

**Supplement to “Political mergers as coalition formation:
An analysis of the *Heisei* municipal amalgamations”**
(*Quantitative Economics*, Vol. 6, No. 2, July 2015, 257–307)

ERIC WEESE
Yale University

APPENDIX C: VOTING MODEL

The variables determined by a political process at the local level, given a certain municipality m , are q_m , τ_m , and θ_m . The national government chooses the transfer function T . Due to the form of the utility function in (3), all individuals agree on the optimal level for the public good, q_m^* , and the tax rate, τ_m^* . For the other political choices, however, different individuals will have different ideal points, and each of these may be multidimensional: the location θ_m of the public good is geographic (latitude and longitude), and the transfer function is chosen from a space of functions that in Section 4 is assumed to be two dimensional. A very simple model of the political process at both the local and national level gives the result that the policy selected is a weighted sum of individual utilities. To obtain this result, make the following assumptions:

ASSUMPTION 1. *There are two identical office-motivated candidates who run on policy platforms to which that they can commit.*

ASSUMPTION 2. *Voting in elections is determined via a probabilistic voting model where vote probabilities are linear in utility difference between the two candidates.*

ASSUMPTION 3. *There is a continuum of voters.*

ASSUMPTION 4. *For all θ, θ' , for all $\gamma \in (0, 1)$, the set of voters i such that*

$$u_i(\gamma\theta + (1 - \gamma)\theta') > \gamma u_i(\theta) + (1 - \gamma)u_i(\theta')$$

has positive measure.

With these assumptions, the unique political equilibrium is for both candidates to propose θ_m^* to maximize the sum of individual utilities: this is Theorem 4 in [Banks and Duggan \(2005\)](#).⁸⁹

Eric Weese: eric.weese@yale.edu

⁸⁹An additional requirement of Theorem 4 of [Banks and Duggan \(2005\)](#) is that for each voter, utility is (weakly) concave with respect to the policy choice. The utility function given in (3) is concave in θ , although it is not strictly concave.

C.1 Japanese context

Municipal politics in Japan involve both a mayor and a municipal council, and thus there is, in reality, more than the one decision-maker assumed above; however, with the exception of very large “designated municipalities,” the council is elected on an entirely at-large basis, without wards or other subdivisions. The mayor has veto power, which can be overruled by a two-thirds vote of the municipal council. Given the lack of wards in the municipal council, it is not entirely clear how or why policies proposed by the council might diverge from policies proposed by the mayor, although examples of this sort of conflict can be found in municipal records. Because this paper’s focus is inter- rather than intramunicipal decision-making, the following assumption will be used: mayors will veto anything other than the policy proposed in their campaign, and less than two-thirds of the council will be opposed to this policy.⁹⁰

National level politics are even more complex and thus diverge even more from this simple model. The candidates in this case would be political parties, which commit to party platforms. Issues with single-member constituencies and multiple houses in the legislature, with different malapportionment in each house, are abstracted away from. Election of representatives is also abstracted away from: individuals in areas that are overrepresented are simply assumed to be able to cast more votes and are thus weighted more heavily.

Assumptions 3 and 4 are not quite satisfied in the data actually used: there is a large but finite number of voters, and there are a few cases (generally in the smallest municipalities) where there are locations θ and θ' such that all voters are indifferent between randomization between the policies versus a convex combination. The argument regarding Assumption 3 is simply that thousands of voters is “close enough” to a continuum. Regarding Assumption 4, violations of this assumption still result in candidates proposing policies that maximize social welfare, only these policies are no longer necessarily unique. For example, with the utility function in (3), consider a municipality with exactly half of its population at one point and exactly half at another point. Then any θ^* between these two points is welfare maximizing and there is not a unique political equilibrium. Empirically, actual population distributions are never this evenly balanced, and a unique θ^* can always be computed. Distance enters the utility function in (3) linearly, and thus these θ^* are points that minimize the sum of distances, points sometimes referred to as *generalized medians*.

C.2 Japanese merger procedures

The general political rules for municipal mergers were the following.⁹¹

⁹⁰Prior to the merger period, mayors were responsible for delivering hundreds of “agency delegated functions” from higher levels of government, making them bureaucrats as well as politicians and making it possible (at least in theory) for central ministries to fire a mayor for not performing a delegated function according to specifications. Agency delegated functions were abolished during the merger period, and municipal policies are thus modeled as being determined by local residents through a political process.

⁹¹This discussion ignores many details, such as the distinction between *hennyuu* municipalities, where bylaws are inherited from one of the merger participants (normally the largest city), and *shinsetsu* mergers, where bylaws and regulations are developed from scratch.

0. Mayors of municipalities can create “voluntary merger committees” and “study committees” to gather information, but there are no regulations regarding these committees and they are not necessary in order to proceed with a merger.

1. A petition for a specified merger from 2% of eligible voters (or the municipal council) in any single municipality forces an official response from all the municipalities included in the proposed merger, based on a debate in their municipal councils. Unanimous “yes” responses result in the creation of an official merger committee. There is no requirement regarding previous voluntary committees or study committees.

2. If a municipal council rejects the proposed merger committee, a petition from one-sixth of eligible voters in the relevant municipality forces a referendum on the creation of the merger committee. A majority vote in the referendum overrides the council's rejection.

3. The merger committee produces reports on the financial situation of the municipalities and proposes some characteristics of the merger (e.g., the name of the merged municipality). A majority vote in each municipal council is required to finalize the merger.⁹²

The existence of an official referendum process during the planning stage but not at the final approval stage suggests that the best strategy for politicians who are opposed to a merger might have been to remain silent during the initial stage, but then prevent the final resolution from passing in the council. Behavior such as this did in fact occur in a small number of municipalities, but does not appear to have been particularly common or successful. First, the process of creating the merger committee generally attracted a considerable amount of attention, particularly in smaller municipalities. In cases where there was controversy, referendum turnout rates could exceed 90%. It was thus difficult for politicians facing a potentially controversial merger to prevent a referendum regarding the creation of the merger committee, and conditional on that referendum passing, it was difficult to then vote against the final proposed merger. Furthermore, in cases where politicians did vote against mergers that appeared to have popular support, a hitherto seldom used recall process was employed to remove them from office via a majority vote in a recall referendum. Whereas there was only one recall referendum during the 1990s, there were at least 41 during the merger period.

