van der Klaauw (2002) or Egger and Koethenbuerger (2010).

[13] Another estimation possibility would be to use only one treatment threshold and define the forcing variable as the distance to the threshold. This approach is discussed in Appendix B.

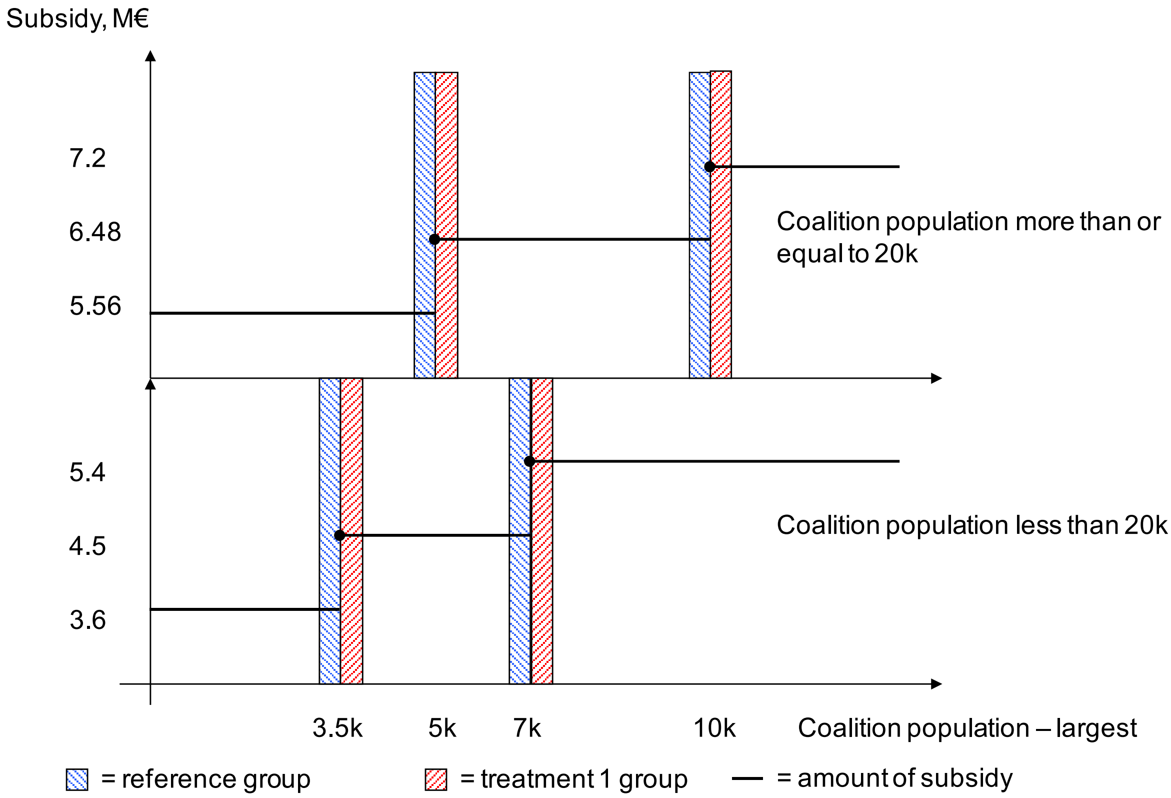

Figure 2. Illustration of the regression discontinuity design.

The decision makers clearly have imperfect control over what side of the discontinuity point they are, because the municipal councilors cannot directly control even their own municipality's population, much less the population composition of the entire coalition. If this assumption holds, the mergers are as good as randomly assigned to the treatment and control groups and the treatment effect has a plausible causal interpretation. We conduct the relevant validity checks for this assumption in Appendix B.

Due to the limited number of observations and the fact that we have multiple discontinuity steps, we conduct the RDD analysis only using parametric regressions. However, we are able to conduct sensitivity analysis with respect to using separate parametric regressions between the different discontinuity thresholds as opposed to using only the pooled regressions that we report in the results. The RDD is incorporated to both logit and Poirier models. Thus, we are able to analyze the causal effect of the government merger subsidy scheme on the merger decisions both on the coalition level and individual municipality level.

4.4 The data and descriptive statistics

We have linked data on municipal mergers from 2008–2009 to municipal characteristics obtained from Statistics Finland and to characteristics of municipal council members obtained from the Ministry of Justice and the Local Government Pensions Institution. The councils were elected in 2004 for a four year term and they made the decision regarding mergers in 2008–2009. Future mergers are decided by councils elected in 2008 and after. The 2009 municipal division was used already in the 2008 elections. By restricting the analysis to a single council and a single subsidy scheme, we can think of this set up as a one period coalition formation game.

The starting point for our empirical specification is that municipalities probably seek economies of scale through merging but face a trade-off in terms of matching service production to more heterogeneous preferences of a larger municipality. On the other hand, large municipalities may even face diseconomies of scale. In the logit model, we include coalition population and its squared term to capture these effects. In the Poirier model, we follow Gordon and Knight (2009) and use a fairly general average cost function. From the perspective of municipality *i* the efficiency gain or loss from merging with municipality *j* can be expressed as

$$\ln\left[\frac{c(N_i)}{c(N_i+N_j)}\right]. \tag{4}$$

where c(.) is the average cost function for local public goods and services and N is population size. If equation (4) is positive there are economies of scale from the merger and *vice versa*. Specifying the function as $c(N)=N^{\beta+\gamma N}$ yields the following measure for efficiency gains

$$\ln\left[\frac{c(N_i)}{c(N_i+N_j)}\right]=\beta\underbrace{\left[\ln(N_i)-\ln(N_i+N_j)\right]}_{\text{economies of scale}} +\gamma\underbrace{\left[N_i\ln(N_i)-(N_i+N_j)\ln(N_i+N_j)\right]}_{\text{diseconomies of scale}}. \tag{5}$$

If $\beta < 0$ and $\gamma > 0$, the cost curve is U-shaped. If both are negative the cost curve is decreasing in N and there are no diseconomies of scale. Thus, β can be interpreted as an estimate of the role of economies of scale and γ as an estimate for diseconomies of scale if there are any.[14] Using this functional form also solves the potential problem of multicollinearity of population size and the RDD forcing variable. If the number of municipalities in the coalition is two, population of the smaller municipality would coincide with the forcing variable exactly.

In addition to economies of scale measures, we include other municipal characteristics, measures of heterogeneity of municipalities in a coalition and characteristics of municipal councils. Variable definitions can be found in Table A2 in Appendix A. Some descriptive statistics of our coalition level data are presented in Table 4. The actual mergers and the randomly sampled control coalitions are surprisingly similar. Even their council characteristics are quite similar. However, the actual mergers are smaller in terms of population. Actual mergers are also clearly more likely to have co-operated already before the merger. Existing cooperation here means that the municipalities in the coalition had a joint authority in producing basic health care services, which is the largest expenditure item for municipalities. The mean distance of population to the largest municipality within a coalition is larger in the control coalitions.[15]

[14] See Gordon and Knight (2009) for details.

[15] The mean distance measure is calculated using Geographic Information System (GIS) techniques in the following way. We have data on population location within 250 m * 250 m grids for the whole of Finland, which we used to calculate a population weighted Euclidean distance from all the grids within a coalition to the centre of the largest municipality in the coalition. The average distance measure for a coalition is simply the average of these distances, which measures the average distance to coalition centre of persons living in a given coalition. The intuition is that if a merger goes through it is likely that some, if not all, municipal services will be concentrated into the largest municipality, and thus, the distance to services will grow for people living in smaller municipalities.

Table 4. Descriptive statistics for coalitions.

