Inefficiency and Self-Determination: Counterfactual Evidence for Meiji Japan

Hayashi Nishikawa Weese

June 2014
(Very Preliminary)

Abstract

Does the exercise of the right of self-determination lead to inefficiency? This paper considers a set of centrally planned municipal mergers during the Meiji period, with data from Gifu prefecture. Quantitative and qualitative evidence suggests that the observed merger pattern can be explained as a social optimum based on a very simple individual utility function. If individual villages had been allowed to choose their merger partners, counterfactual simulations show that the core is always non-empty, but core partitions contain about 80% more (post-merger) municipalities than the social optimum. Substantial inefficiency occurs only due to a combination of vertical and horizontal heterogeneity, and efficiencies of scale and congestion in the provision of public services.

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, and recent world events, such as Russian annexation of the (apparently consenting) Crimean peninsula, suggest policy relevance. Empirical results regarding the efficiency of jurisdiction formation, however, are very limited.[1]

This paper considers a set of municipal mergers that were decided by a central planner, and shows that the observed pattern of boundaries corresponds closely to that of a social planner maximizing a sum of utilities of individuals in an extremely simple model: there is a tradeoff between efficiencies of scale in the provision of public goods, and a geographic distance cost to the point where the public goods are provided. The parameter governing this tradeoff is estimated using a moment-inequalities approach, and an parameter describing

[1]Weese [2014] considers recent Japanese data where the central government provides equalization payments to municipalities. Inefficiency in this environment arises due to the presence of asymmetric information regarding the benefits of potential mergers. 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 efficiencies of scale is calibrated using additional data. The counterfactual case is then considered, where pre-defined subunits are allowed to choose how to arrange themselves into jurisdictions in a coalition formation game without transfers.

The core is used as the solution concept in this decentralized jurisdiction formation game, and the following results are obtained via simulation. First, the core is always non-empty, given the actual data and parameter estimates. This is not obvious theoretically, and there are well-known simple examples Gale and Shapley [1962] where the core is empty in a one-sided matching game. Despite these theoretical difficulties with the concept of a "right to self-determination", however, empirically there are always outcomes that meet the relevant requirements. Furthermore, if there is only horizontal (geographic ideal point) heterogeneity, then the core consists entirely of partitions that are "close" to the optimal partition in social welfare terms. The partition that would be selected by a social planner is also often in the core. Thus, if there is only horizontal heterogeneity in preferences, allowing jurisdiction boundaries to be established by a decentralized process appears to result in minimal or potentially even no inefficiency.

In contrast, in the case where there is also vertical (income) heterogeneity across players, simulations show that decentralized mergers result in substantial inefficiency. This result is due to the form of the cost of public goods for the data used: there is both a fixed cost that generates efficiencies of scale, as well as a marginal cost that reflects high congestibility. In this environment, adding an additional player who is substantially poorer will increase the tax rate for members of a coalition, and thus a decentralized process results in jurisdictions that are substantially smaller than the social optimum. Given the heterogeneity in income observed in the data, and the apparent cost function for the provision of public goods, decentralization of jurisdiction formation would result in substantial inefficiency.

1 Model

Let there be N players, each of whom must receive local government services. Player i has income y_i, requires p_i units of public services, and is located at point θ_i on a two-dimensional plane. A coalition S of players can provide P units of public services at any single point θ_S on the plane at a cost of

$$c(P) = \beta_1 + \beta_2 P \tag{1}$$

where β_1 and β_2 are non-negative. If $\beta_1 > 0$, then average cost declines as more players are added to a coalition. There is no choice regarding the quality of the services provided, and the cost of providing them must be paid via proportional taxation. The budget constraint for municipality S is

$$\beta_1 + \beta_2 \sum_{i \in S} p_i = \tau_S \sum_{i \in S} y_i$$

Other than the proportional taxation at τ_S, municipalities do not perform any redistribution.

The utility function for player i, when facing a tax rate τ_S, and receiving services at point θ_S on the plane, is quasi-linear:

$$u_i(\tau_S, \theta_S) = -||\theta_S - \theta_i||y_i - \tau_S y_i$$

Richer players suffer greater disutility from distance: if the plane represents geographic locations, then this could be due to a greater time cost of travel for those with higher incomes.

For a given municipality S, there is no choice regarding the tax rate τ_S as it is determined by the budget constraint. There is a choice, however, regarding the location on the plane at which to provide the public services. As the set of potential locations is a plane, this is a multi-dimensional decision problem, and has no generally accepted solution. Assume that the political decision-making process within the municipality maximizes the sum of players' utilities: this gives more weight to richer players, which may be appropriate if the political system is biased towards the wealthy. The major advantage of this assumption is that income redistribution becomes irrelevant: transfers within a coalition will not change the sum of the utilities of its members. Let θ^*_S be the location at which the service will be provided by municipality S.

The social planner's objective is also to maximize the sum of players' utilities. Thus, for any municipality S, the location θ^*_S chosen by the municipality is exactly what the social planner would choose for that municipality. There is thus no inefficiency from within municipality political decision-making. Similarly, income redistribution is irrelevant. Inefficiency may arise, however, if players arrange themselves into municipalities via a coalition formation game, as the coalitions that result may not be those that the social planner would have chosen. Assume that players cannot make transfers or bargain over the location choice θ in order to achieve agreement to participate in a given coalition. This could be due to a commitment problem if coalitions are permanent, and the location is chosen only after coalitions are fixed.

