Inefficiency and Self-Determination: Simulation-based evidence from Meiji Japan*

Eric Weese
University of Tokyo†

Masayoshi Hayashi
University of Tokyo

Masashi Nishikawa
Aoyama Gakuin University

October 2020

Abstract

We consider a popular theoretical model of jurisdiction formation where there is a tradeoff between efficiencies of scale and heterogeneity. We develop a maximum score estimation technique to determine the parameters of a central planner's payoff function given the way they partitioned a territory into jurisdictions. We apply this technique to historical data on a set of centralized boundary changes in Japan: walking distance appears to largely determine jurisdiction boundaries, with only small effects from land type and historical feudal ruler, and no effect of religion.

We then assume that local villages shared these preference parameters emphasizing walking distance, and use binary integer programming to calculate core partitions for a decentralized coalition formation game based on this model. Core partitions exist with very high probability. In a counterfactual world in which there are no between-village income differences, these core partitions are extremely close to the partition that would be chosen by a utilitarian central planner. When actual cross-village income differences are used, however, sorting on income results in mergers that are both smaller and geographically discontiguous.

---

*Some ideas in this paper were present in the job market paper version of Weese [2015], but were removed during the publication process. We thank Michihiro Kandori and David Schonholzer for helpful comments. A previous version of this paper that used a moment inequalities estimator benefitted from comments by Steve Berry, Raquel Fernandez, Jeremy Fox, Matias Iaryczower, Hide Ichimura, Yuichi Kitamura, John Londregan, Tsunetoshi Mizoguchi, Masaki Nakabayashi, Kei Okunuki, Larry Samuelson, Motohiro Sato, Junichi Suzuki, Shlomo Weber, Noam Yuchtman, and seminar participants at various seminars. An even earlier version of the estimator in this paper was originally presented as "An Errors-in-Variables Model for Graph Outcome Data" at the 2014 Asian Meeting of the Econometric Society. Julien Clancy provided excellent research assistance with an earlier version of the model. Portions of this research were conducted while visiting Osaka University, Kobe University, and the National Graduate Institute for Policy Studies. Funding was provided by the Japan Society for the Promotion of Science. Initial computations were performed using the FAS High Performance Computing system at Yale University, with final computations using equipment at the Supercomputing Division of the University of Tokyo.

†Corresponding author. Email: weese@iss.u-tokyo.ac.jp.

Does exercise of the right to self-determination result in an efficient arrangement of political boundaries? There is substantial theoretical interest in this issue both inside and outside of economics. Policy relevance is suggested by past votes on independence in Scotland, Catalonia, Quebec, and elsewhere, as well as the recent decision by the United Kingdom to leave the European Union. Empirical results regarding the efficiency of jurisdiction formation, however, are very limited.

This paper considers a historical set of municipal mergers in Gifu, Japan, that were decided by a central planner in the late 19th century. Pre-defined subunits corresponding to feudal villages were merged to create modern municipalities. We use an Alesina and Spolaore [1997] type model, where there is a tradeoff between efficiencies of scale and heterogeneity, and estimate the relative importance of different sorts of heterogeneity to the planner via a variation on the Fox [2007] pairwise maximum score estimator. We then estimate parameters regarding the cost of public good provision via method of moments, combined with calibration based on official government reports and observed spending. We assume that the central planner acted as a Benthamite aggregator of the preferences of the villages, and thus the parameters just estimated also give us the preferences of villages over types of heterogeneity and their importance relative to efficiencies of scale. We find small effects of land type and historical feudal lord, and a much larger effect for geographic distance variables, in particular walking distance.

A counterfactual case is then considered, where villages were allowed to choose how to arrange themselves into jurisdictions in a coalition formation game. The game is without transfers: this is common in political economy [Acemoglu 2003] and matches later actually decentralized mergers [Weese 2015]. In addition to heterogeneity and efficiencies of scale, however, each individual village also cares about the degree to which it – rather than its merger partners – will end up paying the taxes that fund the municipality. The core is used as the solution concept in this decentralized jurisdiction formation game, and results are obtained via simulation. The core partitions in this coalition formation game are automatically Pareto optimal. Payoffs for players, however, are quasi-linear with respect to money, and we consider inefficiency from a utilitarian perspective. We perform simulations with the observed data, as well as considering setups with changed village characteristics or the addition of inter-governmental subsidies.

Our simulations show that decentralized coalition formation results in substantial inefficiency; however, in the case where the players do not differ in per capita tax base core partitions are very close to the central planner's optimal partition. Although from a theoretical perspective inefficiency could arise from a variety of sources, our simulations suggest that the empirically relevant source of inefficiency is the sorting of Farrell and Scotchmer [1988]:

each player would prefer to merge with others that are richer than themselves, while avoiding poorer players. Stratification on income leads to a shortage of mutually acceptable partners, resulting in coalitions that are geographically discontiguous and substantially smaller than those in the central planner's optimal partition. Depending on the exact scenario, back of the envelope calculations suggest that inefficiency could be equivalent to 2.25% of GDP.

This paper makes two methodological contributions. First, we present a method of computing core partitions in coalition formation games that have an Aziz, Brandt, and Harrenstein [2014] fractional hedonic form. Our overall approach follows a method used in the "roommates problem" by Chung [2000], where successive myopic deviations eventually lead to a stable partition. A major issue is that the coalition formation game in Gifu consists of approximately 1000 players, and unlike a standard (pairwise) roommates problem, enumeration of potential coalitions is computationally infeasible. To avoid having to enumerate all potential coalitions, we express each myopic deviation as the solution to a binary integer program. Using this technique, we are able to compute core partitions in a few hours using standard equipment.

Finding core partitions of coalition formation games is known to be an NP-hard problem in general. There is substantial previous applied research using coalition formation games with transfers (e.g. Diermeier, Eraslan, and Merlo 2003) or games with only pairwise coalitions (e.g. Gordon and Knight [2009]). To our knowledge, our results are the first to show solutions for large instances of a coalition formation game without transfers. Although fractional hedonic games have attracted theoretical interest in economics, to date they have not been used in applied research. Our results suggest that many real world instances of fractional hedonic games may be much more easily solvable than theoretical worst case instances.

Second, we present a method of estimating the preferences of a decision maker based on the way they have partitioned a territory. This method is based on the pairwise maximum score estimator of Fox [2007], except we use a relaxed version of the Manski [1975] rank ordering property that only considers alternative partitions that can be created via permutations of player labels. Our method does not rely on any particular structure for the decision maker's deterministic payoffs, and thus could potentially be used to estimate preferences even in cases where payoffs are clearly non-linear, such as in the process of creating gerrymandered districts.

An obvious objection to our approach is that we obtain parameter estimates by assuming that the importance villages place on different sorts of heterogeneity (and the importance of that heterogeneity relative to the cost of running the jurisdiction) is shared by the planners in the national government. If this assumption is incorrect, then the portions of this paper that serve as a program evaluation of a particular set of municipal mergers in 19th century Japan are invalid. The main objective of our analysis, however, is more general. Our model introduces a simple distinction between "horizontal" characteristics such as religion, where similar merger partners are preferred, and "vertical" characteristics such as income, where richer merger partners are preferred. We show that there is an approximate "invisible hand" type result for decentralized jurisdiction formation when players differ only in horizontal characteristics, but not when they also differ in vertical characteristics.

Simulations are required to reach these conclusions because the models we are most interested in have no closed form solutions [Gregorini 2009]. To perform these simulations we must somehow choose plausible parameter values and distributions of characteristics. We base our parameters on a particular instance of boundary changes in Japan because the institutional setup there appears to match a fractional hedonic model and we can thus make it computationally tractable. We use further Monte Carlo simulations of a dramatically simplified game to show that our key results require only that the cost of providing public goods must be increasing in population. The precise degree of congestibility, however, as well as the precise distribution of heterogeneity among players, are not critical to our results. If we have dramatically misunderstood the actual process of mergers in our Japanese data, the general results from our simulations still hold for a hypothetical environment in which there is substantial congestibility of public goods as well as some sort of tradeoff between heterogeneity and efficiencies of scale. We believe that this is an environment of interest because it corresponds to how public goods appear to be provided in many situations.

1.1 Related Literature

This paper is inspired by Desmet et al. [2011], who consider European national boundaries as a coalition formation game. Desmet et al. use exhaustive enumeration for their simulations, and are forced to cut their dataset in half (from 24 down to 11 players) in order to make this computationally feasible. A number of other papers consider similar political coalition formation games. These include Brasington [1999], Gordon and Knight [2009], and Weese [2015]. These papers do not focus on simulating the outcome of the coalition formation game that they study, and the simulations that are performed do not face computational constraints because the object of interest is pairwise mergers.

From a broader perspective, this paper presents a sorting model based on a "strong Tiebout equilibrium" [Greenberg and Weber 1986], one where players arrange themselves into core-stable coalitions based on their income as well as other characteristics. Many Tiebout models involve each player caring about the "average" type of player in their coalition, and a fractional hedonic game provides a natural way of modelling this "average" based payoff. We provide an algorithm to compute the solution to large fractional hedonic games, and we show that empirically this solution exists and features substantial inefficiency. Although the precise model we consider is much simpler than those generally used in the Tiebout literature, our algorithm would also apply to fractional hedonic games where the payoff for players depended on many more variables. It thus may be empirically possible to use strong Tiebout equilibrium models to analyze problems related to public schooling or other local public goods.

Integer programming has been used extensively in the two-sided matching literature for finding chains of kidney donors [Roth, Sönmez, and Ünver 2007]. Our application of integer programming techniques is to a matching game where there is only one type of player: both our algorithm and the underlying theory differ substantially from the two-sided case. The previous literature on integer programming in political economy appears limited to Serafini [2012], who considers a seat allocation problem in the EU parliament. Previous uses in other fields of economics include Gomory [1994] and Pinar and Camci [2012], as well as Elomri et al. [2013] in supply chain management.

Maximum score estimation has been used on a wide variety of empirical topics, including network formation games (e.g. Fox [2018]). From one perspective, the problem we study is simpler than a network game because we have a single decision maker choosing a partition, and thus our estimation matches directly the classic discrete choice framework used in Fox [2007]. On the other hand, simple models of network formation involve pairwise idiosyncratic shocks that can be isolated by considering only one edge on the network, whereas moving a single player to a different coalition in a partition potentially involves losing (and gaining) many coalition partners. The pairwise structure of idiosyncratic shocks we use is closely related to the standard structure in network formation, but the "permutation rank ordering property" we base our objective function on appears to be new to the literature.

This paper appears to be the first quantitative study of municipal mergers in Meiji Japan. Previous work on more recent Japanese municipal mergers includes Hirota and Yunoue [2014], Miyazaki [2014], and Weese [2015]. The techniques used and results obtained in these papers differ substantially from those presented below. Hirota and Yunoue [2014] uses a logit framework to look at political determinants of mergers. Miyazaki [2014] uses data on municipal referenda. Weese [2015] considers recent Japanese data where the central government provides equalization payments to municipalities. The observed equalization payments in this recent data are extremely large (up to 25% of GDP per capita for the smallest municipalities), and counterfactuals where there are no such payments are thus so far out of sample that there would be computational difficulties with any such simulation, as well as theoretical difficulties in interpreting the results.

The remainder of this paper has the following structure: Section 2 presents the theoretical model, Section 3 introduces the data used, Section 4 describes the maximum score estimation approach, Section 5 presents the estimation results, Section 6 shows how we obtain estimates for two additional parameters that are otherwise not identified, and Section 7 covers the counterfactual simulations performed.

2 Model

In a hedonic game [Bogomolnaia and Jackson 2002; Dreze and Greenberg 1980], payoffs to players depend only on the identity of their coalition partners. A *fractional hedonic game* [Aziz, Brandt, and Harrenstein 2014] is a hedonic game where the payoff to player i from being a member of coalition S is $v_i(S) = \sum_{i' \in S} v_i(i')/|S|$, where $|S|$ is the number of players in S. That is, the payoff to a coalition is the average of pairwise payoffs $v_i(i')$.

Games of this sort have long been considered in economics: for example, if we set

$$\begin{aligned} v_i(i) &= y_i - \gamma_1 \\ v_i(i') &= y_{i'} \qquad\qquad \text{for } i' \neq i, \end{aligned} \tag{1}$$

then we have a Farrell and Scotchmer [1988] "partnership" game, where team members generate different incomes y but these are pooled and shared equally by the entire team after paying a fixed cost γ_1. If we instead assign each player a type and set

$$\begin{aligned} v_i(i) &= -\gamma_1 \\ v_i(i') &= 0 \qquad \text{if } i \text{ and } i' \neq i \text{ are the same type} \\ v_i(i') &= -1 \qquad \text{if } i \text{ and } i' \neq i \text{ are different types,} \end{aligned} \tag{2}$$

then we have a setup similar to that used by Alesina, Baqir, and Hoxby [2004] to describe the formation of school districts in the United States. Here players face a tradeoff between efficiencies of scale and a desire to share a district only with others that look like themselves.

In this paper we will describe heterogeneity between players as *vertical* if it is of Farrell and Scotchmer [1988] type, where it leads to all players having the same ranking over potential partners. In contrast, we will call heterogeneity *horizontal* when it leads to each player preferring to join with other players similar to themselves. Situations involving both vertical and horizontal heterogeneity are difficult to analyze theoretically [Gregorini 2009] but occur frequently in the real world. For example students partition into schools, with (almost) everyone preferring higher academic ability peers but often looking to be with others that share their own religious or political views. Civil servants and staff at universities often partition themselves into multiple unions (e.g. "service & maintenance", "clerical & technical", "security", and "police"), where vertical heterogeneity is important because of within-union wage compression and horizontal heterogeneity arises due to the natural differences in the bargaining objectives of different occupations. The most obvious application of fractional hedonic games in economics, however, is to political jurisdictions: these jurisdictions generally group together people that are close to one another (either geographically or in terms of other characteristics), but at the same time a shared pool of tax revenue results in wealthier individuals trying to escape and form more exclusive jurisdictions.

A particularly simple way to model a situation with both vertical and horizontal heterogeneity is to use a *weighted* fractional hedonic game, with the special choice that the weights assigned to players are their incomes y_i. The payoff to player i will then be $v_i(S) = \sum_{i' \in S} y_{i'} v_i(i') / \sum_{i' \in S} y_{i'}$, and we choose

$$\begin{aligned} v_i(i) &= -\gamma_1 - \gamma_2 \\ v_i(i') &= -\frac{y_i}{y_{i'}}\gamma_2 - y_i d(i,i'). \end{aligned} \tag{3}$$

Here γ_1 is the fixed cost of running a jurisdiction, γ_2 is a per player variable cost, and $d(i,i')$ is some sort of distance (geographic or otherwise) between player i and i'. In the geographic data presented in Section 3 it will always be the case that $d(i,i') \geq 0$, $d(i,i) = 0$, and the triangle inequality is satisfied, but these facts are not used anywhere below and for non-geographic variables and in Monte Carlo exercises we will frequently generate $d(i,i')$ entries that reflect "dissimilarity" rather than a true mathematical distance.

To see why this choice of payoffs makes sense, let $Y_S = \sum_{i' \in S} y_{i'}$. The payoff for player i then expands to

$$v_i(S) = -\frac{y_i}{Y_S}(\gamma_1 + \gamma_2 |S|) - y_i \sum_{i' \in S} \frac{y_{i'}}{Y_S} d(i,i'). \tag{4}$$

The first term here is player i's share of the cost of running the jurisdiction under proportional taxation: from the perspective of the club good literature, $\gamma_2 > 0$ corresponds to the case with congestibility. The second term is the weighted distance between i and the players in S. If the distance in question is a geographic distance, then the weighting by y_i could be due to a greater time cost of travel for those with higher incomes. An obvious simplification here is that there is no quality dimension to the services provided. For a given coalition S, the amount that must be paid by each player is determined mechanically by the tax base Y_S and the total cost of providing services. Other than the proportional taxation to pay for these services, there is no redistribution or other transfers within a coalition.

When we perform Monte Carlo simulations, we will use payoffs as given in Equation 4. In the Japanese data that will be presented in Section 3, however, a player will correspond to a feudal village with some population p. We thus need to scale the congestion cost γ_2 by the population, leading to payoffs of

$$v_i(S) = -\frac{y_i}{Y_S}(\gamma_1 + \gamma_2 \sum_{i' \in S} p_{i'}) - y_i \sum_{i' \in S} \frac{y_{i'}}{Y_S} d(i, i'). \tag{5}$$

An obvious objection to the form of the payoffs just presented relates to our choice to weight by y. This weighting is actually not essential, because techniques following Barros [1998] could be used to run a mixed integer program similar to that presented below in Section 2.1 even though payoffs would no longer have the form of a fractional hedonic game. We do not explore this extension, however, because in our historical Japanese dataset there is substantial evidence for the overrepresentation of the elite both in formal and informal institutions and thus it seems appropriate to overweight richer players (see Nishikawa, Hayashi, and Weese [2018] for details). A further advantage is that in our data the players correspond to feudal villages that contain multiple households with different levels of y. The preferences described in Equation 5 lead to all of these households having the same preferences over mergers. Thus, the specific form chosen allows us to ignore within-village heterogeneity in income.

Another potential objection is that we follow Desmet et al. [2011] and have our players experience disutility proportional to the average distance between them and their coalition partners. In recent years the modelling choice used in Alesina and Spolaore [1997] has become popular, where players experience disutility based on how far their ideal point is from a policy decision.[1] We do not use this type of model for two reasons. First, we will show that the most important type of heterogeneity in our data is the walking distance between players, which does not map to a low dimensional euclidean space as would be required by an ideal point model. Second, the loss function in an Alesina and Spolaore [1997] type model involves a quadratic, which means that the program we are about to present would be a computationally costly quadratically constrained program, instead of a linearly constrained program.

[1] The "average distance" functional form comes from the Lieberson [1964] extension of Greenberg [1956]. Earlier work on Alesina and Spolaore [1997] type models includes Greenberg and Weber [1986]: a literature review is provided in Donder, Breton, and Peluso [2012].

2.1 Solution

We will use the core as our solution concept. In a hedonic coalition formation game without transfers, the core is

$$\Pi^* = \{\pi | \forall S' \notin \pi, \exists i \in S' \text{ s.t. } v_i(\pi) \geq v_i(S')\}, \tag{6}$$

where π is a partition of players into coalitions, and $v_i(\pi)$ indicates the utility that player i receives from whatever coalition it belongs to in partition π. There is the possibility that Π^* is empty, and a substantial amount of work has been devoted to finding conditions under which the non-emptiness of Π^* is guaranteed.[2] In general, the results in this literature have been mostly negative: it is difficult to find conditions under which the core is guaranteed to be non-empty, and even more difficult to define these conditions in such a way that they can be easily checked. Brandl, Brandt, and Strobel [2015] discuss how fractional hedonic games also suffer from this problem, and provide a specific six player example with an empty core.

In this paper we will mostly ignore this issue, and instead simply show that core partitions exist given our data and parameter estimates. A potential criticism here is that this "works for me" attitude does not tell us anything about whether core partitions generally exist in fractional hedonic games. On the other hand, in the two-sided matching literature in certain cases empirical results [Roth and Peranson 1999] regarding existence of stable matchings predated theoretical results [Kojima, Pathak, and Roth 2013] explaining why this occurred. In an attempt to generalize our empirical results we will also perform Monte Carlo simulations and show that our results hold with high probability on randomly generated datasets.

To generate core partitions we will follow Roth and Vate [1990] and Chung [2000] and use a series of myopic deviations by blocking coalitions. However, the data that will be presented below in Section 3 involves approximately 1000 players who could combine to form $2^{1000} - 1$ distinct coalitions. It is thus not computationally feasible to enumerate all the coalitions that could potentially form. Instead, we will use a binary integer program that relies on the payoff structure inherent in fractional hedonic games.

Any given partition π may have many blocking coalitions. Suppose that S' is a blocking coalition. Taking the payoffs defined in Equation 5 for a weighted fractional hedonic game,

[2] These conditions include "consecutiveness" [Greenberg and Weber 1986], "intermediate preferences" [Demange 1994], the "top coalition property" [Banerjee, Konishi, and Sönmez 2001], and the "single lapping property" [Pápai 2004], among many others. Banerjee, Konishi, and Sönmez [2001] provides a detailed literature review.

it must be the case that, for every player i in S',

$$v_i(\pi) < \frac{\sum_{i' \in S'} y_{i'} v_i(i')}{\sum_{i' \in S'} y_{i'}}. \tag{7}$$

The intuition behind our strategy is that multiplying both sides of this inequality by $\sum_{i' \in S'} y_{i'}$ creates inequalities that are linear in the membership of S', and the membership of S' can be described by a binary variable for each player indicating whether or not it is a member. Details of this approach are provided in Appendix A.1. The problem of finding a blocking coalition can then be posed as a maximization problem based on binary integer programming with linear constraints. Let w be a vector of weights. The program

$$\begin{aligned} &\underset{S'}{\operatorname{argmax}} \sum_{i \in S'} w_i \qquad\qquad (8) \\ \text{s.t.} \quad \forall i \in S', \quad & 0 < -v_i(\pi) \frac{\sum_{i' \in S'} y_{i'}}{y_i} - \sum_{i' \in S'} y_{i'} d(i, i') - (\gamma_1 + \gamma_2 \sum_{i' \in S'} p_{i'}) \end{aligned}$$

will either return a blocking coalition S', or an infeasibility certificate with a numerical proof that no such coalition exists. We will call the S' that is returned by this optimization a "myopic deviation". If there are multiple possible myopic deviations, the program returns the deviation with the highest possible weight. By choosing different weights for different players, different deviations and paths of deviations can be selected.

We are able to numerically prove the non-existence of a blocking coalition when none exists because the maximization problem in (8) is a binary integer program. Programs of this type are solvable by techniques such as branch and bound [Land and Doig 1960], which use linear relaxations of the integer program to eliminate large parts of the search space. If there is no blocking coalition, the entire search space will eventually be eliminated in this way, resulting in a numerical proof that no blocking coalition exists.

Binary integer programs are NP-hard and there is no theoretical guarantee that a solution (or numerical proof that a solution does not exist) can be obtained in a reasonable amount of time. However, we do not encounter this problem when using the commercial solver CPLEX on modern hardware, and even with $N = 1000$ players the binary integer program usually terminates within a few minutes, and often much faster. Counterintuitively, however, we observe much *worse* computational performance in the special case where there is no horizontal heterogeneity and $d(i, i') = 0$ for all i and i'. This suggests that the speed of the fast solutions that we observe in the case where there is horizontal heterogeneity is in fact due to that heterogeneity. CPLEX is likely able to rule out many potential blocking coalitions because they would have too much horizontal heterogeneity, and quickly restrict

Data: N, v, w, y
Result: A core partition π^* (or loop forever)
Arbitrarily assign players to a starting partition π^1;
Iteration count $j = 1$;
while *there is a blocking coalition* S' *for partition* π^j **do**
    Identify "affected" coalitions, $A = \{S | S \in \pi^j, \exists i \in S \text{ s.t. } i \in S'\}$;
    Identify "residual" players, $R = \{i | i \in S \in A, i \notin S'\}$;
    if $R \neq \emptyset$ **then**
        Recursion: find a core partition π^{R*} using only players R;
        $\pi^{j+1} = (\pi^j \setminus A) \cup \{S'\} \cup \pi^{R*}$;
    else $\pi^{j+1} = (\pi^j \setminus A) \cup \{S'\}$;
    $j = j + 1$;
end
$\pi^* = \pi^j$;

Algorithm 1: Core Partition via Myopic Deviations

the search for blocking coalitions to a much smaller set of such coalitions where the coalition partners are close to each other in terms of distance $d(i, i')$.

Algorithm 1 attempts to find a core partition by successively generating blocking coalitions, where nomenclature and some ideas regarding recursion are taken from Ray and Vohra [1997]. In this algorithm, the binary integer program to find a blocking coalition given in (8) appears in the loop condition.

Although we know that the program given in (8) will always terminate with either a blocking coalition or a proof that that none exists, there is the possibility that Algorithm 1 will not terminate. In particular, Algorithm 1 may loop forever if there is a cycle of myopic deviations. Using our data and parameter estimates, however, the algorithm always terminates so long as the weights w_i are positive for all players i.[3] By running Algorithm 1 multiple times with different weights w, we pick out different paths of myopic deviations, and thus different core partitions. The partition π^1 used to initialize Algorithm 1 does not appear to matter: we obtain qualitatively similar results using the grand coalition, the all-singleton partition, or the central planner's preferred partition. We use the all-singleton partition as our starting partition for the results discussed below, because any stable partition can be obtained via a sequence of myopic deviations starting from the all-singleton partition.[4]

[3]When some weights are negative, such as when $w_i \sim \text{Normal}(0, 1)$, we very occasionally observe cycles. These cycles are short and thus easy to identify. Re-randomizing weights from a non-negative distribution such as $w_i \sim \text{Uniform}(0, 1)$ generally breaks the cycle.

[4]Proof: Let $S' \in \pi^*$ be a non-singleton coalition. Set w_i to some high value for $i \in S'$ and to some very negative value for all other i. Then S' is the coalition that maximizes (8): so long as different coalitions have different payoffs S' must be a deviation from the singletons because otherwise π^* would have some singleton deviation. Repeat for all other coalitions in π^*.

Additional computational details are proviided in Appendix A.2.

We now perform several Monte Carlo exercises to better understand the nature of the core in coalition formation games of the type we are considering. We begin by using the payoffs in Equation 4 and setting $y_i = 1$ for all players, $\gamma_2 = 0$, and randomly drawing $d(i, i') = d(i', i) \sim \text{Normal}(0, 1)$. Appendix Figure B.1 shows the fraction of games of this type that have no core partitions for different values of γ_1. For the 99.9+% of games that have a non-empty core, Appendix Figure B.2 shows the average number of core partitions. We see that for very small or very large fixed cost γ_1 the core has only a single partition, while for intermediate values the core can be quite large. This makes sense, because there is only one all-singleton partition, and only one grand coalition, but there are many intermediate options where some but not all players have chosen to merge together.

For games with non-empty cores, Appendix Figure B.3 shows the average size of coalitions in a core partition. Comparing Appendix Figure B.3 to Appendix Figure B.1, we see that values for γ_1 that sometimes result in games with an empty core are parameter values that lead to many small but non-singleton coalitions forming. This is likely because it is more difficult to create a cycle that involves many singletons, but with small coalitions the total number of coalitions is higher and thus there are more opportunities for a cycle to arise. Low but not tiny values of γ_1 should thus generate the highest probability of cycles forming.

We now show that the probability of an empty core falls dramatically when we impose more structure on the pairwise distances $d(i, i')$. Specifically, suppose that we let $d(i, i') = ||z_i - z_{i'}||$, where z_i is a two-dimensional vector with each dimension drawn from a $\text{Uniform}(0, \sqrt{N})$ distribution. That is, each player is given a location inside a square of side length $\sqrt{N}$, where we increase the size of the square with the number of players in order to keep player density constant. Appendix Figure B.4 shows that games based on this new set of payoffs have dramatically smaller probabilities of having an empty core: less than one game in a million for most values of γ_1, even for games with $N = 8$ players. This very high probability of a non-empty core appears not to have been previously noted in the literature; however, in the case of coalition formation games with transfers, Drèze et al. [2008] consider a game with a similar setup to ours, and show theoretically that the core in their game is "almost" non-empty. This finding matches with our empirical finding here that the core is non-empty with high probability in the case without transfers.

One interesting feature of Appendix Figure B.4 is that the probability of having an empty core does not appear to grow evenly with the number of players. Unfortunately, the very small probabilities in question mean that it is computationally costly to investigate this further, particularly with higher numbers of players. However, in Section 4 we run a Monte Carlo exercise (Appendix Table B.5) with $N = 1000$ players, and we do not observe any empty cores over 1000 replications of this game. This suggests that the dramatic growth in the probability of an empty core between $N = 5$ and $N = 6$ in Appendix Figure B.4 may be a one time occurrence, not recurring at higher values of N.

We now examine whether Algorithm 1 generates core partitions that are "representative" in some sense of the whole set of partitions in the core. To do this we randomly generate 10000 games with $N = 10$, compute the cores of these games using brute force enumeration, and select the game with the largest core. We then run Algorithm 1 10000 times using different random weights $w_i \sim \text{Normal}(0, 1)$ for each run. We consider both the case where the random games are generated using $d(i, i') \sim \text{Normal}(0, 1)$, and where they are generated by choosing player locations in a square. We show the relative frequencies of different core partitions being selected for these two types of random games in Appendix Figure B.6. For both cases, all core partitions were generated by Algorithm 1 at least once in the 10000 runs of the algorithm. However, the frequency with which the different core partitions were generated varied dramatically depending on the partition. We thus might worry that certain types of partitions are more likely to be generated by the algorithm than other types, and that an analysis of a coalition formation game using this algorithm thus may be biased.

In this paper, our analysis below focusses mainly on comparing core partitions to a Benthamite social planner's optimal partition. To see whether this analysis is likely to suffer from bias due to Algorithm 1, we randomly generate 100 games, run Algorithm 1 on each of the games 10000 times with (normally distributed) random weights, and then compare the frequency with which a partition appears as a solution to the sum of payoffs for all players in that partition. Appendix Figure B.7 shows the results from this analysis. We find that while there is a correlation between the frequency of appearance of a partition and the total payoff it offers to players, this appears to be due to the fact that games that have many core partitions are games that have on average higher payoffs for partitions in the core. For a specific game, we do not observe any tendency for Algorithm 1 to be biased towards generating partitions that are particularly good or bad for players overall.