	Merger = 1		Merger = 0	
	Mean	**Std. dev.**	**Mean**	**Std. dev.**
Number of observations	33		132	
population (10,000)	3.44	3.30	5.13	9.04
mean of taxable income (€10,000)	1.23	0.17	1.16	0.18
central government grants (€10,000 per capita)	0.13	0.05	0.15	0.05
dependency ratio	0.36	0.03	0.36	0.03
cooperation	0.52	0.51	0.16	0.37
language	0.06	0.24	0.10	0.30
merger subsidy (€1,000 per capita)	0.33	0.22	0.37	0.28
mean distance to largest municipality (km)	7.94	4.46	13.3	7.85
total land area	0.11	0.06	0.15	0.19
population density	0.80	0.26	0.73	0.29
deficit	0.21	0.48	0.09	0.29
std.dev. population	1.63	2.34	2.16	4.46
std.dev. taxable income	0.17	0.12	0.14	0.09
std.dev. tax rate	0.01	0.004	0.004	0.003
sdt.dev. population density	0.37	0.34	0.32	0.28
std.dev. depency ratio	0.02	0.02	0.03	0.02
reduction in council size	0.45	0.14	0.45	0.14
Herfindahl index	0.31	0.12	0.31	0.09
same largest party	0.73	0.45	0.71	0.45
left-w ing party share	0.31	0.12	0.28	0.09
centre party share	0.34	0.19	0.41	0.17
share of municipal employees in council	0.08	0.03	0.08	0.03
share of municipal employees in population	0.06	0.01	0.06	0.01

Table 5 presents descriptive statistics for municipal level data sorted according to population size as explained earlier. Number of observations is 420 because of the way multi-partner mergers are treated as described in Table 3. Differences are measured in absolute terms. Obviously, the larger merging partners are larger in terms of population. But other than that the smaller and larger merger partners are quite similar in terms of their characteristics. Note that many of the variables are measured at coalition level, and thus, are the same for both municipality types. The smaller municipalities are slightly more concentrated politically as indicated by the Herfindahl index but the difference is small. They also have a higher Centre Party council share, which is to be expected because the support for the Centre Party is highest among rural areas.

Table 5. Descriptive statistics for municipalities sorted according to size.

	Smaller		Larger	
	Mean	**Std. dev.**	**Mean**	**Std. dev.**
Number of observations	420		420	
population (10,000)	0.75	1.37	5.12	6.70
mean of taxable income (€10,000)	1.09	0.18	1.16	0.17
central government grants (€10,000 per capita)	0.16	0.06	0.14	0.05
dependency ratio	0.38	0.02	0.36	0.03
cooperation	0.16	0.37	0.16	0.37
language	0.15	0.36	0.15	0.36
merger subsidy (€1,000 per capita)	0.34	0.24	0.34	0.24
distance to largest municipality (100 km)	0.24	0.20	0.24	0.20
land area	0.05	0.10	0.15	0.20
population density	0.56	0.41	0.80	0.25
deficit	0.22	0.88	0.41	0.73
difference in population	4.37	6.15	4.37	6.15
difference in taxable income	0.14	0.11	0.14	0.11
difference in tax rates	0.005	0.004	0.005	0.004
difference in population density	0.33	0.33	0.33	0.33
difference in depency ratio	0.03	0.02	0.03	0.02
reduction in council size	0.55	0.15	0.55	0.15
Herfindahl index	0.39	0.15	0.31	0.10
same largest party	0.71	0.45	0.71	0.45
left-wing party share	0.25	0.13	0.29	0.10
centre party share	0.46	0.20	0.37	0.19
share of municipal employees in council	0.04	0.04	0.05	0.03
share of municipal employees in population	0.06	0.02	0.06	0.01

5 Econometric results

This section presents our econometric results. First we discuss the results of the coalition level logit analysis. This is followed by the results from the municipality level Poirier analysis. At the end of the section, we discuss numerous sensitivity checks of the results provided by these models.

5.1 The logit model

The results from the coalition level logit models are presented in Table 6. Table 6 includes three models. In the first model, we include variables that describe the characteristics of potential coalitions in terms of population, tax base, financial conditions and cost structure. In the second model, we add variables that measure the heterogeneity of the municipalities involved in potential coalitions. In the third model, we add council characteristic in order to capture the effects of local politics[16]. Although we have a rich set of covariates, there is always the possibility that we have overlooked or do not observe some important variables that drive merging. Since it is unlikely that all the unobservable factors driving the mergers are independent of the observable ones, the results should interpreted cautiously as associations rather than causal effects. One exception is the subsidy where we use regression discontinuity design. The reduction in council size measure could also allow a tentative causal inference because this variable is an exogenous step function based on municipal population and we control for population directly.[17]

The results indicate that existing cooperation through a joint authority in producing basic health care services has a strong positive association with the likelihood of merging. This is an expected, but also a very interesting result because it indicates that merging takes place between municipalities who probably already benefit from economies of scale in basic health care. This casts some doubts on whether these mergers lead to increased cost-effectiveness in service production. Another expected result is that population's average distance to the centre of the largest municipality clearly prohibits merging. Population of

[16] In Table A3 in Appendix A, we present results for running these regressions in reverse order; first only on council characteristics and then adding other covariates. Council characteristics are jointly significant in a model with no other covariates, but not individually.

[17] In principle, also the council size step function would allow us to use RDD. In fact, Egger and Koethenbuerger (2010) use exactly this type of discontinuity. Unfortunately, the allocation to the treatment group is more difficult in our coalition formation setting. In principle, we could allocate coalitions into treatment groups if the individual municipalities were all just below or all just above any discontinuity and the coalition under consideration was just at the opposite side of some discontinuity than all the individual municipalities. Unfortunately, in our data there would not be many observations that could be allocated to conservatively chosen treatment groups. Therefore, we must be satisfied that the exogeneity of the council size function as such grants some reliability on the results concerning the reduction in council size.

the coalition, however, does not seem to matter, whereas heterogeneity in population size increases the likelihood of merging. Interestingly, the more variation there is in the tax rates of the municipalities in a potential coalition the more likely a merger is. This would indicate that tax competition is not particularly fierce among merging municipalities. Heterogeneity in dependency ratio is negatively associated with merging. This variable may capture age related preference heterogeneity concerning public services provided by the municipalities.

Table 6. Results from coalition level logit models.

	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
population	0.81	0.61	0.31	1.07	-0.47	1.60
(population)^2	-0.06**	0.02	-0.06**	0.02	-0.05	0.04
mean of taxable income	3.73	4.31	5.62*	8.54	-9.48	11.0
central government grants	55.5	52.2	60.6	70.4	98.3	50.4
dependency ratio	-12.2	46.1	-30.5	89.0	-143.9	116.0
cooperation	4.08*	2.00	4.56*	2.07	5.08*	2.29
language	-1.76	1.64	-2.40	2.20	-1.35	1.84
RDD treatment 1	5.14**	2.01	4.57	3.08	7.37*	3.11
RDD treatment 2	1.95*	0.88	2.04*	0.91	3.69**	0.69
mean distance	-0.56**	0.25	-0.60	0.38	-0.97*	0.42
total land area	-6.85	7.24	-8.68	15.7	9.9	6.41
population density	1.26	2.75	-0.61	4.39	8.15	5.83
deficit	2.14**	0.39	2.53**	0.84	0.59	0.36
std.dev. population			0.87	0.57	1.34*	0.65
std.dev. taxable income			-1.73	8.86	4.84	7.06
std.dev. tax rate			-37.6	485.9	330.1*	89.9
sdt.dev. population density			2.94*	1.18	6.96*	2.79
std.dev. depency ratio			-54.4	50.1	-52.0**	5.80
reduction in council size					45.5*	18.8
Herfindahl index					62.2**	23.4
same largest party					-0.05	0.70
left-w ing party share					1.40	3.48
centre party share					-33.1*	15.0
share of municipal employees in council					61.1**	17.1
share of municipal employees in population					-188.9*	80.1
forcing variable	-0.15	1.97	0.90	2.49	0.46	1.91
(forcing variable)^2	0.13	0.19	0.09	0.19	0.07	0.13
Number of obs.	165		165		165	
Pseudo Log-L	-83 061		-78 224		-34 256	
Pseudo R^2	0.95		0.95		0.98	

*Notes: The results are from conditional logit models where stratification dummies are included in the estimation. ** and * indicate statistical significance at 1and 5 percent level, respectively. Forcing variable = (coalition population - population of the largest municipality in the coalition).*

The main variables of interest in this study are variables related to local politics and other characteristics of the municipal councils. The importance of local politics is captured with variables that measure political fragmentation, similarity of political power in different municipalities and party composition of the councils. Moreover, we aim to capture the importance of politicians' private incentives in merger decisions with two variables that aim to measure the importance of career concerns in both political and non-political careers.