Let $\pi \in \Pi$ be a partition of players into coalitions, where Π is the set of all possible partitions. The solution concept used will be the core: $\pi \in \Pi^*$ is a solution to the coalition formation game if and only if $\forall S' \notin \pi, \exists i \in S'$ such that $u_i(\tau_{S(i,\pi)}, \theta^*_{S(i,\pi)}) > u_i(\tau_{S'}, \theta^*_{S'})$, where $S(i, \pi)$ indicates the coalition that i is a member of in partition π. Non-emptiness of the core is not guaranteed with the utility functions presented above, but it will be shown below to be true empirically, for the data considered.

Let π^{FB} be the partition that would be chosen by the social planner. This first-best partition may not be in the core. Intuitively, players at the "edge" of a coalition are experiencing high disutility from distance, and at the same time are being charged a tax rate τ determined by the total cost of providing services, rather than the marginal cost. Thus, the incentives faced by players considering a deviation are different from the objective being maximized by the social planner. With no restrictions on the cost functions c, or the distribution of y_i and θ_i, few theoretical results are known. However, results are available for more restricted cases:

1. [Farrell and Scotchmer, 1988] If $\beta_2 = 0$, and there is no variation in θ, then the grand coalition is efficient, and will form in decentralized mergers. This is because the marginal cost of adding an additional player is zero, and thus such an addition will never be opposed.

2. [Farrell and Scotchmer, 1988] If there is no variation in θ, and $y_i \sim F$ for some distribution F, then the inefficiency of decentralized coalitions decreases to zero in per capita terms as additional players are added. The intuition here is that a player with higher income does not want to merge with one that has lower income if the tax collected from the poorer player would not cover the marginal cost of providing services. Thus, richer players will not merge with poorer players. However, the number of players with similar incomes increases with population, and thus the size of coalitions will be increasing in total population.

3. If y_i is constant, and the distribution of θ is uniform, then the efficient partition is a hexagonal tiling of the plane (known at least as far back as the "central place theory" of Christaller [1933]. Drèze et al. [2008] provides extensive citations) 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.

Gregorini [2009] shows that obtaining theoretical results for the more general case where there is variation in both income and ideal points will be difficult because of non-existence problems. The empirical results below will show that, however, for the actually observed data and estimated parameters, there is no non-existence problem. Furthermore, horizontal heterogeneity alone results in only minimal inefficiency; however, substantial inefficiency occurs when there is both horizontal and vertical heterogeneity.

2 Data

The data used come from a set of municipal mergers that were mandated by the central government during the 1890s in Japan. These mergers occurred across the country: data is available is for Gifu Prefecture, and covers 1095 feudal villages (*shizen son*) that were combined to form 288 western-style municipalities.[2]The initial boundaries of the *shizen son* are shown in Figure 1, and the boundaries after the mergers are shown in Figure 2. The major advantage of historical data is that the public services provided were simple and limited, and the number of potential externalities that need to be considered are very limited. Almost all municipal spending was devoted to three items: education, public works, and general administration. Education spending was almost entirely devoted to primary schooling, generally using a single school. Public works spending consisted mainly of spending on roads, irrigation, and flood prevention. Administration consisted mainly of maintaining the population register, collecting taxes, and some shared organizational overhead for the other expenditures. The major source of tax revenue was property tax, which was proportional to land holdings.[3] Decision-making was democratic, but with a restricted franchise, and a two-tier voting system that gave extra weight to the elite.

Japan was overwhelmingly rural during this period: while population density varies substantially depending on terrain, there are only two locations in Gifu that could reasonably be considered to be cities, consisting of less than 10% of the population. The lack of large urban agglomerations also reduces the number of potential externalities that need to be considered. In rural areas, the only plausible spillovers between municipalities are related to public works: irrigation that uses scarce water, levees to protect against flooding, or roads. Below (unfinished), we consider each of these, and argue that they do not offer a plausible explanation for the observed pattern of mergers.

The observed pattern of mergers follows the classic arrangement of administrative units: areas 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. The pattern appears to hold across a wide selection of

[2]The original data source is the *Gifu-ken Chouson Ryakushi* and related documents; the version used is courtesy of Tsunetoshi Mizoguchi and Kei Okunuki, based on an original version at the Skinner Data Archive.

