NBER WORKING PAPER SERIES

THE CAUSES OF POLITICAL INTEGRATION: AN APPLICATION TO SCHOOL DISTRICTS

Nora Gordon

Brian Knight

Working Paper 12047

http://www.nber.org/papers/w12047

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

February 2006

We thank Julie Cullen, Andrew Foster, Vernon Henderson, Sean Campbell, Kim Sau Chung, Jose Alvaro Rodrigues-Neto, Hal White, and seminar participants at several universities for helpful comments. We are grateful to Svetla Vitanova for research assistance, Lynn Carlson for GIS coding, Xiaoping Wang for assistance with administrative data, and Guy Ghan for sharing his extensive institutional expertise. Gordon thanks the Advanced Studies Fellowship program, and Knight's research was supported by a Salomon Award. We are solely responsible for the content of this paper, which does not reflect the views of the Advanced Studies Fellowships program or its funders, the Spencer Foundation and the William and Flora Hewlett Foundation. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Causes of Political Integration: An Application to School Districts
Nora Gordon and Brian Knight
NBER Working Paper No. 12047
February 2006
JEL No. H4, H7, I2, C7

ABSTRACT

This paper examines the forces behind political integration through the lens of school district consolidations, which reduced the number of school districts in the United States from around 130,000 in 1930 to under 15,000 at present. Despite this large observed decline, many districts resisted consolidation before ultimately merging and others never merged, choosing to remain at enrollment levels that nearly any education cost function would deem inefficiently small. Why do some districts voluntarily integrate while others remain small, and how do those districts that do merge choose with which of their neighbors to do so? In addressing these questions, we empirically examine the role of potential economies and diseconomies of scale, heterogeneity between merger partners, and the role of state governments. We first develop a simulation-based estimator that is rooted in the economics of matching and thus accounts for three important features of typical merger protocol: two-sided decision making, multiple potential partners, and spatial interdependence. We then apply this methodology to a wave of school district mergers in the state of Iowa during the 1990s. Our results highlight the importance of economies of scale, diseconomies of scale, state financial incentives for consolidation, and a variety of heterogeneity measures.

Nora Gordon
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
negordon@ucsd.edu

Brian Knight
Brown University
Department of Economics, Box B
64 Waterman Street
Providence, RI 02912
and NBER
brian_knight@brown.edu

1 Introduction

Political disintegration has been the international norm over the past 60 years as the number of nations has risen from 74 in 1946 to 192 in 1995 (Alesina, Spolaore, and Wacziarg, 2000). At the same time, the recent rise of international organizations such as the European Union is often associated with increasingly coordinated economic and fiscal policies. Within the United States, a similar disparity has occurred. While the number of special districts rose dramatically over the 20th century, the number of school districts has declined at a similarly dramatic rate. To better understand the forces behind changing borders, this paper develops an econometric methodology for analyzing jurisdictional merger decisions and then applies this method to a wave of school district mergers in the state of Iowa during the 1990s.

Throughout the twentieth century, bureaucrats, professional educators, and elected officials in the United States encouraged school districts to consolidate. Proponents of consolidation argued that by consolidating, districts would gain from economies of scale: high schools could offer more subjects, elementary schools could separate classes by grade level, and the quality of education could generally be improved at lower costs in larger consolidated schools and districts than in smaller ones. But many school districts resisted: residents consistently voted in favor of retaining their small districts, revealing that they preferred local control over the types of schools their children attended, who their children's classmates would be, and the determination of local tax rates to their own estimation of the potential efficiency gains so touted by consolidation's proponents. Ultimately, many states enacted legislation mandating or providing strong financial incentives for districts to consolidate, prompting sharp drops in the number of school districts (see Hooker and Mueller (1970) for an overview of such legislation), and a vast number of these political battles were resolved in favor of consolidation. As Figure 1 shows, the number of school districts in the United States plummeted from around 120,000 in 1940 to under 15,000 today.