Being a member of the municipal council in Finland is a part-time job that incurs practically no salary even in the largest municipalities. However, some council members are municipal employees (as teachers, nurses, doctors etc.), which may affect their attitude toward mergers, since cost savings in public service production are often put forth as the primary reason for mergers. The central government anticipated some resistance from municipal employees and enacted a transition period spanning five years after a merger during which municipal employees cannot be laid off due to merger-related efficiency reasons.

In order to study whether councilors' municipal employment somehow affected the mergers, we collected data on council members' age and whether they are municipal employees. Age is important here because a council member nearing retirement could, in effect, be sure that he or she is not laid off before retirement if a merger goes through. Using this information we constructed the variable "share of municipal employees in council", which is the share of council members who are employed by their own municipality and are under the age of 58 at the time of the vote (retirement age is 63). These council members have the highest incentives to block a merger if they feel it threatens their employment in the municipal sector. This variable may, of course, also capture voter sentiments because a high share of municipal employees in the population could be correlated with the council share. In order to identify the effect of the potential merger on council member's own career separately from its effect on their voters' careers, we also include the share of municipal employees in the population to the model.

Interestingly, both of these variables are statistically significant but enter with opposite signs. The share of municipal employees in the council is positively associated with merging at the coalition level, whereas the population share gets a negative sign. The rationale for the first result may be that the council members want to guarantee their and possibly their voters' municipal jobs for the next five years or that they anticipate better career opportunities in the new and larger municipality. We are unable to differentiate between competing explanations, but what is important here is that the council share is significant even when population share controlled for. This indicates that municipal mergers are affected in a significant way by council members' self-interests as municipal employees.[18]

Political career concerns are included in the model using the "reduction in council size" variable, which measures the relative amount of council seats lost if the merger goes through. The larger this variable is the more seats are lost in relative terms. This should capture changes in the re-election probability if a merger goes through. Interestingly, the more council seats are lost in the merger the more likely it is. This result is puzzling because the direction of this effect is opposite to what would maximize re-election probabilities. However, it again suggests that private interests of council members play a part in the merger decisions. To explore this further, we need to look at the potential heterogeneity of this effect using the Poirier model.

The coefficient for the Herfindahl index, which measures political fragmentation in councils, is positive and highly significant. This indicates that the more concentrated the political power is at coalition level the more likely the coalition is to form a municipality. Furthermore, the share of the Centre Party is negatively associated with merging while the share of left-wing parties (the Social Democrat Party of Finland and the Left Alliance) is not statistically significant (the reference group consists of the National Coalition Party (conservatives), the Green League and some smaller parties). Somewhat surprisingly though, whether individual members of coalitions all have the same largest party is not important for coalition formation. This could be due to opposite effects that cancel each other out so that municipalities with same elected parties have similar preferences but at the same time council members may face stiffer political competition from the members of their own party than from members of other parties in the new coalition. There is indeed some evidence of this according to the Poirier results reported below.

The RDD results indicate that the merger subsidy has a positive causal effect on merging.[19] This means that subsidies can be used to influence merger decisions, and thus, the subsidy program had the intended effect. Because of our choice based sampling scheme we do not know how the subsidy affected the overall number of municipalities involved in mergers. These aspects of the subsidy scheme are left for further research.

---

[18] The result could also mean that municipal employees are better informed about the costs and benefits of merging and the situation their municipality is in.

[19] This result is fairly robust to different model specifications and validity checks. See Appendix B for details.

5.2 The Poirier model

The results from the Poirier's partial observability bivariate probit are presented in Table 7. For comparison, we have included in the Table 7 also the corresponding coalition level logit analysis. These logit results do not contradict any of the Poirier results and adds confidence to both of the analyses. Because the units of observation are different in the logit and Poirier analyses, some of the variables are not exactly the same, but we have used the closest logical counterparts.

The model specifications in coalition level analysis were based on selecting the most interesting variables from a theory point of view. The Poirier model specification, however, is based on selecting the largest set of interesting variables that survive certain econometric scrutiny related to poor converge properties of the maximum likelihood estimator. These requirements are explained more carefully when we discuss sensitivity analysis. Although this sort of model selection has a flavor of data mining, we believe that it could be rather a signal that this particular specification is a fairly good statistical approximation for this particular data generating process. Despite the fact that the econometric properties of the Poirier model are far from ideal, this model provides such interesting results concerning the possible heterogeneity of our variables of interest, that they deserve detailed discussion here.

In Table 7, the municipalities are sorted according to population so that the larger municipality (or hypothetical municipality) is always in equation (2) and the smaller municipality is in equation (3). Therefore, we can analyze whether the associations are heterogeneous with respect to size. The first result of interest is that the average cost curves are different for small and large municipalities. For small municipalities, β and γ are both negative indicating that they are on a downward sloping segment of the cost curve. However, for large municipalities $\beta < 0$ and $\gamma > 0$, which means that large municipalities are experiencing diseconomies of scale. Again as Gordon and Knight (2009) point out, these estimates of economies and diseconomies of scale should be interpreted as those perceived by the voters or councils when they decide on whether to merge or not. These may differ from actual economies of scale realized through merging.

Table 7. Results from Poirier's bivariate probit model.

	Smaller		Larger		Coalition logit	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
constant	-0.98	7.00	-15.7*	6.47	-33.9*	15.1
economies of scale	-3.73**	1.04	7.28**	2.23	2.56**	0.85
diseconomies of scale	-0.79**	0.25	-0.14	0.13	-0.11**	0.03
mean of taxable income	-2.34	2.54	5.89*	2.53	3.08	3.69
dependency ratio	-31.4*	13.8	38.1*	17.6	26.4	29.1
cooperation	2.52**	0.57	4.15**	0.90	2.65**	0.89
RDD treatment 1	0.31	1.67	1.93*	0.88	3.23*	1.37
RDD treatment 2	-0.71	1.50	2.15*	0.87	1.61	1.02
number of municipalilties	-0.99**	0.31	0.49	0.33	0.22	0.47
distance	-19.8**	4.82	-0.33	1.89	-15.5**	4.75
total land area	-29.0**	7.19	12.9*	5.86	1.34	2.77
deficit	-1.23**	0.40	0.19	0.27	0.57	1.02
difference in population	-1.22*	0.60	-0.31**	0.09	-0.42	0.50
difference in taxable income	-9.58**	2.79	4.75	2.60	3.56	4.30
difference in tax rates	80.8	54.2	61.9	43.1	328**	112
reduction in council size	21.0**	5.29	-20.9*	6.28	10.0*	4.88
Herfindahl index	14.6**	3.82	13.1*	4.07	23.2**	7.27
same largest party	-2.07**	0.70	-0.73	0.50	0.13	0.79
left-wing party share	8.09**	2.72	-3.09	3.00	7.46	4.25
share of municipal employees in council	17.1**	5.25	-10.6	6.36	16.2	12.3
share of municipal employees in population	33.5*	13.9	-15.3	19.5	-2.3	26.0
forcing variable	1.11	0.63	-0.11	1.11	-4.00**	1.55
(forcing variable)^2	-0.36**	0.11	0.31	0.23	0.41*	0.18
Rho (95% Conf. Interval)	-0.9999993 (-1,1)				NA	
Number of obs.	420				165	
Log-L	-52.7				-39.3	