2.2 Centralized Solution

Let $\pi \in \Pi$ be a partition of players into coalitions, where Π is the set of all possible partitions. With some abuse of notation, let the payoff to a Benthamite central planner of choosing partition π be

$$v(\pi) = \sum_i v_i(\pi)$$
$$= -\mathrm{WSS}(\pi) - \gamma_1|\pi| - \gamma_2 \sum_i p_i, \qquad \mathrm{WSS}(\pi) = \sum_{S\in\pi}\left[\frac{1}{Y_S}\sum_{i\in S}\sum_{i'\in S} y_i y_{i'} d(i,i')\right], \tag{9}$$

where WSS(π) is the "within sum of squares" of partition π. At first glance this appears to be unusual terminology, given that nothing in Equation 9 is being squared.

A particularly well-known special case of this problem is where $d(i, i')$ is euclidean distance, calculated based on players' locations on some plane. In the limiting case, where the players are tiny, uniformly distributed on the plane, and otherwise identical, the optimal partition is given by a regular hexagonal tiling. This was first discussed in the "central place theory" of Christaller [1933].[5] In general, however, there is no closed-form solution for the optimal partition of a fractional hedonic game [Aziz, Gaspers, et al. 2015] and it must be computed via some combinatorial optimization technique.

We will break the planner's problem down into two steps. First, find the best partition among those partitions with exactly k coalitions, for all possible values of k. Second, choose the optimal value of k. The advantage of this seemingly-inefficient approach becomes apparent when we write down the first step optimization problem:

$$\pi_k^{\mathrm{FB}} = \underset{\pi\in\Pi \text{ s.t. } |\pi|=k}{\operatorname{argmax}} \; v(\pi)$$
$$= \underset{\pi\in\Pi \text{ s.t. } |\pi|=k}{\operatorname{argmin}} \; \mathrm{WSS}(\pi), \tag{10}$$

where we can drop the fixed cost term $\gamma_1|\pi| = \gamma_1 k$ and the variable cost term $\sum_i p_i$ because they are both constant. The overall best possible partition is

$$\pi^{\mathrm{FB}} = \underset{k\in\{1,\ldots,N\}}{\operatorname{argmin}} \; \mathrm{WSS}(\pi_k^{\mathrm{FB}}) + \gamma_1 k. \tag{11}$$

The immediately apparent tradeoff here is between heterogeneity within each coalition and the fixed cost of running k jurisdictions. This optimal partition π^{FB} is generically unique, and can easily be determined if π_k^{FB} is known for all possible values of k.z

[5] Drèze et al. [2008] provides extensive citations regarding hexagonal tiling results. The efficient partition will be stable in this case because any deviation to a larger hexagon would be opposed by those at the edge of the deviation, and any deviation to a smaller hexagon would be opposed by those at the center of the hexagons in the efficient partition. However, the core is very large here, and thus depending on how it is determined which partition from the core actually occurs, substantial inefficiency could result.

The optimization problem in Equation 10 corresponds to a weighted kernel k-means objective, where $d(i, i')$ gives the kernel distance between players i and i'.[6] In Appendix A.3 we show that a "spread transform" [Hathaway and Bezdek 1994] can be used to transform this weighted kernel k-means problem into a standard weighted k-means problem, and the objective for this problem is (up to the "spread" constant) identical to the objective in Equation 10. (All standard k-means problems can be expressed in terms of the squared distance between points in a euclidean space, explaining why it is appropriate to label the $\text{WSS}(\pi)$ term of Equation 9 as a sum of squares.)

Making use of this transformation, we will compute π_k^{FB} via a weighted version of Hartigan and Wong [1979]. The solution to a single run of Hartigan and Wong [1979] is not guaranteed to be a global optimum, but from a theoretical perspective multiple restarts of the algorithm with different starting configurations will eventually find the global optimum. We use the kmeans++ [Arthur and Vassilvitskii 2007] method of selecting starting configurations with 250 restarts.

3 Data

We use data from Gifu Prefecture in Japan. During the Meiji period, a set of municipal mergers (the *Meiji Daigappei*) were mandated by the central government as part of its modernization policies: prefectural and national officials acted as a central planner for these mergers. Mergers occurred across the country, mainly in the 1880s and 90s, and are described in more detail in the introduction of Nishikawa, Hayashi, and Weese [2018]. We concentrate on Gifu because high quality data on covariates is only available for that prefecture.

The main advantage of historical data is that it better matches the simple theoretical model presented in Section 2. Modern data for Japan (and most other countries) would feature municipalities receiving some sort of transfers from a higher level of government. In the presence of such a transfer system, centrally planned mergers become a complicated tradeoff between redistribution through this transfer system, and redistribution by forcing rich and poor municipalities to merge together. Using data from a period when there was almost no redistribution avoids these issues.

Interior Ministry Order 352 is the main document responsible for the mergers. It states in fairly explicit terms that the mergers are to involve a tradeoff between efficiencies of scale and geographic distance: "... for the purpose of creating independent municipalities, in each

[6] To see this, look at Equation 3 of Roth, Laub, et al. [2003] and note that it corresponds to an unweighted version of Equation 10. Now suppose that we combined $y_{i'}$ identical unweighted players into a single new player i'. Then suppose that we combined y_i identical unweighted players into a single new player i.

Figure 1: Mergers in Gifu Prefecture

(a) Feudal villages, 1881 (b) Municipalities, 1897

municipality an appropriate amount of financial resources are required ... when merging do not make area excessively large and do not disturb convenience of access." Although the words "efficiencies of scale" do not appear in the order, the tax base of each village is fixed, and thus the obvious way a municipality with "appropriate" financial resources could be created out of villages without "appropriate" financial resources would be if the per capita cost of providing services is decreasing with scale.[7] We thus have primary source support for a model that can conveniently be written in the form of a fractional hedonic game. The same source also states that the planner is to "give consideration to the wishes of the villages, and not be antagonistic to the sentiment of the people.": the mergers are thus to be consultative but not democratic, with the planner responsible for considering local opinions before deciding on the final set of jurisdiction boundaries.

The main dataset for covariates is the *Gifu-ken Chouson Ryakushi* (GKCR, "Outline of Towns and Villages of Gifu Prefecture") of 1881 and related documents. This covers the southern portion of Gifu, describing 1111 feudal villages that were combined to form 289 western-style municipalities. The initial boundaries of these villages from the GKCR are shown in Figure 1a, and the boundaries after the mergers are shown in Figure 1b.

To match the model in Section 2 to the data, let the players be feudal villages. In

[7] A potential counterargument would be that the order is not referring to efficiencies of scale, but is instead asking for poorer areas to be deliberately merged with richer areas, thereby creating municipalities with closer-to-average (and thus "appropriate") resources. This alternative interpretation is not consistent with the portion of the order that reads "existing villages that ... have sufficient resources should not be merged or split", because if the objective was to redistribute resources then the richer villages are precisely the ones that should be targeted for merging.

Table 1: Summary Statistics

Variable	Actual (x_{π^0})	Permuted ($x_{\psi\pi^0}$)	Difference ($\times$ 1000)	Units	Description	Source
DIST.WALKING	1.37	1.37	-55.90 (44.81)	1000 sec.	Distance in terms of walking time	CIAS, GIAJ
DIST.STRAIGHT	1.19	1.19	-49.05 (38.87)	km	Straight line distance	CIAS
ADJACENT	0.27	0.27	-1.53 (0.85)	Indicator	Are feudal villages adjacent?	CIAS, GKCR
RELIGION	0.87	0.87	-0.05 (0.22)	Herfindahl	Differences in religious sect identities	GKCR
PRODUCTION	0.50	0.51	-8.54 (45.96)	Herfindahl	Differences in (agricultural) products	GKCR
PROPERTY	0.41	0.42	-6.16 (12.46)	Herfindahl	Differences in types of property	GKCR
LORD	0.31	0.31	-0.99 (1.00)	Herfindahl	Differences in identity of feudal lord	GKCR
FISH	0.50	0.50	-0.02 (0.26)	100*Herf.	Differences in fishing activity	GKCR
WEALTH	0.79	0.79	-0.003 (0.006)	Herfindahl	Differences in land holding distribution	GKCR
HH300	0.90 (0.0007)	0.90	0.0002	Indicator	Are there more than 300 households?	GKCR
HH500	0.44 (0.0008)	0.44	-0.0001	Indicator	Are there more than 500 households?	GKCR
# players	1111	1111				
# partitions	1	616605				
# distinct		612075				

Actual: characteristics of actually observed partition π^0
Permuted: characteristics of permuted partitions $\psi\pi^0$, where ψ flips one player with another
Difference: $(x_{\pi^0} - x_{\psi\pi^0})$ multiplied by 1000.
x is calculated as described following in Equation 12, and the meaning of this calculation is reported in the "Units" column. In most cases x corresponds to a (weighted) Herfindahl index of the characteristic in question, calculated at the coalition level and then summed across all coalitions in the partition. For a more detailed description of the characteristics and the method used, see Appendix C.
Reported numbers have been divided by total tax base $\sum_i y_i$, and thus correspond to the average heterogeneity experienced by a single player. For example, the average player will be located an average straight line distance of 1.19km away from players in its coalition, and will be in a coalition with religious heterogeneity corresponding to a Herfindahl index of 0.87.
distinct: number of partitions actually used in estimation, after removing permutations that do not actually alter the partition structure because they flip two players in the same coalition
Data source abbreviations:
CIAS: Center for Integrated Area Studies, Kyoto Univ.
GIAJ: Geospatial Information Authority of Japan
GKCR: *Gifu-ken Chouson Ryakushi*

Equation 5, p_i will be the population of village i. Income y_i is taken to be the *koku* rating of the village: this is a historical land tax assessment, converting all production in a village into common units of rice. Nishikawa, Hayashi, and Weese [2018] discuss local government funding, and show that land tax was the most important source of revenue during this period: use of the *koku* ratings thus seems appropriate.

For distance $d(i, i')$ there is an overabundance of variables, leading to the parameter estimation problem we consider below in Section 4. For geographic sources of heterogeneity, there is data on the geographic adjacency of villages. In addition, we will calculate a straight line distance (in km) and a walking distance (in minutes) between each pair of villages. The construction of these last two variables relies on being able to represent the population distribution of villages using one or a small number of points. Details are provided in Appendices C.1 and C.2.

For non-geographic sources of heterogeneity, there is data on the specific religious sects present in each village, the types and quantities of agricultural and non-agricultural products produced, the total valuations of the different types of land present, the identity of the feudal lord that had controlled the village prior to the Meiji Restoration, the amount of fishing taking place, and a binned distribution of the wealth of households. More details are provided in Appendix C.3, including a demonstration that d is equivalent to computing a weighted Herfindahl index in the case where there are discrete types of players: this shows that our model could be used in cases where ethnolinguisitic or similar types of fragmentation are relevant, a popular subject of study in political economy.

In Section 2 we developed a model where there is a scalar distance $d(i, i')$ between each pair of players. Our data instead provides nine different distances, three geographic and six non-geographic. To match our theoretical model to the data, let us suppose that $d(i, i')$ is now a *vector* giving nine different distances between players i and i', with each entry in the vector corresponding to a different type of heterogeneity. We then modify the definition of the "within sum of squares" term $\mathrm{WSS}(\pi)$ in Equation 9 to

$$\mathrm{WSS}(\pi) = -x_\pi^{\mathrm{T}}\beta, \qquad x_\pi = \sum_{S\in\pi}\left[\frac{1}{Y_S}\sum_{i\in S}\sum_{i'\in S} y_i y_{i'} d(i, i')\right], \tag{12}$$

where β is a parameter vector with length equal to the number of different types of heterogeneity, with each entry in β indicating the importance of the corresponding type of heterogeneity. Our data vector x_π has the same length as β, with each entry giving the total amount of heterogeneity of a given type experienced by players in partition π. We add a negative sign in Equation 12, thereby making plausible values for entries in β negative, so that more desirable partitions have higher values of $x_\pi^{\mathrm{T}}\beta$. This is necessary to match the convention in the discrete choice literature that $x_\pi^{\mathrm{T}}\beta$ represents the utility from choice π, rather than disutility. The total distance experienced by players in a partition is thus expressed as $-x_\pi^{\mathrm{T}}\beta$.

Summary statistics for these data are provided in Table 1. The first column in Table 1 displays the average amount of each type of heterogeneity experienced by players in the jurisdictions that were actually created by central planner. The remaining columns in Table 1 provide information about the heterogeneity of alternative mergers that could have happened but did not, but we will delay discussion of these columns until Section 4.2, where we describe the alternative mergers that we will consider.

4 Estimation Strategy

The observed data is a partition of players, as chosen by a central planner. Let the payoff to the planner of choosing partition π be

$$u(\pi|\beta) = v(\pi|\beta) + e(\pi), \tag{13}$$

where $v(\pi|\beta)$ is the deterministic payoff of partition π when the parameters are β, and $e(\pi)$ is the idiosyncratic payoff. To estimate β in this general case, in Section 4.1 we will present a variation on the Fox [2007] pairwise maximum score estimator, which is based in turn on Manski [1975]. This estimator requires specific assumptions regarding the form of idiosyncratic shocks $e(\pi)$, but does not place any restrictions on the deterministic component of the payoff $v(\pi|\beta)$.

Substituting Equation 12 into Equation 9, we have

$$v(\pi|\beta,\gamma) = x_\pi^{\mathrm{T}}\beta - \gamma_1|\pi| - \gamma_2 \sum_i p_i. \tag{14}$$

In Section 4.2 we will present an objective function to produce a maximum score estimate for β for this specific functional form. The parameters γ_1 and γ_2 will not be identified by our estimator: we will produce estimates for these parameters in Section 6.

4.1 General Case

Let the idiosyncratic shock for partition π be

$$e(\pi) = \sum_{S\in\pi}\sum_{i\in S} \frac{\sum_{i'\in S}\epsilon_{ii'}}{|S|}. \tag{15}$$

That is, $e(\pi)$ is a sum of the average of pairwise idiosyncratic shocks $\epsilon_{ii'}$ for each player within each coalition. For simplicity we will assume that $\epsilon_{i'i} = \epsilon_{ii'}$ because these shocks enter identically into the decision maker's payoff. The ϵ_{ii} shock gives the idiosyncratic desire of player i to remain alone; this is not essential to the model and $\epsilon_{ii} = 0$ could be imposed without any substantive changes to the discussion below.[8]

In Equation 15 the $e(\pi)$ idiosyncratic shocks are not i.i.d. across choices π. This is desirable given that the choices we are considering are partitions: the idiosyncratic shock for the {{1,2}, {3,4},{5,6},{7,8},{9,10}} partition should naturally be highly correlated with the shock for the {{1,2},{3,4}, {5,6}, {7,10}, {8,9}} partition, and not very correlated with the shock for the {{1,10}, {3,2} {5,4}, {7,6}, {9,8}} partition. The pairwise $\epsilon_{ii'}$ idiosyncratic shock structure we use creates the desired correlation. An estimation strategy that ignored this correlation structure, such as a standard multinomial logit model, would encounter substantial difficulties due to the very large number of potential partitions.

Maximum score estimation relies on the Manski [1975] rank ordering property: a partition must be more likely to be chosen if it has a higher deterministic payoff. Specifically,

Definition 1 (Rank Ordering Property)**.**

$$v(\pi_j|\beta) > v(\pi_{j'}|\beta)$$

if and only if

$$\Pr(\pi_j|v(\beta)) > \Pr(\pi_{j'}|v(\beta)).$$

Here the probability of partition π_j being selected is $\Pr(\pi_j|v) = \int_{\mathbb{R}^{N(N+1)/2}} f(\epsilon) 1\left(\pi_j = \pi(\epsilon|v)\right) d\epsilon$, where $f(\epsilon)$ is the density of the idiosyncratic shocks and $\pi(\epsilon|v)$ is the partition selected by the decision maker when the shocks are ϵ and the deterministic payoffs for all possible partitions are given by the vector v. To simplify notation we suppress the parameters β here writing v instead of $v(\beta)$.

Unfortunately, a data generating process based on Equations 13 and 15 violates the rank ordering property:

Example 1. *Let* $N = 3$ *and* $v(\pi) = 0$ *for all partitions* π*. Let* $\pi_1 = \{\{1,2\},3\}$ *and*

[8] We retain the ϵ_{ii} shock because without it $e(\pi) = 0$ for the all-singleton partition, and the probability of this partition being selected in the null model ($v(\pi) = 0$ for all partitions) is then very small, particularly for larger N.

$\pi_2 = \{\{1, 2, 3\}\}$. *Then*

$$u(\pi_1) = e(\pi_1) = \frac{\epsilon_{11} + \epsilon_{12}}{2} + \frac{\epsilon_{22} + \epsilon_{12}}{2} + \epsilon_{33}$$
$$u(\pi_2) = e(\pi_2) = \frac{\epsilon_{11} + \epsilon_{12} + \epsilon_{13}}{3} + \frac{\epsilon_{22} + \epsilon_{12} + \epsilon_{23}}{3} + \frac{\epsilon_{33} + \epsilon_{13} + \epsilon_{23}}{3}$$

Let the $\epsilon_{ii'}$ idiosyncratic shocks be i.i.d., with $\epsilon_{ii'} \sim N(0, 1)$. Then $\Pr(\pi_1) \simeq 0.223$ *but* $\Pr(\pi_2) \simeq 0.165$*.*

As a small change in deterministic payoffs will generate only a small change in choice probabilities, we will have a violation of the rank order property in Example 1 if we increase the structural payoff to the grand coalition by some small amount.

Our estimation strategy will instead use a weaker version of the rank order property, one that is satisfied by the data generating process described by Equations 13 and 15. Let ψ be a permutation of the player indices such that player 1 is relabeled to be player $\psi(1)$, and so forth. We will write $\psi\pi$ for a permuted partition, where the player indices have been permuted but the coalition structure otherwise left unchanged.[9] We now present our innovation, the *permutation rank ordering property*:

Definition 2 (Permutation Rank Ordering Property)**.** *Either $v(\pi) = v(\psi\pi)$ for all $\pi \in \Pi$ or*

$$\sum_{j=1}^{|\Pi|} \left(\Pr(\pi_j|v) - \Pr(\psi\pi_j|v)\right) \cdot \operatorname{sign}\left(v(\pi_j) - v(\psi\pi_j)\right) > 0 \tag{16}$$

To see the relationship between our permutation rank ordering property and the original Manski [1975] rank ordering property, note that the original rank ordering property requires that either $v(\pi_j) = v(\pi_{j'})$ or

$$\left(\Pr(\pi_j|v) - \Pr(\pi_{j'}|v)\right) \cdot \operatorname{sign}\left(v(\pi_j) - v(\pi_{j'})\right) > 0. \tag{17}$$

We weaken this original rank ordering property in two ways. First, we do not allow partition π_j to be compared to an arbitrary alternative partition $\pi_{j'}$: instead, we restrict consideration to permutations $\psi\pi_j$. This eliminates the problem shown in Example 1, as the grand coalition is not a permutation of the $\{\{1,2\}, 3\}$ partition.

Second, we do not base our inequality on only a single partition π_j and its permutation $\psi\pi_j$; rather, we consider the whole set Π of partitions. To see why this is important, consider the following example:

[9] For example suppose that $N = 3$, $\psi(1) = 2$, $\psi(2) = 1$, $\psi(3) = 3$. That is, ψ swaps the identities of the first two players and leaves the third unchanged. Then if $\pi = \{\{1, 3\}, \{2\}\}$ we will have $\psi\pi = \{\{2, 3\}, \{1\}\}$.

Example 2. *Let $N = 4$, $\pi_1 = \{\{1,2,3\},\{4\}\}$, $v(\pi_1) = 1$, and $v(\pi) = 0$ for all other partitions $\pi \neq \pi_1$. Let the $\epsilon_{ii'}$ idiosyncratic shocks be i.i.d, with $\epsilon_{ii'} \sim N(0,1)$. Then*

$$\begin{aligned} \Pr(\{\{1,2,3\},\{4\}\}) &\simeq 0.275 & \Pr(\{\{1,2,4\},\{3\}\}) &\simeq 0.050 \\ \Pr(\{\{1,3\},\{2\},\{4\}\}) &\simeq 0.038 & \Pr(\{\{1,4\},\{2\},\{3\}\}) &\simeq 0.066 \end{aligned}$$

Here we see that the probability of the decision maker choosing the $\{\{1,3\},\{2\},\{4\}\}$ partition is low because of substitution into the $\{\{1,2,3\},\{4\}\}$ partition. This sort of substitution effect would result in a violation of the original rank ordering property if we increase $v(\{\{1,3\},\{2\},\{4\}\})$ slightly. Our permutation rank ordering property, however, aggregates across the $\{\{1,2,3\},\{4\}\}$ to $\{\{1,2,4\},\{3\}\}$ comparison and the $\{\{1,3\},\{2\},\{4\}\}$ to $\{\{1,4\},\{2\},\{3\}\}$ comparison (see Appendix D for details). This eliminates any problems arising from the substitution effect just described.

A remaining problem is that maximum score estimation generally relies on the exchangeability of idiosyncratic shocks.[10] As defined above, however, the $e(\pi)$ idiosyncratic shocks are not exchangeable: the difficulty in Example 2 arises precisely because $e(\{\{1,2,3\},\{4\}\})$ is positively correlated with $e(\{\{1,3\},\{2\},\{4\}\})$ but less correlated with $e(\{\{1,4\},\{2\},\{3\}\}$. We will instead apply exchangeability to the ϵ pairwise idiosyncratic shocks, otherwise following Assumption 2 of Fox [2007]:

Assumption 1. *The errors ϵ have an absolutely continuous joint distribution with full support on $\mathbb{R}^{N(N+1)/2}$. The associated joint density $f(\epsilon|v)$ exists and is exchangeable.*

In this paper, to obtain our empirical results we will use permutations ψ that flip a single player i with a single other player i'. Such a permutation is self-inverse: $\psi\psi\pi = \pi$ for any partition π. For our theoretical result, we consider the class of all self-inverse permutations:

Proposition 1. *Let ψ be a self-inverse permutation. The permutation rank ordering property then holds under Assumption 1.*

Proof. See Appendix E. □

Under normal circumstances, we would now construct a pairwise objective function following Equation 4 of Fox [2007]. Specifically, if we observed m distinct choices of partitions, we could use the objective function

$$Q_m^{\psi}(\beta) = \frac{1}{m}\sum_{b=1}^{m} 1(v(\pi^b|\beta) > v(\psi\pi^b|\beta)), \tag{18}$$

[10] If $f(e) = f(e(\pi_1), e(\pi_2), ...)$ is the probability density of the idiosyncratic shocks, the distribution $f(e)$ is *exchangeable* if $f(e(\pi_1), e(\pi_2), ...) = f(e(\pi_2), e(\pi_1), ...)$ and likewise for all other such permutations.

where π^b is the partition observed to have been chosen in the bth observation out of the m observations. Here we could choose any single potential permutation ψ in the set of self-inverse permutations: this parallels the situation in Fox [2007] where the objective function in his Equation 4 is based on a pairwise comparison between only two potential choices, even though there may have been many choices available to the decision maker.

A consistency proof for an estimator based on maximizing this objective function appears to be possible via a tedious but relatively straightforward rewriting of the proof in the appendix of Fox [2007]. There is a major problem, here, however: the asymptotics used in Fox [2007], which are standard throughout the discrete choice literature, assume that the econometrician observes an increasing number of distinct choices. In our case, this would correspond to observing the decision maker choosing a partition for one set of players, and then another partition for a separate set of players, and so forth, with an asymptotically increasing number of distinct partitions being chosen. The data we actually have, however, corresponds to exactly one partition being chosen for a very large set of players. The asymptotics that are plausibly associated with our data are thus an increasing number of players N, with the decision maker choosing a single partition that divides those players into coalitions. We now present an objective function based on these asymptotics.

4.2 Application to Municipal Mergers

Suppose that we observe a single partition that consists of an asymptotically increasing number of players. In order to make use of the data provided by the increasing number of players we will increase the number of other partitions we compare the observed partition to. Specifically, we choose a modified objective function,

$$Q^{\pi^0}_{\Psi(N)}(\beta) = \frac{1}{|\Psi(N)|} \sum_{\psi \in \Psi(N)} 1(v(\pi^0|\beta) > v(\psi\pi^0|\beta)), \tag{19}$$

where π^0 is the partition that was actually observed and $\Psi(N)$ is a set of self-inverse permutations of N players. The term inside the summation remains unchanged from Equation 18: for a given β, we increment the objective function whenever the observed choice has a higher deterministic payoff than the alternative choice. The summation itself, however, is non-standard, because we are summing over the permutations in $\Psi(N)$ whereas in Equation 18 we were summing over the m partitions that were actually observed. This change is necessary because the way to incorporate the additional information provided by a newly added Nth player is to compare the actually observed partition with an alternative partition that allocated that player to a different coalition. We thus need to increase the number of permutations we consider as the number of players N increases.

We define $\Psi(N)$ as the set of all pairwise flips, where a single player i is exchanged with another player i'. In Table 1 we show the heterogeneity of the actually observed partition, compared with these permuted partitions. There are a total of $N(N-1)/2$ distinct permutations of this type. Some of these permutations will leave the observed partition π^0 unchanged when applied. For example, if $\pi^0 = \{\{1,2\},3\}$ the permutation ψ that flips players 1 and 2 gives us $\psi\pi^0 = \pi^0$. A permutation ψ of this sort will have no effect on estimation because the indicator variable in Equation 19 will always be set to zero as $v(\pi^0|\beta) = v(\psi\pi^0|\beta)$ for all values of β. Formally we do not want to remove these permutations from $\Psi(N)$ because the set of permutations should not depend on the observed partition π^0. In Table 1, however, we report the number of permutations ψ such that $\psi\pi \neq \pi$, in addition to reporting the nominal number of permutations $|\Psi(N)|$.

Identification of β depends on the difference $v(\pi|\beta) - v(\psi\pi|\beta)$ changing as we vary the choice of β. If we substitute in our specific functional form for v from Equation 14, however, it is always the case that $\gamma_1|\psi\pi| = \gamma_1|\pi|$, because permuting the player labels does not change the number of coalitions in the partition. Thus, when we calculate the difference $v(\pi|\beta,\gamma) - v(\psi\pi|\beta,\gamma)$ for our objective function, the terms involving γ_1 and γ_2 will always cancel. We will thus estimate our parameters sequentially: first we will estimate the vector of parameters β using maximum score estimation, and then we will obtain the cost parameters γ via method of moments and a calibration exercise, as described in Section 6. For pointwise identification the maximum score estimator requires a "special regressor" with a continuous distribution and full support. In our case, we assume that walking distance has these properties.

Substituting Equation 14 into Equation 19 and cancelling the terms involving γ, we see that our objective function for the maximum score estimator has now simplified to

$$Q^{\pi^0}_{\Psi(N)}(\beta) = \frac{1}{|\Psi(N)|} \sum_{\psi \in \Psi(N)} 1(x^{\mathrm{T}}_{\pi^0}\beta > x^{\mathrm{T}}_{\psi\pi^0}\beta). \tag{20}$$

This is a standard objective function for a maximum score estimator (except for the unusual asymptotics), and estimation can be performed using standard techniques. We will compute $\hat{\beta}$ by using differential evolution to find the value of β that maximizes $Q^{\pi^0}_{\Psi(N)}$. This follows the approach in Fox and Santiago [2015]. The scale of β is not identified here because any scalar multiple yields the same value for the objective function.

Proving consistency for an estimator based on the objective function in Equation 20 would likely follow proofs in the mixing literature. The first step would be to show that some players are "far away" from other players, and will almost never end up in the same coalitions. The second step would then be to show that we could use sets of players that are far away from each other as effectively independent "observations". The proof appears to be non-standard, however, because formally describing how to create smaller almost-independent partitions from one large partition is more challenging than the standard problem of creating almost-independent observations from one long time series. We thus leave the theoretical proof of consistency for future research, and offer instead Monte Carlo evidence of performance in finite samples where there are a large number of players divided into a single partition. Specifically, in Appendix Table B.5 we examine the performance of an estimator based on Equation 19 when $N = 10$, 100, and 1000. We find substantial finite sample bias for $N = 10$, but minimal bias for $N = 100$ and 1000. The decrease in mean squared error corresponds to convergence at a rate between $N^{1/2}$ and $N^{1/3}$, but is closer to cube-root convergence. This seems plausible, given that in general there is cube root convergence for maximum score estimators [Kim and Pollard 1990], but Fox [2018] uses square root convergence following Sherman [1993] for a network game that has a somewhat similar form to the partition outcomes that we are considering.

5 Parameter Estimates

As is standard in the discrete choice literature, a normalization is required, and so results will be reported relative to the importance of walking distance, which will always have a coefficient of -1. We choose this normalization because, while a wide variety of variables describing heterogeneity are available in the data, casual inspection of Figure 1 suggests that geography is by far the most important.