What explains the pattern of school district consolidations over this period? Why do some districts voluntarily integrate, while others choose to remain small? How do districts that do merge choose with which of their neighbors to do so? In attempting to answer these and related questions, a theoretical and empirical literature has investigated the role of several factors.[1] First, regarding the role of size, small districts may benefit from any economies of scale associated with consolidation due to the spreading of fixed costs over

[1] We survey the relevant empirical literature in the next section. For an overview of the theoretical literature on endogenous borders, see Alesina and Spolaore (1997 and 2003), Bolton and Roland (1997), and Persson and Tabellini (2000).

more taxpayers. On the other hand, large districts may be discouraged from consolidation due to potential diseconomies of scale. Second, if the potential merger partner has different preferences for publicly-provided goods, the median voter may fear the loss in autonomy associated with consolidation. This heterogeneity in preferences for spending levels, along with related issues, such as preferences for homogenous peer groups, may serve as a repelling force in merger decisions. Finally, higher level governments, U.S. states in particular, may either encourage or discourage consolidations through the form of annexation laws or through state aid formulas.

In empirically evaluating the impact of these factors in consolidation choices, researchers are immediately confronted with several methodological issues. In particular, standard econometric models of discrete choice fail to account for three key features of standard merger protocol. First, mergers must typically be approved by voters in both districts, and the decision-making is two-sided; standard discrete choice models, such as the logit, are designed for single agent decision making. Second, in addition to deciding *whether or not* to merge, districts typically have multiple borders and thus must decide *with whom* to merge. Third, merger decisions are spatially interdependent. That is, if two districts A and B merge, then the choice set is altered for all districts sharing a border with either A or B. While the bivariate Probit model of Poirier (1980) accounts for the first feature and the multinomial logit model accounts for the second feature, we know of no estimators that simultaneously account for all three of these features of merger protocol.

To overcome these limitations of existing estimators, we first develop an econometric model of discrete choice that accounts for these three key features of the merger protocol. This approach is rooted in the economics of one-sided matching and thus allows for two-sided decision making, multiple potential partners, and spatial interdependence. In the context of this model, we show that, under a seemingly reasonable restriction on preferences, which we refer to as *symmetry in match quality*, a unique stable matching exists. Moreover, this stable matching can be calculated via a simple iterative algorithm. Finally, we develop a simulation-based estimator, which uses this iterative algorithm in order to calculate the probability of a merger between any two adjacent districts in stable matchings.

To illustrate its value, we then apply this methodology through an analysis of school district mergers in the state of Iowa, which offered significant financial incentives for mergers during the early 1990s. Due in part to these incentives, over 50 mergers involving more than 100 districts occurred during this period, and, due to these mergers, the number of districts fell from 430 in 1991, the first year included in the analysis, to 371 in 2002, the final year in the analysis. In order to identify all *potential* mergers, which can occur only between adjacent districts, we have obtained a school district map from 1989, just before the start of the sample period. To examine the role of district characteristics in these mergers, we have also collected data on pre-merger district characteristics, such as population, demographics, and property values. Finally, in order to examine the role of the state of Iowa, we have calculated the state-level financial incentives specific to each potential merger. Our results demonstrate the importance of economies of scale as well as diseconomies of scale in explaining the patterns of mergers in Iowa during this time period. We also find an important role for both state financial incentives in encouraging these mergers and various measures of heterogeneity, which serve as a repelling force in merger decisions.

The paper proceeds as follows. In section 2, we describe the methodology and findings of the existing literature. Section 3 and 4 develop the theoretical and econometric framework, which is then applied to school districts mergers in Iowa in Section 5. Finally, section 6 concludes.

2 Existing Literature

Several existing empirical studies shed light on the role of factors underlying political integration. Alesina, Baqir, and Hoxby (2004) examine the number of jurisdictions, including school districts, within U.S. counties over the period 1960-1990 and find evidence for a trade-off between economies of scale and heterogeneity in both race and income. That is, counties with high levels of heterogeneity in these dimensions tend to have more school districts, all else equal. On the other hand, they find little effect of heterogeneity in religion or ethnicity. Regarding the role of state governments, the authors find that the strength of annexation laws matter in determining the number of school districts within a state. In a study analyzing the role of state characteristics in determining the number of school districts within a state, Kenny and Schmidt (1994) find that the decline in the number of school districts between 1950 and 1980 can be explained by the decline in farming and corresponding increase in population density, the increased importance of state aid, and the increased prominence of teacher unions.

Relative to this literature, which examines the number of school districts within larger geographic units, such as states and counties, we are focused on specific individual merger decisions involving adjacent school districts. Our approach thus arguably better accounts for constraints on the availability of potential partners that are imposed by existing boundaries;

variation in these constraints could lead two otherwise identical districts to have different merger patterns. While our approach is most appropriate within a single state, the papers by Kenny and Schmidt (1994) and Alesina, Baqir, and Hoxby (2004) are more naturally suited to an examination of multiple states. Thus, we view our analysis as complementary to this existing line of research.