*Notes: ** and * indicate statistical significance at 1and 5 percent level, respectively. Forcing variable = (coalition population - population of the largest municipality in the coalition). Some variables in the logit are not exactly the same as those in Poirier due to coalition level analysis. The heterogeneity variables are standard deviations instead of differences, actual population is used instead of scale variables and maximum distance is used instead of individual distance.*

Again we find that the existing cooperation is positively and significantly associated with merging for both municipality types. Distance, on the other hand, gets a negative sign but is significant only for smaller municipalities.[20] This is not surprising because larger municipalities in a given merger can be quite sure that services will not be re-located to the other municipality. For larger municipalities, the effect of income is positive, perhaps indicating their ability to bear the costs of merging with the smaller and often poorer neighbors. Larger municipalities' merger decisions are also associated positively and smaller municipalities' negatively with dependency ratios. Municipality land area has

[20] The distance measure here is simply the unweighted Euclidean distance to the largest municipality from the municipality in question. If the municipality in question is the largest municipality, distance is set equal to zero.

intuitive heterogeneous effects. Large municipalities prefer larger land mass while small municipalities prefer small land mass. Both large and small municipalities seem to want partners that are of the same size with them. This is opposite to Brasington's (2003a) findings and Ellingsen's (1998) theoretical predictions. Small partners also dislike income differences.

All the political variables seem to be important determinants of the merger decisions, but they seem to be more important for smaller merging partners. According to the political Herfindahl index it is easier to make merger decisions in a politically concentrated council. If the Herfindahl index captures the effects of population preferences instead of the effects of the council composition, we would expect the sign to be the opposite.[21] Moreover, this is a very robust result. Therefore, this result can be interpreted as capturing political fragmentation effects instead of preferences with a fair amount of confidence.

Smaller councils dislike merging with municipalities where the largest party is the same as in their own council. This could imply that the members of the same party in the other council are perceived as a threat in future elections. In smaller municipalities, a large left-wing share seems to favor merging, but for larger municipalities this effect is not significant. The share of municipal employees in the council is positively associated with merging in small municipalities. This result comes on top of the positive association between the share of municipal workers in population and the propensity to merge. For larger municipalities these variables are not significant. A number of things could explain these results, but again the important thing is that council members' self-interests as municipal employees matter.

Perhaps the most interesting and at the same time puzzling result concerning local politics is the heterogeneous effect of the reduction in council size. The results indicate, that large municipalities' councils prefer mergers where a relatively small number of seats is lost (a negative sign on the "reduction in council size" variable), whereas councils from small municipalities prefer mergers where this number is relatively large (a positive sign). This result is exactly the opposite of what would be intuitive if strategic re-election concerns were important. One would expect that council members in large municipalities who are more likely to be re-elected in the new council would prefer a merger where relatively large number of seats is lost. These council members would not face a huge increase in competition due to a merger but would be council members of a larger municipality with more political power. The opposite would be true for councilors from smaller municipalities so that they would want to

[21] The costs of merging would be higher on average for homogenous (high index value) than for heterogeneous (low index value) populations, because they would more likely lose more ability to tailor public services to their preferences. Thus, we would expect the sign to be negative if the Herfindahl index captured voter preferences.

maximize their thin chances of re-election by preferring mergers where only a small number of seats are lost.

This result could possibly be explained by different selection processes of councilors in municipalities of different sizes. In large municipalities, there is much more competition for council seats than in small municipalities. We indeed observe that the share of elected councilors to all candidates in the election is negatively correlated with the municipal population. In large municipalities, the candidates who really want to be councilors are elected. Therefore they should also be concerned about re-election as our result shows. However, in small municipalities being a councilor could be more akin to a public service. Social or peer pressure may force some inhabitants to become councilors. Municipal mergers could provide a convenient way for them to get rid of these responsibilities.

The following can be said to sum up the results concerning local politics. Local politics and councilors' self-interest clearly matter, although some of the heterogeneous results are puzzling. Even though we cannot explain all of these results, it is clear that the political environment and the decision makers' private incentives may hinder the formation of optimal coalitions. This naturally raises the question of which government level should make the decisions concerning mergers. One option is that the central government should force coalitions even without the approval of the municipality councils. However, there are some caveats to this. First, information requirements for the central government to form optimal coalitions are large. Second, it may not be politically feasible. Third, it is not clear that the central government decision makers do not have private incentives at stake.

Similar to the coalition level analysis, these results again indicate that the government subsidy has a positive causal effect on merging for both large and small municipalities, although the effect is statistically significant only for large merger partners. Therefore, with a clever design of subsidies or other monetary incentives, the central government could in principle offset the potentially adverse effects that local politics or self-interest have on coalition formation. Although technically challenging, this might be politically more feasible than, for example, forcing merging against individual municipalities' wishes. Based on our analysis, we cannot say anything about the welfare effects of this scheme, i.e. whether it is welfare enhancing to collect distorting taxes and pay merger subsidies.

5.3 Sensitivity analysis for the Poirier model

Our Poirier analysis is not very reliable for a number of reasons. First, we have a corner solution in terms of the correlation coefficient ρ. Second, we have not used the WESML weights in the reported results, and third, we have not included the strata dummies. Moreover, the Poirier estimator, in general, has very poor convergence properties. Most of the model specifications that we attempted did not converge at all or in the best cases resulted in similar corner solutions as the in the reported model. The poor convergence properties also limit the extent to which we can conduct robustness analysis. However, we are able to conduct several different sensitivity analyses for the Poirier model to assess the reliability of these reported results.

We use the WESML weights also for the Poirier to take into account choice based sampling when conducting sensitivity analysis. We also weight the analysis based on multi-partner mergers. These latter weights are inversely proportional to the number of municipalities in a coalition. Unfortunately, the standard errors could not be calculated when these weights were used. Standard errors cannot be calculated using a bootstrap either because in some of the bootstrap samples the coefficients explode. The best we can do is to compare the coefficients from these weighted estimations to the reported ones using the standard errors from the reported estimation. When we impose only the stratification dummies, but not the WESML weights, the model allows for standard errors. This resulted in some changes in the Poirier results. The Poirier analysis is subjected to many other robustness checks as well and we discuss which of the reported results are robust to all of the checks at the end of this section.

In estimating the Poirier model, we use STATA's biprobit command and the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm to conduct the numerical optimization. BFGS seemed to work the best out of the standard options in STATA. We conduct sensitivity analysis also with respect to the algorithm used. We are also concerned that local convergence could drive the result and check whether the results change when using a global maximization routine instead of a local routine like BFGS. The global optimization was implemented with the Rgenoud package in R that uses an evolutionary optimization algorithm. Again, the resulting Hessian did not allow for calculation of the standard errors.

Our main concern with the Poirier model is that, similar to Brasington (1999, 2003a and 2003b), we find a corner solution in a sense that the error term correlation is close to minus unity as reported in Table 7. Brasington disregards this serious problem in all his studies. We conduct sensitivity analysis with respect to the correlation coefficient (ρ) by imposing a grid on the correlation and conduct the estimation within this grid. In other words, we fix the value of ρ to a given level and then estimate the model and repeat the estimation for different values of ρ. This analysis reveals whether the corner solution drives the results. It also shows the relative importance of unobserved factors. If the results are robust to different values of ρ, it is likely that unobserved heterogeneity is not a major issue. We suggest using this type of sensitivity analysis even in models where the parameter ρ gains an interior value, because our bootstrap analysis revealed that even in such models where the original sample ρ is in the interior (we found some such models with our data), in less than 1 percent of the bootstrapped samples it remained in the interior. This again indicates that the Poirier model is highly unstable and requires careful sensitivity analysis. Overall, this sensitivity analysis casts doubts on the reliability of the Poirier model in general, and we have shown how to assess some of the potential pitfalls when using this model.