[3]The other major source of tax revenue was a household tax that was related to income, but was not strictly proportional. There appears to be debate in the literature regarding the exact nature of this tax Takayose [2000]. The assumption used in this paper is that taxation is overall proportional, and thus that the non-proportionality in the household tax is due to the fact that poorer households derive less of their income from property, and thus are taxed more heavily on the household tax.

countries, and varieties of jurisdictions Stephan [1984].[4] In terms of economic theory, the observed pattern is consistent with the model presented above, in which there are efficiencies of scale in the provision of services that are provided at a single point, combined with some distance cost. The pattern is difficult to explain, however, using a model that explains the structure of jurisdictions by claiming that the benefit of larger jurisdictions is the internalization of externalities. As with other goods, externalities would generally be considered to have decreasing marginal benefit (or increasing marginal damage). Externalities that are generally plausible, such as pollution, traffic, or the exhaustion of common pool resources, should generally be observed at higher levels in areas with higher population density. Thus, the benefit of internalization should be higher, and larger jurisdictions should be observed in these areas. In areas with low population density, the damage from the externalities is limited, and so there is no need for large jurisdictions. Thus, the observed pattern of jurisdictions is opposite of that which would be expected in the case where jurisdiction structure was determined by externalities.

To match the model to the data, let the players be the feudal villages (*shizen son*), and let p_i be the population of village i. Income is taken to be the *koku* rating of the village: these ratings converted all production in a village into common units of rice, for the purpose of taxation. Let the ideal point θ_i be represented by the geographic location of the village.

With this setup, the variable cost β_2 is not identified using the data regarding the partition chosen by the central planner, as for any value of β_2 the rankings of partitions in terms of social welfare are the same. For an estimate of β_2 additional data must be used. During the mergers, the government produced a document describing the cost of providing public services for municipalities of three sizes:[5]

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

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 in Equation 1, with a fixed cost $\beta_1 = c_1/3$, and a variable cost $\beta_2 = \frac{2c_1/3+c_2}{3165}$, despite the fact that statistically there is an extra degree of freedom.

[4] Suzuki [1999] provides additional citations along with an application to Japanese data.

[5] 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 a briefly-used *ku* (ward) system.

There is no explanation offered of how the government arrived at these estimates, 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.

There is no documentation available describing how these numbers were arrived at, but some 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 4, along with the government cost functions. 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. Finally, 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. 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. 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 4.[6]

3 Estimation

The observed outcome is a partition of N players into coalitions. Estimation of the parameters of a structural model where the outcome takes this sort of partition form presents substantial difficulties. As any structural model proposed is extremely unlikely to explain the observed outcome perfectly, an error term of some sort needs to be included as part of the data generating process. The number of possible coalitions, however, is very large: with N players, there are 2^N potential coalitions. The total number of possible partitions of players into coalitions is given by the Bell numbers, and grows even more quickly in N.

A model that includes an idiosyncratic term for each of these possible mergers will lead to some potential mergers having extremely positive idiosyncratic shocks. This causes problems similar to those discussed in Berry and Pakes [2007], where an increasing number of products results in consumer utility increasing without bound if there is an idiosyncratic shock representing "taste for products". The situation here, however, has an additional feature that makes a "taste for products" approach particularly unattractive: the number of

[6] 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.

potential coalitions is a function of the size of the coalition: there are N coalitions of size 1, $N(N-1)$ coalitions of size 2, and $\binom{N}{k}$ coalitions of size k. If idiosyncratic shocks are uncorrelated, some of the larger coalitions will have positive shocks much larger than those of any of the smaller coalitions.

Without careful attention to the correlation structure of the error term, estimating such a model will generally lead to parameter estimates that suggest that only the smallest mergers are attractive from a structural perspective, and that the larger observed configurations occurred only because these were mergers with very positive idiosyncratic shocks. From a theoretical perspective, it might be possible to estimate a correlation structure for the idiosyncratic shocks such that the model does not have this undesirable feature. From a practical perspective, estimators of this sort tend to involve high dimensional integrals and are not computationally feasible. This paper proposes an alternative solution to this problem, based on an errors-in-variables framework.

Specifically, let π be the observed partition of municipalities into mergers. Let the payoff for merger S be

$$u(S|\beta) = X_S\beta_0 + g(Z_S|\beta)$$

There are some covariates Z that enter the restriction through an arbitrary function g, but there are also covariates X that enter the restriction linearly. The parameter vector can be normalized, as is standard in the discrete choice literature: let $\beta_0 = 1$.

If the partition π was chosen by the social planner over the alternative partition π', then it must be that

$$\Sigma_{S\in\pi} u(S|\beta) - \Sigma_{S\in\pi'} u(S|\beta) \geq 0$$

Suppose that this restriction always holds for the true covariates, but that there is measurement error in the X:

$$u(S|\beta) = X_S^*\beta_0 + g(Z_S|\beta) \tag{2}$$

$$X_S = X_S^* + \eta_S, \qquad E(\eta|X^*, \pi, Z) = 0 \tag{3}$$

Here η is measurement error, and the only endogenous variables are the X. A standard moment inequalities approach, following Pakes [2010], can then be applied. Specifically, assume that the restriction holds in expectation, and proceed to estimate β using sample moments.[7]

[7] In the case where mergers were decided via a decentralized process, rather than a social planner, the set of restrictions becomes more complicated but could still be analyzed in the same framework. Specifically, in the case where there are disjunctive restrictions (at least one out of a set must hold) the maximum can be used. In the case where all the restrictions in a set must hold, then these can be used directly.