The only studies of which we are aware that examine the decisions of adjacent school districts to consolidate are a series of papers by Brasington. Brasington (1999) identifies 298 pairings of Ohio communities that either do or potentially could jointly provide education services through a single school district. He then estimates a bivariate Probit model developed by Poirier (1980); this model allows for both communities to have veto power over the merger decision and thus a merger is observed only if it is supported by both districts. Using this econometric methodology, he finds that small and large districts tend to jointly provide education services, while medium-sized communities do not enter such arrangements. Neither racial heterogeneity nor income levels explain these patterns. In two follow-up papers, Brasington uses the same dataset from Ohio but allows for the coefficients to vary between the larger and smaller potential merger partner (Brasington, 2003b), between the richer and poorer community (Brasington, 2003a), and between the more and less white community (Brasington, 2003a).

Relative to these papers by Brasington, our paper provides several contributions. First, while all of Brasington's papers account for the two-sided nature of mergers, they do not account for the two other key features described above: districts must choose from one of several potential partners and merger decisions are spatially interdependent.[2] A failure to account for these features of merger decisions may lead to specification errors. For example, the bivariate probit model does not restrict the merger probabilities for a given district with multiple potential partners to be less than one. Similarly, this failure to account for multiple potential partners and spatial interdependence may lead to incorrect inference due to the statistical dependence across borders. Our approach, by contrast, restricts merger probabilities for a given district to be less than one and accounts for statistical dependence. Second, while Brasington uses school district characteristics, such as enrollments, test scores, and property values, from the early 1990s to explain consolidation decisions in Ohio, many of which occurred during the 1930s and 1960s, we better model the timing of the merger

[2] To be clear, Brasington's analysis does account for each district having multiple borders, and he does correctly account for all of the possible pairwise combinations. The difference is that our approach restricts districts to merge with at most one other district, whereas Brasington's bivariate probit model assumes independence in merger decisions across potential merger partners for a given district.

decisions. The failure to account for these timing considerations could lead to problems in interpretation. For example, if district characteristics tend to converge post-merger, then Brasington's analysis may incorrectly interpret similarities in district characteristics between merger partners to a preference for homogeneity, rather than the true source of these similarities: the ex-post convergence in district characteristics. In our empirical application, by contrast, we measure school district characteristics during the year in which the merger decisions were made, allowing us to separately identify the causes of mergers from their subsequent effects.[3] While we have provided several methodological contributions to this literature, Brasington's specification is somewhat more general in other dimensions. In particular, it allows for an imperfect correlation between the unobserved preferences for consolidation between the two merger partners and, in the two follow-up papers, allows the coefficients to vary across the two potential merger partners. Thus, we again view our approach as complementary to this existing line of research.

3 Theoretical framework

In empirically analyzing the determinants of mergers between jurisdictions, the analyst is immediately confronted by three methodological challenges. First, in order to take place, mergers must be approved by both districts, and the problem is thus two-sided. Second, in addition to deciding whether or not to merge, districts typically have multiple borders and thus must decide with whom to merge out of this set of potential partners. Finally, merger decisions are spatially interdependent across districts. In order to overcome these methodological challenges, we develop a simulation-based estimator that is rooted in the economics of matching and thus accounts for two-sided decision making, multiple potential partners, and spatial interdependence. We first describe the matching environment and the associated equilibrium concept of stability before deriving the econometric estimator in the next section.

3.1 Matching model

Consider a set of school districts and the following merger protocol. First, mergers can occur only between two adjacent districts, and, for reasons of tractability, we rule out mergers

[3] In separate work (Gordon and Knight, 2005), we are examining the effects of these mergers on subsequent school district fiscal outcomes.

involving three or more districts. Second, we do not allow for one district to dissolve into multiple districts, again for reasons of tractability. Third, mergers must be approved by voters in both districts, and the decision-making is thus two-sided. Fourth, districts may choose not to merge with any adjacent districts; that is, districts may remain unmerged. Finally, given our empirical motivation, we assume that districts have strict preferences and are thus not indifferent between their potential merger partners. The role of these and other key assumptions will be described more completely at the end of this section.