The reported Poirier results in Table 7 are the ones with the largest set of explanatory variables that we found while being fairly robust to the sensitivity analysis with respect to the error term correlation. The results that survive all the above robustness checks are for the smaller municipalities the left-wing share, the Herfindahl index and distance. For larger municipalities, robust results are mean of taxable income, the Herfindahl index and reduction in council size.

6 Conclusions

This paper analyzed empirically the coalition formation of local governments. We introduced a novel econometric strategy involving choice based sampling from a network to handle multi-partner mergers and applied the method to recent municipality mergers in Finland. We used two different estimation strategies that provide answers to different questions. First, we analyzed coalition formation using a simple logit analysis where a coalition consisting of two or more municipalities is used as the observation unit. Second, in order to allow for heterogeneity in the decision making of individual municipality councils, we followed Brasington (2003a) and estimated a Poirier bivariate probit model with population sorting. Our main interest was the association of local politics with merger decisions, a clear gap in both the empirical and theoretical literature on local government coalition formation. If local politics matter, the optimal coalition formation may be hindered further from what majority voting process would imply. Here optimality refers to maximizing municipal utility given the trade-off between economies of scale in the provision of public good and tailoring the public goods to heterogeneous population preferences.

In addition, using a regression discontinuity design we studied the effectiveness of a central government merger subsidy scheme aimed at promoting municipal mergers. This allows us to evaluate a policy that can be potentially used to solve the problems caused by local politics. We also studied how different municipality characteristics and heterogeneity both within a potential coalition and between different merger partners are associated with merging.

The main result from our analyses is that the local political environment has an important role in merger decisions. Councilors seem to act strategically indicating concerns over both re-election and own municipal employment. Furthermore, political fragmentation seems to hinder coalition formation and party composition also makes a difference so that a larger Centre Party share decreases the likelihood of merging. This means that local political environment may hinder the creation of optimal coalitions. Our results imply that theoretical and empirical work on coalition formation that abstracts away from political decision making process and decision makers' characteristics may be too simplistic.

Furthermore, regression discontinuity analysis reveals that the government subsidy scheme has had a positive effect on merging activity. This result does not mean that the subsidy scheme in itself is a good policy but it shows that subsidies can be used to achieve desired goals, for example to correct for the effects of local politics. The welfare effects of the scheme depend on whether the mergers induced by the subsidy enhance welfare more than the increase in taxes that need to be collected to finance the scheme. These aspects of the subsidy scheme are left for further research. We also find that existing cooperation through a joint authority in producing basic health care has a strong positive effect on the likelihood of merging and that distance prohibits merging.

The results in this paper are interesting in their own right, but they also serve as a first step in evaluating whether municipality mergers are an effective way of meeting the ultimate goal of the central government, which is to make municipal service production more efficient. For example, the finding that existing cooperation in basic health care is an important factor driving merger decisions could mean that a merger does not produce large efficiency gains because economies of scale are already exhausted through cooperation. However, we leave these issues for further research.

References

Alesina, A. – Spolaore, E. (1997): On the Number and Size of Nations. Quarterly Journal of Economics 112(4), 1027–1056.

Alesina, A. – Spolaore, E. (2003): The Size of Nations. Cambridge: MIT Press.

Alesina, A. – Baqir, R. – Hoxby, C. (2004): Political Jurisdictions in Heterogeneous Communities. Journal of Political Economy 112, 348–396.

Bolton, P. – Roland, R. (1997): The Breakup of Nations: A Political Economy Analysis. Quarterly Journal of Economics 112(4), 1057–1090.

Bradford, D. F. – Oates, W. E. (1974): Suburban Exploitation of Central Cities and Governmental Structure. In H. Hochman and G. Peterson (eds.): Redistribution through Public Choice. New York: Columbia University Press.

Brasington, D. (1999): Joint Provision of Public Goods: The Consolidation of School Districts. Journal of Public Economics 73, 373–393.

Brasington, D. (2003a): Size and School District Consolidation: Do Opposites Attract? Economica 70, 673–690.

Brasington, D. (2003b): Snobbery, Racism, or Mutual Distaste: What Promotes and Hinders Cooperation in Local Public-Good Provision? Review of Economics and Statistics 85, 874–883.

Breslow, N. E. (1996): Statistics in Epidemiology: The Case-Control Study. Journal of the American Statistical Association 91(433), 14–28.

Butler, J. S. (1996): Estimating the Correlation in Censored Probit Models. Review of Economics and Statistics 78(2), 356–358.

Cram, D. – Karan, V. – Stuart, I. (2010): Three Threats to Validity of Choice-based and Matched-Sample Studies in Accounting Research. Contemporary Accounting Research 26, 477-516.

Egger, P. – Koethenbuerger, M. (2010): Government Spending and Legislative Organization: Quasi-Experimental Evidence from Germany. American Economic Journal: Applied Economics 2, 200–212.

Ellingsen, T. (1998): Externalities vs Internalities: A Model of Political Integration. Journal of Public Economics 68, 251–268.

Gordon, N. – Knight, B. (2009): A Spatial Merger Estimator with an Application to School District Consolidation. Journal of Public Economics 93, 752–765.

Imbens, G. (1992): The Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling. Econometrica 60, 1187-1214.

Imbens, G. – Lemieux, T. (2008): Regression discontinuity designs: A guide to practice. Journal of Econometrics 142(2), pp. 615–635.

King, G. – Zeng, L. (2001): Logistic Regression in Rare Events Data. Political Analysis 9, 137-163.

Lee, D.S. – Lemieux, T. (2010): Regression Discontinuity Designs in Economics. Journal of Economic Literature 48, pp. 281–355.

Manski, C. F. (1995): Identification Problems in the Social Sciences. Cambridge MA: Harvard University Press.

Manski, C. F. and Lerman, S. R. (1977): The Estimation of Choice Probabilities from Choice Based Samples. Econometrica 45(8), 1977–1988.

Miceli, T. J: (1993): The Decision to Regionalize in the Provision of Education: An Application of the Tiebout Model. Journal of Urban Economics 33(3), 344–360.

Moisio, A. – Loikkanen, H. A. – Oulasvirta, L. (2010): Public Services at the Local Level – The Finnish Way. VATT Policy Reports 2. Helsinki.

Oates, W. E. (1972): Fiscal Federalism. Harcourt Brace Jovanovich, New York.

Oates, W. E. (1999): An Essay on Fiscal Federalism. Journal of Economic Literature, Vol. XXXVII, pp. 1120–1149.

Poirier, D. (1980): Partial Observability in Bivariate Probit Models. Journal of Econometrics 12, 209–217.

Tiebout, C. M. (1956): A Pure Theory of Local Expenditures. Journal of Political Economy 64(5), 416–424.

Van der Klaauw, W. (2002): A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment. International Economic Review 43(4), pp. 1249–1287.

Weese, E. (2009): Political Mergers as Coalition Formation: Evidence from Japanese Municipal Amalgamations. MIT mimeo.

Wernicke, S. (2005): A Faster Algorithm for Detecting Network Motifs. In S. Istrail, P. Pevzner, and M.Waterman (eds.): Algorithms in Bioinformatics. Berlin/Heidelberg: Springer.

Wernicke, S. – Rasche, F. (2006): FANMOD: a tool for fast network motif detection. Bioinformatics 22(9), 1152–1153.

Appendix A. Additional figures and tables.