A formal interpretation of these rules is somewhat difficult; however, a common element in all mergers is that they were approved by all municipalities in question, either via local referendum or in the municipal council.⁹³ As council resolutions were subject to veto by the mayor, this paper will assume that the binding constraint on the behavior of a municipality is the ability of its residents to recall the mayor. Suppose that there

⁹²In general, the division of a municipality was prohibited. In one case, such a split did occur, but both of the resulting municipalities were immediately merged with different neighbors.

⁹³In about one-third of cases, referenda were held. Most of these were nominally consultative, but there is only one instance in which a municipal council voted opposite to a referendum result. This case was complicated due to multiple referenda with conflicting results as well as a number of other procedural irregularities, and finally resulted in a recall of the mayor and a request to the prefectural governor to reverse the merger. The request for reversal was denied.

is perfect information regarding what mergers are feasible (i.e., will be approved by all other participants). The mayor proposes a merger for the municipality to participate in or proposes remaining independent. A single challenger then appears and similarly proposes a policy. With the probabilistic voting model presented above, the challenger will run on the same platform as the mayor and the selected merger will maximize the sum of the utilities of residents.⁹⁴

A potential objection here is that the costs of organizing a recall election could be large, and thus the mayor's incumbency advantage could be significant enough to allow merger proposals that are far from the optimal to be enacted.⁹⁵ There are two responses to this objection: first, the merger period was sufficiently long that at least one regularly scheduled election occurred during the merger period, and during this election the merger issue was particularly salient; second, the cost of organizing a recall does not appear to be as large as might be supposed. Specifically, in 4 of the 41 recalls, a majority voted against the recall in the referendum, and in another 6 of the recalls, the mayor was reelected in the special election following the recall process (usually after resigning voluntarily to avoid the recall referendum). Thus, a full quarter of the organized recall referenda did not succeed in removing the mayor. If the costs of organizing a recall referendum were very high, one would expect that they would be organized only when the mayor would not have majority support in the recall referendum or the subsequent election. Thus, this paper will use the assumption that the municipal merger selected by each municipality maximized the utility of local residents, given the possible alternative mergers.⁹⁶

APPENDIX D: DATA AND INSTITUTIONAL DETAILS

Data for municipal characteristics X_m come from the Statistical Information Institute, which aggregates a variety of government sources. Where 1995 data were not available, the year closest to 1995 was used. Data regarding municipal mergers come from the Japan Geographic Data Center. The 1995 Japanese national census provides population data at the kilometer grid square level. Information on municipal boundaries is taken from shape files produced by ESRI Japan, also for 1995. Other data and formulae are generally taken from Ministry of Internal Affairs and Communications documents.

⁹⁴The assumption here that mayors do not have a large effect on mergers might still seem suspicious. Kawaura (2010) investigates the effect of a mayor's length of tenure on merger configurations, and finds effects that are small and not statistically significant at the 95% level. While there is certainly anecdotal evidence that certain mayors may have obstructed certain mergers, there is no immediately obvious relationship in the aggregate data. The private incentive for municipal politicians to maintain the independence of their municipality so as to preserve their own employment is not as strong as might be anticipated. This is due to central government policies: for example, the length of service required to receive a pension were reduced for politicians in a municipality participating in a merger, and following the merger period, the pension system was abolished, with a (disadvantageous) one-time payment to those who did not meet the standard 12 year length of service requirement.

⁹⁵A recall referendum required a petition by between one-sixth and one-third of residents.

⁹⁶The specific definition of "alternative" used here is that they would be deviations as discussed in Appendix A.

D.1 *Tax base*

The tax base Y_m is determined from *standard fiscal revenue* figures produced by the Ministry. While the model in Section 1 has each municipality choosing a single tax rate τ_m , actual municipal tax revenues come from several taxes, with “fixed asset” taxes (land, housing, and some business assets) and personal income tax being the two most important types. For each type of tax, the Ministry sets a reference tax rate, and then calculates the total revenue each municipality would receive if it charged these reference rates on its tax base. That is, if $\bar{\tau}^k$ is the reference rate for tax type k and Y_m^k is the tax base for this tax for municipality m , the Ministry Standard Fiscal Revenue estimate for municipality m is $\sum_k \bar{\tau}^k Y_m^k$. To convert this to the single tax base that is assumed in this paper, suppose that the single tax base is income and set $\bar{\tau} = 0.12$, which is total municipal Standard Fiscal Revenue as a fraction of total income. Then calculate Y_m for each municipality so as to satisfy

$$\bar{\tau} Y_m = \sum_k \bar{\tau}^k Y_m^k. \quad (46)$$

That is, Y_m is calculated so that $\bar{\tau} Y_m$ is exactly equal to the Standard Fiscal Revenue for that municipality, as reported by the Ministry. The tax rate actually observed in the municipality, τ_m^* , is defined as taxes as a fraction of Y_m .

In general, collapsing multiple tax bases into a single tax base would be problematic, but in the Japanese case, although municipalities are *de jure* allowed to choose a tax rate different from the standard rate, the amount of actual variation is very low. For example, in the extreme case of Yuubari City, effectively bankrupt with a debt of over ¥3 million per capita, the income tax rate was raised from 6.0% to 6.5%, but almost all other municipalities charge the standard 6.0%. The standard fixed asset rate of 1.4% is levied by about 9 out of 10 municipalities, with the remaining tenth mostly charging 1.5% or 1.6%.⁹⁷ Thus, the observed tax data that the model is attempting to explain involve all municipalities charging very similar tax rates, which are equal or very close to the Ministry’s reference rate $\bar{\tau}$. In particular, there are no cases where a municipality chose to charge a very high rate on a particular tax base for which the reference rate was much lower, a situation which could lead to high and nonsensical calculated values for τ_m^* .