This merger environment can be modeled as a one-sided matching game. In particular, a matching is defined as a set of merger assignments, in which each district is assigned either a single merger partner or is assigned to remain alone.[4] Following the literature on matching, we use stability as the equilibrium concept. A *stable matching* is a set of merger assignments in which 1) no district prefers to remain unmerged over merging with their assigned partner, and 2) no two districts prefer to merge with each other over their respective merger assignments, which may include remaining unmerged.

Unfortunately, in one-sided matching situations such as this one, stable matchings do not generally exist, and when they do exist, are not necessarily unique. Consider, for example, three districts $1, 2,$ and 3 all of which border each other. Suppose that all three districts prefer any merger over remaining unmerged and further that district 1 prefers 2 over 3 ($2 \succ_1 3$), 2 prefers 3 over 1 ($3 \succ_2 1$), and 3 prefers 1 over 2 ($1 \succ_3 2$), and denote this odd cycle as 123. In this case, no stable matching exists since any merger between two districts can be broken by the unmerged district. On the other hand, with a four-district case and an even cycle such as 1234, multiple stable matchings may exist.[5]

Given our objective of developing an empirical methodology, which requires a comparison of mergers observed in the data to those predicted by the econometric model, the problems of non-existence and multiplicity clearly create significant hurdles to be overcome. Fortunately, a simple restriction on preferences guarantees both existence and uniqueness. Before introducing such a restriction, we define the utility, or gains, to district i from a merger with

[4] In one-sided matching games, any agent can match with any other agent. In two-sided matching games, such as the marriage model of Becker (1973), agents can be sorted into two groups, say men and women, and matches can occur across, but not within, groups. Roth (1990) provides additional treatment of the theory of two-sided matching games.

[5] That is, if all districts prefer any merger over remaining alone, 1 merging with 2 and 3 with 4 is a stable matching so long as there is no profitable deviation involving a merger between districts 2 and 4 (that is, $1 \succ_2 4$ or $3 \succ_4 2$). However, 1 merging with 4 and 2 with 3 is also a stable matching so long as there is no profitable deviation involving a merger between districts 1 and 3 (that is, $4 \succ_1 3$ or $2 \succ_3 1$).

district j as follows:

$$U_{ji} = A_j + I_i + Q_{ji}$$

where A_j represents the attractiveness of district j as a partner and is valued equally by all of j's potential partners, I_i represents district i's inclination to merge with any of its potential partners, relative to remaining unmerged, and Q_{ji} represents the quality of the match between districts i and j, as valued by district i.[6] Utility from remaining unmerged is normalized to zero ($U_{ii} = U_{jj} = 0$). As noted above, we assume throughout that districts have strict preferences over their potential merger partners.

It should be clear that this specific formulation of utility places no restrictions on preferences, as we can always manipulate the qualities of the match to generate the cycles described above. To eliminate this cycling problem, we next introduce the restriction of *symmetry in match quality*:

$$Q_{ji} = Q_{ij}$$

That is, conditional on the attractiveness of a district, which is equally valued by each potential partner, and the inclination of a district to merge with any of its partners, the quality of the match is equally valued by the two districts i and j; we again defer a discussion of the role of this assumption to the end of this section.[7] Using this restriction, we have established the following result:

Proposition: *Under the assumptions of symmetry in match quality and strict preferences, there exists a unique stable matching.*

Proof: See Appendix.

Intuitively, the restriction of symmetry in match quality places enough regularity on preferences over merger partners in order to rule out cycling, which is the underlying source of the problems of non-existence and multiplicity.[8] To see this, consider again the three district cycle 123 described above; this cycle can be equivalently represented by the following

[6] We do not explicitly model geographic constraints here. However, these constraints can be easily incorporated into preferences by setting $Q_{ji} = -\infty$ for non-adjacent districts.

[7] This result is similar in flavor to that in Sorensen (2005), who shows that the condition of aligned preferences, which is equivalent to symmetry in utility, leads to a unique stable matching in two-sided matching games. He then develops a Bayesian estimator and applies this to the market for venture capital.