Table 2 gives the results of the maximum score estimator. In Columns I-VIII we consider three types of geographic distance (walking distance, straight line distance, and adjacency), and six other types of heterogeneity. We expect that coefficient estimates will be negative, because the types of heterogeneity being considered are undesirable. In Column I we see that the coefficient estimate on straight line distance is positive and statistically insignificant. We thus conclude that the planner seems to have followed the Ministry instructions to consider the convenience of transportation: they did not take the dramatic shortcut of assuming that Gifu was flat, and that all points that appeared to be equidistant on a map were in fact equally easy to reach. We will thus not consider straight line distance in further specifications.

In Column II we see that geographic adjacency is statistically significant. Looking at Table 1, however, we see that magnitude of the geographic adjacency variable is much smaller than the walking distance variable. Thus, while geographic adjacency is important in determining the partition, walking distance appears to be much more important. Columns

Table 2: Maximum Score Parameter Estimates

	I	II	III	IV	V	VI
DIST.WALKING	-1	-1	-1	-1	-1	-1
DIST.STRAIGHT	0.07					
	$[-0.51, 0.65]$					
ADJACENT		-0.24	-0.34	-0.25	-0.19	-0.37
		$[-0.31, -0.17]$	$[-0.47, -0.21]$	$[-0.41, 0.09]$	$[-0.24, -0.15]$	$[-0.52, -0.21]$
RELIGION			-0.02			
			$[-0.04, 0.01]$			
PRODUCTION				-0.03		
				$[-0.05, -0.01]$		
PROPERTY					-0.10	
					$[-0.12, -0.08]$	
LORD						-0.05
						$[-0.08, -0.02]$
FISH						
WEALTH						
HH300						
HH500						
# Inequalities violated	696	378	376	375	356	363

A total of 612075 permuted partitions are considered (see Table 1). 95% symmetric confidence intervals are shown. These are generated via subsampling [Politis, Romano, and Wolf 1999] following Fox and Santiago [2015], with a 10% subsample and cube root asymptotics.

No confidence interval is reported for walking distance because it is fixed to -1.

Table 2: Maximum Score Parameter Estimates, cont.

	VII	VIII	IX	X	XI
DIST.WALKING	-1	-1	-1	-1	-1
DIST.STRAIGHT					
ADJACENT	-0.24	-0.35	-0.35	-0.46	-0.32
	$[-0.39, -0.08]$	$[-0.47, -0.22]$	$[-0.48, -0.23]$	$[-0.70, -0.21]$	$[-0.80, 0.16]$
RELIGION				0.01	-0.01
				$[-0.04, 0.07]$	$[-0.07, 0.04]$
PRODUCTION				-0.02	-0.01
				$[-0.05, 0.01]$	$[-0.06, 0.05]$
PROPERTY				-0.10	-0.11
				$[-0.20, -0.00]$	$[-0.24, 0.01]$
LORD				-0.07	-0.06
				$[-0.13, -0.02]$	$[-0.13, 0.02]$
FISH	-0.01			-0.01	0.00
	$[-0.02, 0.01]$			$[-0.03, 0.00]$	$[-0.02, 0.03]$
WEALTH		-0.03			-0.04
		$[-0.08, 0.01]$			$[-0.10, 0.02]$
HH300			-0.01		
			$[-0.02, 0.01]$		
HH500			-0.02		
			$[-0.04, 0.00]$		
Inequalities violated	378	366	366	341	334

A total of 612075 permuted partitions are considered (see Table 1). 95% symmetric confidence intervals are shown. These are generated via subsampling [Politis, Romano, and Wolf 1999] following Fox and Santiago [2015], with a 10% subsample and cube root asymptotics.

No confidence interval is reported for walking distance because it is fixed to -1.

III-VIII show that, while other types of heterogeneity are often also statistically significant, they have even smaller effects on the partition that is chosen. We thus conclude that geography is the principal determinant of the partition that was chosen. Perhaps not coincidentally, geography is the only type of heterogeneity explicitly mentioned in Interior Ministry Order 352. In Column VIII we include information on the (binned) wealth distribution of villagers. We see that this is statistically insignificant, which is in accordance with the model in Section 2, where the central planner does not care about income differences within mergers.

We might wonder whether there is any evidence that the central planner actually chose boundaries following the sort of model outlined in Section 2. One alternative hypothesis is that the planner simply followed a set of instructions it was given. The instructions in Interior Ministry Order 352 stated that municipalities should have between 300 and 500 households. In Column IX, we include two additional variables to take into account these instructions: Appendix C.4 provides details regarding the construction of these variables. These variables are not statistically significant, and furthermore at the estimated coefficients they are not economically important regardless of their statistical significance.[11]

In Columns X and XI we consider specifications including all of the non-geographic types of heterogeneity together. In general the results here are the same as Columns III-VIII. We continue to find that religious heterogeneity and differences in fishing activity have no statistically significant effect. The effect of heterogeneity in the types of production (crop types, etc.) drops slightly in magnitude and becomes statistically insignificant when it is included alongside heterogeneity in types of property (paddy field, dry field, etc.). This may be because these two types of heterogeneity are highly correlated and average production may have been difficult to measure accurately (the GKCR reports a single year).

Comparing the coefficients reported in Column XI to the summary statistics in Table 1, we see that all the non-geographic types of heterogeneity have not only small coefficient estimates, but the corresponding variables also have a small magnitude relative to walking distance. This is very convenient because village residents generally live in the same physical location (for details, see Appendix C.1), and given the form of Equation 5 this means that all residents of a village will have the same preferences over potential merger partners: we can thus ignore potential political issues with within-village decision-making.

[11]Matsuzawa [2013] presents at least one case where they argue that central planners attempted to force a village to merge with specific reference to the lower population limit. The point of our analysis here is not to argue that Matsuzawa [2013] is incorrect: our point is that from a *quantitative* perspective, geographic distance explains the observed pattern of mergers, and Column IX shows that this is not due to collinearity between geographic distance and the 300/500 household thresholds.

6 Estimation of Additional Parameters

The parameter estimates provided in Table 2 are arbitrarily scaled so that the payoff to 1000 seconds of walking distance is -1, and the importance of other types of heterogeneity is relative to walking distance. We now need to determine the correct scaling of these heterogeneity costs in order to compare them with the costs of running a jurisdiction. Below, we will first choose values (in 1880s yen) for γ_1 and γ_2 by calibrating to Japanese government estimates. We will then establish how important distance is relative to these costs by using an optimality condition from the central planner's problem in Section 2.2.

One might worry that, because we are choosing values for γ_1 and γ_2 based on auxiliary data, different choices might lead to very different outcomes when we perform counterfactual simulations in Section 7. The situation here is actually quite different for the fixed cost γ_1 than for the variable cost γ_2. The fixed cost γ_1 appears as part of the planner's problem in Section 2.2, and below in Section 6.2 we will show how we can use this to determine the relative weight the planner placed on fixed cost versus heterogeneity. Our use of government estimates to assign a yen value to γ_1 thus serves only to provide units of measurement for our calculation, and will not otherwise affect the simulation results in Section 7.

On the other hand, the variable cost γ_2 is constant in the planner's problem, and our only information about it comes from the government estimates discussed below. The size of γ_2 relative to the fixed cost will turn out to be an important determinant of the amount of inefficiency in core partitions, and thus we might worry that important results are being determined by a calibration exercise. In Section 7.2 we run Monte Carlo simulations that the precise value of γ_2 is actually not important to our results, so long as it is not tiny.

6.1 Cost of public goods γ

During the merger period, government bureaucrats produced a document describing the cost of providing public services for municipalities of three sizes: these costs are shown in Table 3.[12] Here $c_1 = ¥545.668$ was a constant that corresponded to administrative costs that exhibited efficiencies of scale, and $c_2 = ¥1467.931$ was a constant corresponding to costs that did not exhibit efficiencies of scale. At the three points provided, the costs correspond exactly to a cost function of the form used in Equation 5, with a fixed cost $\gamma_1 = c_1/3$, and a variable cost $\gamma_2 = \frac{2c_1/3+c_2}{3165}$, despite the fact that statistically there is an extra degree of freedom. There is no explanation offered of how the government arrived at these estimates,

[12] A population of 3165 appears to have been used as the base because it corresponded to the average population served by an administrative office under the briefly-used "ward" system, described in Nishikawa, Hayashi, and Weese [2018].

Table 3: Cost of providing local government services

	population	cost
“large”	3165	$c_1 + c_2$
“medium”	$\frac{3165}{2}$	$\frac{2}{3}c_1 + \frac{c_2}{2}$
“small”	$\frac{3165}{4}$	$\frac{1}{2}c_1 + \frac{c_2}{4}$

Source: government document reproduced in *Niigata-ken Shichouson Gappei Shi* (“History of Municipal Mergers in Niigata Prefecture”).

and thus it is unclear whether they believed that a cost function with only a fixed cost and a variable cost was particularly appropriate, or whether at the sizes that they chose to examine the efficiencies of scale happened to fit this pattern.

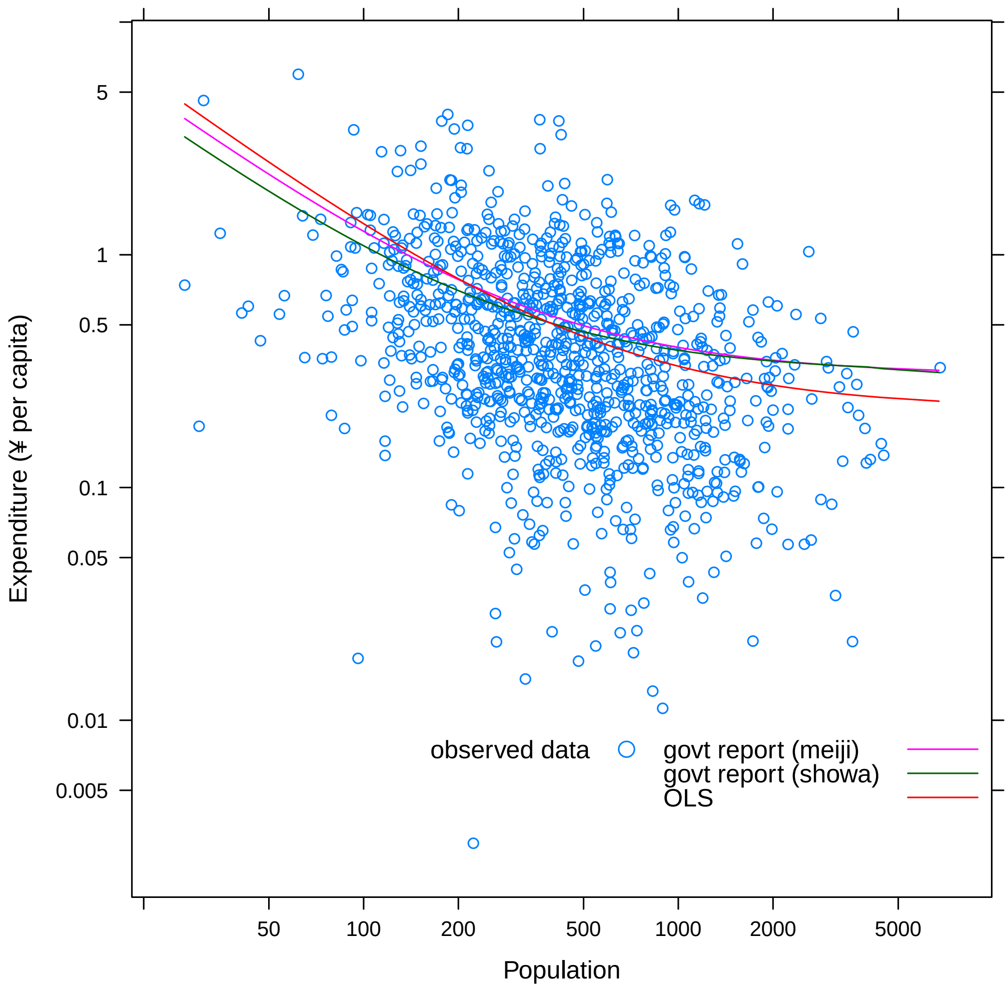
While there is no documentation available describing how these numbers were arrived at, verification of other sorts is available. Data on actual municipal expenditures is available for 1881, before the implementation of the new municipal system. This is shown in Figure 2, along with the cost function based on Table 3. Reiter and Weichenrieder [1997] survey the existing literature and conclude that there has been limited success in using actual expenditure data to estimate efficiencies of scale. For comparison purposes, however, a bivariate regression is provided in the figure.

The strongest confirmation for the validity of the numbers in Table 3 comes from much later sources. In 1950, when roughly the same municipal structure was still in place, a government document describing the efficiencies of scale in the provision of public services was produced.[13] This document provides a detailed breakdown of efficiencies of scale by service, for 20 public services, with the cost of each service described by a spline function with 6 knots: Figure 19 in Weese [2015] shows these spline functions for later data that uses a greater number of knots but the same basic structure. Despite the gap of 60 years and a substantial expansion in the number of public services provided, the estimates match the 1890 figures very closely, as shown in Figure 2.[14] We thus use values of γ_1 and γ_2 corresponding to Table 3 (roughly $\gamma_1 = 182$ and $\gamma_2 = 0.6$).

[13] This is the *kijun zaisei juyou gaku* (“standard fiscal need”) portion of the calculation for the *heikou koufukin* equalization payments.

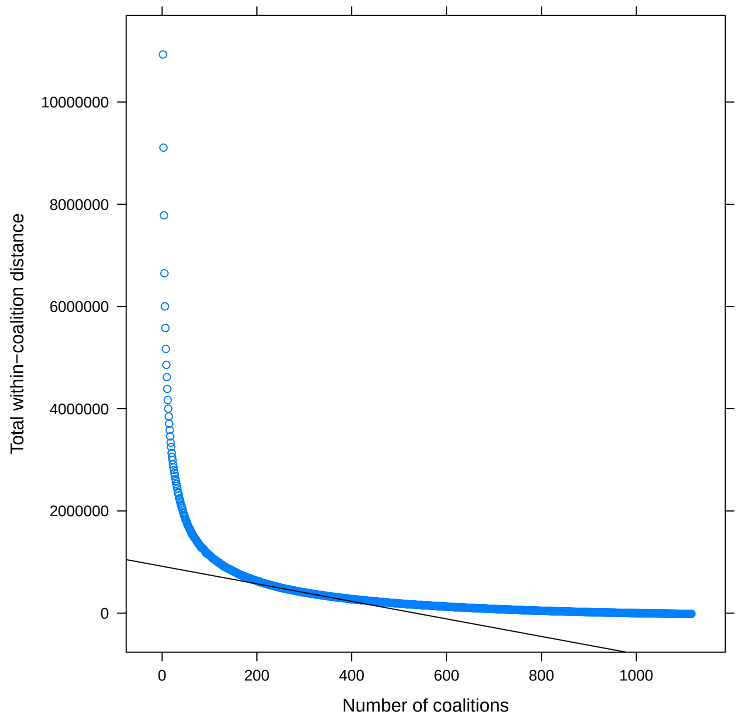
[14] An additional difference is that the 1890 figures appear to have been produced to describe the appropriate size for municipalities during the reorganization, while the 1950 figures were for use in an equalization transfer program. The different purpose suggest that any political biases in the reported figures would also be quite different.

Figure 2: Public good spending per capita



Data points indicate actual spending by precursors to final municipalities: this data is from 1881, when the final municipal system was still under development. The "OLS" line gives predicted expenses from a bivariate regression (not in logs), predicting total spending based on population and intercept: the line is curved as a result of transformation to log scale. The R^2 for this regression is 0.22 (an equivalent regression in logs also has an R^2 of 0.22). The Meiji "govt" line is based on the fixed cost plus variable cost for the points in Table 3 (populations 904, 1806, and 3615), with an adjustment to take into account that some services were not paid for by municipalities when the 1881 data was collected, and some revenue and associated expenses appears not to have been included. The Showa "govt" line is exactly the functional form provided in 1950 government documents describing the efficiencies of scale in the provision of local public goods, but has been normalized such that it is equal to "govt (meiji)" at a population of 3165 (the reference population for the Meiji government document).

Figure 3: Distance in Partitions with k Coalitions



6.2 Relative importance of distance and fixed cost

Let $\hat{\beta}^{\mathrm{u}}$ be the unscaled estimates reported in Column XI of Table 2, where the first entry is -1. We wish to produce correctly scaled estimates $\hat{\beta} = \alpha\hat{\beta}^{\mathrm{u}}$, such that we have distance costs $x_\pi^{\mathrm{T}}\hat{\beta}$ and the cost of running the municipalities $\gamma_1|\pi| + \gamma_2 \sum_i p_i$ on the same scale. This could not be done using our maximum score estimator because we could only consider alternative partitions that were permutations of the actually observed partition. To determine the appropriate scaling factor α we will employ a method of moments approach.

Specifically, with $N = 1111$ players, there can be anywhere between 1 and 1111 coalitions in the central planner's optimal partition. Let us begin by computing $\mathrm{WSS}(\pi_k^{\mathrm{FB}})$ from Equation 10, which gives us the distance experienced by players in the best possible partition with exactly k coalitions. The blue dots in Figure 3 show $\mathrm{WSS}(\pi_k^{\mathrm{FB}})$ for all possible values of k. These results are obtained by substituting our unscaled estimates $\hat{\beta}^{\mathrm{u}}$ into Equation 12, and then running weighted kernel k-means as described in Section 2.2 for all values of k.

In the clustering literature, the line formed by the blue dots in Figure 3 is referred to as a "scree plot". We see that this line is a decreasing convex function of k. Kinks in this scree plot, or transformations of it, are frequently used to try to identify the "natural" number of clusters present in the data (see Tibshirani, Walther, and Hastie [2001] for further discussion). Most of the population of Gifu during this period consisted of subsistence farmers, and it appears unlikely that the mountain ranges of Gifu were designed with the intent that farmland should be naturally clustered into municipalities with a certain fixed cost γ_1. We would thus expect that no kink should be present in Figure 3, as there should not be natural clusters in our data (c.f. Trebbi and Weese [2019], where natural clusters are expected).

Suppose that we scale distance cost by α. The benefit $\alpha(\text{WSS}(\pi_k^{\text{FB}}) - \text{WSS}(\pi_{k+1}^{\text{FB}}))$ of adding an additional coalition is decreasing in k. The cost of adding an additional coalition is fixed at γ_1. Optimality thus requires that

$$\text{WSS}(\pi_{k^0}^{\text{FB}}) - \text{WSS}(\pi_{k^0+1}^{\text{FB}}) < \gamma_1/\alpha < \text{WSS}(\pi_{k^0-1}^{\text{FB}}) - \text{WSS}(\pi_{k^0}^{\text{FB}}), \tag{21}$$

where $k^0 = 289$ is the number of coalitions actually observed in the data. So long as Figure 3 does not have a kink, as the number of players (and coalitions) becomes large the inequality on the left and the inequality on the right will converge, and α will be point identified relative to the fixed cost γ_1.

As might be expected, there are computational issues that prevent using this approach exactly as presented. Specifically, while the Hartigan and Wong [1979] algorithm has good performance it will not return exactly the globally optimal partition, and thus in general there will be some noise in the calculation of $\text{WSS}(\pi_{k^0}^{\text{FB}}) - \text{WSS}(\pi_{k^0+1}^{\text{FB}})$ and other such differences. Appendix Figure B.8 shows this calculation for all possible values of k, giving some idea of the magnitude of the noise introduced by non-globally optimal solutions from Hartigan and Wong [1979].

To reduce the effect of this computational noise, rather than simply examine $\text{WSS}(\pi_k^{\text{FB}})$ for $k \in \{288, 289, 290\}$, we suppose that there is a smooth decreasing convex function $g(k) = \text{WSS}(\pi_k^{\text{FB}})$, and then model this function using the Liao and Meyer [2019] implementation of a Meyer [2013] constrained generalized additive model, and data from $k \in \{89, ..., 489\}$. Appendix Figure B.9 shows the results of this analysis. As confirmation, we graph the resulting estimate $\alpha = 0.11$ (jackknife s.e. 0.02) as the black line on Figure 3 and observe that it appears tangent to the blue dot curve at around $k^0 = 289$, as would be desired.

7 Counterfactual Simulations

Our interest is in comparing π^{FB} with partitions that would form in the case of a decentralized coalition formation game. It is not theoretically obvious what the result will be here. A general theme in the literature on local public good coalition formation games is that while games with only a small number of players may feature undesirable behaviour, these problems disappear as the game becomes "large". This has been formalized by Kaneko and Wooders [1986] along with many others, and is essential to the sorting results in the Tiebout model.

We might think that having $N = 1111$ players would make our game "large", but this will in fact turn out not to be the case. Substantial inefficiency arises because although there are many players, it is difficult to find a willing merger partner that is attractive from both a horizontal and a vertical perspective: as a result, core partitions generally feature small and geographically discontiguous mergers. We then check whether this inefficiency persists in the case where there is *only* horizontal heterogeneity. We find that – as predicted by Bogomolnaia, Breton, et al. [2007] and others – discontiguous coalitions continue to emerge even with only horizontal heterogeneity. However, from an empirical perspective, there are only a tiny number of these coalitions in this case and the resulting inefficiency is minimal.

We then show via additional simulations that the inefficiency in the model with both vertical and horizontal heterogeneity only emerges because the public good is congestible. If the variable cost of local public good provision is subsidized, inefficiency falls to close to zero. This type of subsidy was in fact present in Japan during a set of decentralized mergers in the 1950s, but not during a later set of decentralized mergers in the 2000s. As predicted by our model, the former set of mergers was generally regarded as achieving its goals while the latter set was criticized for having mergers that were too small and geographically discontiguous.

To obtain all these results, we use parameter estimates that were obtained under the assumption that the observed partition was chosen by a benevolent central planner. We thus cannot use the model estimates to show that the central planned Meiji mergers were socially optimal: this is an assumption rather than a conclusion of the model. What we can investigate is whether decentralized mergers would have matched the outcome desired by our central planner. We find that this is the case only when there is either no vertical heterogeneity or no congestibility.

7.1 Basic results

Results are shown in Tables 4 and 5. The rows of these tables show two different values of α, used to scale the importance of the heterogeneity parameters $\hat{\beta}$ relative to the costs of running a jurisdiction. Our estimate from Section 6.2 is $\alpha = 0.11$. We also consider the case where there is no heterogeneity cost, that is, $\alpha = 0$. This latter case corresponds to the model in Farrell and Scotchmer [1988]: from the social planner's perspective, the grand coalition is optimal, but differences in tax base will lead to partitions with multiple coalitions

Table 4: Simulation Results: Number of Municipalities

α	I (social planner)	II (actual tax base)	III (equal tax base)	IV (no variable cost)	V (subsidized fixed cost)
0	1	60	1	1	99
0.11	288	576.2	430.9	420.0	913.4
		(3.3)	(3.3)	(2.4)	(1.3)

Estimated standard deviations of partitions within solution set are given in parentheses were there are multiple solutions to the coalition formation game. The social planner's optimal partition is unique, as is the decentralized outcome where there is no distance cost.

Table 5: Simulation Results: Inefficiency (¥1000)

α	I (social planner)	II (actual tax base)	III (equal tax base)	IV (no variable cost)	V (subsidized fixed cost)
0	0	10.9	0	0	18.0
0.11	0	60.4	3.8	1.7	67.2
		(0.7)	(0.1)	(0.02)	(0.2)

Estimated standard deviations of partitions within solution set are given in parentheses were there are multiple solutions to the coalition formation game. The social planner's optimal partition is unique, as is the decentralized outcome where there is no distance cost.

in the decentralized case. Below, when we refer to columns of Tables 4 and 5 we are referring to the results in the $\alpha = 0.11$ row unless we specify otherwise.

Column I in Table 4 describes the partition that would be chosen by the social planner, computed based on the technique given in Section 2.2. Columns II-V give the characteristics of core partitions in a decentralized coalition formation process under various conditions. These are generated by running Algorithm 1, with $\hat{\beta}$ based on Column XI of Table 2. In order to discover different partitions in the core, we rerun Algorithm 1 a total of 100 times, each time changing the weights w that appear in Equation 8. For each run, we randomly generate weights by drawing $w_i \sim \text{Normal}(0, 1)$ for each player i. We do not include any idiosyncratic error term in these simulations: this follows other applied work that uses maximum score estimates including Baccara et al. [2012] and Fox and Bajari [2013].

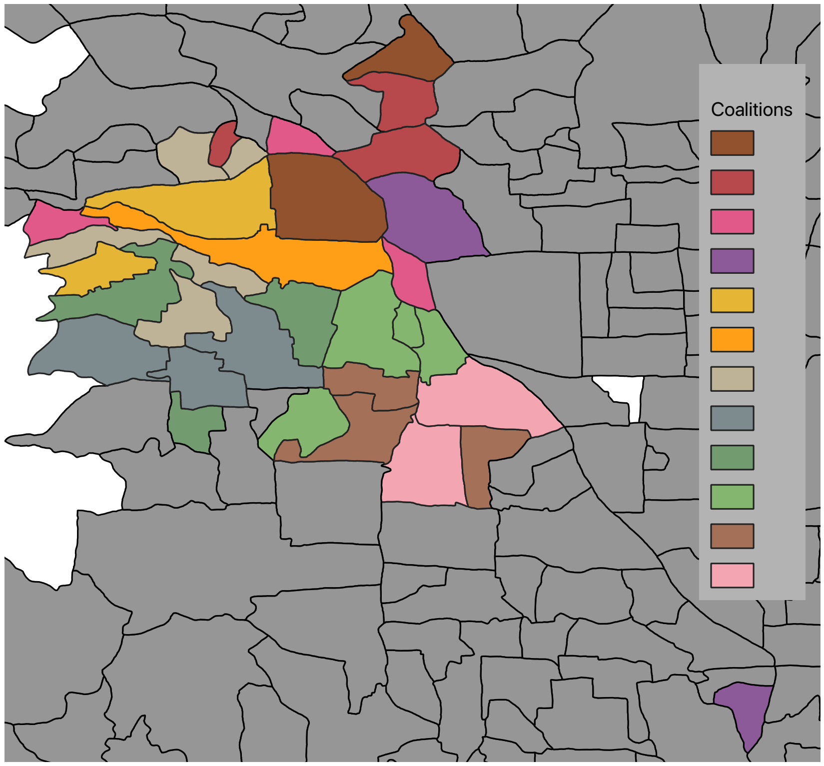
Table 5 compares the partitions obtained in each case considered in Table 4 to the central planner's preferred partition: Column I in Table 5 thus reports zero inefficiency by definition. A potential difficulty here is that the Hartigan and Wong [1979] algorithm that we use to generate the central planner's preferred partition gives a partition that is a local optimum, but is not guaranteed to be a global optimum. To the extent that the partition we compute is not the true global optimum, the differences reported in Columns II-V of Table 5 will be underreported by some constant. We do not believe this source of error is substantial for two reasons. First, increasing the number of random restarts of the Hartigan and Wong [1979], even up to 10000, does not substantially change the results. Second, recent research such as Slonim, Aharoni, and Crammer [2013] shows good performance for the Hartigan and Wong [1979] algorithm, particularly when using Arthur and Vassilvitskii [2007] "kmeans++" starting locations.

Column II of Table 4 shows that in the decentralized case there are twice as many coalitions compared to the partition selected by the central planner. Figure 4 gives an example of the type of coalition structure that arises in this decentralized coalition formation game. We see that, with the exception of a singleton coalition and one pair, all coalitions depicted are geographically discontiguous. Column II of Table 5 reports average inefficiency for this case of approximately ¥60,000, which is equivalent to 13.5% of total spending on local public goods in our model.

Column III of Tables 4 and 5 shows the characteristics of partitions that arise if the per capita tax base, y/p, were set to be the same for all players in the game.[15] This is the case where there is horizontal heterogeneity, but no vertical heterogeneity. We see that even in

[15] It is easiest to make y/p equal by changing p rather than changing y. This is because y appears in the central planner's optimization problem in Equation 11 but p does not, and thus changing y would change the central planner's solution, whereas even after a change in p Column I describes the social planner's optimum. We set $p_i = y_i \sum_{i'=1}^{N} p_{i'} / \sum_{i'=1}^{N} y_{i'}$.

Figure 4: Decentralized mergers with actual per capita tax base



this case the partitions that arise as a result of the decentralized coalition formation process are not exactly the same as the one that would be chosen by the social planner. We also verify that the social planner's partition is not part of the core.