D.2 *Geographic distance*

Surveys conducted around the time of the mergers reveal a widespread perception that there were unexploited efficiencies of scale in the provision of public services: the most popular response to questions regarding the potential benefits of municipal mergers was “avoid duplication of facilities/avoid useless capital expenditures” in Kyoto, “reduce expenditures by improving administrative efficiency, eliminating duplicate facil-

⁹⁷A fixed asset rate of more than 1.7% requires Ministry approval, but few municipalities are at this cap. While there are Ministry caps on taxes, they are rarely binding. The sole exception is for taxes on corporations, where a sizeable number of municipalities do charge at the upper bound. These corporate taxes are a small percentage of total taxes, and thus issues with this upper bound are not considered in this paper.

ities, and reducing personnel” in Yamanashi, “reduce personnel expenses” in Akita, and “reduce personnel and other expenditures and improve efficiency” in Okinawa.⁹⁸

It was thus generally understood that a municipal merger involving a smaller municipality and a larger one would result in the closure of city hall and some other facilities in the smaller municipality, and that this would result in substantial cost savings.⁹⁹ This led to a concern among residents of smaller municipalities that after a merger, public services would only be provided at a location much farther away than was the case currently. As an increase in geographic distance to public facilities was perceived as a major disadvantage of a municipal merger, it seems appropriate to use geography as the metric for heterogeneity in the structural part of the model.

By combining census grid square data and municipal boundary shape files, the location of individuals in municipalities can be known to the kilometer grid square level. To calculate distances, first, the population of grid squares that are on a boundary between two municipalities is divided between the municipalities in proportion to the area of each grid square in each municipality. Then, for any θ_m , the distance $\ell(i, \theta_m)$ can be calculated as the great-circle distance from the physical longitude–latitude location of individual i to θ_m . For computational simplicity, all individuals within a given census grid square are assumed to live at the center of the square. The distances in question are small relative to the curvature of the earth, so this is effectively a straight-line distance calculation.

The location θ_m^* chosen by a municipality will minimize the sum of these individual distances due to the assumption that the local political process is as described in Appendix C. These θ_m^* are calculated via standard optimization techniques. Although, as discussed in Appendix C, there are cases where θ_m^* might not be unique, a unique value is, in fact, obtained for all municipalities. For each coalition S , the optimal location θ_S^* is calculated via exactly the same process. The value of $\ell_m(\theta_S^*)$ in (4) can then be calculated by averaging over distances $\ell(i, \theta_S^*)$ for all individuals in m . This process is computationally intensive, but $\ell_m(\theta_S^*)$ depends neither on ϵ nor on β , and thus the calculation of these distances only needs to be performed once.

APPENDIX E: COMPUTATIONAL DETAILS

E.1 *Idiosyncratic shocks*

The standard relationship for sample variances from the analysis of variance literature (e.g., [Scheffé \(1959\)](#)) will hold:¹⁰⁰

$$s_S^2 = \frac{\sum_{m \in S} (N_m - 1) s_m^2 + \sum_{m \in S} \sum_{m' \in S} \frac{N_m N_{m'} (\bar{\omega}_m - \bar{\omega}_{m'})^2}{2N_S}}{N_S - 1}. \quad (47)$$

⁹⁸Unfortunately, these surveys are difficult to analyze quantitatively, as they were only conducted in a few prefectures and a different questionnaire was used in each prefecture.

⁹⁹Although many services at city hall could be accessed via mail, telephone, or the Internet, it is common and in some cases required to visit in person.

¹⁰⁰This is immediate from the “textbook” variance decomposition formula because $\sum_m \sum_{m'} \frac{N_m N_{m'}}{N_S} (\bar{\omega}_m - \bar{\omega}_{m'})^2 = \sum_m \sum_{m'} \frac{N_m N_{m'}}{N_S} ((\bar{\omega}_m - \bar{\omega}) - (\bar{\omega}_{m'} - \bar{\omega}))^2 = \sum_m \sum_{m'} \frac{N_m N_{m'}}{N_S} ((\bar{\omega}_m - \bar{\omega})^2 + (\bar{\omega}_{m'} - \bar{\omega})^2)$.

The function $\epsilon(\cdot, \cdot)$ is exactly as given by (20) and (47), and maps from the $2^{|M|}$ dimensional space of $\bar{\omega}_M$ and s_M to the $2^{|M|}$ dimensional space of ϵ . The elements of both $\bar{\omega}_M$ and s_M have known distributions, which will be helpful when computing moment inequalities.

With this construction of ϵ , for any guess $\hat{\beta}$ for the parameter vector, any observed partition can be rationalized: simply choose s_S^2 sufficiently close to zero if S is in the observed partition, and large otherwise, and then choose $\bar{\omega}_S$ and $\bar{\omega}_{S'}$ such that $(\bar{\omega}_S - \bar{\omega}_{S'})^2$ is sufficiently large to discourage any coarsenings into larger coalitions.

A linear approximation is important for implementation, as the moment inequality that will be described immediately below requires a search over ϵ , constrained by the stability of the observed partition. Using the approximation

$$\begin{aligned} \log s_S^2 &\simeq \log(1 + \delta_S), \quad \delta_S \sim N\left(0, \frac{2}{N_S - 1}\right) \\ &\simeq \delta_S \end{aligned} \tag{48}$$

makes the stability constraints linear and, thus, makes the optimization problem computationally feasible.¹⁰¹

The correlation of ϵ_S and $\epsilon_{S'}$ in this case will be approximately $\frac{N_{S \cap S'}}{\sqrt{N_S} \sqrt{N_{S'}}}$, where $N_{S \cap S'}$ is the number of individuals in common between the two coalitions. Thus, ϵ are uncorrelated when there is no overlap and approach perfect correlation as the overlap becomes total. A covariance matrix with this structure will not, in general, have full rank: this rank deficiency makes the estimation approach that follows feasible.

E.2 First type of moment inequality

Specifically, for an observed partition π , consider the two types of deviations

$$\mathcal{S}^c(\pi) = \{S' | S' \text{ is a merger of two geographically adjacent singletons in } \pi\},$$

$$\mathcal{S}^r(\pi) = \{S' | S' \subset S \in \pi\}.$$

Here $\mathcal{S}^c(\pi)$ are coarsenings of the partition π and $\mathcal{S}^r(\pi)$ are refinements. To retain computational feasibility, restrict \mathcal{S}^r to contain only coalitions that are geographically contiguous and do not cross more than two county boundaries.¹⁰² Here, counties are

¹⁰¹More specifically, Bartlett and Kendall (1946) show that $\delta_S \sim N(-\frac{1}{N_S} - \frac{1}{3N_S^2}, \frac{2}{N_S - 1})$. Multiplying gives $\sqrt{\frac{N_S - 1}{2}} \delta_S \sim N(-\frac{\sqrt{N_S - 1}}{\sqrt{2}N_S} - \frac{\sqrt{N_S - 1}}{3\sqrt{2}N_S^2}, 1)$, which has a mean of approximately zero for large N_S .