[8] In independent work, of which we became aware after developing our theoretical results, Rodrigues-Neto (2005) showed that, under symmetric utilities ($U_{ij} = U_{ji}$) and strict preferences, there is always a unique stable matching. While our restriction of symmetric match quality appears to be more general at first glance, these two restrictions turn out to be theoretically equivalent. In particular, if $U_{ji} = A_j + I_i + Q_{ji}$, where $Q_{ji} = Q_{ij}$, then preferences can be represented equivalently by $V_{ji} = U_{ji} - I_i + A_i$ and thus $V_{ji} = V_{ij}$.

inequalities:

$$\begin{aligned} U_{21} &> U_{31} \\ U_{32} &> U_{12} \\ U_{13} &> U_{23} \end{aligned}$$

Inserting our utility specification and imposing symmetry in match quality, we have that:

$$\begin{aligned} A_2 + Q_{12} &> A_3 + Q_{13} \\ A_3 + Q_{23} &> A_1 + Q_{12} \\ A_1 + Q_{13} &> A_2 + Q_{23} \end{aligned}$$

Summing across these three inequalities, we thus have a contradiction ($A_1 + A_2 + A_3 + Q_{12} + Q_{23} + Q_{13} > A_1 + A_2 + A_3 + Q_{12} + Q_{23} + Q_{13}$), and it should be clear that cycling cannot occur under our assumption of symmetry in match quality.

While these existence and uniqueness properties are interesting from a theoretical perspective and, as noted above, are clearly important from an empirical perspective, the proposition is incomplete as we also need to characterize this unique stable matching in order to complete the development of our econometric estimator. Fortunately, under the assumptions of the proposition, symmetry in match quality and strict preferences, this unique stable matching can always be computed using the following simple algorithm:

Step A: Match mutual 1st choices (including option to remain unmerged)

Step B: Remove matched districts from map

Step C: Re-rank from remaining borders and return to Step A

That is, each district ranks its potential merger partners, and mutual first choices are matched with one another. After removing these matched districts from the map, each district re-ranks from its remaining borders, and the process continues until all districts are either matched with another district or remain unmerged. Again, the restriction of symmetry in match quality rules out cycles and thus guarantees that at least one border with two districts that are mutual first choices can be found in Step A. Our ability to calculate the stable matching via this simple iterative algorithm suggests that a simulation approach may be productive from an econometric perspective. After describing the role of the assumptions underlying these theoretical results, we next turn to the development of such an empirical approach.

3.2 Role of the assumptions

Several assumptions were required in order to generate these results of existence and uniqueness. In this section, we discuss the role of the key assumptions, both explicit, such as symmetry in match quality, and implicit, such as no side payments. Regarding the merger protocol, the key assumptions were no mergers involving more than two districts and no separations. Allowing mergers involving more than two districts would certainly complicate the problem as it would require consideration of many more combinations of merger partners. Given the difficulty of empirical implementation with only two-district mergers, we leave the issue of multiple merger partners for future work. Separations would also significantly complicate the analysis as it would require the researcher to identify borders within districts along which such separations may occur. As will be noted below, in our empirical application, we have only one case of a merger involving more than two districts and only one separation. Thus, at least in the context of our empirical application, we view these cases as exceptions to the rule.

In our model, side payments were not allowed, and thus mergers that might increase overall surplus may be blocked by one of the two potential partners. In models with side payments, by contract, the mergers that generate the largest joint surplus are the most likely to occur.[9] While introducing such side payments would clearly alter mergers that occur in a stable matching, we feel that this assumption is reasonable in our empirical application, as will be described more completely below.

Another implicit assumption is myopic decision-making: that is, we do not allow districts to consider how a merger today might alter the pool of potential merger partners in the future. In the context of U.S. states, for example, our framework would not allow Rhode Island to merge with Massachusetts in order to fulfill an ultimate objective of merging with New Hampshire in future years. While these dynamic considerations are certainly interesting, they would significantly complicate the analysis, and we thus leave them for future work. In our empirical application, however, we do update borders following mergers and allow these new districts to subsequently re- merge; as will be described below, however, these subsequent mergers were rare in practice.

Perhaps the most crucial assumption is symmetry in match quality. Given that our specification allows for attractiveness and inclination, however, it is important to note that this assumption does not require symmetry in utility. Suppose, for example, that all districts

[9] In a model with side payments, Fox (2005) develops a non-parametric estimator for two-sided matching games using these efficiency properties of stable matchings.

prefer to merge with richer districts and that this gain can be simply captured by $U_{ji} = y_j - y_i$, where y_i and y_j represent a summary measure of income, such as average or median income, in district i and j, respectively. Then, it is clear that, while utilities are not necessarily symmetric ($U_{ji} \neq U_{ij}$), income in district j can be incorporated into attractiveness ($A_j = y_j$) and income in district i can be incorporated into inclination ($A_i = y_i$), and thus the quality of the match is symmetric ($Q_{ij} = Q_{ji} = 0$), albeit uninteresting, in this case. More generally, as will be shown in the next section, even under the assumption of symmetry in match quality, the researcher can estimate a variety of econometric specifications as it allows us to control for own-district characteristics, characteristics of the other district, interactions of these characteristics across the two districts, and symmetric differences, such as squared or absolute differences, between the characteristics of the two districts.