Figure 5 shows a randomly selected core partition from Column III. From a qualitative point of view, the coalitions displayed in Figure 5 appear reasonable from the social planner's perspective; however, we see that one coalition (orange) on the southern side of the area of interest with three members is not geographically contiguous and appears somewhat inexplicable. In the theoretical literature coalitions of this type are described as "nonconsecutive" [Bogomolnaia, Breton, et al. 2007]. Further investigation reveals that the members of this coalition in Figure 5 appear to be "leftovers": each player in the coalition would like to instead join one of the "regular" coalitions that it is geographically adjacent to, but there is at least one player in each of these coalitions that is opposed to this move. The small amount of inefficiency reported in Column III of Table 5 appears to be due in part to the very low payoffs experienced by a small number of "leftover" players. In Appendix Figure B.12, however, we display an entire core partition, and show that there appear to be only two coalitions of this type in the entire partition. Inefficiency reported in Column III of

Figure 5: Decentralized mergers with equal per capita tax base

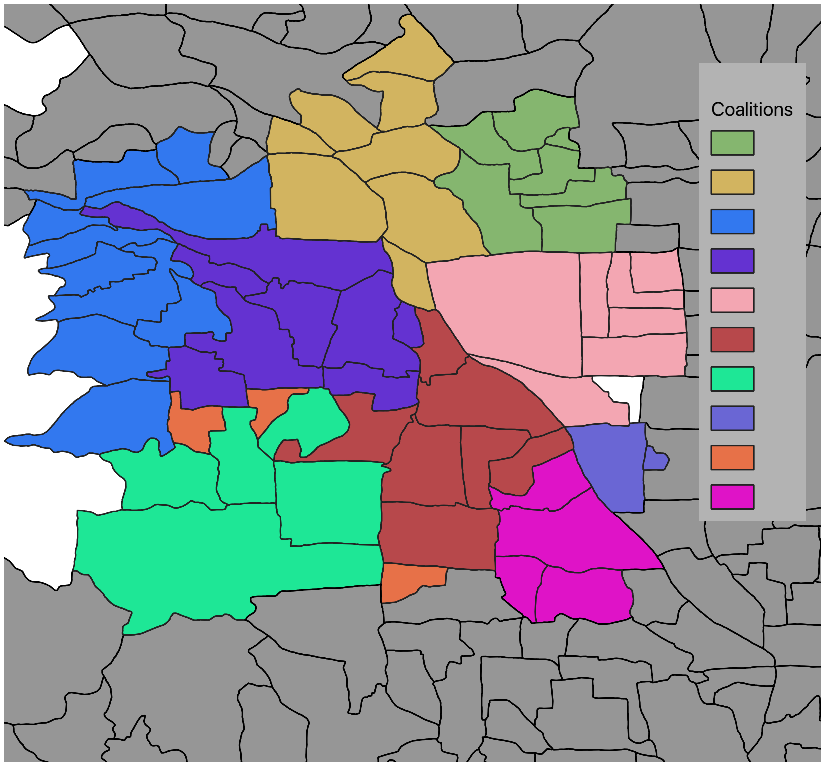


Table 5 is approximately 1/16th the inefficiency reported in Column II, equivalent to less than 1% of total spending on local public goods. We thus see that in the case with only horizontal heterogeneity, discontiguous coalitions are a theoretical concern but turn out to not be empirically important.

An alternative quantification of the difference between the coalitions in Columns II and III can be obtained by reusing the permutation approach previously employed to estimate $\hat{\beta}$. Consider swapping the coalition membership of players i and i' in a core partition, and calculating the central planner's payoff for the new partition. If the starting partition is π^{FB}, then this swap will never improve the planner's payoff. For other partitions, however, there may sometimes be an improvement. We consider all potential pairwise swaps for 1111 players in the core partitions shown in Figures 4 and 5. For the partition in Figure 4, 58% of players have such a swap, and for those players there is an average of 3.2 such swaps. For the partition in Figure 5, only 15% of the players have any swap that would improve the planner's payoff, and for those players there is only an average of 1.8 such swaps. Moreover, the welfare improvement from such a swap in this case is on average 10.5 times higher for the cases in Figure 4 than for those in Figure 5.

Comparing Columns II and III in Table 5, we see that the main source of inefficiency appears to be that described by Farrell and Scotchmer [1988]. When players differ in a verti-

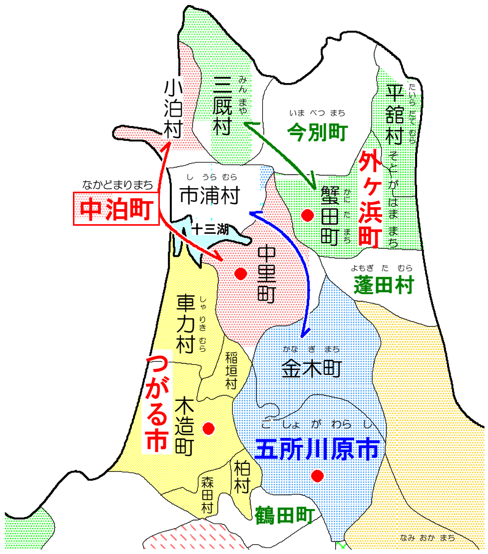


Figure 6: Actual decentralized merger outcome, Aomori Prefecture, c. 2006.

Map colours indicate members of the same coalition. Japanese text gives the names of pre-merger municipalities (in black) and post-merger municipalities (in colour). Arrows join members of non-contiguous coalitions.
Source: Geospatial Information Authority of Japan

cal characteristic their preferences regarding merger partners are determined (partly, in our case) by this vertical characteristic. Players thus trade off heterogeneity in the horizontal characteristic in exchange for similarity in the vertical characteristic. In the model in Section 2, the same amount in total is being spent on providing services for any configuration involving the same number of coalitions. A willingness to accept coalition partners that are a worse match on the horizontal characteristics in exchange for being a better match on the vertical characteristic therefore creates inefficiency from a (utilitarian) social perspective. The small but non-zero amount of inefficiency present in Column III suggests why simulations using actual data and parameter estimates were required in order to reach this conclusion: a small amount of inefficiency is still present even when there is no vertical differentiation, and thus without actual data a theoretical model is unlikely to clarify the importance of vertical heterogeneity to inefficiency.

At this point, a potential objection is that the above discussion is based on the simulated mergers shown in Figure 4, but these mergers are unreasonable and in fact only serve to show that the model of Section 2 is an inappropriate model of municipal mergers. In particular, many of the mergers predicted are not geographically contiguous. Should any realistic model of municipal mergers really be predicting these sorts of configurations?

Figure 6 shows a set of decentralized municipal mergers that actually occurred in Aomori Prefecture around 2006. These mergers were part of the *Heisei daigappei* set of municipal mergers. The figure shows three geographically discontiguous mergers, involving in total eight geographically contiguous municipalities. In the centralized mergers of the Meiji period, these sorts of arrangements are not observed in the data. Thus, bizarre configurations of the sort shown in Figure 4 can and do occur in reality, and they do so only during decentralized mergers (Heisei) and not centralized ones (Meiji). Furthermore, the qualitative explanation offered for the outcome shown in Figure 6 is also in line with the model in Section 2: the municipalities shown on the map differed substantially in levels of indebtedness and revenue sources, and the observed mergers were the result of an attempt to avoid matching with undesirable but neighbouring municipalities.

Another potential objection at this point is that the above results are due to the fact that our coalition formation model uses feudal villages as the players, but these are relatively large: in the central planner's preferred partition, a municipality on average consists of only 3.8 feudal villages. In contrast, results in the literature on Tiebout sorting generally assume that sorting can occur at the household level. In Appendix F we illustrate a simple way to change parameters in order to check whether our results regarding inefficiency are driven by the small number of players per coalition. We find that they are not, which suggests that our findings may be generally applicable to games where there is Tiebout sorting with both horizontal and vertical heterogeneity. We now investigate what particular aspects of our game are essential to generating the inefficiency results that we have just reported above.

7.2 Importance of Congestibility

Column IV of Table 4 shows results where there is both vertical and horizontal heterogeneity but $\gamma_2 = 0$. In this case the cost of running a jurisdiction is the same regardless of the number of residents. We see that here core partitions in the decentralized game are very close to the partition that would have been chosen by the central planner: inefficiency reported in Column IV of Table 5 is less than half of the already tiny amounts reported in Column III.

The intuition for this result is that, with only a fixed cost of providing services, adding any member to a coalition will decrease the contribution required from existing members. If there were no horizontal differentiation, this case would correspond exactly to one considered in Farrell and Scotchmer [1988] where there is no inefficiency. The presence of horizontal heterogeneity creates a theoretical potential for inefficiency, but empirically this turns out not to be important. Thus, in an environment in which public goods can be provided with only a fixed cost, it appears that decentralized mergers are unlikely to cause any particular problems. Congestibility of the public good is key to our inefficiency results, further underscoring the importance of obtaining parameters based on data. A pure public good yields very different simulation results; however, even the most casual empiricism suggests that the vast majority of local government spending is not on pure public goods.

A potential objection at this point is that the different results we obtain for $\gamma_2 = 0$

might suggest that our simulation results in general are sensitive to changes γ_2, which is not identified in the main econometric model of Section 4 but was rather calibrated to auxiliary spending data in Section 6. To deal with this concern we perform a Monte Carlo exercise where we randomly generate games with $N = 7$ players, and then compute the core of these games at different parameter values. Appendix Figure B.10 presents inefficiency over a two dimensional parameter space of potential values for γ_1 and γ_2. We see there is little inefficiency when either γ_1 or γ_2 is small. If γ_1 is small then there is little cost to forming additional jurisdictions, and the all-singleton partition is both the decentralized outcome and the central planner's optimum. If γ_2 is small then even if a potential merger partner is poor the post-merger per capita cost of running the jurisdiction will still be lower, and thus vertical stratification is largely avoided: this replicates the result in Column IV of Table 5. The values of γ_1 and γ_2 corresponding to our choice in Section 6.1 are shown in Figure B.10 by the black triangle. We see that at this point, the inefficiency gradient is such that γ_1 is much more important in determining total inefficiency than γ_2. This is reassuring, because γ_1 is part of the planner's problem in Equation 11 and can thus be estimated as part of the model in Section 6.1. In contrast, γ_2 is calibrated from auxiliary data. Figure B.10 shows us that it is important that γ_2 is not zero (or tiny), but that beyond that, at our value of γ_1 the precise value of γ_2 is not particularly important. Appendix Figure B.11 shows that this result does not appear to depend on the number of players N used in our Monte Carlo simulation.

We chose to analyze data from the Meiji period in part because there was no transfer scheme from the national government to municipalities, and this lack of transfers allowed for a particularly simple model in Section 2. An intergovernmental transfer system was developed later, however, in the post-war period. At the risk of oversimplification, this scheme originated mainly as a subsidy on the variable cost of providing public services, reducing γ_2 from the perspective of municipalities. Over several decades, this scheme mutated into one that instead subsidized the *fixed* cost of providing public services, reducing γ_1 (instead of γ_2) from the perspective of municipalities. These two periods correspond with the two post-Meiji waves of municipal mergers in Japan: the "Showa" mergers occurred when variable cost was subsidized, and the "Heisei" mergers, fixed cost. The transfer situation in the Showa period corresponds roughly to the simulations shown in Column IV. We perform a final set of simulations, shown in Column V, to see how the change in subsidy scheme would have changed the pattern of decentralized mergers. The number of coalitions is far higher than in any of the other columns, and indeed reaches the resolution of the data, in the sense that most municipalities remain as singletons, with only the smallest participating in mergers at all. Appendix G provides details about the transfer schemes during these periods, and the simulations performed.

The intuition for the result shown in Column V is straightforward: a transfer payment scheme equivalent to part of the fixed cost γ_1 reduces the incentive to merge from the perspective of the municipality. In contrast, a transfer payment scheme equivalent to part of the variable cost γ_2 reduces the importance of the differences in per capita tax base. Thus, the former case results in municipalities choosing a decentralized merger pattern that is even less desirable, from the perspective of the social planner, while the latter case results in mergers that are quite close to the social optimal. Qualitative evidence from the merger waves in question supports these results: In the official government evaluation of the Heisei mergers,[16] the reluctance of municipalities to merge is noted. In contrast, in the Showa mergers, the number of municipalities was reduced by 6152, very close to the targeted reduction of 6273 [Yoshitomi 1960].[17] It thus appears that, in contrast to the equalization payments offered during the Showa period, the type of subsidy provided during the Heisei period results in substantial problems when considering decentralized mergers.

7.3 Economic Significance

A quantitative interpretation of the amount of inefficiency displayed in Table 5 is challenging, because local government during the Meiji period began as a very small portion of GDP but then grew quickly as Japan industrialized. Table 1 in Nishikawa, Hayashi, and Weese [2018] suggests that the units used in Table 5 correspond roughly to the annual salary of a senior government official. The inefficiency reported in Column V of Table 5 is equivalent to over 20% of total administrative expenses throughout Gifu, or alternatively about 150% of the amount spent on the police by Gifu prefecture.[18] The calculated amounts of inefficiency are thus important when considered in the context of the size of Meiji-era local governments, even though they are not large in absolute terms.

An alternative interpretation could be obtained by scaling up the inefficiency reported in Table 5 by the difference in expenditures on public services between the Meiji period and the present. The fixed costs γ_1 for 289 municipalities represent 11.75% of the total expenditures on local public services in Gifu predicted by our model, and the inefficiency reported in Column II of Table 5 is 1.2 times as large as this (equivalent to 13.5% of local government spending). Local government spending as a percentage of GDP in Japan is difficult to

[16] *"Heisei no Gappei" no Hyouka-Kenshou-Bunseki* ("Evaluation, Inspection, and Analysis of the 'Heisei Mergers"').

[17] This result also suggests why there it is still unresolved whether the Showa mergers were centralized or decentralized: it does not matter, because the outcome would have been the same. See Appendix G for discussion.

[18] See Tables 9 and 5 of Nishikawa, Hayashi, and Weese [2018], respectively.

calculate due to the extremely large transfer system. The OECD reports that subnational government spending was 17% for Japan in 2014, coincidentally equal to the average across all OECD countries. Taking this number as given, multiplication then suggests annual inefficiency of 2.25% of GDP.

8 Conclusion

In this paper we showed that a decentralized process of jurisdiction formation in an environment with both vertical and horizontal heterogeneity results in smaller coalitions as well as (socially) suboptimal matching on the horizontal characteristics. On the other hand, a pattern of boundaries very similar to what would be chosen by a utilitarian social planner emerges when players differ only in horizontal characteristics. We obtained these results using historical data on Japanese municipal mergers, a model that expresses the merger process as a fractional hedonic game, a new method for estimating the preference parameters of a single decision maker choosing a partition, and a new method that makes it possible to calculate solutions for large fractional hedonic games through binary integer programming.

The external validity of simulations regarding 19th century Japanese farming villages is open to debate, and we did not attempt a formal extension of our results to the international arena. Consider very briefly, however, the stance of various countries regarding the existence of a "right to self-determination". In the United Kingdom, it is generally accepted that Scotland can hold an independence referendum. In Canada, a supreme court decision and federal legislation has established that a successful Quebec independence referendum would force the rest of Canada to negotiate a separation agreement. On the other hand, the Spanish government denies the existence of any such rights for the autonomous communities of Spain. Perhaps not coincidentally, the potentially separatist regions of the United Kingdom and Canada are by most calculations not particularly rich, while the Basque Country and Catalonia are substantially better off than the remainder of Spain. Our results suggest that the positions regarding self-determination held by each of these countries may be welfare-maximizing: decentralized decision-making results in partitions that are very close to the social optimum when inequality between regions is low, but not when inter-regional inequality is high.

We believe that fractional hedonic games are an appropriate way of modelling mergers and splits of political jurisdictions. Fractional hedonic games have attracted interest both inside and outside of economics, but suffer from the potential non-existence of stable partitions in the same way as many other hedonic games. In this paper we considered a particular form of fractional hedonic game, and found that empty cores are observed at a less than five per million rate in random games on a plane, even when the parameters for these games are most favourable for non-existence. This suggests that the particular type of fractional hedonic game that we are considering, while still theoretically at risk of having an empty core, has a stable partition with probability very close to one.

Thus, despite potential theoretical difficulties with the "right to self-determination" as operationalized in this paper empirically there is almost always a non-empty core. Furthermore, core partitions can be reached by a sequence of myopic deviations. The ongoing legal and philosophical debates regarding self-determination are thus relevant, as the existence of such a right would not result in instability and endless cycling, but would instead lead to stable partitions.

Using the techniques developed in this paper, it may now be possible to analyze quantitatively some open questions that previously seemed amenable only to theoretical analysis. For example, it is frequently asserted that because Londoners are not a "people", London and other such urban agglomerations should not possess a right to self-determination. Consider a model in which at the beginning of time a constitutional rule must be set for what units are allowed to exercise a right to self-determination: one possibility is that only units corresponding to a historical geographic region (or ethnic group) have this right, while another possibility is that any arbitrary geographic unit can exercise this right. We could then use historical data on changes in population and per capita income across space and time, and consider how these two different rules would perform across time. Based on the results presented above, it seems likely that areas that become wealthy would frequently want to secede from their surroundings, and that more reasonable boundaries (and thus higher welfare) would be obtained by restricting the right to self-determination to predefined large geographic regions.

In addition to political jurisdictions, with some modification it may also be possible to analyze other phenomena using the fractional hedonic model of this paper. Ethnic or linguistic groups, for example, could be considered in this framework, and have recently received extensive theoretical analysis (e.g. Ginsburgh and Weber [2011]). Empirical analysis of choice of language or ethnicity has often focussed on individual decisions,[19] but reduced-form results such as those in Michalopoulos [2012] suggest that analysis of identity formation at the group level might also be informative. Weese [2016] provides an early attempt at this sort of analysis.

Even further afield, fractional hedonic games might also be used to model the formation of students into schools or classes, workers into unions, or public employees into different pension funds. In these cases substantial changes to the model presented in this paper would

[19] See Clingingsmith [2014] for example, or Jia and Persson [2015] and the references therein.

likely be required, but the basic approach presented should still be applicable.

References

Acemoglu, Daron (Dec. 2003). "Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics". *J. Comp. Econ.* 31 (4):620–652.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby (Apr. 2004). "Political Jurisdictions in Heterogeneous Communities". *J. Polit. Econ.* 112 (2):348–396.

Alesina, Alberto and Enrico Spolaore (Nov. 1997). "On the Number and Size of Nations". *Q. J. Econ.* 112 (4):1027–1056.

Arthur, David and Sergei Vassilvitskii (Jan. 2007). "k-means++: the advantages of careful seeding". *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. SODA '07. USA: Society for Industrial and Applied Mathematics, pp. 1027–1035.

Aziz, Haris, Felix Brandt, and Paul Harrenstein (May 2014). "Fractional Hedonic Games". *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. Paris.

Aziz, Haris, Serge Gaspers, et al. (June 22, 2015). "Welfare Maximization in Fractional Hedonic Games". *Twenty-Fourth International Joint Conference on Artificial Intelligence*. Twenty-Fourth International Joint Conference on Artificial Intelligence.

Baccara, Mariagiovanna et al. (Aug. 2012). "A Field Study on Matching with Network Externalities". *American Economic Review* 102 (5):1773–1804.

Banerjee, Suryapratim, Hideo Konishi, and Tayfun Sönmez (Jan. 8, 2001). "Core in a Simple Coalition Formation Game". *Soc. Choice Welf.* 18 (1):135–153.

Barros, Ana Isabel (1998). *Discrete and Fractional Programming Techniques for Location Models*. Dordrecht; Boston: Kluwer Academic Publishers.

Bogomolnaia, Anna, Michel Le Breton, et al. (Jan. 1, 2007). "Stability under Unanimous Consent, Free Mobility and Core". *Int J Game Theory* 35 (2):185–204.

Bogomolnaia, Anna and Matthew O. Jackson (Feb. 2002). "The Stability of Hedonic Coalition Structures". *Games Econ. Behav.* 38 (2):201–230.

Brandl, Florian, Felix Brandt, and Martin Strobel (2015). "Fractional Hedonic Games: Individual and Group Stability". *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. AAMAS '15. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1219–1227.

Brasington, David M. (1999). "Joint Provision of Public Goods: The Consolidation of School Districts". *J. Public Econ.* 73 (3):373–393.

Christaller, Walter (1933). *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Wissenschaftliche Buchgesellschaft. 352 pp.

Chung, Kim-Sau (2000). "On the Existence of Stable Roommate Matchings". *Games and Economic Behavior* 33 (2):206–230.

Clingingsmith, David (Jan. 1, 2014). "Industrialization and Bilingualism in India". *J. Human Resources* 49 (1):73–109.

Demange, Gabrielle (Jan. 1994). “Intermediate Preferences and Stable Coalition Structures”. *J. Math. Econ.* 23 (1):45–58.

Desmet, Klaus et al. (Sept. 1, 2011). “The Stability and Breakup of Nations: A Quantitative Analysis”. *J Econ Growth* 16 (3):183–213.

Diermeier, Daniel, Hülya Eraslan, and Antonio Merlo (2003). “A Structural Model of Government Formation”. *Econometrica* 71 (1):27–70.

Donder, Philippe De, Michel Le Breton, and Eugenio Peluso (July 1, 2012). “On the (Sequential) Majority Choice of Public Good Size and Location”. *Soc Choice Welf* 39:457–489 (2-3).

Dreze, J. H. and J. Greenberg (May 1980). “Hedonic Coalitions: Optimality and Stability”. *Econometrica* 48 (4):987–1003.

Drèze, Jacques et al. (Nov. 2008). ““Almost” Subsidy-Free Spatial Pricing in a Multi-Dimensional Setting”. *Journal of Economic Theory* 143 (1):275–291.

Elomri, A. et al. (Oct. 17, 2013). “Stability of Hedonic Coalition Structures: Application to a Supply Chain Game”. *Applications of Multi-Criteria and Game Theory Approaches: Manufacturing and Logistics*. Ed. by Lyes Benyoucef, Jean-Claude Hennet, and Manoj Kumar Tiwari. Springer Science & Business Media, pp. 337–364.

Farrell, Joseph and Suzanne Scotchmer (May 1988). “Partnerships”. *Q. J. Econ.* 103 (2):279–297.

Fox, Jeremy T. (2007). “Semiparametric estimation of multinomial discrete-choice models using a subset of choices”. *The RAND Journal of Economics* 38 (4):1002–1019.

— (2018). “Estimating matching games with transfers”. *Quantitative Economics* 9 (1):1–38.

Fox, Jeremy T. and Patrick Bajari (Feb. 2013). “Measuring the Efficiency of an FCC Spectrum Auction”. *American Economic Journal: Microeconomics* 5 (1):100–146.

Fox, Jeremy T. and David Santiago (Feb. 2015). *A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Subsampling Inference*.

Ginsburgh, Victor and Shlomo Weber (2011). *How Many Languages Do We Need?: The Economics of Linguistic Diversity*. Princeton: Princeton University Press. 232 pp.

Gomory, Ralph E. (Apr. 1994). “A Ricardo Model with Economies of Scale”. *Journal of Economic Theory* 62 (2):394–419.

Gordon, Nora and Brian Knight (June 2009). “A Spatial Merger Estimator with an Application to School District Consolidation”. *J. Public Econ.* 93:752–765 (5-6).

Greenberg, Joseph and Shlomo Weber (Feb. 1986). “Strong Tiebout Equilibrium under Restricted Preferences Domain”. *Journal of Economic Theory* 38 (1):101–117.

Greenberg, Joseph H. (Jan. 1956). “The Measurement of Linguistic Diversity”. *Language* 32 (1):109–115.

Gregorini, Filippo (2009). *Political Geography and Income Inequalities*. Working Paper 152. University of Milano-Bicocca, Department of Economics.

Hartigan, J. A. and M. A. Wong (1979). “Algorithm AS 136: A K-Means Clustering Algorithm”. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1):100–108.

Hathaway, Richard J. and James C. Bezdek (Mar. 1, 1994). “Nerf c-means: Non-Euclidean relational fuzzy clustering”. *Pattern Recognition* 27 (3):429–437.

Hirota, Haruaki and Hideo Yunoue (Oct. 2, 2014). “Municipal Mergers and Special Provisions of Local Council Members in Japan”. *The Japanese Political Economy* 40:96–116 (3-4).

Jia, Ruixue and Torsten Persson (Nov. 29, 2015). *Individual vs. Social Motives in Identity Choice: Theory and Evidence from China*.

Kaneko, Mamoru and Myrna Holtz Wooders (Oct. 1986). "The core of a game with a continuum of players and finite coalitions: The model and some results". *Mathematical Social Sciences* 12 (2):105–137.

Kim, Jeankyung and David Pollard (1990). "Cube Root Asymptotics". *The Annals of Statistics* 18 (1):191–219. Publisher: Institute of Mathematical Statistics.

Kojima, Fuhito, Parag A. Pathak, and Alvin E. Roth (Nov. 1, 2013). "Matching with Couples: Stability and Incentives in Large Markets*". *Q J Econ* 128 (4):1585–1632.

Land, A. H. and A. G. Doig (1960). "An automatic method for solving discrete programming problems". *Econometrica* 28 (3):497–520.

Liao, Xiyue and Mary C. Meyer (May 10, 2019). "cgam: An R Package for the Constrained Generalized Additive Model". *Journal of Statistical Software* 89 (1):1–24.

Lieberson, Stanley (Oct. 1964). "An Extension of Greenberg's Linguistic Diversity Measures". *Language* 40 (4):526–531.

Manski, Charles F. (1975). "Maximum Score Estimation of the Stochastic Utility Model of Choice". *J. Econom.* 3 (3):205–228.

Matsuzawa, Yūsaku (2013). *Chōson gappei kara umareta Nihon kindai : Meiji no keiken*. Kōdansha sensho mechie 563. [Municipal Mergers and the birth of Modern Japan: the Meiji experience]. Tokyo: Kabushiki Kaisha Kōdansha.

Meyer, Mary C. (Sept. 2013). "Semi-parametric additive constrained regression". *Journal of Nonparametric Statistics* 25 (3):715–730.

Michalopoulos, Stelios (June 2012). "The Origins of Ethnolinguistic Diversity". *Am. Econ. Rev.* 102 (4):1508–39.

Miyazaki, Takeshi (Nov. 2014). "Municipal Consolidation and Local Government Behavior: Evidence from Japanese Voting Data on Merger Referenda". *Econ Gov* 15 (4):387–410.

Nishikawa, Masashi, Masayoshi Hayashi, and Eric Weese (Feb. 2018). "Meiji Era Local Government". *Journal of Economics & Business Administration* 217 (2):101–125.

Pápai, Szilvia (Aug. 2004). "Unique stability in simple coalition formation games". *Games and Economic Behavior* 48 (2):337–354.

Pinar, M. C. and A. Camci (Jan. 2012). "An Integer Programming Model for Pricing American Contingent Claims under Transaction Costs". *Computational Economics* 39 (1):1–12.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf (1999). *Subsampling*. Springer Series in Statistics. New York: Springer-Verlag.

Ray, Debraj and Rajiv Vohra (Mar. 1997). "Equilibrium Binding Agreements," *J. Econ. Theory* 73 (1):30–78.

Reiter, Michael and Alfons Weichenrieder (1997). "Are Public Goods Public? A Critical Survey of the Demand Estimates for Local Public Services". *Finanz. Public Finance Anal.* 54 (3):374–408.

Roth, Alvin E. and Elliott Peranson (Sept. 1999). "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design". *American Economic Review* 89 (4):748–780.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (June 2007). "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences". *Am. Econ. Rev.* 97 (3):828–851.

Roth, Alvin E. and John H. Vande Vate (Nov. 1990). "Random Paths to Stability in Two-Sided Matching". *Econometrica* 58 (6):1475–1480.

Roth, V., J. Laub, et al. (Dec. 2003). "Optimal cluster preserving embedding of nonmetric proximity data". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (12):1540–1551.

Serafini, Paolo (Mar. 2012). "Allocation of the EU Parliament Seats via Integer Linear Programming and Revised Quotas". *Mathematical Social Sciences* 63 (2):107–113.

Sherman, Robert P. (1993). "The Limiting Distribution of the Maximum Rank Correlation Estimator". *Econometrica* 61 (1):123–137.

Slonim, Noam, Ehud Aharoni, and Koby Crammer (Aug. 2013). "Hartigan's K-means versus Lloyd's K-means: is it time for a change?" *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie (2001). "Estimating the number of clusters in a data set via the gap statistic". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2):411–423.

Trebbi, Francesco and Eric Weese (2019). "Insurgency and Small Wars: Estimation of Unobserved Coalition Structures". *Econometrica* 87 (2):463–496.

Weese, Eric (July 2015). "Political Mergers as Coalition Formation: An Analysis of the Heisei Municipal Amalgamations". *Quantitative Economics* 6 (2):257–307.

— (Sept. 2016). *European Political Boundaries as the Outcome of a Self-Organizing Process*. Discussion Paper 1629. Graduate School of Economics, Kobe University.

Yoshitomi, Shigeo (1960). *Chihō jichi : jittai to tenbō*. [Local government autonomy: Current condition and prospects]. Tokyo: Keisō Shobō.

Online Appendices – Not For Publication

A Simulation Details

A.1 Derivation of Equation 8

For coalition S' to be a blocking coalition for the partition π, it must be the case that $v_i(\pi) < v_i(S')$ for all i in S'. For a given π, we are interested in finding such an S', or proving that none exists and thus π is in the core. Leaving $v_i(\pi)$ as is, and using Equation 5 to expand $v_i(S')$, we obtain

$$v_i(\pi) < -\frac{y_i}{Y_{S'}}(\gamma_1 + \gamma_2 \sum_{i' \in S'} p_i) - y_i \sum_{i' \in S'} \frac{y_{i'}}{Y_{S'}} d(i, i'). \tag{22}$$

Both terms on the right in this inequality have a $y_i/Y_{S'}$ term. We divide both sides by this, obtaining

$$v_i(\pi)\frac{Y_{S'}}{y_i} < -(\gamma_1 + \gamma_2 \sum_{i' \in S'} p_i) - \sum_{i' \in S'} y_{i'} d(i, i'), \tag{23}$$

which (after rearrangement) is the constraint in (8).