For estimation, alternative mergers that could have occurred but did not need to be constructed, and these should include mergers that are both smaller and larger than those that actually occurred. For the mergers that are smaller, consider the three *shizen son* at each three-way intersection of actual mergers. Suppose that these three subunits had instead been broken off and formed into their own municipality. Such a change would increase the total number of municipalities by one, and decrease the distance to the location where services are provided in the affected municipalities.[8] The moment inequalities approach states that on average, such an increase in the number of municipalities should not lead to a better outcome.

As a second set of alternative potential mergers, consider the case where a municipality that was actually observed was instead eliminated, and its component *shizen son* distributed to adjacent municipalities on the basis of adjacency and total population.[9] This sort of alternative partition reduces the total number of municipalities by one, and thus will tend to be attractive when the parameter estimate for the relative importance of distance is low.

To obtain a reasonable identified set, a number of moment inequalities need to be used. Figure 3 shows the tradeoff between distance and cost savings for these alternative potential mergers. The inequalities just described should hold for municipalities regardless of their population, but savings due to efficiencies of scale are different at different sizes. Consider small (< 800), medium (1600 - 2400), and large (2400-) municipalities separately. Similarly, the inequalities should also hold for various levels of wealth: consider municipalities with a low koku rating (<2000) and a high koku rating (>2000) separately. Finally, consider those municipalities in particularly rural areas, as indicated by the presence of mountains, separate from those municipalities in the flatter central areas.

Using these moment inequalities, a relatively small identified set is obtained, corresponding approximately to $\beta_1 \in [2.2, 2.3]$ (not exactly right). The 95% confidence set is quite a bit larger, at about $[1.8, 2.4]$ (also not exactly right). β_2 is not identified, as player i always results in a variable cost of $\beta_2 p_i$, regardless of partition. β_0 is normalized to a cost of 1 per kilometer.

To evaluate model fit, the optimal partition for the social planner can be calculated, based on the parameter estimates. The optimization problem here is NP difficulty (cite?), and thus an approximation algorithm is necessary. Figures 5 and 6 show the results of an

[8] Ignore those cases where one or more of the municipalities consists only of a single *shizen son*, as in this case the total number of municipalities is not increased (the other changes in boundaries are unlikely to be informative regarding the parameter).

[9] The model assumes that distance is measured with error, and thus it would be inappropriate to choose alternative partitions on the basis of distance. Geographic contiguity is assumed to not be measured with error. This implies that the measurement error in distance takes the form of a diffeomorphism, where topology is the same before and after the distortion.

approximation based on local optimization: divide the players into geographic subregions with about 40 players each, and calculate the optimal partition of each subregion.[10] Then, taking all these partitions together, look for any "local improvements" that could be obtained by moving a player from coalition to another, or merging two coalitions.[11] This is similar to the "k-means" algorithm commonly used in the analysis of clustering.

4 Counterfactual Simulations

Suppose that, instead of centralized mergers, the individual feudal villages had been allowed to decide independently how to arrange themselves into municipalities. With N players, this leads to a potential 2^N coalitions. Restrictions thus need to be imposed so that the coalition formation game is computationally feasible. Specifically, consider only coalitions that are geographically contiguous, and consist of 9 players or less.[12] Furthermore, rather than considering the coalition formation game as a whole, with about a thousand players, break the game into geographically contiguous units with about 40 players each. This corresponds to the standard IO setup of considering geographic "markets" for goods. To ensure that these restrictions themselves do not bias the results, the socially optimal partition will be computed using this same restricted set of coalitions.

To compute the core of each coalition formation game, an approach based on random myopic deviations will be used. As the number of potential partitions is finite, a properly randomized approach will eventually find all partitions in the core. Ensuring reasonable performance is difficult, however, because of the possibility of cycles of myopic deviations. The coalition formation game considered in this paper contains "minimal inaccessible coalition configurations" as described in Papai [2003], and thus the no-cycle results from the two-sided matching literature, such as Roth and Vate [1990], do not apply. A brute force approach to avoid cycles is thus used. The algorithm is provided in Appendix A.

The core of the decentralized coalition formation game that results is never empty, and while occassionaly a singleton, it generally consists of a small number of partitions, all corresponding to about the same level of social welfare. The difference between the best

[10]These subregions are the same as those used below for the simulations of the decentralized mergers.

[11]Splits of coalitions are not considered here, as the main problem with starting with the optimal partition of subregions is that this will lead to too many coalitions, relative to the optimal partition, rather than too few.

[12]This covers the vast majority of actual mergers. As an additional modification for computational simplicity, eliminate the smallest players: iteratively merge each *shizen son* with a population of less than 250 (340?) with the neighbouring *shizen son* that has the closest koku rating per capita. This reduces the number of players to about 780, instead of 1095, but more importantly reduces the number of mergers with a large number of players, such that considering mergers with up to 9 players covers virtually all of the potentially stable coalitions.

and worst element of the core is minimal, corresponding to less than the fixed cost of one additional municipality, in games where twenty municipalities are forming on average. The partitions in the core, however, are substantially different than the partition that would be selected by the social planner: on average there are a total of 518 municipalities, rather than 288. This results in a higher cost of providing public services, only partially offset by the lower distance cost: in welfare terms, the average decentralized partition is worse by an amount equivalent to the fixed cost of 132 municipalities.