¹⁰²For the purposes of determining geographic adjacency, islands with only a single municipality on them are treated as being connected to the closest municipality on the mainland (i.e., Hokkaidō, Honshū, Shikoku, or Kyūshū) if it is within 50 km. Using this definition, almost all observed mergers are geographically contiguous. More specifically, not counting island municipalities there are 13 observed mergers that are not geographically contiguous. This is usually because one of the participants dropped out late in the merger process.

defined using county definitions from the Meiji era.¹⁰³ For some larger observed coalitions S , the number of such subsets is still very large. In cases where there are more than 1000 potential subsets $S' \subset S$ for a given $S \in \pi$, use only the following:

1. Subset S' is a singleton.
2. The remainder $S \setminus S'$ is a singleton (i.e., subsets that involve leaving out only one municipality).
3. A random sample of other subsets such that the number of subsets examined totals 1000.

The following conditions are necessary for π to be in the stable set $\Pi^*(\epsilon(\bar{\omega}_M, s_M) | X, \beta)$:

$$\begin{aligned} \forall S' \in \mathcal{S}^c(\pi) \quad & \text{either } v_{mm}(\beta) + \epsilon_m(\bar{\omega}_M, s_M) > v_{mS'}(\beta) + \epsilon_{S'}(\bar{\omega}_M, s_M) \\ & \text{OR } v_{m'm'}(\beta) + \epsilon_{m'}(\bar{\omega}_M, s_M) > v_{m'S'}(\beta) + \epsilon_{S'}(\bar{\omega}_M, s_M), \end{aligned} \quad (49)$$

where $S' = \{m, m'\}$ is the potential merger of two singletons, and

$$\begin{aligned} \forall S' \in \mathcal{S}^r(\pi) \quad & \exists m \in S' \\ & \text{s.t. } v_{mS}(\beta) + \epsilon_S(\bar{\omega}_M, s_M) > v_{mS'}(\beta) + \epsilon_{S'}(\bar{\omega}_M, s_M), \end{aligned} \quad (50)$$

where S is the coalition that m is a member of π . As $S' \subset S$, this condition can be expressed in a computationally simpler form:

$$\forall S' \in \mathcal{S}^r(\pi) \quad \max_{m \in S} (v_{mS}(\beta) - v_{mS'}(\beta)) > \epsilon_{S'}(\bar{\omega}_M, s_M) - \epsilon_S(\bar{\omega}_M, s_M). \quad (51)$$

Here \mathcal{S}^c considers only singletons in π , and \mathcal{S}^r considers only the nonsingletons in π . Equations (20) and (47) show that the calculation of ϵ for the coalitions involved in these two sets of potential deviations depends on the $\bar{\omega}$ and s of disjoint sets of municipalities. This is because checking the necessary conditions for the coarsenings requires examining only those municipalities that remained as singletons, while checking the necessary conditions for refinements requires checking only those municipalities that participated in mergers. Furthermore, within the latter set, each actually observed merger can be checked separately. That is,

$$\begin{aligned} h^{**}(\pi | \beta, X) = & \min h(\bar{\omega}_M, s_M | X) \quad \text{s.t. (49) is satisfied} \\ & + \sum_S \min h(\bar{\omega}_M, s_M | X) \quad \text{s.t. (50) is satisfied for } S, \end{aligned} \quad (52)$$

¹⁰³In particular, the county boundaries used are from 1878 for eastern Japan and from 1896 for western Japan. Counties are statistical divisions and have not had any political function since the 1920s. Counties in Tokyo and Nagano prefectures are anomalously large; thus, in those prefectures only, the restriction is to one and two counties, respectively, rather than three. Two actual mergers violate the restriction on number of counties that is imposed: one size 12 merger in Shizuoka and one size 11 merger in Niigata. This represents 0.3% of all observed mergers.

where conditions (49) and (50) depend on π , β , and X .

Next, note that for a given $\hat{\beta}$, the problem of calculating h^{**} can be represented as a minimization problem where the variables are $(\bar{\omega}_m - \bar{\omega}_{m'})^2$ for adjacent m and m' , and s_m^2 for all m . The constraints in (49) and (50) are linear in these variables, and the objective function h is a quadratic of them. Thus, the problem of computing h^{**} can be expressed as a quadratic program or as a sum of the solutions to several quadratic programs as shown in (52). This latter form is computationally feasible and can be solved quickly using commercially available quadratic programming libraries such as CPLEX.

E.3 Second type of moment inequality

To see how (26) can be used to create a moment inequality, first let $F_Q(\beta)$ be the distribution of the number of municipalities that would have participated in mergers during the 1999–2010 period if the government had not implemented any new merger promotion policies. By assumption (and after making an appropriate adjustment for the fact that the 1979–1998 period is longer than the 1999–2010 period), $F_Q(\beta)$ is stochastically dominated by $F_Q^{79}(\beta)$, the distribution of the number of municipalities that participated in mergers during the 1979–1999 period. The F_Q is difficult to calculate directly: not only is the true equilibrium selection rule unknown, but, as discussed at the beginning of Section 3.1, the precise membership of \mathcal{S} is both unknown and likely very large. Thus, instead consider a bound μ_Q^* , such that

$$\begin{aligned} E_Q g_2(Q^{79}, \beta^0 | X) &= \mu_Q^{79} - \mu_Q^*(\beta^0 | X) \\ &\geq \mu_Q^{79} - \mu_Q, \end{aligned} \tag{53}$$

which is greater than zero because the expected number of mergers under no policy change has been assumed to not be greater than the number of mergers that occurred before the merger period.