Now let z be a vector of binary variables of length N, with $z_i = 1$ if $i \in S'$, and zero otherwise. In Inequality 23, the expression $Y_{S'} = \sum_{i' \in S'} y_{i'}$ can then be rewritten as $\sum_{i'} z_{i'} y_{i'}$. Similarly, $\sum_{i' \in S'} p_i = \sum_{i'} z_{i'} p_{i'}$ and $\sum_{i' \in S'} y_{i'} d(i, i') = \sum_{i'} z_{i'} y_{i'} d(i, i')$. The only variable in Inequality 23 that is not known is then z. Consider the following set of (pairs of) disjunctive constraints:

$$\forall i \text{ either Inequality 23 holds or } z_i = 0. \tag{24}$$

This is one pair of constraints per player, for a total of N pairs of constraints.

The problem of finding a coalition S' that is a blocking coalition for the partition π is thus equivalent to the problem of finding a vector z that satisfies the restriction given in (24). By adding a vector of weights w to Inequality 23 and maximizing the sum $\sum_{i \in S'} w_i = \sum_i z_i w_i$, we turn the problem of finding a blocking coalition into an optimization problem. If there are multiple potential blocking coalitions, different blocking coalitions can be selected by varying the choice of w: a very negative value for w_i will select a blocking coalition that does not include i, whereas a very positive value for w_i will select a blocking coalition that does include i.

A.2 Computation Notes for Algorithm 1

To run Algorithm 1, we use CPLEX to compute solutions for the binary integer program given in 8: CPLEX considers this problem to be a binary program with "indicator constraints". From a computational perspective, presolving the problem with about 1000 players is computationally costly. We thus first check sub-problems for each county in Gifu. Only if none of these have a blocking coalition do we check the whole problem. This reduces dramatically the amount of time spent in presolve.

A further reduction in time required can be obtained by stopping the optimization problem at the first solution. Fully solving the optimization problem dramatically lengthens the time required to find the next blocking coalition, and does not appear to improve the time required to find a core partition. Stopping upon finding the first (integer feasible) solution results in different runs of the algorithm producing different solutions, as desired.

The generalizability of the computational approach employed might appear limited because of the specific form required. The use of linear inequalities, however, is not as restrictive as it may first appear. Barros [1998] and others describe methods for expressing more complex restrictions in linear form by generating additional variables. Although using this approach may currently be computationally challenging, integer programming algorithms have seen enormous improvements in speed in recent years [Bixby 2012]. Combined with improvements in hardware performance, simulation of at least approximate versions of models with non-linear payoff structures may soon become feasible. The easiest of these would likely be an Alesina and Spolaore [1997] type model involving quadratic loss from some ideal point.[20]

One might wonder what the "best" core partitions look like, where the planner can select a partitions out of those that are a decentralized solution. Unfortunately, it is difficult to modify Algorithm 1 to produce partitions that are desirable from a social perspective, because choosing the "best" myopic deviation in this case involves a non-linear objective and is thus computationally costly. Using the Dinkelbach [1967] approach is possible, but it appears that the increase in computation time is not rewarded by partitions that are particularly good from the social planner's perspective. This provides suggestive evidence that there may be no partitions that are anomalously good.

[20]Early results using a dataset based on the *Heisei* municipal mergers suggest that games of up to 100 players can be simulated using quadratically constrained mixed integer programs. Although the increased numerical difficulty of the quadratic model reduces the number of players by an order of magnitude, 100 players would still be sufficient for many applied models.

A.3 Computation of Centralized Solution $\pi_k^{\mathbf{FB}}$ (Equation 10)

In order to compute π_k^{FB}, the central planner's optimal partition conditional on there being exactly k coalitions, we must have already obtained estimates $\hat{\beta}$ for the importance of different types of heterogeneity. We then combine all the types of distance present in our data:

$$s(i,i') = a + \sum_{\text{type}} \hat{\beta}_{\text{type}} d^{\text{type}}(i,i'), \tag{25}$$

where $s(i,i')$ is *similarity* between i and i', a is an arbitrary constant, and the types of distance are listed in the rows of Table 2. We use $\hat{\beta}$ from Column XI of Table 2. In order to run the Hartigan and Wong [1979] algorithm, however, we must convert our problem into a standard k-means problem. We do so via the following steps, using the pairwise similarities between players that we just computed.

A k-means problem can be described by giving the coordinates of each player in euclidean space. Alternatively, the problem can be described by giving a Gram matrix that allows the calculation of distances between players. The relation between these two forms is given by the definition of a Gram matrix: if z_1, z_2, and z_3 are the locations of three points in a euclidean space of any dimension, then the associated Gram matrix is

$$\begin{bmatrix} z_1 \cdot z_1 & z_1 \cdot z_2 & z_1 \cdot z_3 \\ z_2 \cdot z_1 & z_2 \cdot z_2 & z_2 \cdot z_3 \\ z_3 \cdot z_1 & z_3 \cdot z_2 & z_3 \cdot z_3 \end{bmatrix},$$

where $z_1 \cdot z_2$ is the dot product. The distance between any two points can easily be computed by using the entries of this matrix: for example, the (squared) distance between points 1 and 2 is $z_1 \cdot z_1 + z_2 \cdot z_2 - 2z_1 \cdot z_2$. Any symmetric positive semi-definite matrix is a Gram matrix, and gives the distance between a set of points in euclidean space in the way just described.

If the matrix

$$G = \begin{bmatrix} s(1,1) & s(1,2) & \cdots & s(1,N) \\ s(1,2) & s(2,2) & & \vdots \\ \vdots & & \ddots & \vdots \\ s(1,N) & \cdots & \cdots & s(N,N) \end{bmatrix} \tag{26}$$

is a Gram matrix (i.e. if it is positive semi-definite), then we are done, and we use G as the input to Hartigan and Wong [1979]. The choice of the constant a has no effect on the distances described by G: the (squared) distance between players 1 and 2, for example, is given by $s(1,1) + s(2,2) - 2s(1,2)$ which does not depend on a. The choice of a, however, does change the eigen decomposition of G and thus is important to algorithmic performance.

We choose

$$a = \frac{-3}{N^2} \sum_i \sum_{i'} \sum_{\text{type}} \hat{\beta}_{\text{type}} d^{\text{type}}(i, i'), \tag{27}$$

where the negative sign is necessary because the entries in $\hat{\beta}$ are negative, but we wish a to be a positive constant.

If G is not positive semi-definite, then we perform a Hathaway and Bezdek [1994] "spread transform" of G. If all players have the same income y_i then we can perform this as follows. G must have negative eigenvalues because otherwise it would be positive semi-definite. Let λ be the most negative eigenvalue of G. Now construct $G' = G + |\lambda| I_N$, where I_N is the identity matrix. By construction G' will be positive semi-definite, and the (squared) distances between players have all increased by exactly $2|\lambda|$. Since *all* pairwise distances have increased by this amount, the optimal solution to the k means problem described by G' is the same as the original problem described by G. The difference is that G' is a Gram matrix, and thus describes a clustering problem in euclidean space that can be solved via the Hartigan and Wong [1979] algorithm.

For the case where y_i is different for different players, construction proceeds as follows. Let λ be the most negative eigenvalue of $D(y)^{1/2} G D(y)^{1/2}$, where $D(y)$ is the diagonal matrix with entries y on the diagonal. Then let

$$G' = D(y)^{-1/2} \left(D(y)^{1/2} G D(y)^{1/2} + |\lambda| I_N \right) D(y)^{-1/2}. \tag{28}$$

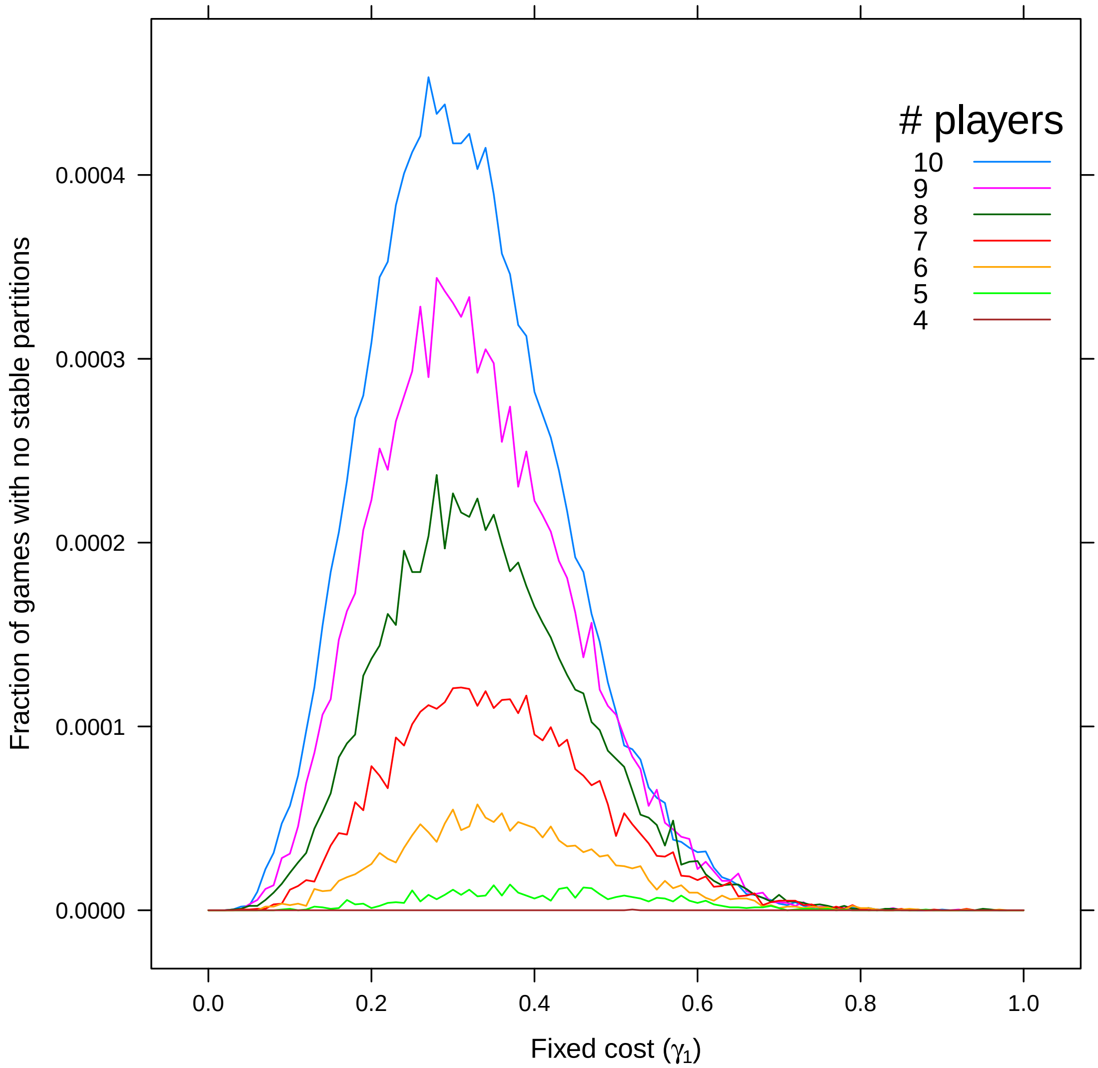
To see why this construction makes sense, consider the case where we begin with three players all with $y_i = 1$, labeled 1a, 1b, and 2. Let these players have locations in euclidean space given by a spread transform away from the origin by $|\lambda|$: that is, $z_{1\text{a}} = (\sqrt{|\lambda|}, 0, 0)$, $z_{1\text{b}} = (0, \sqrt{|\lambda|}, 0)$, $z_2 = (0, 0, \sqrt{|\lambda|})$. Now suppose we wish to create a new game by combining players 1a and 1b so that now the game consists of only two players. Due to the properties of k-means we only need to keep track of the centroids of the players. The centroid of the new player 1 is now located at $z_1 = (\sqrt{|\lambda|}/2, \sqrt{|\lambda|}/2, 0)$. The Gram matrix associated with this new game will thus be $\begin{bmatrix} |\lambda|/2 & 0 \\ 0 & |\lambda| \end{bmatrix}$, illustrating why the spread transform must be scaled by the inverse of the weight of the players.

Some versions of the Hartigan and Wong [1979] require that input be provided in euclidean form, rather than as a Gram matrix. This is not problematic: if G' is a Gram matrix, then it has a decomposition of the form $G' = AA^{\mathrm{T}}$, where A provides the locations of points in euclidean space. This decomposition is not unique, but the objective function of the algorithm is defined in terms of distances between points and these are as specified in G'. The precise choice of A is thus of limited importance: we create A via an eigen decomposition.

The solution to a k-means problem is generically unique. On the other hand, the decentralized coalition formation game can in general have multiple partitions in the core. We thus see that the central planner's solution is qualitatively different from the decentralized solution: it is not possible to apply some modified version of k-means to generate core partitions, because we would be generating a unique solution for a game that does not have that property.

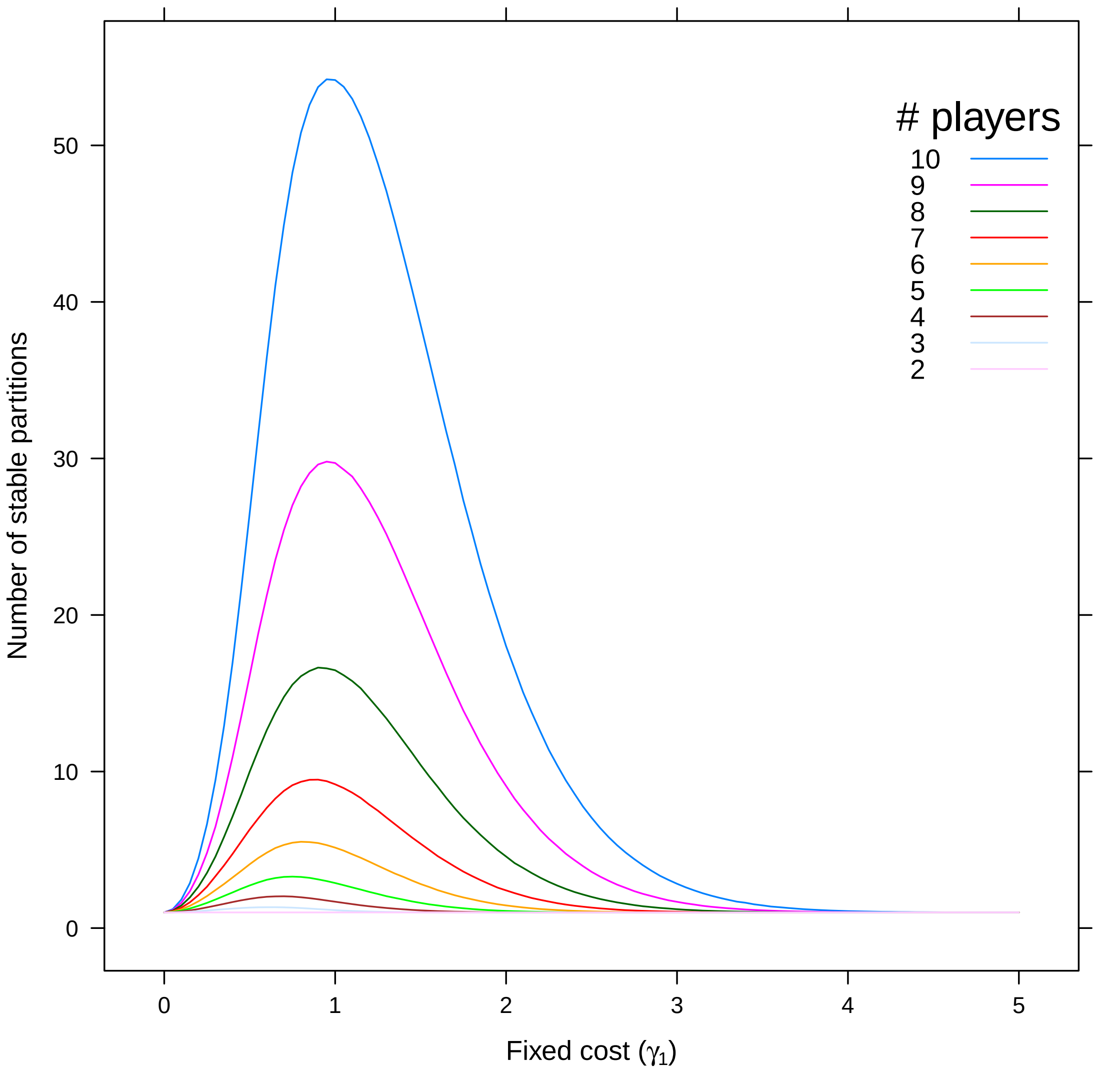
B Additional Tables and Figures

Appendix Figure B.1: Probability of Empty Core (Random Pairwise Distances)



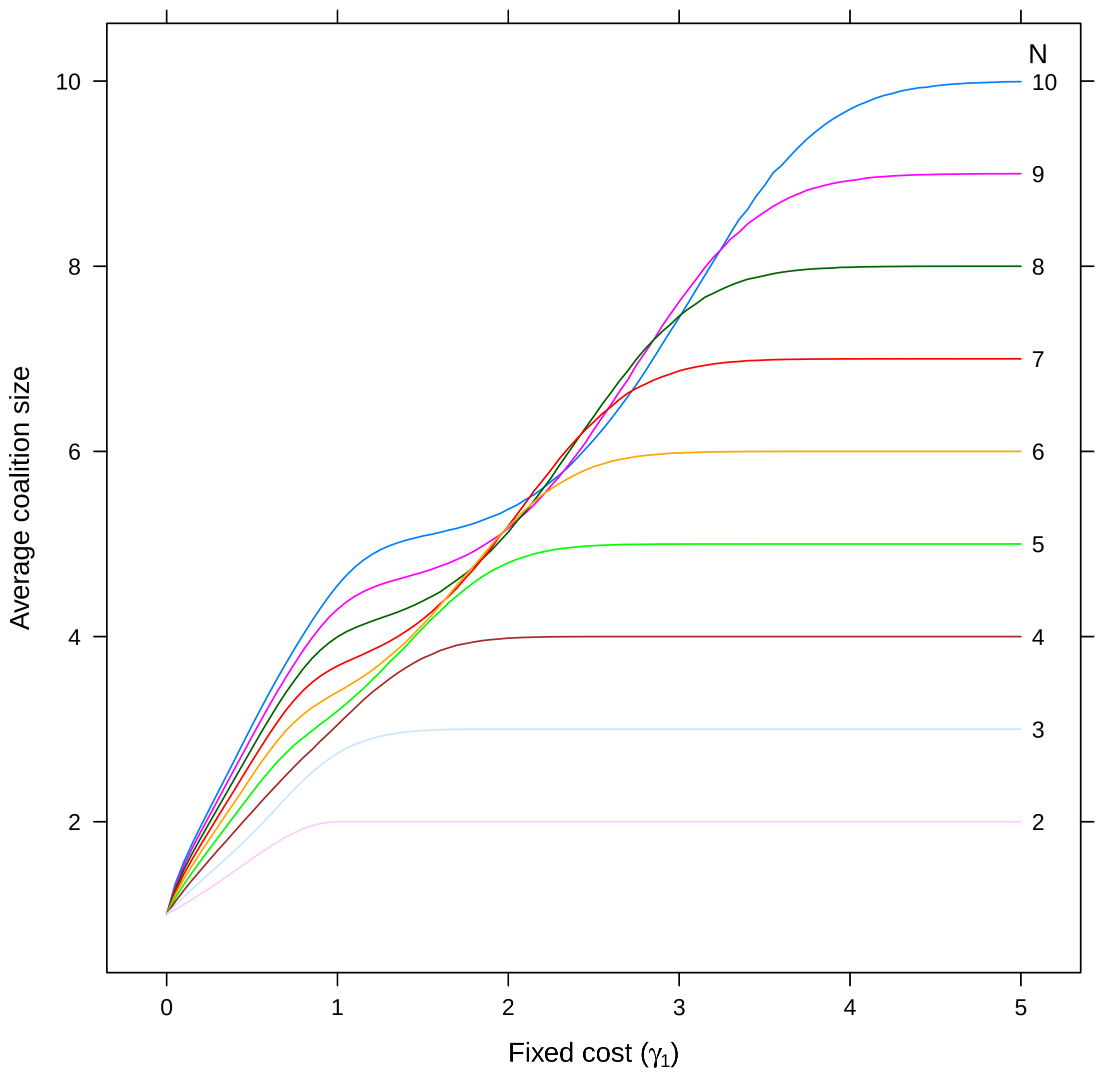
Each line consists of 101 grid points, corresponding to values of γ_1 in the [0,1] interval. Each grid point on each line is based on 1 million randomly generated games.

Appendix Figure B.2: Average Number of Partitions



Each line consists of 101 grid points, corresponding to values of γ_1 in the [0,5] interval. Each grid point on each line is based on 100,000 randomly generated games where $d(i, i') = d(i', i) \sim \text{Normal}(0, 1)$. Core partitions are counted via brute force enumeration (rather than Algorithm 1) to avoid any potential for undercounting.

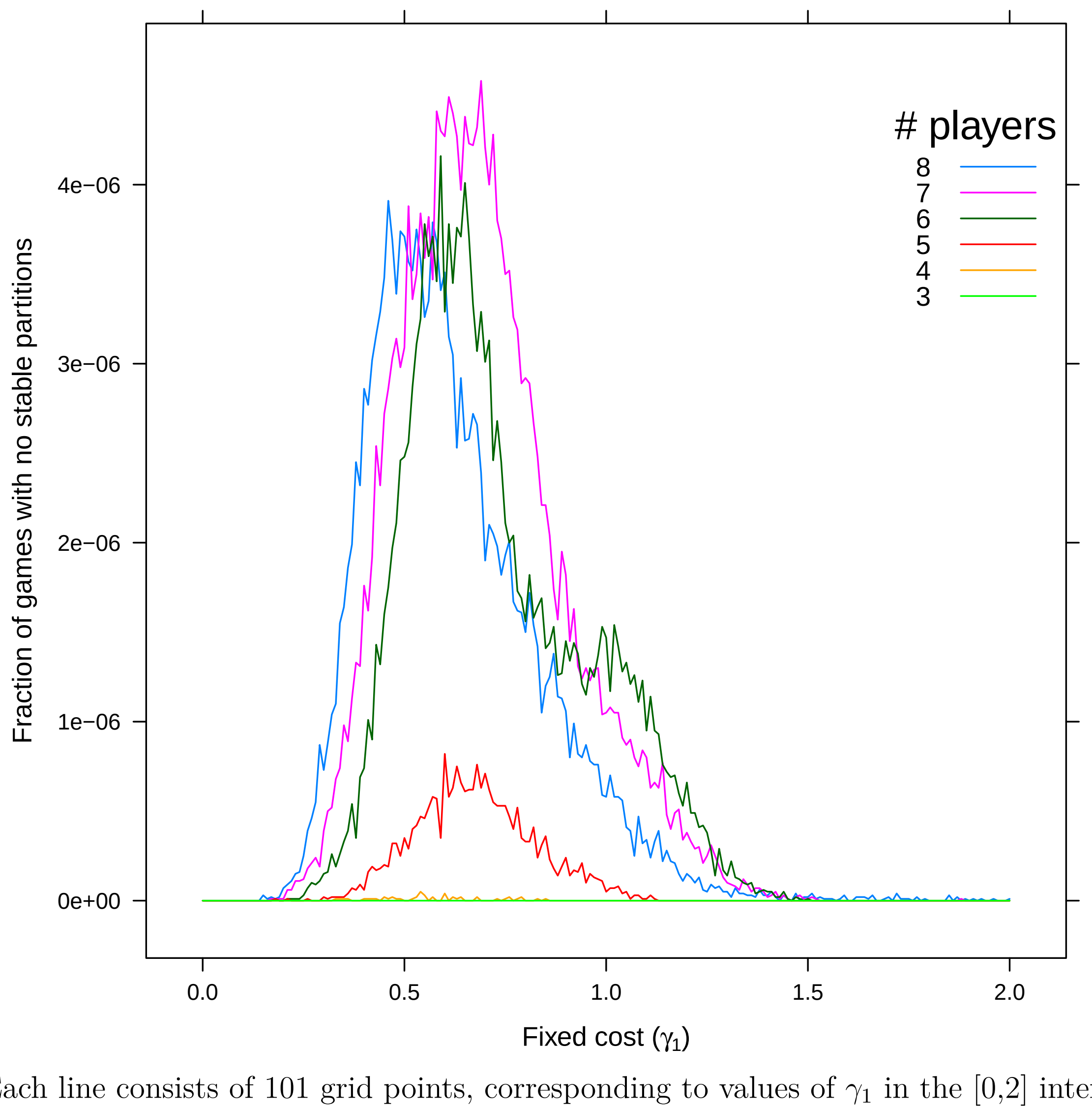
Appendix Figure B.3: Average Size of Coalitions



Each line consists of 101 grid points, corresponding to values of γ_1 in the [0,5] interval. Each grid point on each line is based on 100,000 randomly generated games. The average size of coalitions is calculated by using brute force enumeration to generate all core partitions for a given game, calculating the average coalition size in each of these partitions, averaging across all partitions, and then averaging across all randomly generated games. In this way random games that have more core partitions are not overweighted relative to games that have fewer.

The "plateau" around $\gamma_1 = 1.5$ corresponds to a partition with two (usually evenly sized) coalitions. The grand coalition begins appearing with high probability only at higher values of γ_1.

Appendix Figure B.4: Probability of Empty Core (Random Euclidean Distances)



Each line consists of 101 grid points, corresponding to values of γ_1 in the [0,2] interval. Each grid point on each line is based on 100 million randomly generated games.

Appendix Table B.5: Monte Carlo Simulation Results

	$N = 10$	$N = 100$	$N = 1000$
True value	-0.3	-0.3	-0.3
Mean	-0.60	-0.31	-0.33
Bias	-0.30	-0.01	-0.03
Standard Deviation	0.48	0.26	0.11
RMSE	0.57	0.26	0.11
RMSE ratio	2.19	2.36	
$\sqrt[3]{10}$	2.15	2.15	
$\sqrt{10}$	3.16	3.16	

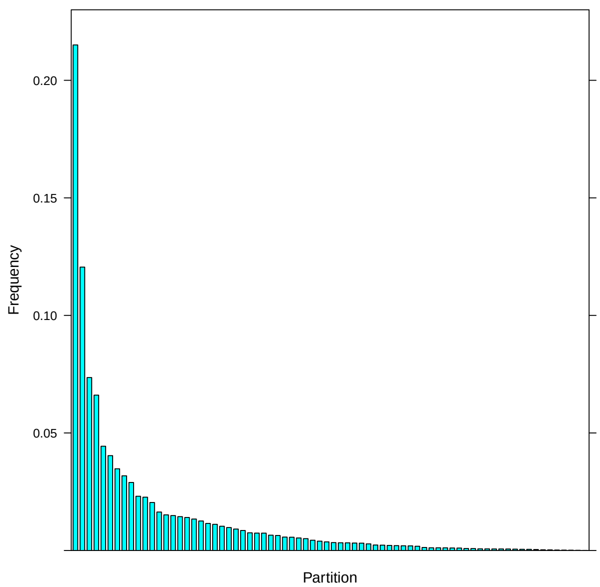
For each choice of number of players N, we construct 1000 random datasets via the following technique. First, randomly choose locations for N players, uniformly in a square of side length $\sqrt{N}$. Compute the "geographic" distance between each pair of players. Next, create a second type of heterogeneity ("religion") by randomly assigning each player a 0-1 variable with probability 0.5 of either. Let $\beta = (-1, -0.3)$ give the weight on geographic vs. religious heterogeneity.

Next, we compute the optimal k-means partition following Equation 10, for $k = N/3$. We do not need to choose a value of γ_1 and compute the "correct" socially optimal partition, because γ_1 does not enter into the estimation process for β: all we need is that the players are divided correctly into a reasonable number of coalitions, and that, conditional on the number of coalitions, this partition is optimal for the central planner.

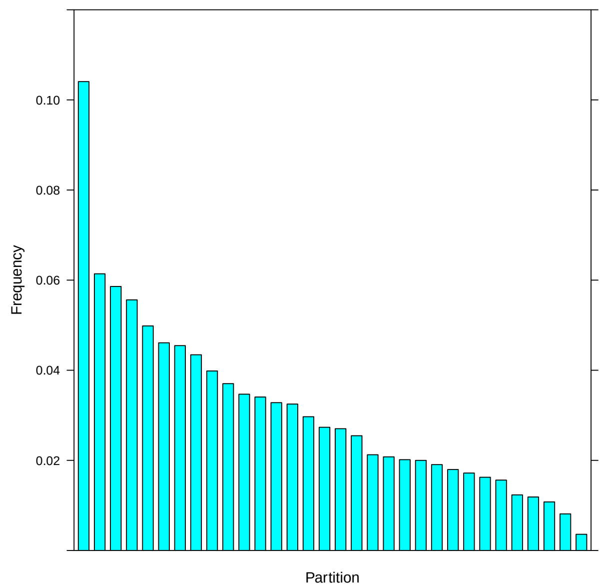
Next, treat this optimal partition as the actually observed partition (like Column I in Table 1), and generate a set of alternative partitions corresponding to the set of all pairwise permutations (like Column II in Table 1). Run maximum score estimation on this artificial dataset in exactly the way described in Section 4.2. Estimates are for the relative importance of religious heterogeneity, with the importance of geographic heterogeneity fixed at -1. There is thus only one parameter being estimated, with a true value of -0.3.