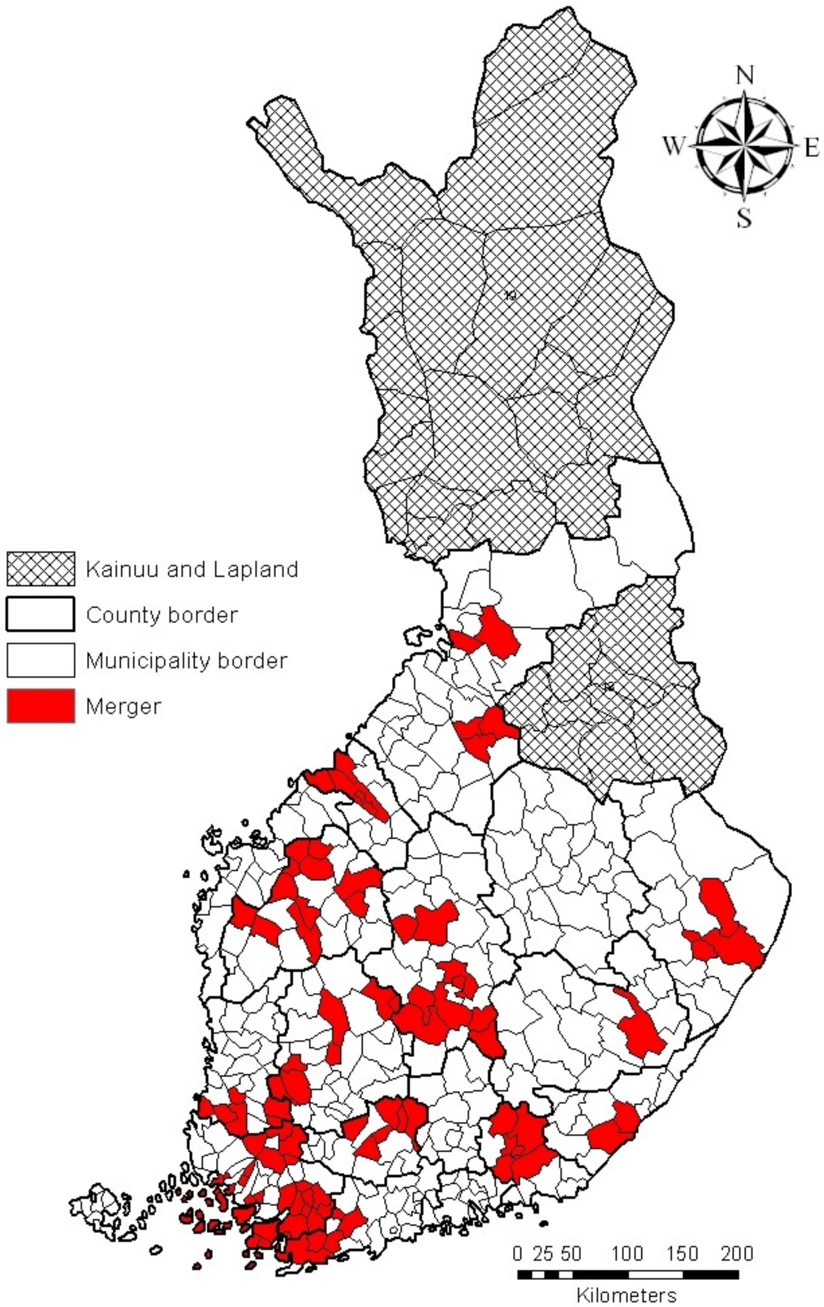

Figure A1. Map of municipality mergers in 2008–2009.

Table A1. Council size as a function of municipal population.

Population	Council size
Less or equal to 2,000	13, 15 or 17
2,001–4,000	21
4,001–8,000	27
8,001–15,000	35
15,001–30,000	43
30,001–60,000	51
60,001–120,000	59
120,001–250,000	67
250,001–400,000	75
Over 400,000	85

Table A2. Variable definitions.

population	Coalition population in the logit model.
mean of taxable income	Coalition level income in the logit model, municipal level income in the Poirier model.
dependency ratio	Share of 0-15 and over 65 year-olds in the population.
cooperation	Dummy that equals 1 if municipalities in the coalition organize basic health care through a joint authority, zero otherw ise.
language	Dummy that equals 1 if at least one municipality in the coaltion is classified in a different language than others in the coalition, zero otherw ise. There are four different classes: unilingual Finnish, unilingual Sw edish, bilingual w ith a Finnish speaking majority and bilingual w ith a Sw edish speaking majority.
distance	Tw o distance measures are used. In the logit model the distance measures the average population Euclidean distance to the largest municipality in the coalition. In the Poirier model the measure is the unw eighted maximum Euclidean distance from other municipalities to largest municipality.
total land area	Total land area of the coalition.
popualtion density	Population density measured using the focalsum method based in 1 km grids.
central government grants	Central government grants.
deficit	The number of municipalities in the coalition w ith a 4-year running deficit of more than €1,000 per capita in the logit model. This is a limit set by the ministry of finance. If a municipality is below this limit it has to give the ministry of finance a briefing on how to deal w ith the deficit. In the Poirier model the deficit is in Euros per capita.
std.dev. population	Standard deviation of municipal population in the coalition.
std.dev. taxable income	Standard deviation of municipal taxable income in the coalition.
std.dev. tax rate	Standard deviation of municipal tax rates in the coalition.
sdt.dev. population density	Standard deviation of municipal population density in the coalition.
std.dev. depency ratio	Standard deviation of municipal dependcy ratios in the coalition.
difference in population	Absolute difference in municipal population.
difference in taxable income	Absolute difference in municipal taxable income.
difference in tax rates	Absolute difference in municipal income tax rates.
reduction in council size	Relative reduction in overall council seats if the merger goes through.
Herfindahl index	Standard Herfindahl index based on party shares in the council.
same largest party	Dummy that equals 1 if municipality councils in the coalition all have the same largest party, zero otherw ise.
left-w ing party share	The share of left-w ing parties (Social Democrats and Left Alliance).
Centre Party share	The share of the Centre Party.
share of municipal employees in council	The share of council members w ho are employed by the municipality they live in.
share of municipal employees in population	The share of municipal employees in the population.

Table A3. Additional coalition level logit results.

	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
population			0.65	0.86	-0.47	1.60
$(\text{population})^2$			-0.05	0.03	-0.05	0.04
mean of taxable income			-6.47	7.33	-9.48	11.0
central government grants			61.8	33.5	98.3	50.4
dependency ratio			-71.4	57.7	-143.9	116.0
cooperation			3.54	2.09	5.08*	2.29
language			-0.16	1.43	-1.35	1.84
RDD treatment 1			5.47**	1.77	7.37*	3.11
RDD treatment 2			2.40**	0.71	3.69**	0.69
mean distance			-0.60*	0.24	-0.97*	0.42
total land area			3.78	5.64	9.9	6.41
population density			7.4*	3.64	8.15	5.83
deficit			1.02**	0.33	0.59	0.36
std.dev. population					1.34*	0.65
std.dev. taxable income					4.84	7.06
std.dev. tax rate					330.1*	89.9
sdt.dev. population density					6.96*	2.79
std.dev. depency ratio					-52.0**	5.8
reduction in council size	-0.43	13.1			45.5*	18.8
Herfindahl index	106.0	64.9			62.2**	23.4
same largest party	6.53	4.23			-0.05	0.70
left-w ing party share	12.7	9.42			1.40	3.48
centre party share	-67.0	39.0			-33.1*	15.0
share of municipal employees in council	50.8	29.7			61.1**	17.1
share of municipal employees in population	-295.7	156.0			-188.9*	80.1
forcing variable			-1.58	2.54	0.46	1.91
$(\text{forcing variable})^2$			0.18	0.24	0.07	0.13
Number of obs.	165		165		165	
Pseudo Log-L	-495 593		-47 586		-34 256	
Pseudo R^2	0.68		0.97		0.98	

*Notes: The results are from conditional logit models where stratification fixed effects are included in the estimation. ** and * indicate statistical significance at 1and 5 percent level, respectively. Forcing variable = (coalition population - population of the largest municipality in the coalition).*

Appendix B. Robustness checks for the regression discontinuity analysis.