This bound μ_Q^* can be computed by examining only size 2 mergers. Specifically, if $S = \{m, m'\}$ and $u_{mS} > u_{mm}$, $u_{m'S} > u_{m'm'}$, then any stable partition must have at least one of m and m' participating in a merger, because S is a blocking coalition for all other partitions. Let \mathcal{S}_a be the set of size 2 mergers where the municipalities are geographically adjacent and both municipalities prefer the merger to remaining as a singleton. This set can then be used to construct an easily computable minimal number of municipalities that must be involved in mergers. Consider the following variable, which is random because the membership of \mathcal{S}_a depends on the draw of $\bar{\omega}_M$ and s_M :

$$Q^* = \underset{Q \subset M}{\operatorname{argmin}} \#Q \quad \text{s.t.} \quad \forall S \in \mathcal{S}_a, S \cap Q \neq \emptyset. \tag{54}$$

That is, Q^* is a minimal hitting set for \mathcal{S}_a : for each potential geographically contiguous size 2 merger where both participants prefer the merger relative to not merging at all, at least one of those municipalities is in Q^* . For a given random draw of shocks, Q^* can be computed via a linear program. Let $F_Q^*(\beta)$ be the distribution of $\#Q^*$. The F_Q stochastically dominates F_Q^* , because \mathcal{S}_a is a subset of all mergers whose participants prefer the merger to remaining as a singleton, and thus any stable partition must include at least

$\#Q^*$ municipalities participating in mergers regardless of equilibrium selection rule. Examining potential pairwise mergers thus gives a lower bound on the total number of municipalities participating in mergers of any kind.

The calculation of Q^* involves a high dimensional integral because there are two idiosyncratic shocks for each municipality and many municipalities in a prefecture. However, as only size 2 mergers between geographically adjacent municipalities are considered and very few of these mergers actually occur, the interactions between municipalities that are far away are very weak. Thus, in a large prefecture such as Hokkaidō, with 210 municipalities, simulation error in the southern portion of the prefecture will have little effect on mergers of northern municipalities, and thus the law of large numbers will apply. In addition, there are multiple prefectures, and there are assumed to be no interactions between them.¹⁰⁴

E.4 Identified set and confidence sets

The data consist of 47 prefectures, which are treated as independent coalition formation games. As in Table 1, prefectures are classified as metropolitan, mixed, and rural, depending on the percentage of municipalities with a population of less than 10,000. Let these sets of prefectures be J^{metro} , J^{mixed} , and J^{rural} , respectively. Let $\bar{g}_1^{\text{metro}}(\beta)$ be the sample moment of g_1 with prefectures J^{metro} ,

$$\bar{g}_1^{\text{metro}}(\beta) = \frac{1}{\#J^{\text{metro}}} \sum_{j \in J^{\text{metro}}} g_1(\pi_j^0, \beta | X), \quad (55)$$

where π_j^0 is the actually observed partition in prefecture j .¹⁰⁵ Construct \bar{g}_2 similarly: for example, \bar{g}_2^{mixed} would be

$$\bar{g}_2^{\text{mixed}}(\beta) = \frac{1}{\#J^{\text{mixed}}} \sum_{j \in J^{\text{mixed}}} g_2(Q_j^{79}, \beta | X), \quad (56)$$

where Q_j^{79} is the actually observed number of municipalities participating in mergers in prefecture j during the 1979–1998 period. There are thus three sample moments calculated from g_1 (metro, mixed, and rural), and another three calculated in the same way for g_2 . On the other hand, g_3 is only calculated for metro (it will not bind for any value of β for mixed or rural, due to the large number of mergers in these types of prefectures),

$$\bar{g}_3^{\text{metro}}(\beta) = \frac{1}{\#J^{\text{metro}}} \sum_{j \in J^{\text{metro}}} g_3(Q_j^{99}, \beta | X), \quad (57)$$

where Q_j^{99} is the actually observed number of municipalities participating in mergers in prefecture j during the 1999–2010 period.

¹⁰⁴An adjustment is made to both g_2 and g_3 to take into account that mergers are infrequent, and thus are not well approximated by a normal distribution: quantiles of the actual distribution of the lower bound on the number of mergers are estimated, and these are then used to perform a normalization.

¹⁰⁵In the actual computations, the weights on individual prefectures differ to account for the differing number of municipalities in each prefecture.

Due to the distributional assumption already made regarding $\bar{\omega}$ and s , the first six of these terms are uncorrelated. However, \bar{g}_3^{metro} could well be correlated with \bar{g}_1^{metro} , as π^0 affects both of these sample moments. Similarly, \bar{g}_4 could plausibly be correlated with other moments. These correlations are estimated using the subsample approach given in Andrews and Guggenberger (2009), with critical values of $\mathbf{T}(\beta)$ computed accordingly. The identified set is the set of $\hat{\beta}$ such that

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \mathbf{T}(\beta). \quad (58)$$

Optimization is performed via Nelder–Mead, using the base implementation in R. For the second and third types of moment inequalities, 1000 simulation draws are used to calculate the moment. The optimization algorithm used finds only a local minimum, and thus there is the possibility that this differs from the global minimum. The calculation of the confidence set for each parameter, however, involves rerunning the optimization routine with the relevant parameter fixed at a certain value. On average, about 100 values were used: some of these values were very close to the estimated parameter, while others were farther away. These additional starting points do not result in any objective function values lower than the one at the reported optimum.

Following convention, the 95% confidence set will be

$$\{\beta | \mathbf{T}(\beta) < \mathbf{T}(\hat{\beta}) + \mathbf{T}_{0.95}(\hat{\beta})\}, \quad (59)$$

where $\mathbf{T}_{0.95}(\hat{\beta})$ is the 0.95 quantile of the distribution of the test statistic under the hypothesis that $\beta = \hat{\beta}$. Constructing a 95% confidence set for β is simplified because assumptions regarding the distribution and most of the correlation structure of the error terms have already been necessary to develop the model.

The coverage probability of this confidence set is checked via the Monte Carlo method. One-hundred simulations of the municipal merger process are performed using the estimated parameters as the true parameters. The 95% confidence set has a coverage probability of 1.00 and an 80% confidence set has a coverage probability of 0.98.¹⁰⁶

E.5 Example with two municipalities per prefecture

The moment inequality based on g_1 has not previously appeared in the literature. A simple example illustrates the identified set produced: notably, it will not exclude $\beta = 0$.