RMSE Ratio: to assess the convergence rate, we report $\text{RMSE}_{N=10}/\text{RMSE}_{N=100}$ and $\text{RMSE}_{N=100}/\text{RMSE}_{N=1000}$.

Appendix Figure B.6: Probability of Partitions under Algorithm 1



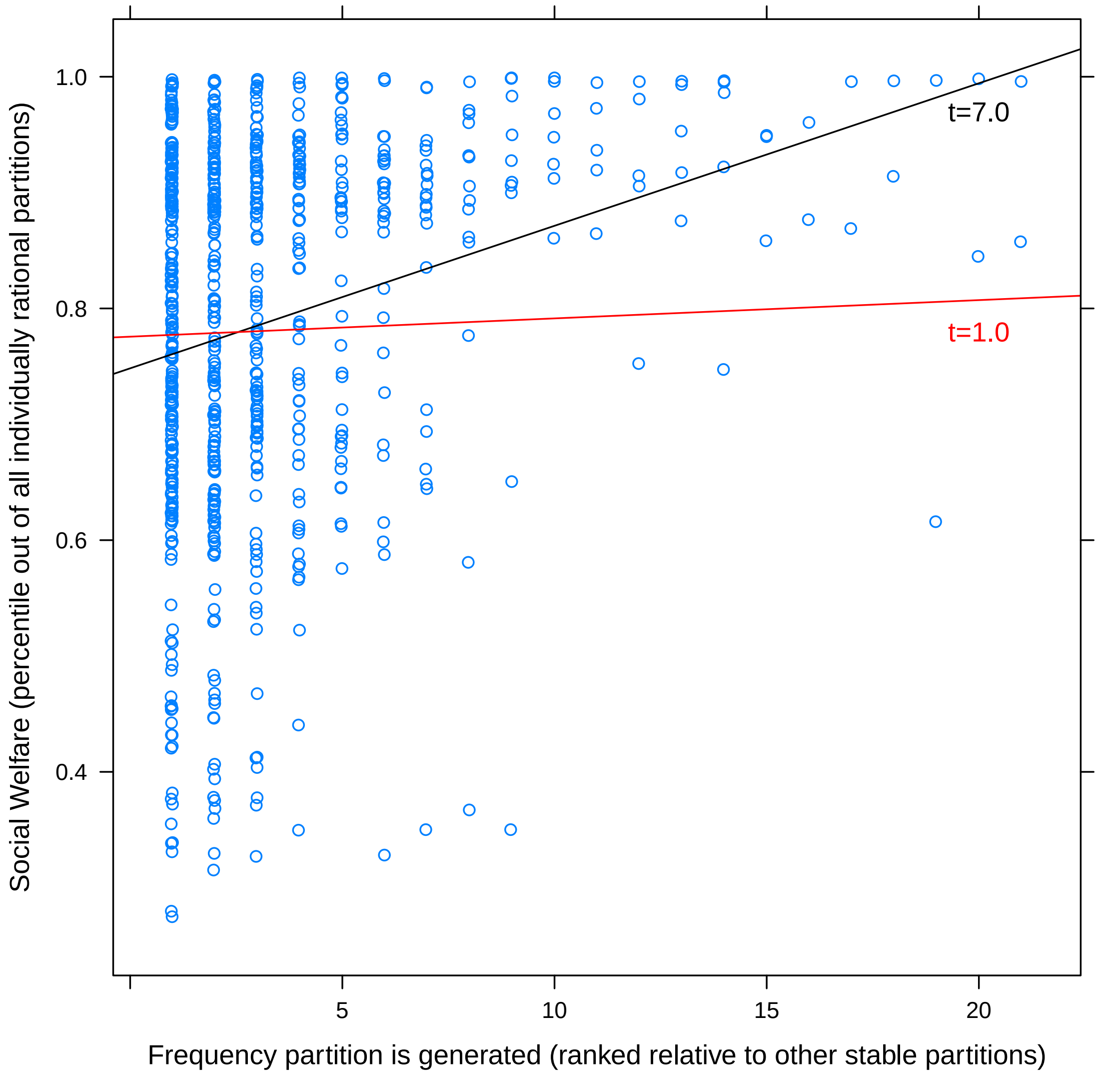
(a) Random Pairwise Distances



(b) Random Euclidean Distances

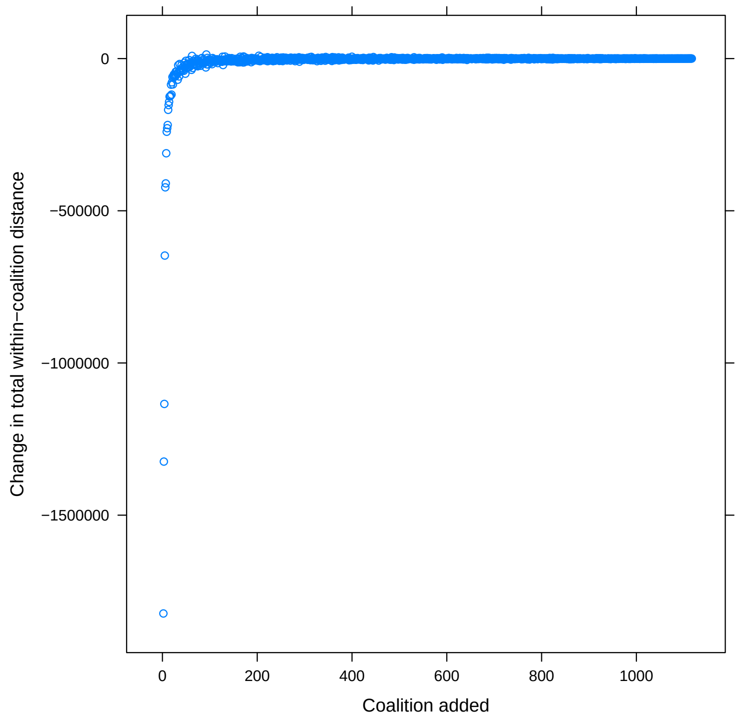
Probabilities that various core partitions will be generated by Algorithm 1 in a single random game (this is the random game with the greatest number of partitions out of 10000 games). We run Algorithm 1 10000 times on this random game, with random weights $w_i \sim \text{Normal}(0, 1)$. All core partitions are generated at least once, but some are generated with much higher probability than other ones. Figure B.6a shows the case where $d(i, i') \sim \text{Normal}(0, 1)$, and Figure B.6b shows the case where players are randomly located in a square.

Appendix Figure B.7: Welfare vs. Frequency of Partition Being Generated by Algorithm 1

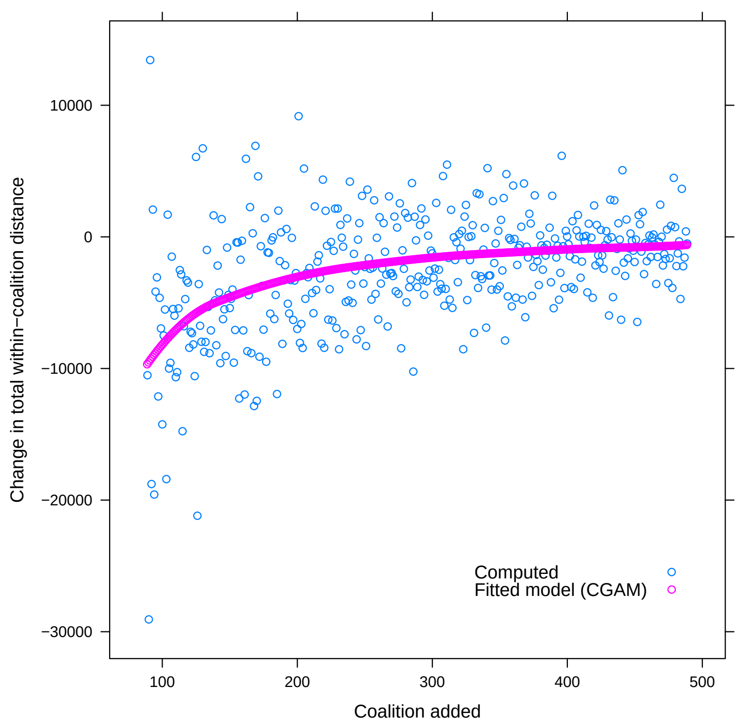


Each point represents a single partition in a single randomly generated game, where $N = 10$ players are randomly assigned locations on a square. Each partition is assigned a rank based on how frequently it is generated by Algorithm 1. Each partition is also given a percentile relative to the social welfare across all individually rational partitions: this is an attempt to standardize across games, because some games may have more distant players, and thus lower payoffs, compared to other games. When we run a bivariate regression comparing the frequency rank of the partition with its social welfare percentile, we find that partitions that have a higher numeric rank have a higher social welfare (black line, t=7.0). However, when we add a dummy variable identifying each distinct game to this bivariate regression, we find that the statistical significance of the relationship disappears (red line, t=1.0). It thus appears that games that have many partitions have higher payoffs, but *within* a single given game, partitions that are more likely to be generated by Algorithm 1 do not systematically have higher or lower payoffs than those less likely to be generated.

Appendix Figure B.8: Change in Distance: $\text{WSS}(\pi_k^{\text{FB}}) - \text{WSS}(\pi_{k+1}^{\text{FB}})$

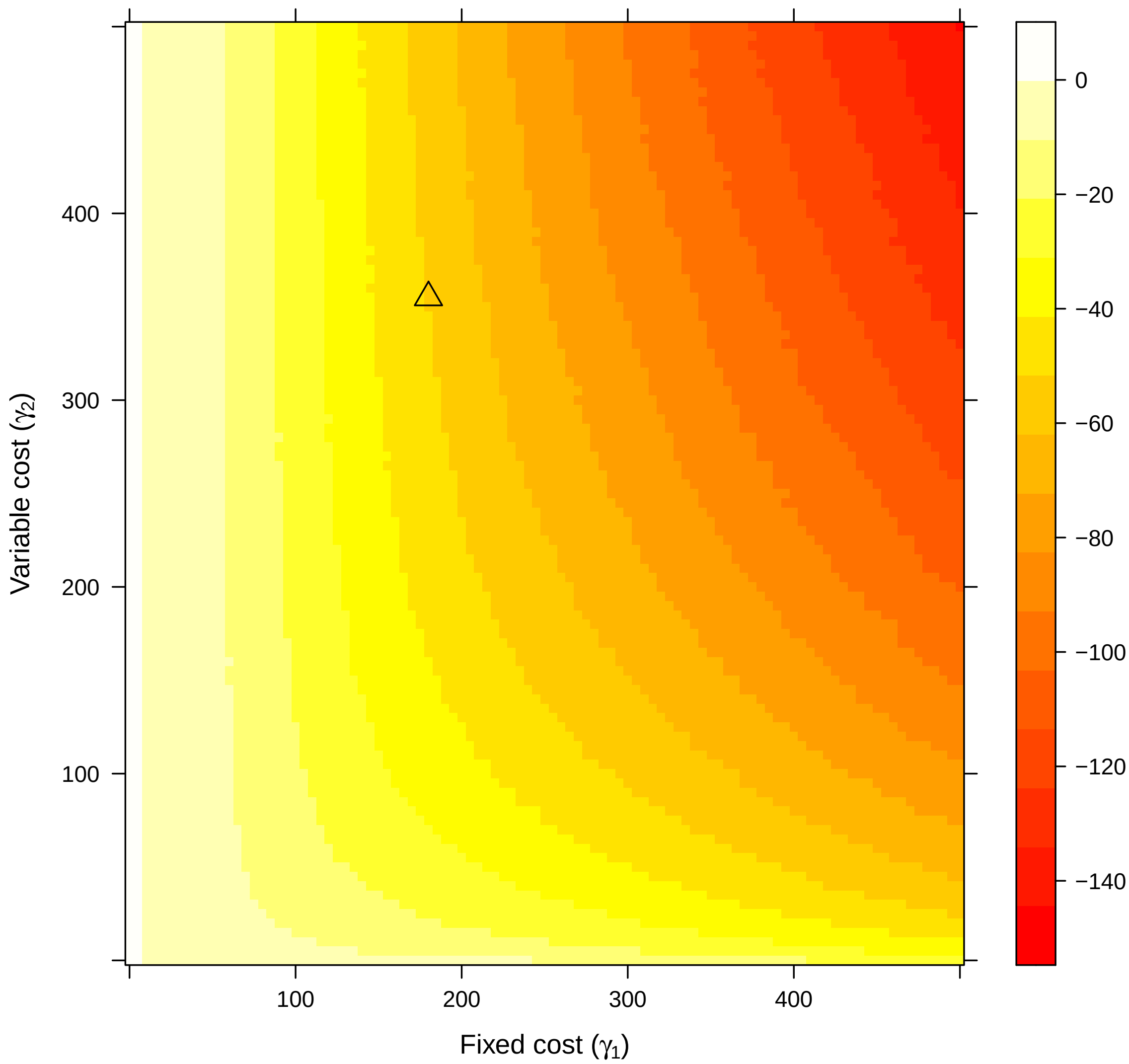


Appendix Figure B.9: Constrained Generalized Additive Model Estimates



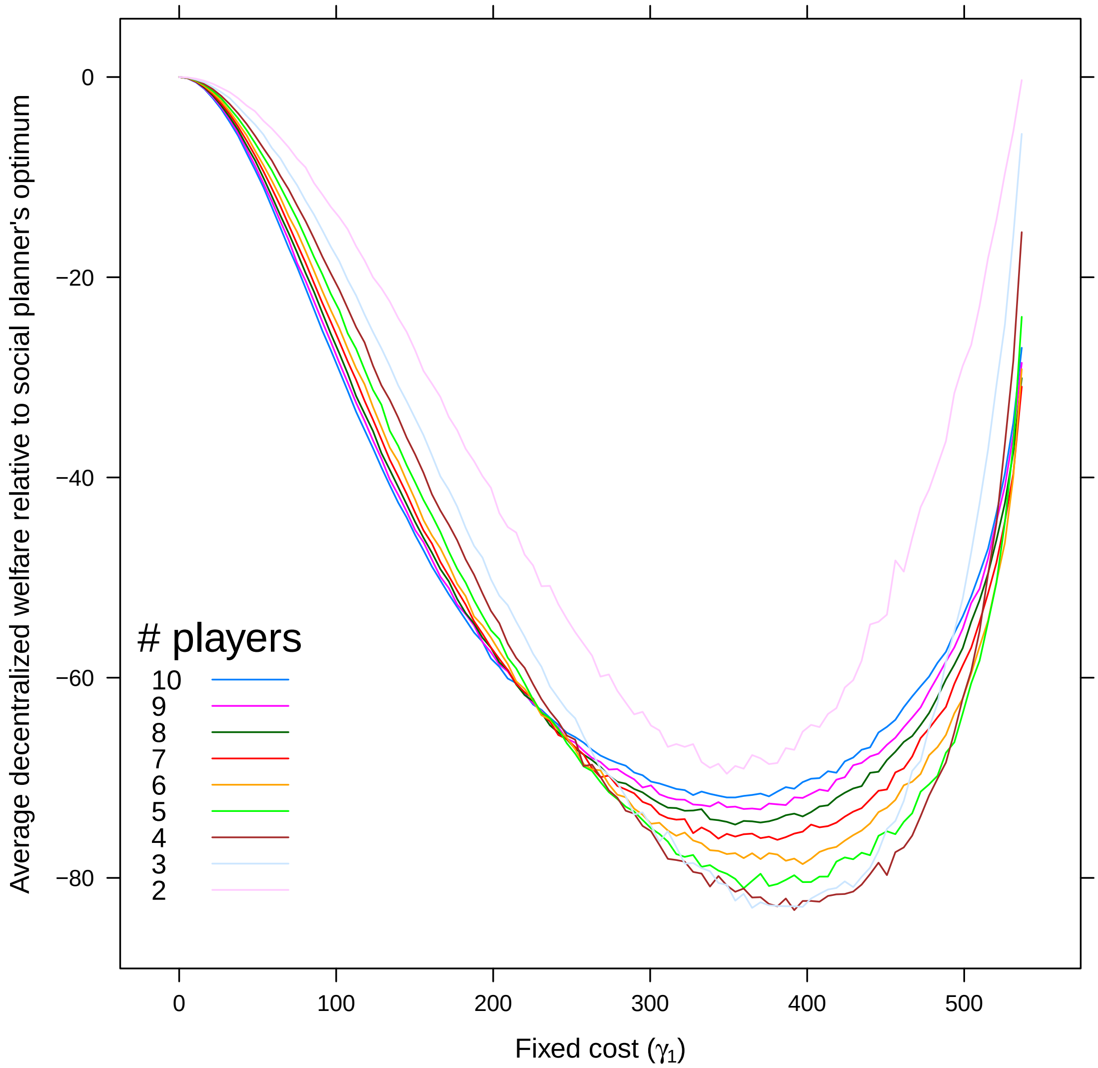
Blue dots are the computed differences $\text{WSS}(\pi_k^{\text{FB}}) - \text{WSS}(\pi_{k+1}^{\text{FB}})$ taken from Figure B.8. Pink dots are the smooth increasing function that best fits the blue dots, computed via a Liao and Meyer [2019] constrained generalized additive model.

Appendix Figure B.10: Ratio of Welfare as a function of both γ_1 and γ_2



Average inefficiency in games with $N = 7$ players for a grid of 101×101 grid points, evenly spaced at parameter values between 0 and 500 for both fixed cost γ_1 and variable cost γ_2. We randomly generate 10000 games per grid point.
We begin by randomly generating player locations from a uniform density on a square with side length $\sqrt{6.8 \times 7}$ km. This matches the density in our Meiji data: one player per 6.8 sq. km. We do not generate different populations for players; instead, we let γ_2 be the per *player* variable cost of running a jurisdiction. In our Meiji data, the average population of a village is 613 people, and we calibrated $\gamma_2 \simeq 0.58$ in Section 6.1. This means that in this figure $\gamma_2 = 0.58 \times 613 \simeq 355$ has the same per player variable cost. A per jurisdiction fixed cost of $\gamma_1 \simeq 182$ corresponds to our choice in Section 6.1. We place a black triangle in the figure at this (γ_1, γ_2) pair. We do not attempt to match the income dispersion in the Meiji data, but instead draw $y_i \sim \text{Uniform}(0, 2 \times 613)$, which matches the Gifu average of roughly 1 *koku* per capita.

Appendix Figure B.11: Ratio of Welfare to Planner's Optimum, $\gamma_1 + \gamma_2 = 537$



Average inefficiency in games with different numbers of players, for 101 points, with 10000 random games per point. Each game is generated the same way as described in Appendix Figure B.10.

The points used correspond to (γ_1, γ_2) pairs that satisfy the equation $\gamma_1 + \gamma_2 = 537$. This is a line that passes through the black triangle at $(182, 355)$ and intersects the axes near then end of the plot area of Appendix Figure B.10.

Appendix Figure B.12: Decentralized mergers with equal per capita tax base

Edges indicate members of the same coalition. Colours are the same within each coalition, but random across coalitions. This is a different core partition than that shown in Figure 5. There appear to be only two coalitions of “leftovers”, both involving only a pair of players.

C Data Construction

This appendix describes the exact method used to construct each of the variables listed in Table 1. The main data source is the *Gifu-ken Chouson Ryakushi*. The GKCR was originally digitized by Skinner [1988], but he does not appear to have used it in published work. The GKCR version used is courtesy of Tsunetoshi Mizoguchi and Kei Okunuki, based on an original version at the Skinner Data Archive. A bilingual codebook is available in Mizoguchi [2004].

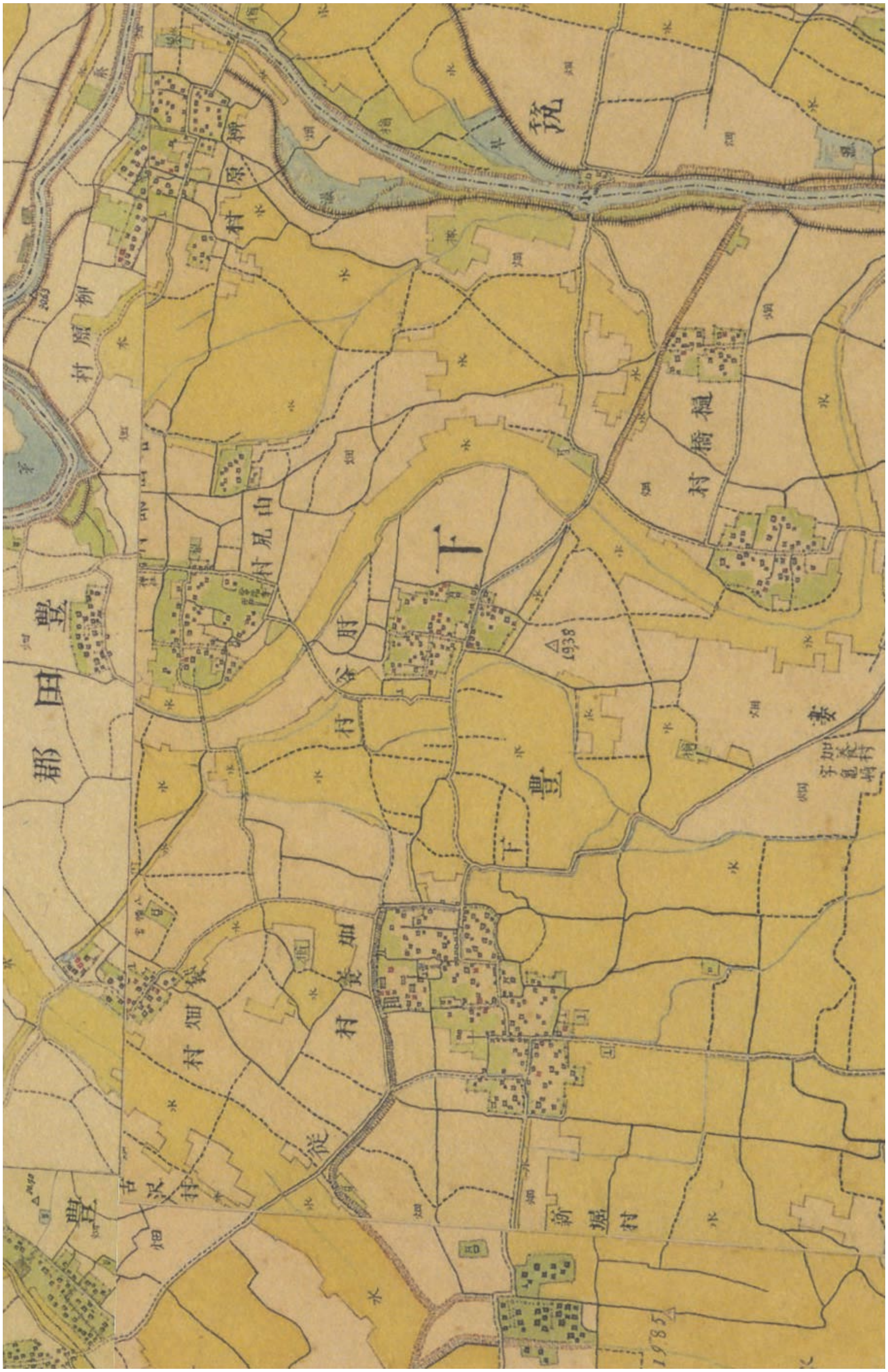
The GKCR describes the feudal villages (*shizen son*) present in Minou province, which contains most of the population of Gifu. It omits villages in Hida province, further to the north. There is a lengthy debate in the domestic Japanese literature regarding the exact definition of a *shizen son*: Yamaoka [1977] provides detailed examples. We do not participate in this debate, as our definition of a *shizen son* is imposed on us by the data source. Each line in the GKCR becomes one player in our coalition formation game. This decision is defensible because the GKCR was collected for administrative purposes, and data is available for the units in question only because they were administratively important: the line items in the GKCR were the base level at which taxes were collected during the feudal period.

Municipal mergers are not recorded directly in the GKCR. Our post-merger boundaries are based on official 1919 municipal boundaries provided by the Ministry of Land, Infrastructure, Transport, and Tourism. Two mergers and one split that occurred in 1903-05 were reversed (the next boundary change did not occur until 1921). The boundaries shown in Figure 1b are thus those of 1897, except for one minor boundary adjustment that does not affect any calculations.[21]

C.1 Points

The boundaries shown in Figure 1a describe each feudal village as a polygon. However, in most cases, the villages in the GKCR actually correspond to one or more clusters of houses, surrounded by the land worked by these households. Although a household level map is not available for Gifu, the *Jinsoku Sokuzu* ("Rapid Map") of 1880 shows the precise location of households for part of eastern Japan. Appendix Figure C.13 shows a representative rural area: the households are clearly clustered, and the location of households could be reasonably accurately described using only a small number of points. We take advantage of this feature of the population distribution of this period, and create a new dataset based on points rather

[21] The most comprehensive list of municipal changes appears to be that maintained by M. Higashide at http://uub.jp/. For Gifu, we cross-checked these with the *Gifu-ken Chouson Gappeishi* ("History of Town and Village Mergers in Gifu Prefecture") and verified that they were correct.



Appendix Figure C.13: Toyota County, Ibaraki Prefecture than polygons. In most cases only one point is required to describe a village: this means that there is basically no within-village heterogeneity in terms of geography, and will allow for us to ignore potential within-village politics when we perform counterfactual simulations in Section 7.

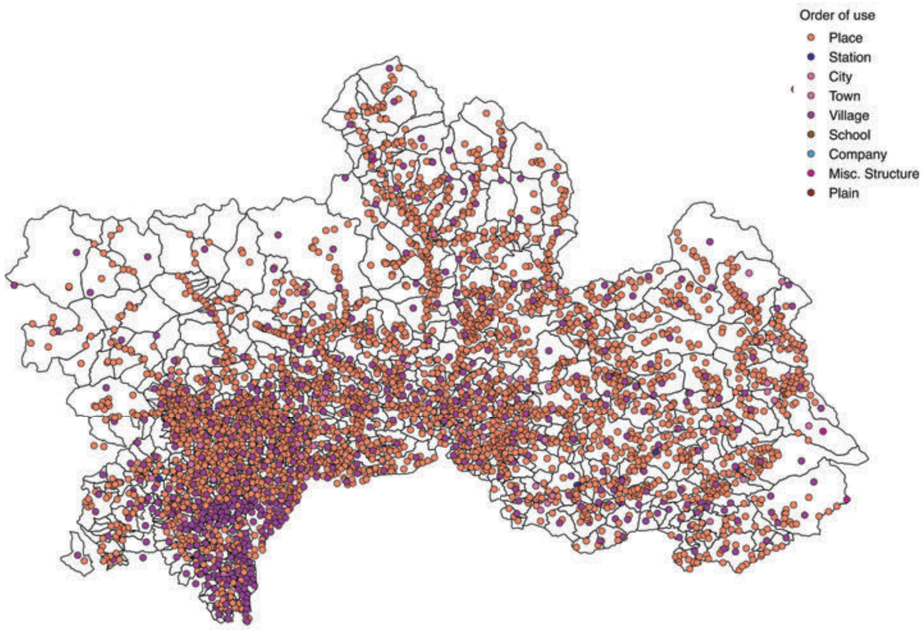
The specific method we use to construct this point data is as follows. We begin with geocoded gazetteer data on Meiji locations, courtesy of the Center for Integrated Area Studies at Kyoto University. This data is based on the 1891 and subsequent official maps of Gifu Prefecture. As with more modern maps, a variety of data was presented on these historical maps. There are two particularly important types of data for our purposes. The first of these is the "place" data: here a name was written on the map and an associated dot indicated the exact place on the map where the settlement with that name was located. The second type is where the name of a city, town, or village was written on the map without a dot indicating its exact location. These two different types of data, along with many other types, are coded separately in the gazetteer. The relevant subset of this data is shown in Appendix Figure C.14. The legend in this figure shows different types of points that are provided by the gazetteer.

We reviewed a subset of this data to establish the relative geographic accuracy of these different types of points. The "place" points appear to have excellent geographic precision. In contrast, the "city", "town" and "village" points appear to be based only on the location at which the name of the city, town, or village was written on the underlying 1891 paper map. In general, this does not appear to correspond to the physical location of houses, and thus these types of points are of substantially lower accuracy than the "place" points. About half of the feudal village polygons in our data have exactly one "place" point within their boundaries. For these villages, we use this point as their geographic location. If a feudal village has more than one "place", then we keep track of all of these points, and will use them all (with equal weighting) for distance calculations.