In this appendix, we report additional robustness checks for our RDD analysis concerning the merger subsidy. We start with a visual inspection of the jump in the propensity to merge in the discontinuity. In Figure B1, we have pooled all the four thresholds presented in Figure 2 together. This single threshold is located at zero on the horizontal axis of Figure B1. We calculated the distance in the original forcing variable to the nearest threshold for each observation so that each observation has a location on the horizontal axis. The purpose of Figure B1 is to compare the propensity to merge on both sides of the thresholds. This is a standard RDD graph intended as the first evaluation of the treatment effect of interest except that now four discontinuity points are pooled into one.

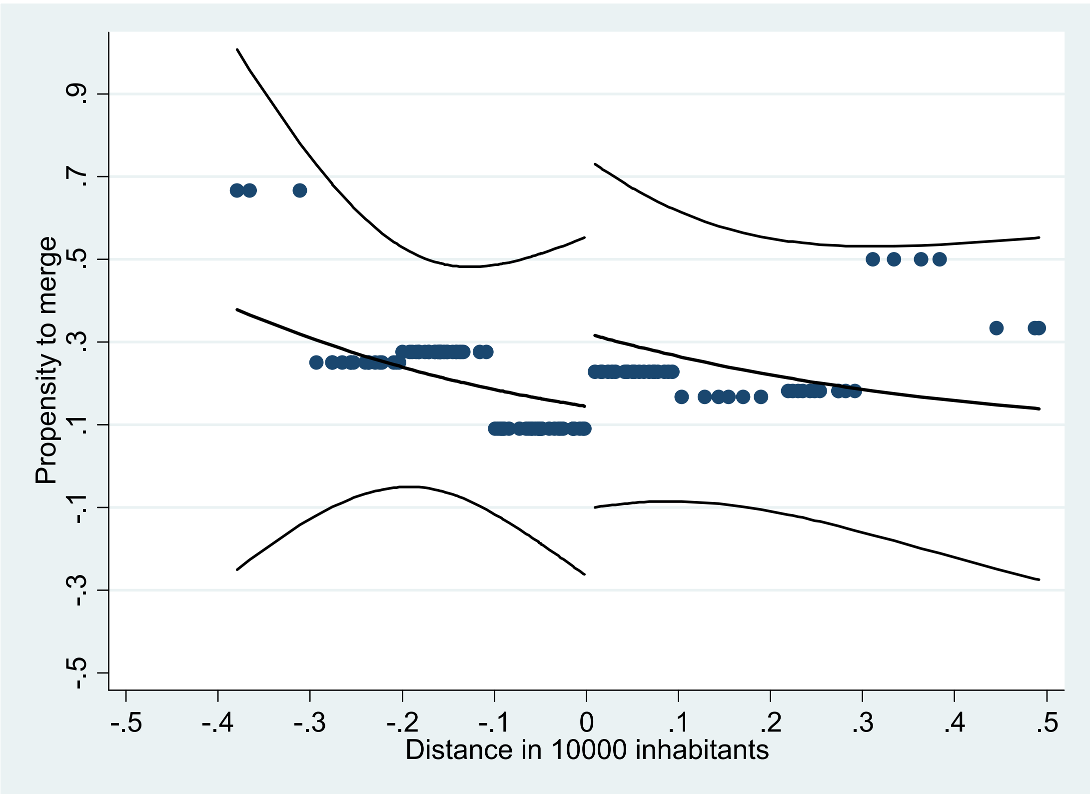

Figure B1. The propensity to merge around thresholds.

The fitted lines and their confidence intervals are based model (5) in Table B1, where the forcing variable is controlled using a parametric fourth order polynomial. The model does not contain control variables. The fit is drawn using the entire data set. The dots correspond to the average values of the outcome (merger) within bins of 1,000 inhabitants. The amount of the dots gives the number of observations within each bin, but all the observations are given the average value in the graph. Based on both the average values and the fitted line, there is a clear jump at the discontinuity, but the jump is not statistically significant in this unconditional model. Moreover, there are jumps of similar size in other parts of the distribution, but the bins involving these jumps have only a few observations. It is likely that these are just small sample problems and therefore we proceed with the RDD analysis.

To complement Figure B1, in Table B2 we report the RDD results using only the single discontinuity as the determinant of treatment and the distance to the threshold as the forcing variable. We run the model with parametric linear regressions separately on each side of the threshold while changing the number of polynomials from zero to four and including the same control variables as in largest model in Table 6.

Table B1. RDD robustness with the pooled treatment and distance as the forcing variable.

	(1)	(2)	(3)	(4)	(5)
treatment	-0.12	0.051	0.069	1.043	1.030
s.e.	0.389	0.441	0.467	0.644	0.705
forcing variable	no	linear	quadratic	cubic	quartic
controls	no	no	no	no	no
strata FE	no	no	no	no	no
WESML	no	no	no	no	no
	(6)	(7)	(8)	(9)	(10)
treatment	3.402	3.310	6.08*	9.228***	9.432***
s.e.	2.274	2.119	3.478	3.302	3.393
forcing variable	no	linear	quadratic	cubic	quartic
controls	yes	yes	yes	yes	yes
strata FE	yes	yes	yes	yes	yes
WESML	yes	yes	yes	yes	yes

****, ** and * indicate statistical significance at 1, 5 and 10 percent level, respectively.*

In order to test the robustness of our reported results in Table 6, we present two different models that differ in the flexibility of the forcing variable. In Table B2, the forcing variable is estimated separately for each of the six different areas in Figure 2. In other words, for each area separated by thresholds we allow for different parameter estimates for all the polynomials of the forcing variable.

Table B2. RDD robustness of the reported results with flexible forcing variable estimation.

	(1)	(2)	(3)	(4)	(5)
treatment 1	1.20	2.61**	2.56	3.16	NA
s.e.	0.800	1.140	1.986	2.194	NA
treatment 2	0.95	2.08	2.40	2.42	NA
s.e.	0.648	0.895	1.654	2.003	NA
forcing variable	no	linear	quadratic	cubic	quartic
controls	no	no	no	no	no
strata FE	no	no	no	no	no
WESML	no	no	no	no	no
	(6)	(7)	(8)	(9)	(10)
treatment 1	7.311**	12.74***	NA	NA	NA
s.e.	2.971	2.774	NA	NA	NA
treatment 2	3.710***	7.66***	NA	NA	NA
s.e.	0.699	1.504	NA	NA	NA
forcing variable	no	linear	quadratic	cubic	quartic
controls	yes	yes	yes	yes	yes
strata FE	yes	yes	yes	yes	yes
WESML	yes	yes	yes	yes	yes

****, ** and * indicate statistical significance at 1, 5 and 10 percent level, respectively.*

Due to the combination of flexible specification and a small sample only up to first order polynomials with controls or third order polynomials without controls can be included between the thresholds. The two regressions that include control variables are robust but in the unconditional regressions the standard errors increase as the number of polynomials increase thus making the results less robust.

In Table B3, we force the effect of the forcing variable to be the same in all the areas, as is done in the reported results in the main text. In Table B3, model (8) corresponds to the largest model in Table 6. These results are very robust to different number of polynomials. Both the standard errors and the coefficients remain very similar when the number of polynomials is adjusted. As is typical with RDD, control variables are needed for significant results in some of the models. What is atypical is that this significance is not achieved through smaller standard errors but through larger coefficients.

Table B3. RDD robustness of the reported results with restricted forcing variable estimation.