Let J be a set of prefectures each containing only two municipalities, A_j and B_j , with a potential merger $S_j = \{A_j, B_j\}$. For simplicity, let there be only a single idiosyncratic shock ϵ_j in each prefecture,

$$u_{A_j A_j} = u_{B_j B_j} = \beta, \quad (60)$$

$$u_{A_j S_j} = u_{B_j S_j} = 2\beta + \epsilon_j, \quad (61)$$

and make the distributional assumption $\epsilon_j \sim N(0, 1)$. Here, setting the variance to 1 normalizes the scale for β . Let π^0 be the observed partition: the only options for prefecture

¹⁰⁶In these simulations, an average of 67 municipalities participated in mergers under the original government transfer policies. This corresponds to about 2% of the 3255 original municipalities.

j are the singletons $\{A_j\}$ and $\{B_j\}$ or the merger $\{A_j, B_j\}$. Define the stability restriction

$$R(\epsilon, \pi|\beta): \quad \forall j \in J, \quad \epsilon_j \leq -\beta \quad \text{if } \{A_j\}, \{B_j\} \in \pi, \quad (62)$$

$$\epsilon_j \geq -\beta \quad \text{if } \{A_j, B_j\} \in \pi. \quad (63)$$

Now choose $h(\epsilon) = \sum_{j \in J} \epsilon_j^2$. Thus, $E(h) = |J|$. Then define

$$\epsilon^*(\beta) = \underset{\epsilon}{\operatorname{argmin}} h(\epsilon) \quad \text{s.t.} \quad R(\epsilon, \pi^0|\beta). \quad (64)$$

That is, $\epsilon^*(\beta)$ is the vector of idiosyncratic shocks that generate the least extreme value from h while still rationalizing π^0 . Let that value of h be

$$h^*(\beta) = \min_{\epsilon} h(\epsilon) \quad \text{s.t.} \quad R(\epsilon, \pi^0|\beta). \quad (65)$$

Let ϵ^0 be the actual ϵ s that were drawn and resulted in π^0 being observed. Let β^0 be the true value of β . Then

$$h^*(\beta^0) \leq h(\epsilon^0) \quad (66)$$

because ϵ^0 being drawn resulted in partition π^0 occurring, and thus $R(\epsilon^0, \pi^0|\beta^0)$ must be satisfied. If the inequality is always satisfied, it is satisfied in expectation,

$$E(h(\epsilon)) - E(h^*(\beta^0)) \geq 0, \quad (67)$$

where $E(h^*(\beta^0))$ indicates the expected value of h^* for a partition generated from a random draw of ϵ . In this particular example, for any draw of ϵ , there will only be one stable partition, but neither this uniqueness nor any particular assumptions regarding an equilibrium selection rule is required for the above inequalities to hold.

The expected fraction of prefectures with a merger is $\Phi(\beta^0)$. The expected value of $h^*(\beta)$ is

$$\Phi(\beta^0) \min(0, \beta)^2 + (1 - \Phi(\beta^0)) \max(0, \beta)^2. \quad (68)$$

Using only the above moment inequality for identification, we will have the identified set

$$\{\beta | 1 - \Phi(\beta^0) \min(0, \beta)^2 + (1 - \Phi(\beta^0)) \max(0, \beta)^2 \geq 0\}, \quad (69)$$

which corresponds to the interval $[\frac{-1}{\sqrt{\Phi(\beta^0)}}, \frac{1}{\sqrt{1-\Phi(\beta^0)}}]$. This interval contains zero, which is a general property of this type of moment: the “entirely idiosyncratic” null hypothesis will never be rejected. To reject $\hat{\beta} = 0$, some additional moments of some other type must be used. In the estimator in the main paper, these correspond to the moments comparing the expected number of mergers if the government had not changed any transfer policies to the actual number of mergers observed during the period in which the old transfer policies were in effect. Considering the specification used in the main

analysis, at $\hat{\beta} = 0$, there would have been a large number of random mergers, and thus this null hypothesis is easy to reject using these other moments.¹⁰⁷

APPENDIX F: OPTIMAL NATIONAL TRANSFER POLICY

F.1 Theory

The national government's objective function depends on ϵ , which it does not know. Expressed as an expectation, its objective is

$$\begin{aligned} E[W(T)] &= \Pr(\pi = \{S\}|T)(w_m + w_{m'})(v_{mS}(T_S) + E[\epsilon_S|\pi = \{S\}, T]) - bT_S \\ &\quad + (1 - \Pr(\pi = \{S\}|T))(w_m(v_{mm}(T_m) + E[\epsilon_m|\pi = \{m, m'\}, T]) \\ &\quad + w_{m'}(v_{m'm'}(T_{m'}) + E[\epsilon_{m'}|\pi = \{m, m'\}, T]) - b(T_m + T_{m'})). \end{aligned} \quad (74)$$

For a decrease in T_m , the marginal merger that would occur is the one where $v_{mm}(T_m) + \epsilon_m = v_{mS}(T_S) + \epsilon_S$; thus, the partial derivatives of (74) simplify to

$$\begin{aligned} \frac{\partial E[W(T)]}{\partial T_m} &= \frac{\partial \Pr(\{S\}|T)}{\partial T_m} (b(T_m + T_{m'} - T_S) + w_{m'}e') \\ &\quad + (1 - \Pr(\{S\}|T)) \left(w_m \frac{\partial v_{mm}(T_m)}{\partial T_m} - b \right), \end{aligned} \quad (75)$$

¹⁰⁷The above estimator may appear to be somewhat similar to other estimation approaches, such as maximizing the probability that the observed partition is stable. Estimators in this latter set, however, are not in general consistent, and will thus not necessarily be inside the identified set based on the moment inequality used above. To take an extreme example, suppose that there are K municipalities in each prefecture and that the only possible mergers are $\{m_1, m_2\}, \{m_1, m_2, m_3\}, \dots, \{m_1, m_2, \dots, m_K\}$. Preferences are determined by

$$u_{m_k S} = \frac{\#S}{K} + \sqrt{K}\epsilon_j, \quad S \neq \{m_k\}, \quad (70)$$

$$u_{m_k m_k} = k - \frac{K-1}{2}, \quad (71)$$

where $\#S$ is the number of municipalities in S , and j indexes prefectures. Thus there is again only one idiosyncratic shock per prefecture and only one stable partition: if $\epsilon = 0$, for example, municipalities up to $K/2$ will merge. The probability of any given merger being stable is thus small if K is large. Now consider the more general model

$$u_{m_k S} = (1 - \beta) \left(\frac{\#S}{K} + K^{1/3}\epsilon_j \right) + \beta \left(k - \frac{K+10}{2} + \epsilon_j^2 \right), \quad S \neq \{m_k\}, \quad (72)$$

$$u_{m_k m_k} = k - \frac{K-1}{2}. \quad (73)$$

When $\beta = 1$, then, most of the time no mergers are stable, but when there is a very extreme ϵ , then all mergers are stable (indifferent municipalities prevent deviations in the definition of stability used in this paper). Now suppose that the true value of β is $\beta^0 = 0$, and consider a pseudo-likelihood estimator that maximizes the probability that the observed partition is stable. Then for sufficiently large K , $\beta^{\text{PMLE}} = 1$, because all mergers are stable when any merger is stable, unlike the situation at β^0 . The identified set using the technique outlined above, however, would exclude $\hat{\beta} = 1$, as this would involve ϵ^* that are very extreme, and the moment inequality would be violated. Thus, the (inconsistent) maximizer of the pseudo-likelihood function based only on stability is not always contained in the identified set.