One question is what the core would look like if the cost function were different. Suppose that there were no variable cost: $c(P) = \beta_1$ for all populations P. This is a situation often considered in the theoretical literature. In this case, the core is much larger, often consisting of hundreds of possible partitions. The difference among these partitions in welfare terms, however, is small, corresponding to the fixed cost of one or two additional municipalities. When comparing the core to the partition that would have been selected by the social planner, the situation is now very different than the case in which there was also a variable cost of providing public services: the core now often includes the planner's preferred partition, and when it does not, the difference in welfare terms between the parititions in the core and the planner's preferred partition is very small, usually corresponding to less than the fixed cost of one additional municipality. Thus, while theoretically there may be inefficiency due to decentralized mergers in this environment, from an empirical perspective this inefficiency, when present, is very small.

The intuition for this result is that, with only a fixed cost of providing services, adding any member to a coalition will decrease the tax rate that needs to be charged. If there were no horizontal (geographic) differentiation, this case would correspond exactly to that of Farrell and Scotchmer [1988], where there is no inefficiency. The presence of geographic heterogeneity creates the potential for inefficiency, but empirically this is not severe. Given that the core in theoretical models is often quite large, it is not particularly surprising that the best partition in the core is either close to or the same as the socially optimal partition. The more surprising aspect of the empirical result is that there are no elements of the core that are particularly bad in welfare terms. 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.

An identical result, where all partitions in the core are close to the socially optimal partition, in welfare terms, is also obtained for the case in which the cost function contains a variable cost but income per capita is assumed to be the same for all players. The explanation for this result is immediate given the above: if there is no variation in income, then tax per capita is equal to the variable cost plus the average fixed cost. Adding an additional member to a coalition thus always decreases the tax rate, and the environment is identical to the case where there is no variable cost.

A separate question is what results would occur if there were vertical (income) heterogeneity, but no horizontal (geographic) heterogeneity. This case corresponds exactly to the model in Farrell and Scotchmer [1988]. Here the socially optimal result is for the grand coalition to form, but the presence of the variable cost means that players will minimize their tax rate by avoiding forming coalitions with poorer players. Empirical estimates here give a total of about 60 coalitions forming. Thus, in one sense the inefficiency resulting from vertical heterogeneity only is dramatically larger than that resulting from horizontal heterogeneity only. However, given the cost function, if the number of players is increased, with the new players having per capita incomes drawn from the same income distribution, the number of municipalities will increase at a slower rate than the number of players. Thus, asymptotically, the per capita inefficiency will fall to zero as the number of players becomes large. In contrast, suppose that the asymptotics considered for the case of horizontal heterogeneity are that additional players are always added on the "edge", and the game thus expands across the plane while maintaining a constant density of players. In this case the level of inefficiency will not decrease as more players are added. Thus, with this asymptotic interpretation, the inefficiency resulting from only vertical heterogeneity eventually disappears, while the inefficiency resulting from only horizontal heterogeneity is low but constant in the number of players.

In order to obtain substantial amounts of inefficiency empirically, then, both horizontal and vertical heterogeneity are required. In addition, the cost function must include a component that looks like a fixed cost (to provide a incentive to form coalitions), and a component that looks like a variable cost (to provide an incentive to keep relatively poorer players out of the coalition). This situation corresponds to the situation that appears to actually be the case empirically, and thus it appears that decentralized mergers, had they been implemented, would have resulted in substantial amounts of inefficiency. However, as theoretical models with both vertical and horizontal heterogeneity appear to be very difficult to solve, the empirical results suggest that it may be difficult to capture the empirically relevant sources of inefficiency in a tractable theoretical framework.

5 Conclusion

Empirical results show that in the case with both vertical and horizontal heterogeneity, decentralized jurisdiction formation results in substantial inefficiency, but this is not the case in the situation with only horizontal heterogeneity. This suggests that allowing political subunits to exercise a "right to self-determination" may result in jurisdictions that are close to the optimal partition if there is little heterogeneity in income, but could result in substantial inefficiency as income heterogeneity increases. Considered in the international context, then, the case of Crimea (poorer than the rest of the Ukraine) choosing to join Russia may actually be less worrisome than, for example, relatively richer areas (such as Catalonia) seeking to leave relatively poorer areas (such as the rest of Spain). The degree of inefficiency that would result from these sorts of deviations depends on the cost function for providing national public goods, for which no good estimates are available, and thus it is difficult to draw firm conclusions about the severity of the potential inefficiency.[13] If the cost function for national public goods is similar to the cost function for local public goods, though, the inefficiency may be economically significant.

[13] Also, the international case may allow for transfers, which leads to a substantially different model.

References

S. Berry and A. Pakes. The pure characteristics demand model*. *International Economic Review*, 48(4):1193–1225, 2007. ISSN 1468-2354. doi: 10.1111/j.1468-2354.2007.00459.x. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2354.2007.00459.x/abstract`.

W. Christaller. *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Wissenschaftliche Buchgesellschaft, 1933. ISBN 9783534044665.