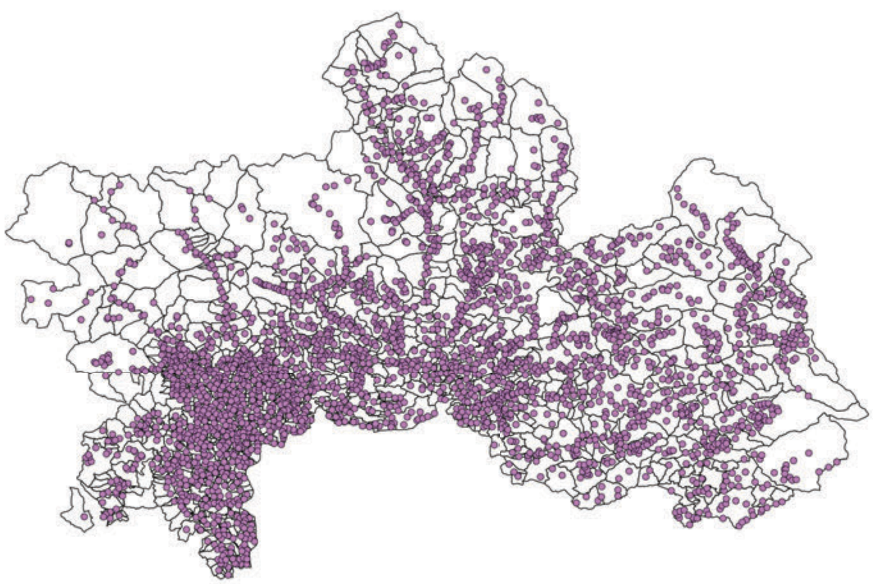
If a village has no "place" points within its boundaries, then we proceed to the next lower type listed in the "Order of use" legend in Figure C.14. This is "station", which gives the location of a train station for the very small number of cases where one exists.[22] If a village still has no points, we continue this process, looking for gazetteer points related to municipality names ("city", "town", "village"). We continue in the same fashion for schools, companies, structures, and plains. In all cases, only the first type shown in the "order of use" legend in Figure C.14 to have any associated points inside a feudal village polygon is

[22] We place these lower in our order of use because train stations were generally built *after* the municipal merger period, and only show up in the gazetteer because some of the maps on which it is based come from the early 20th century. There are only 9 train stations in the gazetteer.

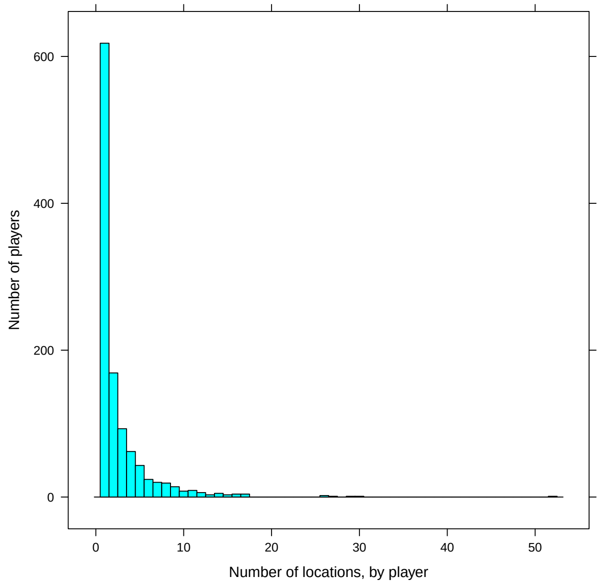
Appendix Figure C.14: Gazetteer Point Data



Appendix Figure C.15: Points Used from Gazetteer



Appendix Figure C.16: Points per Feudal Village



used as the location of that village. If this type has multiple associated points within the feudal village boundaries, then we keep track of all of these.

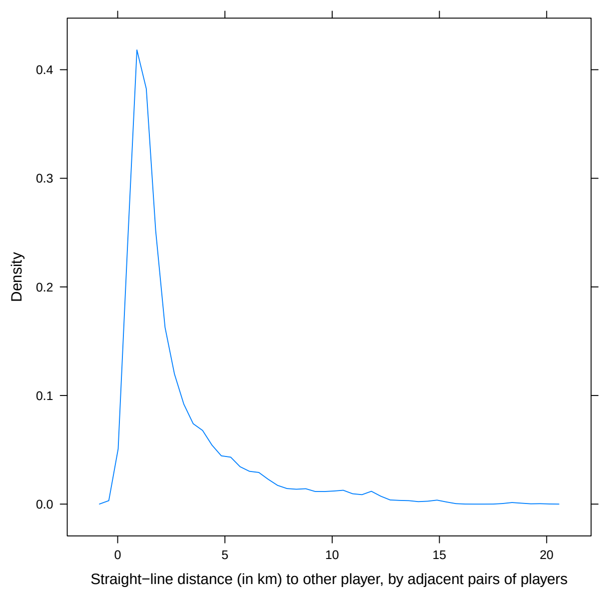
There are 59 feudal villages for which no points are found via this method. After manual inspection, 29 of these are cases where an appropriately named point is located just outside of the village polygon: we assign these points to the appropriate village. In the remaining 30 cases, we use the geographic centroid of the polygon for the location.

Figure C.16 shows a histogram of the number of points per feudal village. For a majority of feudal villages, we have exactly one point associated with the village. In a small number of cases, there are a dozen or more points associated with a single feudal village. Appendix Figure C.15 shows the points that are used.

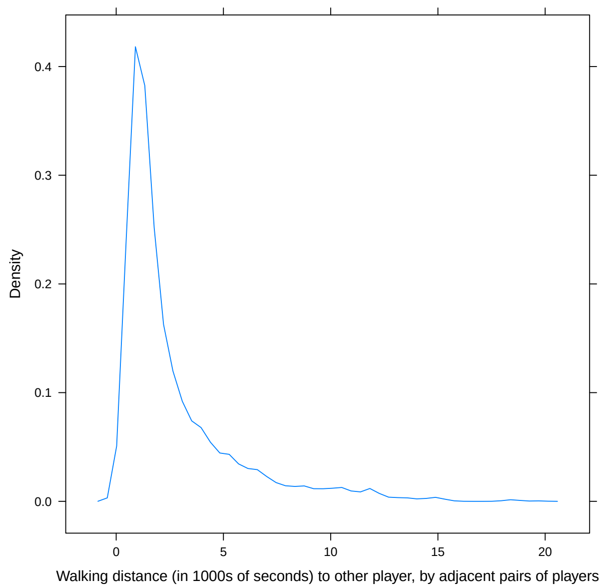
In Appendix Figure B.12 we report simulation results graphically using a single point for each player. Here the location of this point is equal to the mean latitude and longitude for the points associated with that feudal village. These "display points" are shown in Appendix Figure C.17. This is purely for ease of presentation, however, and all calculations are done over multiple points for the feudal villages that have them.

Appendix Figure C.17: Players and Adjacencies

Appendix Figure C.18: Straight-line Distance



Appendix Figure C.19: Walking Distance



C.2 Distance

Straight-line distance is calculated using the great-circle distance formula, assuming that the earth is a sphere. For feudal villages represented by more than one point, this distance is calculated as the average distance across all relevant points. That is, if J_i is the set of points associated with feudal village i, and $J_{i'}$ the set of points associated with feudal village i', then the straight-line distance $d^{\mathrm{SL}}(i, i')$ is calculated as

$$d^{\mathrm{SL}}(i, i') = \frac{1}{|J_i||J_{i'}|} \sum_{j \in J_i} \sum_{j' \in J_{i'}} d^{\mathrm{SL}}_{\mathrm{pt}}(j, j') \tag{29}$$

where $d^{\mathrm{SL}}_{\mathrm{pt}}(j, j')$ is the straight line distance between point j and point j'. The result of repeated application of Equation 29 is a distance matrix d^{SL} containing a straight line distance $d(i, i')$ for each pair of feudal villages i and i'.

For expositional purposes, we calculate a histogram of the straight line distances between adjacent feudal villages: these are $d^{\mathrm{SL}}(i, i')$, with the (i, i') pairs used corresponding to the edges shown in Figure C.17. Figure C.18 reports these distances: the modal distance is bit less than 2km, but some distances are substantially longer.

Straight line distance is likely inappropriate in the case of Gifu, however, because much of the prefecture is mountainous, and thus the actual path used to travel between two villages would not be a straight line, but rather a more complicated route that minimizes elevation changes. For these calculations, we do not consider data on the road network in place during this period: this network was relatively primitive, and we assume that there are walking tracks located wherever our algorithm calculates that people will be travelling. Although trains were being introduced during this period, they were used for longer distance journeys, and are not relevant for distance calculations between a feudal village and its neighbour, or the next village over. Thus, we consider walking as the only mode of transport.

Given the mountainous nature of the prefecture, we assume that the major determinant of walking time is elevation change.[23] We use digital elevation data from the Geospatial Information Authority of Japan, at 10 meter grid square resolution. Figure C.20 shows this elevation data.

To calculate walking time between two points based on this elevation data, we use the Fontanari [2000] implementation of Dijkstra's shortest-path algorithm, applied to a raster version of the elevation data.[24] The walking time returned is anisotropic: the cost of walking uphill is not simply equal to the benefit of walking downhill. Thus, the shortest path to a destination may be different from the shortest path returning from it. We use the roundtrip distance, following both of these paths, divided by two.[25] In the case where feudal villages are associated with multiple points, the approach in Equation 29 is used.

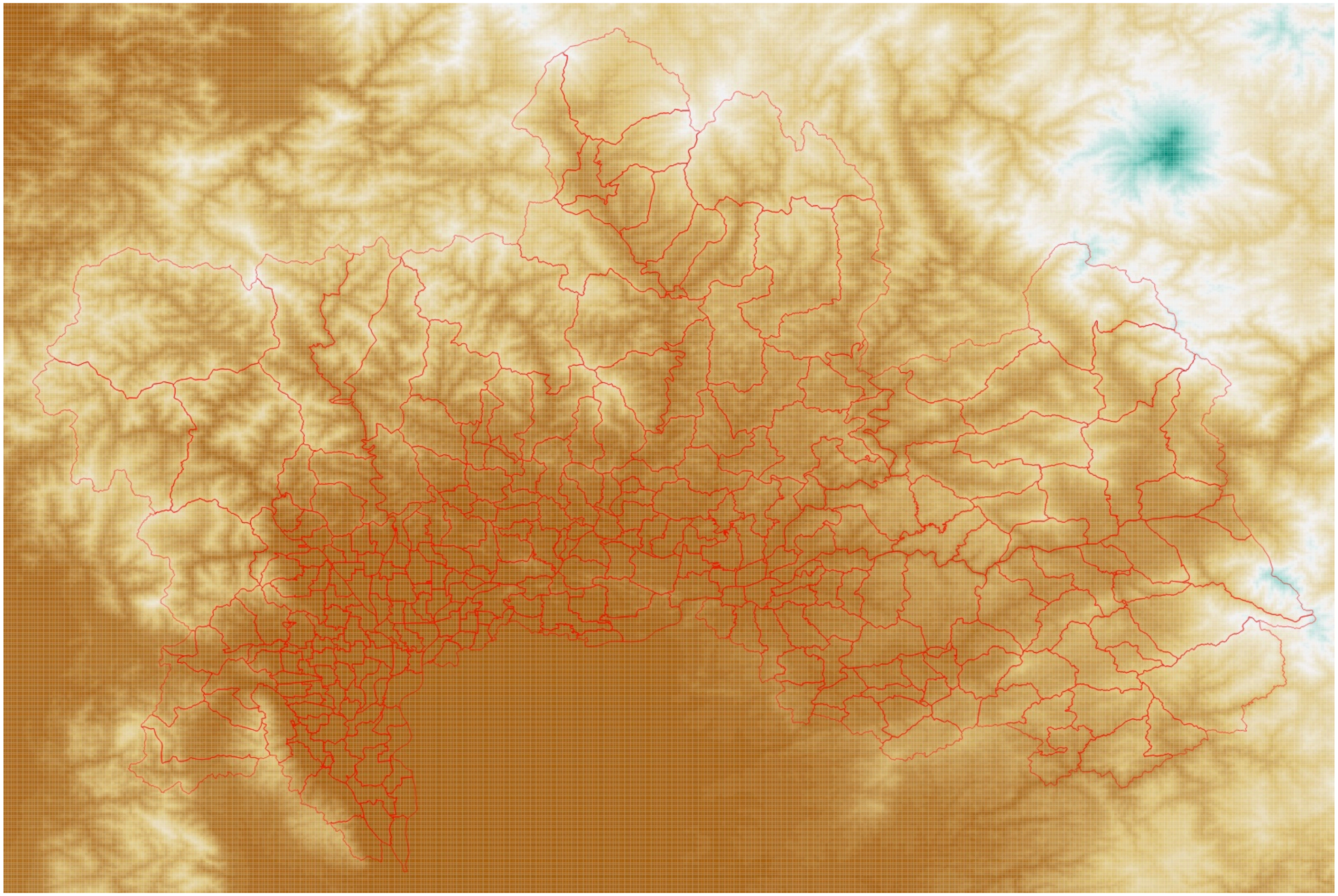
Figure C.19 reports the walking distances for adjacent feudal villages. The units used here are 1000s of seconds, because walking speed is approximately 1km per 1000sec, and thus with these units coefficient estimates will be of roughly comparable magnitudes when using the straight line distance data and the walking distance data. According to the figure, the modal walking time to an adjacent village is a bit less than two hours.

[23] Rivers in Gifu are often seasonal, and generally small. A qualitative inspection of boundaries shows that they do not appear to follow rivers in a systematic way, and a preliminary quantitative analysis using GMM and the actual location of the boundary between two adjacent municipalities appeared to confirm this. We do not report these results because the standard errors were very high, and the econometric model used differs substantially from that presented in Section 4.

[24] This is the "walking distance" function `r.walk` in the GRASS GIS package. Parameters are left at their default values, which correspond to the Aitken [1977] and Langmuir [1984] adjustments to the Naismith [1892] walking time function.

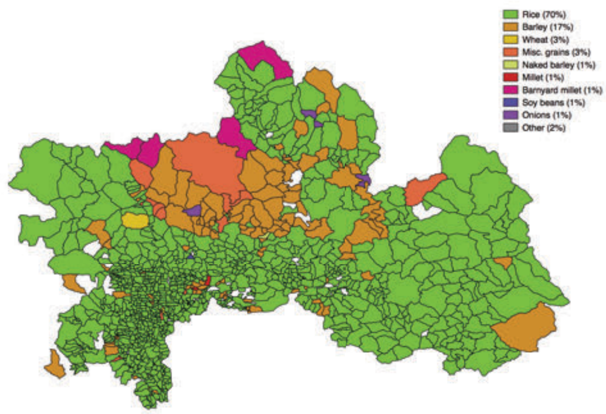
[25] One is reminded of hiking trails that split at certain points, offering both a steeper and a gentler route.

Appendix Figure C.20: Elevation

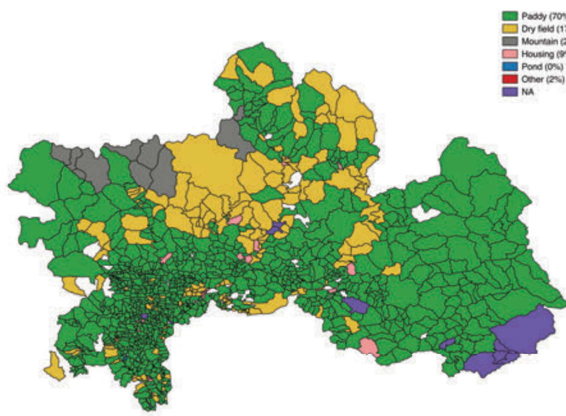


Dark brown indicates low elevation, white higher elevations, and green highest.
In mountainous regions, low elevations tend to correspond to river valleys.

Appendix Figure C.21: Most important product (by value)



Appendix Figure C.22: Most important land type (by value)

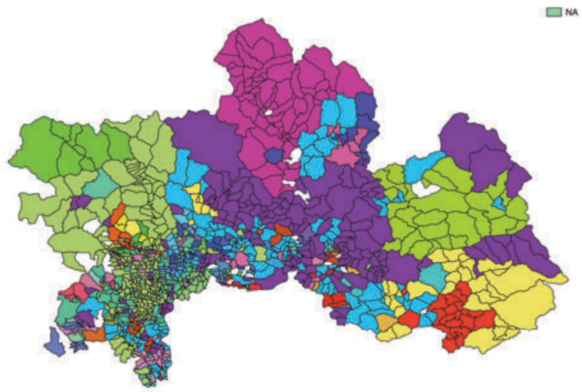


C.3 Other Covariates

For non-geographic variables we will we use a discrete version of Equation 29. Within a given feudal village we may have villagers producing different types of products, farming different types of land, or practicing different religions. For exposition, consider the case where different villagers produce different products (rice, wheat, etc.). Let J_i be the set of villagers in feudal village i, and let $J_{i'}$ be the set of villagers in feudal village i'. The distance $d^{\text{prod}}(i, i')$ is then calculated as

$$d^{\text{prod}}(i, i') = \frac{1}{|J_i|}\frac{1}{|J_{i'}|}\sum_{j\in J_i}\sum_{j'\in J_{i'}} d_{\text{v}}^{\text{prod}}(j, j') \tag{30}$$

Appendix Figure C.23: Most important feudal lord



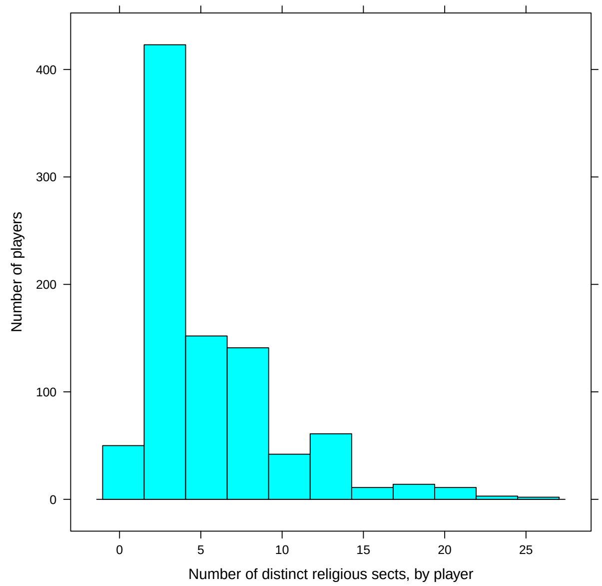
Feudal lords (*ryoushu*) were a basic part of the tax collection and administration system during the feudal period. The lords are reported in the GKCR as historical data: they were removed early in the Meiji restoration, and thus cannot have any direct effect on municipalities during the Meiji period. Their historical legacy, however, could plausibly include cultural differences across feudal villages controlled by different lords.

where $d_{\mathrm{v}}^{\mathrm{prod}}(j, j')$ is equal to 0 if villager j and villager j' are producing the same product, and equal to 1 if they are producing different products. Now let $t(j)$ denote the type of product produced by villager j, and let s_{ti}^{prod} denote the share of product of type t in village i. We can then rewrite the above summations in terms of product shares rather than individual villagers:

$$\begin{aligned} d^{\mathrm{prod}}(i, i') &= \frac{1}{|J_i|} \frac{1}{|J_{i'}|} \sum_{j \in J_i} \sum_{j' \in J_{i'}} (1 - 1(t(j) = t(j'))) \\ &= \frac{1}{|J_i|} \sum_{j \in J_i} \left(1 - s_{t(j)i'}^{\mathrm{prod}}\right) \\ &= \sum_t s_{ti}^{\mathrm{prod}} \left(1 - s_{ti'}^{\mathrm{prod}}\right) \\ &= 1 - \sum_t s_{ti}^{\mathrm{prod}} s_{ti'}^{\mathrm{prod}}. \end{aligned} \tag{31}$$

In the case where $i = i'$ the summation becomes $\sum_t s_{ti}^2$, the Herfindahl index for the concentration of types, and thus $1 - \sum_t s_{ti}^2$ gives an index of the heterogeneity of types. This index is frequently used in political economy, for example in work on ethnic fragmentation following Easterly and Levine [1997]. We thus see that our functional form for non-geographic

Appendix Figure C.24: Religious sects



distance is equivalent to this popular measure of heterogeneity.

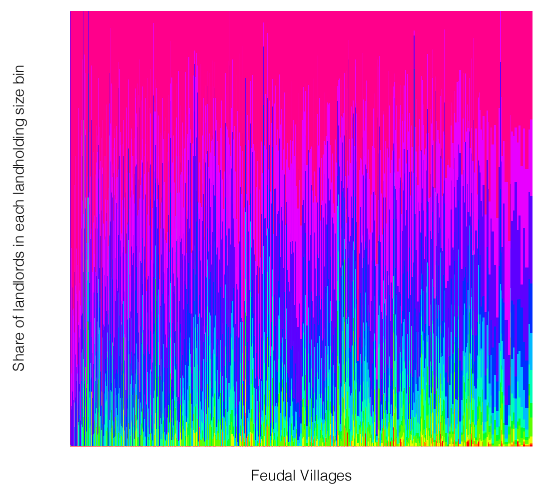
Unless all production in village i is of a single type, we will have $d^{\text{prod}}(i, i) > 0$, but we do not need d to be a true mathematical distance for any of the computations in our paper: it is sufficient that $d^{\text{prod}}(i, i')$ represents the dissimilarity in production between villages i and i'.

In our data, we do not have micro-level information on the specific product being produced by each villager in each village. Instead, we have data on total production for each type of product in a village. Thus, when we compute s_{ti}^{prod} for each product type t village i, we do so by calculating the share of production value that product t accounts for in village i. There are 37 different types of products in the GKCR data, but even approximate price data only appears to be available for 21 of these, and thus only 21 products are used in our calculations: all of these 21 are agricultural, although some of the omitted 16 products are not. Figure C.21 shows the most important product in each feudal village by value, with the legend including data on the total share of each item.

We use a similar method to calculate distances regarding the types of land present in each feudal village, the identity of the lord of the village during the feudal period, the identity of religious sects in the villages, and the distribution of land among landlords. Figures C.22 - C.25 provide a summary of these data.

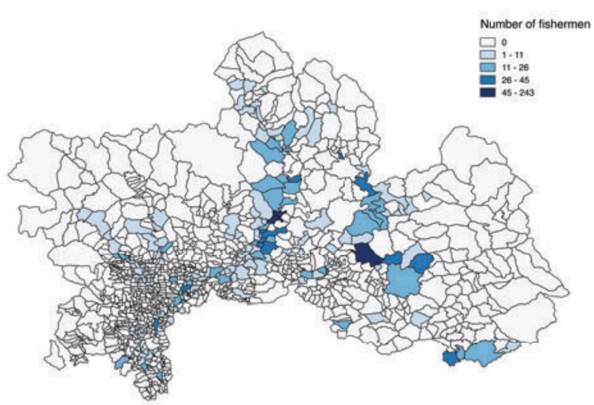
In addition to agriculture, some areas of Gifu engaged in fishing. This was mainly along

Appendix Figure C.25: Landlord size

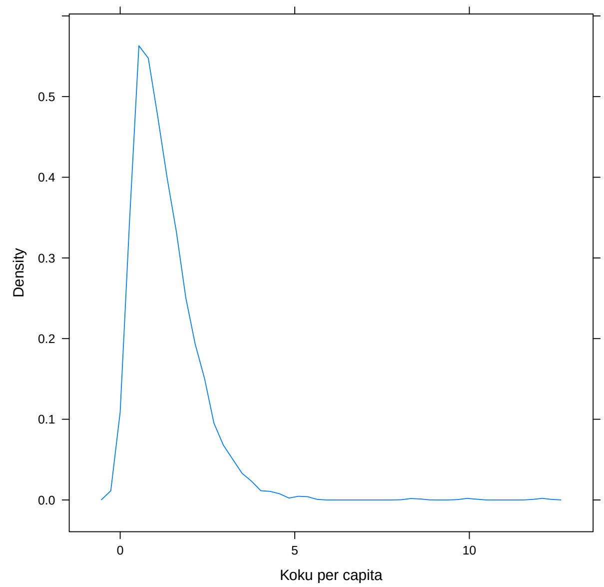


Landlords in each village are counted by size of landholdings: greater than ¥100, 200, 300, 500, 700, 1000, 1500, 2000, 5000, 10000, and 20000. Violet corresponds to small landlords, and red to large landlords. Villages are sorted by amount of total holdings. Because of data quality issues, no "less than ¥100" category is used.

Appendix Figure C.26: Fishing



Appendix Figure C.27: Koku per capita, by Feudal Village

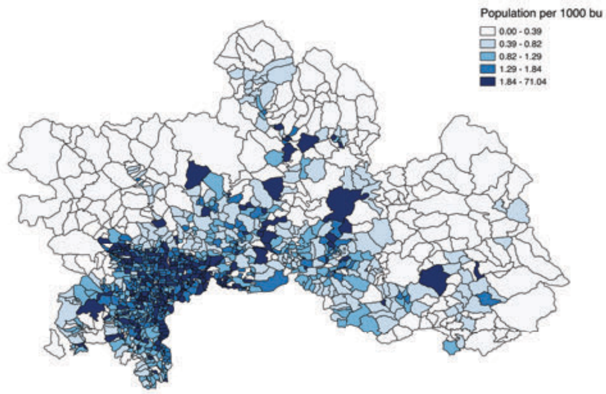


the banks of rivers, although some ponds and lakes appear to also have been used. Figure C.26 shows the distribution of fishermen in Gifu. We calculate a distance measure for fishing, using two types: “fisherman” and “other”.

Two important final variables of interest are income, y, and population, p. Population density data is shown in Figure C.28. We treat income as equivalent to tax base per capita. In our data, tax base is measured in *koku*: Figure C.27 shows the distribution of *koku* per capita in the data. The modal *koku* rating is close to 1 *koku* per capita. This is a plausible value, as the *koku* unit of measure was originally defined as the amount of rice required to feed a man for one year, and production in rural Japan during this period was close to subsistence levels. Gifu, like most of Japan during this period, was predominantly rural and agricultural.

The GKCR contains many other variables that we do not use in the analysis. Most of these are ignored because they appear to be at best only tangentially related to municipal mergers: for example, there are very detailed reports of the calendar day when each of many crops are traditionally planted and harvested in each village. One additional variable that is of interest is that dealing with migration. In the GKCR, only 0.8% of individuals were classified as migrants. This classification is based on *honseki* (“registered domicile”), which should capture all individuals that migrated, and possibly also children of migrants.

Appendix Figure C.28: Population Density



Villages with higher population density have municipalities with smaller surface area, while those areas with lower population density have jurisdictions with larger area. This pattern follows the "size density hypothesis" of Stephan [1977], which has been subjected to considerable study outside of economics. In Gifu, the pattern continues to hold for the post-merger municipalities. The pattern also holds across a wide selection of countries, and varieties of jurisdictions [Stephan 1984]. Suzuki [1999] provides additional citations along with an application to Japanese data. The pattern does not appear to be due to measurement error: this possibility was discussed in Vining, Yang, and Yeh [1979] and Stephan [1979].

Thus, migration does not appear to have been that substantial in Gifu during the period in question, although substantial urbanization was occurring elsewhere in Japan. This justifies the omission of migration in our model.

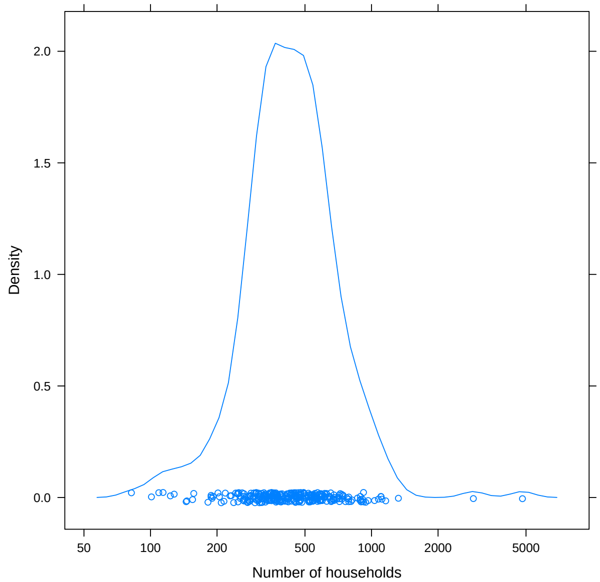
C.4 Official Household Target Size (Column IX, Table 2)

The official orders that resulted in the Meiji municipal mergers occurring explicitly mentioned a target size for amalgamated municipalities: they were supposed to be between 300 and 500 households. Let $\mathtt{HH}_S$ be the number of households in municipality S. Then

$$\begin{aligned} x_{\pi,\mathtt{HH300}} &= \sum_{S\in\pi} 1(\mathtt{HH}_S > 300) \\ x_{\pi,\mathtt{HH500}} &= \sum_{S\in\pi} 1(\mathtt{HH}_S > 500). \end{aligned} \tag{32}$$

Here mergers will be the "right" size when $x_{\pi,\mathtt{HH300}}$ is high and $x_{\pi,\mathtt{HH500}}$ is low. This means that $x_{\pi,\mathtt{HH500}}$ has the same direction (high is worse) as the many other distance variables, but the sense of $x_{\pi,\mathtt{HH300}}$ is reversed. As would then be expected, the sign on $x_{\pi^0,\mathtt{HH300}} - x_{\psi\pi^0,\mathtt{HH300}}$

Appendix Figure C.29: Households (post-merger)



is the opposite of all other variables in Table 1. This variable is presented graphically in Figure C.29.

The definition in Equation 32 only makes sense if all the partitions considered contain the same number of municipalities. It might be possible to rewrite Equation 32 as an average rather than a sum; however, we do not pursue this, and it is not obvious that it would lead to reasonable results. We thus do not include this variable in Columns X and XI of Table 2, because it is not defined in a way that could be used for counterfactual simulations.