$$\begin{aligned} \frac{\partial E[W(T)]}{\partial T_S} &= \frac{\partial \Pr(\{S\}|T)}{\partial T_S} (b(T_m + T_{m'} - T_S) + \eta w_{m'} e' + (1 - \eta) w_m e) \\ &+ \Pr(\{S\}|T) \left((w_m + w_{m'}) \frac{\partial v_{mS}(T_S)}{\partial T_S} - b \right). \end{aligned} \quad (76)$$

There is also a third equation, (75'), which is identical to (75) except with m and m' flipped.

These derivatives show that when there is the potential for mergers, T^{FB} is no longer the optimal transfer policy. The fixed-boundary optimum maximizes the objective from (9), and thus at T^{FB} , the second term in (75) and (76) is zero.¹⁰⁸ Because there are efficiencies of scale, however, $T_m^{\text{FB}} + T_{m'}^{\text{FB}} > T_S^{\text{FB}}$. The first terms of (75) and (76) are thus negative and positive, respectively, at the optimal fixed-boundary transfer policy. This suggests that the transfer policy that maximizes (74) will have a higher T_S and lower T_m than the fixed-boundary optimum.

To characterize the optimal policy, first note that $\Pr(\{S\}|T)$ depends only on the differences $v_{mS} - v_{mm}$ and $v_{m'S} - v_{m'm'}$. Thus, $\frac{\partial \Pr(\{S\}|T)}{\partial v_{mS}} = -\frac{\partial \Pr(\{S\}|T)}{\partial v_{mm}}$ and likewise for m' . For a change in transfers, then,

$$\frac{\partial \Pr(\{S\}|T)}{\partial T_S} \bigg/ \frac{\partial v_{mS}}{\partial T_S} = -\frac{\partial \Pr(\{S\}|T)}{\partial T_m} \bigg/ \frac{\partial v_{mm}}{\partial T_m} - \frac{\partial \Pr(\{S\}|T)}{\partial T_{m'}} \bigg/ \frac{\partial v_{m'm'}}{\partial T_{m'}}.$$

The first right-hand side term of (76) is thus equal to the sum of the first terms of (75) and (75'). As these equations equal zero at the optimal T , the second terms must also equal: after rearrangement this gives (33). An alternative form of (33) expresses $\Pr(\{S\}|T)$ in terms of transfers:

$$\Pr(\{S\}|T^*) = \frac{(T_m^{\text{FB}} - T_m^*) + (T_{m'}^{\text{FB}} - T_{m'}^*)}{(T_m^{\text{FB}} - T_m^*) + (T_{m'}^{\text{FB}} - T_{m'}^*) + (T_S^* - T_S^{\text{FB}})}. \quad (77)$$

Furthermore, $\frac{\partial v_{mm}}{\partial T_m} = \frac{\beta_0 + \beta_1}{T_m - Y_m + \beta_3 c(X_m)}$ by substituting (5) into (7) and differentiating. Combining with (10) leads to

$$\frac{1}{\frac{\partial v_{mm}}{\partial T_m} \bigg|_{T_m^{\text{FB}}}} - \frac{1}{\frac{\partial v_{mm}}{\partial T_m}} = \frac{1}{\beta_0 + \beta_1} (T_m^{\text{FB}} - T_m^*). \quad (78)$$

At the fixed-boundary optimum, it will also be the case that $\frac{\partial v_{mm}}{\partial T_m \big|_{T_m^{\text{FB}}}} = \frac{b}{w}$. Thus,

$$\left(w_m \frac{\partial v_{mm}(T_m)}{\partial T_m} - b \right) = b \frac{\partial v_{mm}(T_m)}{\partial T_m} \frac{1}{\beta_0 + \beta_1} (T_m^{\text{FB}} - T_m^*). \quad (79)$$

Equation (34) can then be obtained by substituting (77) and (79) into the sum of (75) and (75'). Finally, dividing (75) by $\frac{\partial \Pr(\{S\}|T)}{\partial T_m}$ and (75') by $\frac{\partial \Pr(\{S\}|T)}{\partial T_{m'}}$ leads to an initial term that is the same in both cases. The remaining terms thus must be equal, yielding (35).

¹⁰⁸The second term is exactly the first order condition from (9).

F.2 Empirical analysis

One potential explanation for the weakness of the actual incentives is that the national government may have faced technical restrictions on the type of transfers it could offer. Specifically, (33) shows that the optimal transfers depend on the probability that the municipalities in question will merge. However, the optimal fixed-boundary transfers of (9)–(11) depend only on municipal characteristics relevant to public good provision and taxation. Thus, the optimal incentive policy requires incorporating new types of data into the transfer function, related to the attractiveness of potential mergers. Specifically, in the model in Section 1, the distance term ℓ does not appear in (9)–(11) but does enter into (33) through the probability of a merger.

To provide the merger incentives described in Section 2, the Japanese government modified its existing transfer policy. This policy is quite complex: the *Chihō Kōfuzei Seido Kaisetsu* explaining the 2009 formula consists of 600 pages of legal text, 460 pages of formulae, and 240 pages of reference values. However, in line with (11), essentially all of these details are related to either the cost of public goods or the amount of taxes that can potentially be collected. The merger incentives were provided by modifying these formulae, with the modifications also being functions only of these characteristics. For example, two very distant municipalities that merged received the same transfers as two very close municipalities, despite the fact that if the probabilities of these mergers happening were different, a different set of incentives would have been optimal.