J. Drèze, M. Le Breton, A. Savvateev, and S. Weber. "Almost" subsidy-free spatial pricing in a multi-dimensional setting. *Journal of Economic Theory*, 143 (1):275–291, Nov. 2008. ISSN 0022-0531. doi: 10.1016/j.jet.2008.01.006. URL `http://www.sciencedirect.com/science/article/pii/S0022053108000343`.

J. Farrell and S. Scotchmer. Partnerships. *The Quarterly Journal of Economics*, 103(2): 279–297, May 1988. ISSN 00335533. URL `http://www.jstor.org/stable/1885113`.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, Jan. 1962. ISSN 00029890. URL `http://www.jstor.org/stable/2312726`. ArticleType: primary_article / Full publication date: Jan., 1962 / Copyright Â© 1962 Mathematical Association of America.

F. Gregorini. Political geography and income inequalities. Working Paper 152, University of Milano-Bicocca, Department of Economics, 2009. URL `http://ideas.repec.org/p/mib/wpaper/152.html`.

A. Pakes. Alternative models for moment inequalities. *Econometrica*, 78(6):1783–1822, 2010. ISSN 1468-0262. URL `http://dx.doi.org/10.3982/ECTA7944`.

S. Papai. Random paths to stability in hedonic coalition formation. Technical report, University of Notre Dame, June 2003.

M. Reiter and A. Weichenrieder. Are public goods public? a critical survey of the demand estimates for local public services. *FinanzArchiv: Public Finance Analysis*, 54(3):374–408, 1997.

A. E. Roth and J. H. V. Vate. Random paths to stability in two-sided matching. *Econometrica*, 58(6):1475–1480, Nov. 1990. ISSN 0012-9682. doi: 10.2307/2938326. URL `http://www.jstor.org/stable/2938326`.

G. E. Stephan. Territorial division: The least-time constraint behind the formation of subnational boundaries. *Science*, 196(4289):523–524, Apr. 1977. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.196.4289.523. URL `http://www.sciencemag.org/content/196/4289/523`. PMID: 17837082.

G. E. Stephan. Territorial subdivision. *Social Forces*, 63(1):145–159, Sept. 1984. ISSN 0037-7732. doi: 10.2307/2578862. URL `http://www.jstor.org/stable/2578862`.

T. Suzuki. Size-density hypothesis in p-median problems. Technical Report no. 836, Institute of Policy and Planning Sciences discussion paper series, 1999.