D Example 2, continued

Let ψ be the permutation that exchanges players 3 and 4 while leaving players 1 and 2 unchanged. The grand coalition and the all-singleton partition are left unmodified by ψ, and thus we exclude them from calculations below. Additional partitions left unchanged and thus excluded are $\{\{1,3,4\},\{2\}\}$, $\{\{2,3,4\},\{1\}\}$, $\{\{1,2\},\{3,4\}\}$, $\{\{1\},\{2\},\{3,4\}\}$, and $\{\{1,2\},\{3\},\{4\}\}$. We label the remaining partitions as follows, highlighting the players to be permuted:

$$\pi_1 = \{\{1,2,3\},\{4\}\} \qquad \pi_2 = \{\{1,2,4\},\{3\}\}$$
$$\pi_3 = \{\{1,3\},\{2,4\}\} \qquad \pi_4 = \{\{1,4\},\{2,3\}\}$$
$$\pi_5 = \{\{1,3\},\{2\},\{4\}\} \qquad \pi_6 = \{\{1,4\},\{2\},\{3\}\}$$
$$\pi_7 = \{\{1\},\{2,3\},\{4\}\} \qquad \pi_8 = \{\{1\},\{2,4\},\{3\}\}.$$

We have set $v(\pi_1) = 1$ and $v(\pi_j) = 0$ for $j \neq 1$. This yields the following probabilities that the decision maker will select each partition:

$$\Pr(\pi_1) \simeq 0.275 \qquad \Pr(\pi_2) \simeq 0.050$$
$$\Pr(\pi_3) \simeq 0.070 \qquad \Pr(\pi_4) \simeq 0.070$$
$$\Pr(\pi_5) \simeq 0.038 \qquad \Pr(\pi_6) \simeq 0.066$$
$$\Pr(\pi_7) \simeq 0.038 \qquad \Pr(\pi_8) \simeq 0.066.$$

We see that $\Pr(\pi_1) > \Pr(\pi_2)$ because the deterministic payoff for π_1 is higher. Although the deterministic payoffs for all other partitions are the same, $\Pr(\pi_5) < \Pr(\pi_6)$ and $\Pr(\pi_7) < \Pr(\pi_8)$ because if ϵ_{13} or ϵ_{23} is positive then π_1 will likely be chosen instead of π_5 or π_6.

We now verify that the permutation rank ordering property holds. Expanding the summation in Inquality 16, we see that the first few terms are

$$(\Pr(\pi_1) - \Pr(\pi_2)) \cdot \mathrm{sign}(v(\pi_1) - v(\pi_2)) + (\Pr(\pi_2) - \Pr(\pi_1)) \cdot \mathrm{sign}(v(\pi_2) - v(\pi_1)) + \ldots$$

but $\mathrm{sign}(v(\pi_1) - v(\pi_2)) = -\mathrm{sign}(v(\pi_2) - v(\pi_1))$ and thus we can collect these, giving us

$$\begin{aligned}
&2\,(\Pr(\pi_1) - \Pr(\pi_2)) \cdot \mathrm{sign}(v(\pi_1) - v(\pi_2)) + \\
&2\,(\Pr(\pi_3) - \Pr(\pi_4)) \cdot \mathrm{sign}(v(\pi_3) - v(\pi_4)) + \\
&2\,(\Pr(\pi_5) - \Pr(\pi_6)) \cdot \mathrm{sign}(v(\pi_5) - v(\pi_6)) + \\
&2\,(\Pr(\pi_7) - \Pr(\pi_8)) \cdot \mathrm{sign}(v(\pi_7) - v(\pi_8)) > 0
\end{aligned}$$

for Inequality 16. We then see that $\Pr(\pi_3) = \Pr(\pi_4)$, so this term vanishes. Replacing

$\Pr(\pi_1)$, $\Pr(\pi_2)$, and $\text{sign}(v(\pi_1) - v(\pi_2))$ with their values, we obtain

$$\begin{aligned}
&2(0.275 - 0.050)+\\
&2\left(\Pr(\pi_5) - \Pr(\pi_6)\right) \cdot \text{sign}(v(\pi_5) - v(\pi_6))+\\
&2\left(\Pr(\pi_7) - \Pr(\pi_8)\right) \cdot \text{sign}(v(\pi_7) - v(\pi_8)) > 0,
\end{aligned}$$

but $\Pr(\pi_5) - \Pr(\pi_6) = \Pr(\pi_7) - \Pr(\pi_8) = 0.038 - 0.066 = -0.028$. Regardless of any small changes to $v(\pi_5)$,$v(\pi_6)$,$v(\pi_7)$, and $v(\pi_8)$, then, Inequality 16 will be satisfied.

E Proof of Proposition 1

Proof. Begin by letting $\psi\epsilon$ be the permuted version of the idosyncratic shocks. It is easy to visualize this if we express ϵ as a matrix, with $\psi\epsilon$ flipping the relevant indices. For example, if $N = 3$ then we would have

$$\epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{31} & \epsilon_{33} \end{bmatrix}$$

and

$$\psi\epsilon = \begin{bmatrix} \epsilon_{\psi^{-1}(1)\psi^{-1}(1)} & \epsilon_{\psi^{-1}(1)\psi^{-1}(2)} & \epsilon_{\psi^{-1}(1)\psi^{-1}(3)} \\ \epsilon_{\psi^{-1}(2)\psi^{-1}(1)} & \epsilon_{\psi^{-1}(2)\psi^{-1}(2)} & \epsilon_{\psi^{-1}(2)\psi^{-1}(3)} \\ \epsilon_{\psi^{-1}(3)\psi^{-1}(1)} & \epsilon_{\psi^{-1}(3)\psi^{-1}(1)} & \epsilon_{\psi^{-1}(3)\psi^{-1}(3)} \end{bmatrix} = \begin{bmatrix} \epsilon_{\psi(1)\psi(1)} & \epsilon_{\psi(1)\psi(2)} & \epsilon_{\psi(1)\psi(3)} \\ \epsilon_{\psi(2)\psi(1)} & \epsilon_{\psi(2)\psi(2)} & \epsilon_{\psi(2)\psi(3)} \\ \epsilon_{\psi(3)\psi(1)} & \epsilon_{\psi(3)\psi(1)} & \epsilon_{\psi(3)\psi(3)} \end{bmatrix},$$

where the later equality follows from ψ being self-inverse.

Now let $\psi\pi(\epsilon|v)$ be the permuted version of the chosen partition $\pi(\epsilon|v)$. We will show that the inequality

$$\int f(\epsilon) \sum_{\pi^0 \in \Pi} \left(1(\pi^0 = \pi(\epsilon|v)) - 1(\psi\pi^0 = \pi(\psi\epsilon|v))\right) \cdot \text{sign}\left(v(\pi^0) - v(\psi\pi^0)\right) d\epsilon > 0 \tag{33}$$

holds unless $v(\pi) = v(\psi\pi)$ for all partitions π. We will do this by examining every potential value for ϵ and its permuted version $\psi\epsilon$.

1. First, consider a "symmetric" value of ϵ such that $\epsilon = \psi\epsilon$. We are interested in the partition $\pi^1 = \pi(\epsilon|v) = \pi(\psi\epsilon|v)$, and its permutation $\psi\pi^1$. Either $\psi\pi^1 = \pi^1$ or it does not. We consider these two cases separately:

(a) If $\psi\pi^1 = \pi^1$, then

$$1(\pi^1 = \pi(\epsilon|v)) - 1(\psi\pi^1 = \pi(\epsilon|v)) = 0.$$

(b) On the other hand, if $\psi\pi^1 \neq \pi^1$, then it must be that $v(\pi^1) - v(\psi\pi^1) \geq 0$ because the idiosyncratic components $e(\pi^1, \epsilon)$ and $e(\psi\pi^1, \epsilon) = e(\psi\pi^1, \psi\epsilon)$ are equal. Thus

$$\left(1(\pi^1 = \pi(\epsilon|v)) - 1(\psi\pi^1 = \pi(\epsilon|v))\right) \cdot \text{sign}\left(v(\pi^1) - v(\psi\pi^1)\right) \geq 0,$$

and we will have the equivalent when considering the partition $\psi\pi^1$:

$$\left(1(\psi\pi^1 = \pi(\epsilon|v)) - 1(\pi^1 = \pi(\epsilon|v))\right) \cdot \text{sign}\left(v(\psi\pi^1) - v(\pi^1)\right) \geq 0.$$

Here the inequalities will be strict when $v(\pi^1)-v(\psi\pi^1) > 0$. If $v(\pi^1) = v(\psi\pi^1)$ for all partitions π^1 then the partition rank ordering property is satisfied immediately and we are done. If not, then there will be some π^1 such that $v(\pi^1) - v(\psi\pi^1) > 0$ and the inequalities are strict. We have thus shown that the inequality in the rank ordering property holds strictly so long as it holds weakly in the remaining cases.

2. Now, consider a value of ϵ such that $\epsilon \neq \psi\epsilon$. Let $\pi^1 = \pi(\epsilon|v)$. There are two possibilities:

(a) Suppose that $v(\pi^1) - v(\psi\pi^1) \geq 0$. If so, then it follows immediately that

$$\left(1(\pi^1 = \pi(\epsilon|v)) - 1(\psi\pi^1 = \pi(\epsilon|v))\right) \cdot \text{sign}\left(v(\pi^1) - v(\psi\pi^1)\right) \geq 0$$

and the equivalent when considering the partition $\psi\pi^1$:

$$\left(1(\psi\pi^1 = \pi(\epsilon|v)) - 1(\pi^1 = \pi(\epsilon|v))\right) \cdot \text{sign}\left(v(\psi\pi^1) - v(\pi^1)\right) \geq 0.$$

(b) Suppose that $v(\pi^1) - v(\psi\pi^1) < 0$. We are then interested in the nature of the partition $\pi(\psi\epsilon|v)$. There are two possibilities:

i. If $\pi(\psi\epsilon|v) = \psi\pi^1$ then

$$1(\pi^1 = \pi(\epsilon|v)) - 1(\psi\pi^1 = \pi(\epsilon|v)) = 0.$$

ii. If $\pi(\psi\epsilon) \neq \psi\pi^1$ then we need to again consider two different subcases. Let $\pi^a = \pi(\psi\epsilon|v)$ be the alternative partition that was actually selected by the decision maker instead of $\psi\pi^1$. Compare $v(\pi^a)$ to $v(\psi\pi^a)$:

A. If $v(\pi^a) > v(\psi\pi^a)$ then

$$\big(1(\pi^1 = \pi(\epsilon|v)) - 1(\psi\pi^1 = \pi(\epsilon|v))\big) \cdot \text{sign}\left(v(\pi^1) - v(\psi\pi^1)\right) = -1$$

but

$$(1(\pi^a = \pi(\psi\epsilon|v)) - 1(\psi\pi^a = \pi(\psi\epsilon|v))) \cdot \text{sign}\left(v(\pi^a) - v(\psi\pi^a)\right) = 1$$

and thus this ϵ and $\psi\epsilon$ pair will cancel to zero in (33) because $f(\epsilon) = f(\psi\epsilon)$ by exchangeability.

B. If $v(\pi^a) \leq v(\psi\pi^a)$ then we would be in trouble. We will show that this cannot occur by considering the choice between $\psi\pi^a$ and π^1 with idiosyncratic shocks ϵ. Begin by noting that

$$e(\psi\pi^a, \epsilon) + v(\psi\pi^a) = e(\pi^a, \psi\epsilon) + v(\psi\pi^a)$$

because in Equation 15 permuting both the player labels in π and the labels in ϵ results in exactly the same sum as leaving both unpermuted, by the self-inverse property of ψ. Then

$$e(\pi^a, \psi\epsilon) + v(\psi\pi^a) \geq e(\pi^a, \psi\epsilon) + v(\pi^a)$$

because of the subcase we are in, and

$$e(\pi^a, \psi\epsilon) + v(\pi^a) \geq e(\psi\pi^1, \psi\epsilon) + v(\psi\pi^1)$$

because π^a was chosen over π^1. But then

$$\begin{aligned} e(\psi\pi^1, \psi\epsilon) + v(\psi\pi^1) &= e(\pi^1, \epsilon) + v(\psi\pi^1) \\ &> e(\pi^1, \epsilon) + v(\pi^1), \end{aligned}$$

and combining all of this we are left with

$$e(\psi\pi^a, \epsilon) + v(\psi\pi^a) > e(\pi^1, \epsilon) + v(\pi^1),$$

which is a contradiction because π^1 was chosen over $\psi\pi^a$. Thus this subcase cannot occur.

□

F Smaller Players

In Section 7.1, we find substantial inefficiency in cases where there is both horizontal and vertical heterogeneity. This is in contrast to the standard result in Kaneko and Wooders [1986], that games with a "large" number of players should see the efficient sorting that is assumed in Tiebout models. One might wonder whether our result is due to the fact that our players form coalition in relatively small groups: if the size of the players were smaller (with correspondingly more players per coalition) , would the partition that emerged have different sorting properties? It turns out that this appears to not be the case.

Suppose that we wished to consider the "same" coalition formation game, except with smaller players. One way to do this, albeit in only an approximate sense, would be to start by halving the distance cost (e.g. $\alpha = 0.055$ instead of $\alpha = 0.11$), which is equivalent to halving all the distances $d(i, i')$. Then, in order to retain population density at its original value, we reduce all populations p_i and *koku* ratings y_i of players by $(1/2)^2 = 1/4$. This produces a "zoomed in" version of the original configuration.

If the number of players is large, and their distribution fractal, then this "zoomed in" configuration will have the same characteristics as the original configuration. Relative to the social optimum, the resulting number of coalitions in core partitions should thus be the same, except in the case where the core partitions in the initial game were anomalously large because the initial player sizes were too big. We find, however, that this effect is in fact not particularly important: "zoomed in" core partitions have 105% more coalitions than the social optimum. Thus, it appears that dividing players more finely does not change the amount of inefficiency in the coalition formation game.

G Post-Meiji Mergers (Col. IV and V, Tables 4 and 5)

The comparison of Figure 4 and Figure 6 suggests that perhaps the model developed in this paper, although intended for the data from the Meiji period, could be used to perform a more general analysis of municipal mergers. Since the beginning of the modern municipal system described in Nishikawa, Hayashi, and Weese [2018], there have been three waves of mergers in Japan: Meiji, Showa, and Heisei. The Meiji mergers were centralized. The Heisei mergers were decentralized, with the national government offering a specific set of financial incentives (including both a "carrot" and a "stick") in order to encourage municipalities to merge.[26]

The Showa mergers, however, appeared to have characteristics of both centralized and decentralized mergers. There was substantial involvement of higher levels of government, both national and prefectural. At the prefectural level, committees drew up formal merger plans for the entire prefecture: in the case of Aomori Prefecture, these plans were actually based on a formal matching model, complete with a payoff structure and an approximation algorithm. On the other hand, the final mergers had to be approved at the municipal level. One interpretation of this is that the municipal approval was mere window dressing, and municipalities could not refuse to participate in the national government plan. Our simulation results suggest another potential explanation for the seemingly contradictory "both centralized and decentralized" nature of the Showa mergers.

As documented in Nishikawa, Hayashi, and Weese [2018], during the Meiji period there were effectively no transfers from the national government to municipalities, and only extremely limited assistance from prefectures. This allows for the simple model developed in Section 2. After World War II, however, the Shoup Report commissioned by occupation forces recommended revisions to the municipal finance system. This resulted in the *Heikou Koufukin* (later renamed the *Chihou Koufuzei*), a transfer system that would eventually grow to enormous size. Two stylized facts regarding this transfer system are important: during the Showa period, it was widely considered to provide inadequate equalization for capital costs, while in the Heisei period capital costs were seen as overstated, and the system was seen instead as providing inadequate equalization for larger cities.

A stylized version of the transfer system in place throughout the post-war period is that it is based on lump-sum transfers.[27] Previous analysis shows that these transfers can be modelled as a fixed plus variable amount: this is directly related to the fact that the cost of providing services can be modelled in this way. With these transfers, from the local perspective the cost of providing services becomes $(\gamma_1 - T_1) + (\gamma_2 - T_2)\sum_{i \in S} p_i$, where T_1 is the fixed transfer amount, and T_2 the variable transfer amount, depending on the population to which services need to be provided.

A transfer scheme of this sort, from the perspective of the municipalities, is equivalent to

[26] For a recent discussion of the Heisei mergers in the domestic literature, see Machida [2006]. The general consensus appears to be that mergers occurred democratically, subject to the financial incentives. The "stick" portion of the national government funding formula change in particular appears to have led many small municipalities to participate in mergers. This qualitative discussion agrees with the theoretical results in Weese [2015] regarding when the "stick" should be used.

[27] The details of the system are quite complicated, and subsidies were also present. The appendix of Weese [2015] provides a brief English-language description.

Appendix Table G.30: Dependent variable is cost of providing services ('96-'97 fiscal year)

	I	II	III	IV	V
(Intercept)	1294.6	808.4	834.3	792.2	902.7
	(23.0)	(24.4)	(25.2)	(27.1)	(21.2)
POPULATION	136.4	136.0	136.6	142.3	142.5
	(0.3)	(0.3)	(0.3)	(1.7)	(1.3)
AREA		4.3	3.6	3.8	2.9
		(0.1)	(0.1)	(0.1)	(0.1)
INCOME.INEQ		0.04	0.03	-20.9	-12.4
		(4.8)	(4.9)	(4.3)	(3.3)
INCOME		-1070.4	-779.8	-164.9	-483.4
		(69.0)	(104.3)	(69.1)	(79.8)
IS.CITY		324.1	369.8	-16.2	295.4
		(54.9)	(54.2)	(59.2)	(48.1)
POP * INCOME.INEQ				1.1	0.2
				(0.1)	(0.1)
POP * INCOME				-30.5	-8.6
				(1.0)	(1.5)
POP * IS.CITY				5.4	-1.7
				(2.2)	(1.7)
PREFECTURE			X		X
N	3220	3216	3216	3216	3216

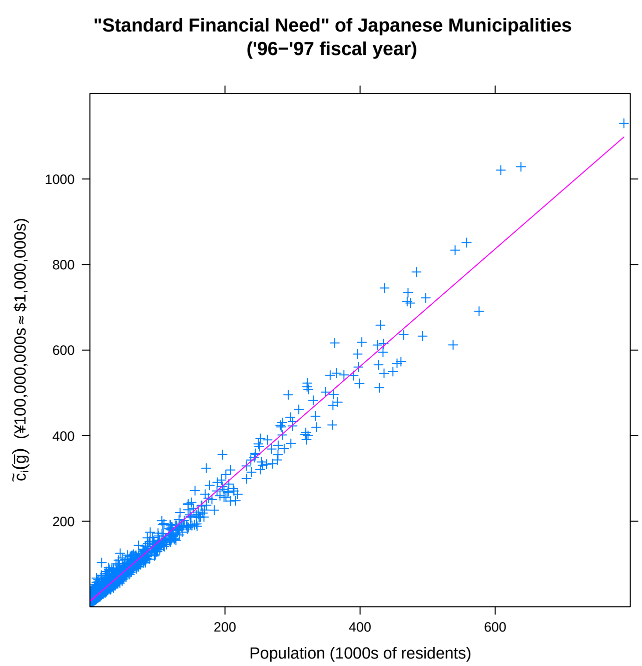
Units: ¥1,000,000 (roughly $10,000) per year. POPULATION is in thousands of residents, AREA is in square kilometers, INCOME is in ¥1,000,000 per capita per year, INCOME.INEQ is the coefficient of variation of income, IS.CITY is a dummy variable coded as 1 if the municipality in question is a city, and zero if it is a village or town. PREFECTURE is a set of dummy variables for each of the 47 prefectures, with the restriction that the sum of the coefficients on these variables must equal zero. Designated cities and special wards are excluded from the regression because they have additional responsibilities devolved from the prefectural governments, and thus have higher (and non-comprable) expenditures per capita.

Appendix Table G.31: Dependent variable is cost of providing services

	96-97	06-07
(Intercept)	899.9	582.2
	(43.9)	(59.5)
POPULATION	129.4	131.5
	(0.5)	(0.6)
AREA	4.6	4.6
	(0.2)	(0.2)
N	1194	1194

Units: ¥1,000,000 (roughly $10,000) per year. POPULATION is in thousands of residents, AREA is in square kilometers, designated cities and special wards are excluded as in Table G.30. The sample is further restricted to those municipalities that did not participate in a merger in order to have the same sample in both periods. Thus, the change in coefficients represents a change in national government transfer policy on the same group of municipalities during the period in question. Inflation during this period was negligible.

Appendix Figure G.32: Predicted and actual Standard Fiscal Need, Heisei data



a change in the fixed and variable costs to $\tilde{\gamma}_1 = \gamma_1 - T_1$ and $\tilde{\gamma}_2 = \gamma_2 - T_2$, respectively. Thus, to simulate the results that would occur in a decentralized set of mergers with a certain equalization transfer scheme, we need only to change the parameters γ_1 and γ_2 to $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ when running the decentralized simulations, but then retain the original γ_1 when calculating the optimal partition from the perspective of the social planner. We can thus ask what the effect of an equalization system of various sorts might have been on municipal mergers, by asking what the core of the decentralized coalition formation game would look like under different parameters for the cost function.

We first consider the case where the variable cost is fully subsidized, with $T_2 = \gamma_2$. This is a situation often considered in the theoretical literature, and corresponds very roughly to the case of the equalization payments in the Showa period. Column IV of Tables 4 and 5 show that in this case the number of mergers is very close to that of the social planner's optimal partition, and there is very little inefficiency. This result provides a potential explanation for the contradictory nature of the Showa mergers, which were apparently both centralized and decentralized at the same time.[28] With an equalization system of the sort in place during the Showa period, decentralized coalitions were very similar to those that the social planner would have selected. Thus, during the Showa period, there was close to no contradiction between the centrally planned merger pattern, and one that municipalities would have wanted to carry out.[29]

Now, consider a transfer scheme of the sort in use in the Heisei period. As documented by DeWit [2002] and others, the transfer system by this point had changed into one that was exceedingly generous with respect to capital investments, which were large for smaller municipalities, but did not really adequately compensate municipalities for the variable costs incurred. The extreme case of such a transfer scheme would be one in which the fixed cost is completely subsidized, in which case no players would be interested in any mergers in the decentralized case. In order to cause decentralized mergers to occur, the national government offered a set of financial incentives, both rewarding those municipalities that chose to participate in mergers, and penalizing those that did not. In an extremely rough

[28] Tanaka and Kadotami [1963] summarizes the view of the Showa mergers as centralized, citing Kawaguchi [1960b], Kawaguchi [1960a], Kawaguchi [1961], Oshima [1958], Oshima [1959], and Fukutake [1959]. Steiner [1965], however, gives a case study where local politics played a key role in the merger process. More recent research includes Ichikawa [2011], who emphasizes the importance of both local governments and the national government in the merger process, and Arakaki [2010], who argues that the large number of municipal splits is evidence of central control.

[29] One might imagine, for example, the central planner began by creating a partition. This was either then modified to ensure that it corresponded to a core partition, or (more likely, given the anecdotal evidence) the small number of reluctant municipalities were bludgeoned into accepting the proposed partition, even though blocking coalitions existed. An explanation of this sort does not appear anywhere in the literature, but instead arises from the simulations performed based on the model in this paper.

sense, these incentives correspond to municipalities becoming responsible for a portion of the fixed cost.

We calculate the fraction of the fixed cost based on a simplification of the calculations in Weese [2015]. Appendix Table G.30 shows that the transfers during this period correspond roughly to the fixed cost plus variable cost system in place since the Showa period. Appendix Table G.31 shows that the changes in the transfer system during the merger period, designed to provide incentives for merging, correspond roughly to a decrease in transfers of a lump sum of about ¥300 million per municipality, but the transfer corresponding to the variable cost component did not change. The merger "incentive" thus corresponds to roughly a third of the fixed cost, as the remaining two thirds was subsidized by the transfer scheme. Column V of Table 4 and 5 thus reports the case with municipalities considering a cost equal to only one third of the fixed cost.

References

Aitken, Robert (1977). "The Wilderness Areas in Scotland." PhD Thesis. University of Aberdeen.

Alesina, Alberto and Enrico Spolaore (Nov. 1997). "On the Number and Size of Nations". *Q. J. Econ.* 112 (4):1027–1056.

Arakaki, Jiro (Mar. 25, 2010). "The formation process of the municipal amalgamation policy". 16:65–80. [In Japanese].

Barros, Ana Isabel (1998). *Discrete and Fractional Programming Techniques for Location Models*. Dordrecht; Boston: Kluwer Academic Publishers.

Bixby, Robert E. (2012). "A Brief History of Linear and Mixed-Integer Programming Computation". *Doc. Math.* Extra Volume ISMP:107–121.

DeWit, Andrew (Oct. 2002). "Dry Rot: The Corruption of General Subsidies in Japan." *J. Asia Pac. Econ.* 7 (3):355–378.

Dinkelbach, Werner (Mar. 1967). "On Nonlinear Fractional Programming". *Management Science* 13 (7):492–498.

Easterly, William and Ross Levine (Nov. 1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions". *The Quarterly Journal of Economics* 112 (4):1203–1250.

Fontanari, Steno (2000). "Sviluppo di metodologie Gis per la determinazione dell'accessibilità territoriale come supporto alle decisioni nella gestione ambientale". Università degli Studi di Trento.

Fukutake, Tadashi (1959). *Nihon sonraku no shakai kōzō*. Tōkyō: Tōkyō Daigaku Shuppankai.

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1):100–108.

Hathaway, Richard J. and James C. Bezdek (Mar. 1, 1994). "Nerf c-means: Non-Euclidean relational fuzzy clustering". *Pattern Recognition* 27 (3):429–437.

Ichikawa, Yoshitaka (June 2011). "Japan's Municipal Mergers in Showa and Heisei Era : A Comparison of Two Government-Led Merger Promotion Policies". *Doushisha Hougaku* 63 (1):331–353. [In Japanese].

Kaneko, Mamoru and Myrna Holtz Wooders (Oct. 1986). "The core of a game with a continuum of players and finite coalitions: The model and some results". *Mathematical Social Sciences* 12 (2):105–137.

Kawaguchi, Akira (Jan. 1960a). "Chouson Gappei to Nouson-1-". *Q. J. Agric. Econ.* 14 (1):1–27. [Municipal Mergers and Farming Villages].

— (Oct. 1960b). "Chouson Gappei to Nouson-2-". *Q. J. Agric. Econ.* 14 (4):117–166.

— (July 1961). "Chouson Gappei to Nouson-3-". *Q. J. Agric. Econ.* 15 (3):35–77.

Langmuir, Eric (1984). *Mountaincraft and Leadership: A Handbook for Mountaineers and Hillwalking Leaders in the British Isles*. Edinburgh: Scottish Sports Council.

Liao, Xiyue and Mary C. Meyer (May 10, 2019). "cgam: An R Package for the Constrained Generalized Additive Model". *Journal of Statistical Software* 89 (1):1–24.

Machida, Toshihiko, ed. (2006). *Heisei daigappei no zaiseigaku*. [Public finance of the Heisei mergers]. Tokyo: Kōjinsha.

Mizoguchi, Tsunetoshi (Mar. 2004). *Edo-Meijiki Ni Okeru Chishi No Zuzouka Ni Yoru Souzouteki Chiiki Ron*. Kenkyuu Seika Houkokusho 14580084. [Creative Regional Studies from Visualization of Edo-Meiji Period Topography]. Nagoya University.

Naismith, W. W. (Sept. 1892). "Excursions: Cruach Ardran, Stobinian, and Ben More". *Scott. Mt. Club J.* 2 (3):135.

Nishikawa, Masashi, Masayoshi Hayashi, and Eric Weese (Feb. 2018). "Meiji Era Local Government". *Journal of Economics & Business Administration* 217 (2):101–125.

Oshima, Tarou (1958). "Chouson Gappei No Ronri to Mondaisei-1-". 34 (10):84–96. [Logic and Problems of Municipal Mergers].

— (1959). "Chouson Gappei No Ronri to Mondaisei-2-". 35 (1).

Skinner, George William (1988). "Nobi as a Regional System". Second Workshop of the Nobi Regional Project. Nagoya.

Steiner, Kurt (1965). *Local Government in Japan*. Stanford, Calif: Stanford University Press. 564 pp.

Stephan, G. Edward (Apr. 29, 1977). "Territorial Division: The Least-Time Constraint Behind the Formation of Subnational Boundaries". *Science* 196 (4289):523–524.

— (July 13, 1979). "Political Subdivision and Population Density". *Science* 205 (4402):219–220.

— (Sept. 1984). "Territorial Subdivision". *Social Forces* 63 (1):145–159.

Suzuki, Tsutomu (1999). *Size-Density Hypothesis in P-Median Problems*. no. 836. Institute of Policy and Planning Sciences discussion paper series.

Tanaka, Issei and Masaaki Kadotami (Oct. 1963). "A Cooperative Research Project on Educational Structure of Changing Rural and Mining Areas in Kyushu". *J. Educ. Sociol.* 18:123–137. [In Japanese].

Vining, Daniel R., Chung-Hsin Yang, and Shi-Tao Yeh (July 13, 1979). "Political Subdivision and Population Density". *Science* 205 (4402):219–219.

Weese, Eric (July 2015). "Political Mergers as Coalition Formation: An Analysis of the Heisei Municipal Amalgamations". *Quantitative Economics* 6 (2):257–307.

Yamaoka, Yoshiaki (1977). "An Historico-Geographical Study about the Dissolution of Communal Forest:In the Case of Takahashi-Yama in the North of Nara Prefecture". *Jpn. J. Hum. Geogr.* 29 (3):313–332. [In Japanese].