Figure 20 compares the actual incentives to a set of “restricted” optimal incentives, where the national government cannot use distance information as part of its transfer policy.¹⁰⁹ The blue points are pairs where the optimal incentive has the form given in (33)–(35). The pink points are pairs where forcing the municipalities to merge gives a better outcome than this interior solution.¹¹⁰ The slope for the points involving an interior solution is close to 1, but the actual incentives are still lower than the optimal incentives: on average, the actual incentive for a pair of municipalities in this group is about 60% of the computed optimal incentive. Thus, an inability to include distance information in the incentive policy does not appear to explain the relative weakness of the incentives offered.

Calculating the per capita value of the incentives in question suggests another potential explanation. As shown in Figure 7, reported cost \tilde{c}^0 declined dramatically in per capita terms for mergers of municipalities with small populations: for pairs where both municipalities had a population between 2500 and 5000, the average decline in \tilde{c}^0 after a merger is ¥64,300 per capita. This corresponds to more than 25% of average transfers to the municipalities in question, and is roughly half of the estimated cost per capita for the larger municipalities shown in Figure 7. A merger policy with an incentive \mathcal{I} close to 1

¹⁰⁹Specifically, suppose that for each pair of municipalities the national government must choose a transfer policy based only on the overall distribution of distances ℓ across all pairs, rather than the specific distance changes $\ell_S - \ell_m$ and $\ell_S - \ell_{m'}$ for the pair in question. The points in the figure are calculated following the numerator of (38).

¹¹⁰These are cases where the horizontal externality e is large relative to the information gains from allowing the municipalities to choose. The pink points in the figure are located where the probability of a merger becomes numerically close to 1, but any location further to the right would be equally correct.

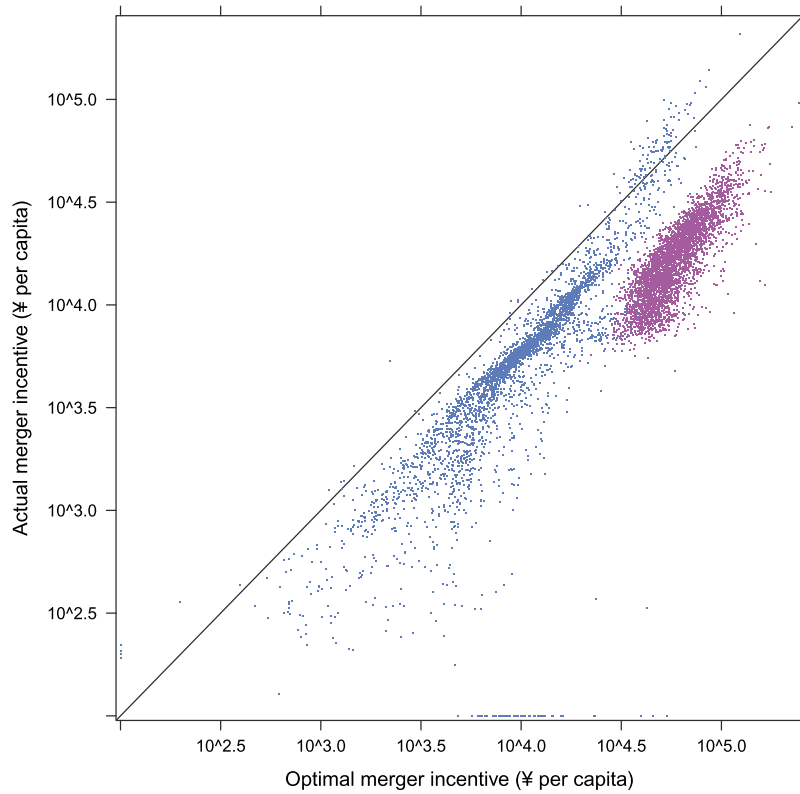


FIGURE 20. Actual and optimal no-distance incentives (adjacent pairs).

would thus lead to a situation where, after the mergers were completed, amalgamated municipalities could be receiving 25% more forever than municipalities with identical observable characteristics that were not the result of an amalgamation. Discrepancies in the transfer system of this magnitude and without (current) justification would likely be difficult to maintain, and post-merger the national government has an incentive to revert to T^{FB} .

The actual incentive policy involved merging municipalities receiving special treatment for only 10 years and then being treated the same as nonamalgamating municipalities. This is a length of time where most of the bureaucrats implementing the policies, as well as many of the local and national politicians, would likely remain in place. The national government might be able to commit to offering special treatment to merging municipalities for an initial period, while being unable to commit forever.

One alternative to permanent transfer differences would be for the national government to structure the merger incentives as large upfront payments. At the low interest rates prevailing in Japan, however, these lump sum payments would need to be very large: at 2% interest, the appropriate per capita lump sum payment for the small municipalities discussed above would be about 75% of GDP per capita. Payments of this size might not be politically feasible and could themselves suffer from a risk of future expropriation.

Figures 10 and 11 show that the form of the incentives offered roughly matches the model in Section 1. The strength of these incentives, however, appears to be too low given the estimated $\hat{\beta}$, and this discrepancy cannot be explained by the importance of the idiosyncratic shock. Given the size of the transfers in question, though, it does appear plausible that the government may have faced a commitment problem.

REFERENCES

- Andrews, D. W. K. and P. Guggenberger (2009), "Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities." *Econometric Theory*, 25 (3), 669–709. [11]
- Banks, J. and J. Duggan (2005), "Probabilistic voting in the spatial model of elections: The theory of office-motivated candidates." In *Social Choice and Strategic Decisions* (D. Austen-Smith and J. Duggan, eds.), Studies in Choice and Welfare, 15–56, Springer, Berlin. [1]
- Bartlett, M. S. and D. G. Kendall (1946), "The statistical analysis of variance-heterogeneity and the logarithmic transformation." *Supplement to the Journal of the Royal Statistical Society*, 8 (1), 128–138. [7]
- Kawaura, A. (2010), "Self-serving mayors and local government consolidations in Japan." Working Paper 10-14, Department of Economics, University of Hawaii. [4]
- Scheffé, H. (1959), *The Analysis of Variance*. Wiley, New York. [6]

Submitted March, 2014. Final version accepted August, 2014.