S. Takayose. *Meiji chihō zaiseishi.* Keisō Shobō, Tōkyō, dai 1-han edition, 2000. ISBN 4326951133.

Figure 1: Initial arrangement of *shizen son*

Figure 2: Political boundaries after mergers

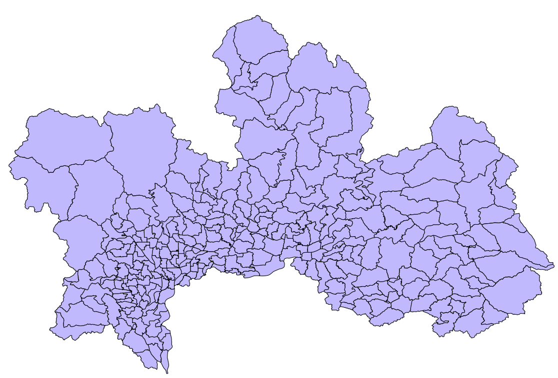


Figure 3: Moment Inequalities

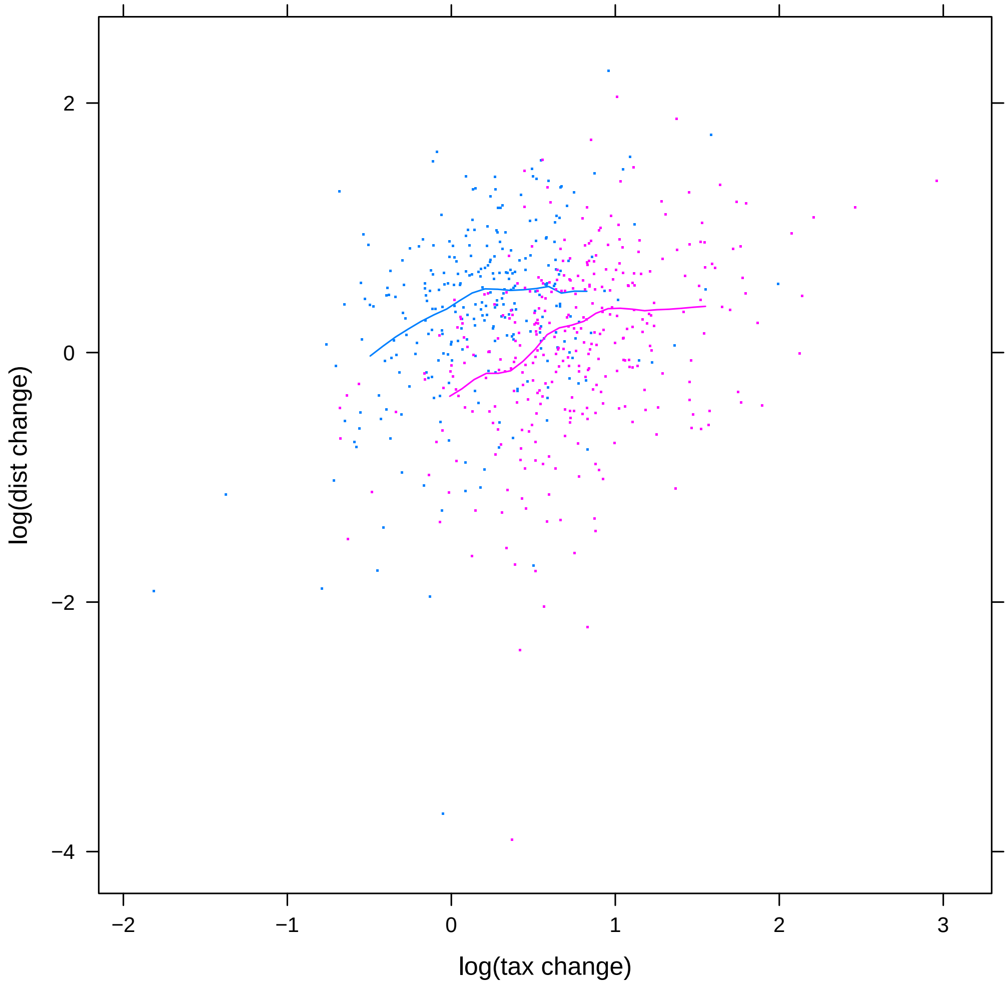
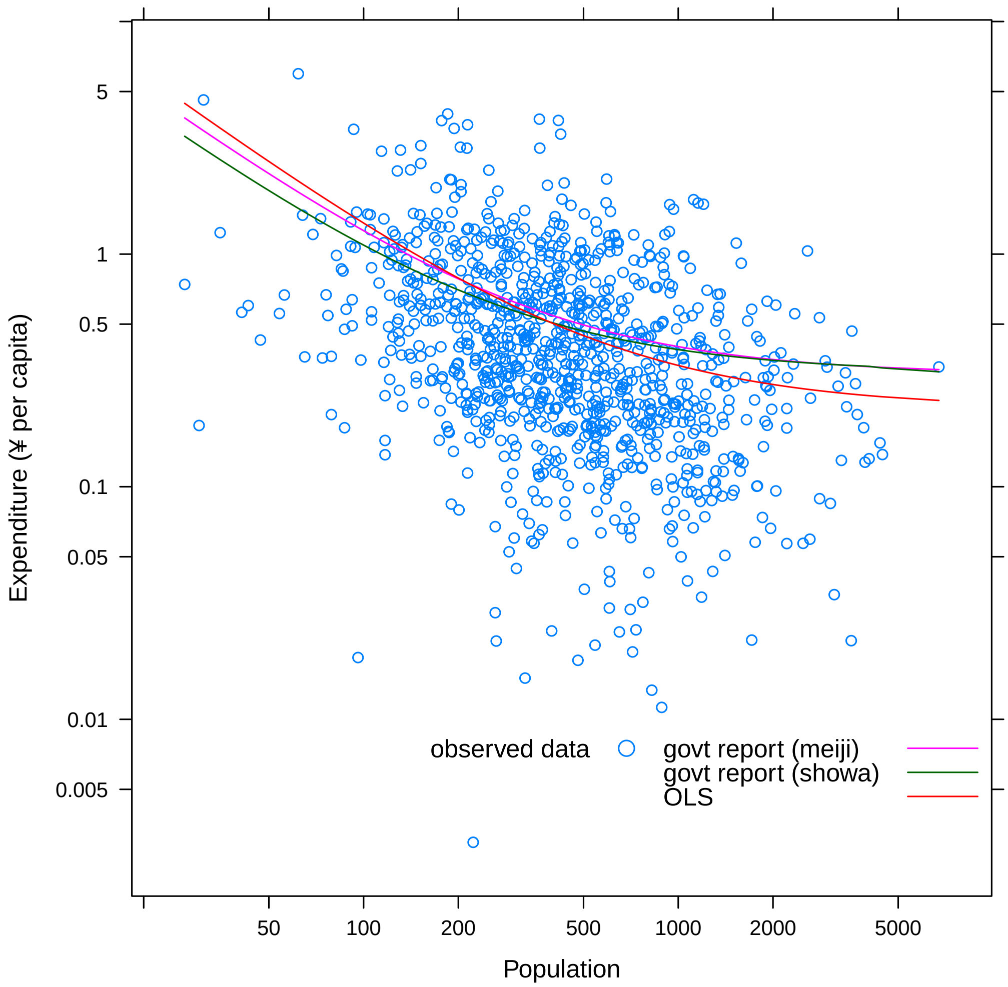
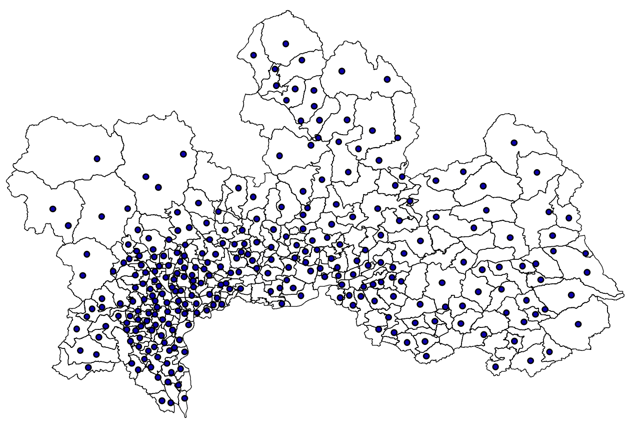