	(1)	(2)	(3)	(4)	(5)
treatment 1	1.20	1.25	1.27	1.44*	1.34*
s.e.	0.800	0.801	0.802	0.811	0.814
treatment 2	0.95	1.16*	1.18*	1.25*	1.33**
s.e.	0.648	0.662	0.664	0.666	0.669
forcing variable	no	linear	quadratic	cubic	quartic
controls	no	no	no	no	no
strata FE	no	no	no	no	no
WESML	no	no	no	no	no
	(6)	(7)	(8)	(9)	(10)
treatment 1	7.311**	7.322**	7.366**	7.321**	8.136*
s.e.	2.971	3.016	3.114	3.148	4.213
treatment 2	3.710***	3.666***	3.694***	3.651***	4.756**
s.e.	0.699	0.643	0.689	0.689	1.884
forcing variable	no	linear	quadratic	cubic	quartic
controls	yes	yes	yes	yes	yes
strata FE	yes	yes	yes	yes	yes
WESML	yes	yes	yes	yes	yes

****, ** and * indicate statistical significance at 1, 5 and 10 percent level, respectively.*

In Table B4, we check whether the reported results are robust to the distance criterion that we use to define whether an observation is "just below" or "just above" the thresholds. We try both absolute population distance and a distance that is relative to the threshold value. This selection involves a trade-off between having more faith in random assignment to the treatment and reference groups (small distance) and having predictive power (large distance). It seems that a distance of 500 inhabitants is too small to get significant results at 5 % level but the coefficients are of the same size as with larger distances. The parameter estimate in the largest 30% inclusion criterion decreases. This somewhat worrying observation is probably due to not being able to use a more flexible specification. Overall, these results imply that our reported results are not driven by treatment and control group selection criterion.

Table B4. RDD robustness of the reported results with respect to the selection criterion of inclusion to the treatment.

	(1)	(2)	(3)	(4)	(5)	(6)
treatment 1	7.809*	7.366**	6.041**	5.514	7.507*	4.767
s.e.	4.352	3.114	2.432	4.047	4.135	2.955
treatment 2	2.792*	3.694***	2.877***	1.420***	3.775***	2.212***
s.e.	1.428	0.689	0.670	0.523	1.240	0.601
forcing variable	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
controls	yes	yes	yes	yes	yes	yes
strata FE	yes	yes	yes	yes	yes	yes
WESML	yes	yes	yes	yes	yes	yes
treatment criterion	500	1000	1500	10%	20%	30%
share treatment 1	4.2%	11.5%	13.3%	9.1%	10.9%	16.4%

****, ** and * indicate statistical significance at 1, 5 and 10 percent level, respectively.*

Finally, we test the validity of our identification assumption by comparing forcing and control variables “just above” and “just below” the threshold in Table B5. First, we check whether any of the other observed characteristics get different values just above and below the discontinuity points. There are no statistically significant differences, which gives credibility to our identification assumption. Second, we should look for discontinuity in the density of the forcing variable at the threshold in order to assess the potential manipulation of the forcing variable in order to get or to avoid the treatment. Because we have more observations “just below” the threshold than “just above” the threshold, manipulation is not a concern in our beneficial treatment case. In summary, our regression discontinuity design seems valid.

Table B5. RDD robustness with one treatment and the distance as a forcing variable.

	Just (1000) below threshold			Just (1000) above threshold		
	Obs.	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
merger	31	0.097	0.301	19	0.263	0.452
forcing variable	31	0.412	0.215	19	0.622	0.270
population	31	1.501	1.231	19	3.205	5.025
mean of taxable income	31	1.095	0.148	19	1.129	0.165
central gov. grants	31	0.169	0.048	19	0.156	0.046
dependency ratio	31	0.375	0.017	19	0.367	0.028
cooperation	31	0.323	0.475	19	0.211	0.419
language	31	0.097	0.301	19	0.000	0.000
nro. of municipalities	31	2.452	0.961	19	2.737	0.653
mean distance	31	11.46	4.365	19	14.36	7.943
total land area	31	0.107	0.064	19	0.120	0.054
population density	31	0.664	0.289	19	0.675	0.286
deficit	31	0.129	0.341	19	0.000	0.000
std. dev. population	31	0.670	1.059	19	1.590	3.209
std. dev. taxable income	31	0.117	0.070	19	0.107	0.068
std. dev. tax rate	31	0.005	0.004	19	0.005	0.004
std. dev. pop. density	31	0.379	0.288	19	0.318	0.260
std. dev. dependency ratio	31	0.019	0.012	19	0.021	0.015
reduction in council size	31	0.386	0.114	19	0.454	0.090
Herfindahl index	31	0.340	0.094	19	0.319	0.086
same largest party	31	0.806	0.402	19	0.789	0.419
left-w ing party share	31	0.247	0.096	19	0.252	0.072
centre party share	31	0.453	0.148	19	0.417	0.167
share mun. emp. in council	31	0.080	0.031	19	0.078	0.033
share mun. emp. in population	31	0.060	0.011	19	0.060	0.009

IN VATT WORKING PAPERS SERIES PUBLISHED PUBLICATIONS

1. Tomi Kyyrä – Pierpaolo Parrotta – Michael Rosholm: The effect of receiving supplementary UI benefits on unemployment duration. Helsinki 2009.
2. Tuomas Pekkarinen – Roope Uusitalo – Sari Kerr: School tracking and development of cognitive skills. Helsinki 2009.
3. Essi Eerola – Niku Määttänen: The optimal tax treatment of housing capital in the neoclassical growth model. Helsinki 2009.
4. Sanna-Mari Hynninen – Aki Kangasharju – Jaakko Pehkonen: Matching inefficiencies, regional disparities and unemployment. Helsinki 2009.
5. Jari Ojala – Jaakko Pehkonen: Technological changes, wage inequality and skill premiums: Evidence over three centuries. Helsinki 2009.
6. Elina Berghäll: R&D, investment and structural change in Finland: Is low investment a problem? Helsinki 2009.
7. Jarkko Harju: Voluntary pension savings: the effects of the Finnish tax reform on savers' behaviour. Helsinki 2009.
8. Katariina Nilsson Hakkala – Fredrik Heyman – Fredrik Sjöholm: Multinational firms and job tasks. Helsinki 2009.
9. Seppo Kari – Hanna Karikallio – Jukka Pirttilä: The impact of dividend taxation on dividends and investment: New evidence based on a natural experiment. Helsinki 2009.
10. Sara Ayllón: Modelling state dependence and feedback effects between poverty, employment and parental home emancipation among European youth. Helsinki 2009.
11. Juha Honkatukia – Jouko Kinnunen – Kimmo Marttila: Distributional effects of Finland's climate policy package – Calculations with the new income distribution module of the VATTAGE model. Helsinki 2009.
12. Juha Tuomala: The threat effect of mandatory programme participation in Finland – A study of activation reform of the labour market support system. Helsinki 2010.
13. Ilja Kristian Kavonius: Fiscal policies in Europe and the United States during the Great Depression. Helsinki 2010.
14. Miikka Rokkanen – Roope Uusitalo: Changes in job stability – evidence from lifetime job histories. Helsinki 2010.
15. Heléne Lundqvist: Granting public or private consumption? Effects of grants on local public spending and income taxes. Helsinki 2010.
16. Ossi Korkeamäki – Tomi Kyyrä: Institutional rules, labour demand and retirement through disability programme participation. Helsinki 2010.
17. Tomi Kyyrä: Early retirement policy in the presence of competing exit pathways: Evidence from policy reforms in Finland. Helsinki 2010.
18. Tuomas Kosonen: What was actually cut in the barbers' VAT cut? Helsinki 2010.

VALTION TALOUDELLINEN TUTKIMUSKESKUS
STATENS EKONOMISKA FORSKNINGSCENTRAL
GOVERNMENT INSTITUTE FOR ECONOMIC RESEARCH

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
P.O.Box 1279
FI-00101 Helsinki
Finland

ISBN 978-951-561-953-2
ISSN 1798-0283