Figure 4: Public good spending per capita



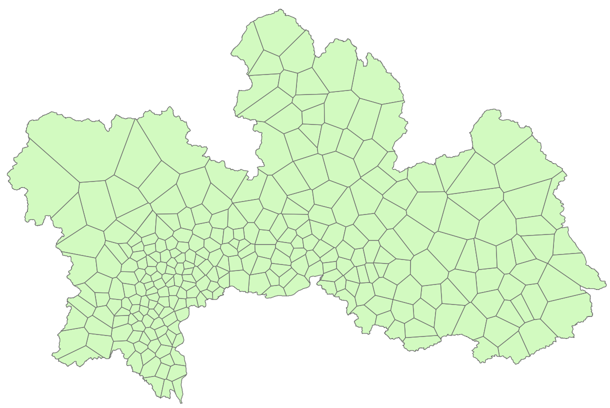
Data points indicate actual spending by precursors to final municipalities: this data is from 1881, when the final municipal system was still under development. OLS is bivariate, predicting total spending based on population and intercept (the line is curved as a result of transformation to log scale). The Meiji "govt" line is based on the fixed cost plus variable cost associated with the three data points discussed in the main text, with an adjustment to take into account that some services were paid not paid for by municiapalities 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 5: Predicted mergers vs. actual mergers



The dots indicate the centroids of the mergers the social planner is predicted to select, given the estimated parameters. The lines indicate the boundaries of the mergers that were actually observed. A good fit of the model with the observed data is indicated by one centroid per observed merger. Areas where there are mulitiple centroids per observed merger, or fewer centroids than observed merger, are areas where the model fit is not as good.

Figure 6: Predicted boundaries



Voronoi cells calculated based on the centroids of the optimal partition, based on parameter estimates. If a new player with very small population were added to the data, the cells indicate which municipality the social planner would assign the new player to. The model predicts substantial variation in the size of municipality between the densely populated plain in the middle of the prefecture, and the more mountainous areas at the edges.

A Core Algorithm

The following algorithm returns a random partition in the core, if one can be found in T iterations.[14] Running the algorithm a large number of times thus provides an increasingly accurate enumeration of partitions in the core. Strict preferences are assumed.

The algorithm is based on myopic deviations, which leads to the possibility of cycles.[15] The approach used is that when the same partition has occurred a multiple of k times, for some constant k, the partition will be modified by substituting in a randomly selected coalition.

Recursion is used to rearrange subsets of players, as this results in substantial speed improvements. It is possible that the core might be empty on a subset of players, even though it is not empty on the full set of players: this case is handled by terminating after T iterations.

Index iterations of the algorithm with t. Let M be the set of players in the coalition formation game. Let $\mathscr{S}$ be the set of coalitions formed by players M that are not blocked by a subset of the coalition itself. Considering the coalitions in $\mathscr{S}$ is sufficient, because any partition that is blocked by a coalition outside of $\mathscr{S}$ is also blocked by a coalition in $\mathscr{S}$ (any coalition not in $\mathscr{S}$ has a subset that is a blocking coalition, and that subset will block anything that the larger coalition would have blocked).

0. Begin at $t = 0$, with the singleton partition $\pi_0 = \{S | S = i, i \in M\}$.[16]

1. Calculate the set of deviations $\mathscr{D} = \{S' | S' \in \mathscr{S} \text{ s.t. } \forall i \in S', u_{iS'} > u_{i\pi_t}\}$

2. If $\mathscr{D} = \emptyset$ then stop: π_t is in the core for the set of players M

3. If π_t has appeared nk times in $\{\pi_0, ..., \pi_{t-1}\}$, for a positive integer n, then randomly choose a deviation S' from $\mathscr{S}$. Otherwise, randomly choose a deviation $S' \in \mathscr{D}$.

3. Identify the coalitions affected by this deviation: $\mathscr{Q} = \{S | S \in \pi, S \cap S' \neq \emptyset\}$

4. Identify the "residual" players: $R = \{i | i \in S \in \mathscr{Q}, i \notin S'\}$

5. If $R \neq \emptyset$, then calculate a rearrangement $\mathscr{R}$ of the residual via recursion: run the algorithm from the beginning with only players R.

6. If $t < T$ then repeat from 1, with $\pi_{t+1} = (\pi_t \setminus \mathscr{Q}) \cup \{S'\} \cup \mathscr{R}$

The algorithm has found a partition in the core if it returns at an iteration $t < T$, and has failed to do so if it returns at $t = T$.

[14] $T = 10000$ is used for the analysis in this paper.

[15] Empirically, these cycles of myopic deviations are observed frequently in the data.

[16] It is computationally faster to begin the algorithm with an empty partition, and assign players extremely negative utility when they do not belong to any coalition in the partition. This approach, however, makes notation slightly more complicated.