4 Econometric implementation

Consider an empirical version of the above utility function defined over merger partners:

$$U_{ji} = X_j \theta_x + Z_i \theta_z + f(W_i, W_j)\theta_w + \varepsilon_{ji}$$

where X_j represents observed measures of the attractiveness of district j as a partner and Z_i represents observed measures of district i's inclination to merge with any of its potential partners, relative to remaining unmerged. The observed quality of the match is given by $f(W_i, W_j)$, while the unobserved quality is given by ε_{ji}; this unobserved match quality is assumed to be distributed type I extreme value and independently across borders. The vector $\theta = (\theta_x, \theta_z, \theta_w)$ represents parameters to be estimated. It is clear that symmetry in match quality is satisfied whenever $f(W_i, W_j) = f(W_j, W_i)$ and $\varepsilon_{ji} = \varepsilon_{ij}$, and we impose these conditions throughout the remainder of the paper. As discussed above, this former restriction $[f(W_i, W_j) = f(W_j, W_i)]$ is fairly general as it allows for a variety of econometric specifications, such as interactions between the two districts $[f(W_i, W_j) = W_i \times W_j]$ and measures of the differences between the two districts, including squared differences $[f(W_i, W_j) = (W_i - W_j)^2]$ and absolute differences $[f(W_i, W_j) = |W_i - W_j|]$.

Given the two-sided nature of the problem, multiple potential partners for each district, and the interdependence of merger decisions, it is clear that no closed form solution exists for the probability of a merger between any two districts. Said differently, the probability of a merger between any two districts depends upon the characteristics of all districts, even non-adjacent ones. As an alternative to analytically expressing the probability of a merger between any two adjacent districts, one can use the simulation methods for discrete choice models first developed by Lerman and Manski (1981). In particular, for replication $r = 1, 2, .., R$, a symmetric unobserved match quality ($\varepsilon_{ji}^r = \varepsilon_{ij}^r$) can be drawn randomly from the type-I extreme value distribution for each border, and, given a set of parameters (θ), the iterative algorithm described above can be applied in order to calculate the unique stable matching assignments. Unobserved match qualities can then be re-drawn R times, and the proportion of replications in which i and j merge in a stable matching serves as an estimate of the probability of a merger between i and j. This leads to the well-known frequency simulator:

$$\widehat{\Pr(i,j)} = \frac{1}{R}\sum_{r=1}^{R} y_{ij}^r$$

where $y_{ij}^r \in \{0, 1\}$ is a dummy variable indicating a merger between districts i and j in the stable matching associated with simulation r. In practice, however, a smoothed simulator, which calculates the probability of a merger in each replication $\Pr(y_{ij}^r = 1|\tau)$, where τ is the smoothing parameter, is preferred.[10] Importantly, as the smoothing parameter goes to zero, the smooth simulator approaches the frequency simulator [$\lim_{\tau\to 0}\Pr(y_{ij}^r = 1|\tau) = y_{ij}^r$]. But, for any positive value of the smoothing parameter, the smooth simulator is bounded between zero and one $\Pr(y_{ij}^r = 1|\tau) \in (0, 1)$, and the average probability across all replications then serves as the estimate of the probability of a merger between i and j :

$$\widehat{\Pr(i,j)} = \frac{1}{R}\sum_{r=1}^{R} \Pr(y_{ij}^r = 1|\tau)$$

We describe one possible smoothed simulator, which we use in the empirical application to follow, in Appendix 2.

For estimation purposes, we use the method of simulated moments due to McFadden (1989).[11] Under this method, parameters are chosen in order to minimize a measure of the

[10] As is known in the literature on simulation, a smoothed simulator is preferred to the frequency simulator for at least three reasons. First, mergers are relatively rare in practice, and thus, even with a large number of replications, borders may experience zero mergers across all replications and the frequency simulator will thus be zero. This creates problems for the GMM estimator, which places infinite weight on this hypothetical border observation given that the simulated variance is zero for this observation. Second, the theory underlying statistical inference for the GMM estimator assumes that the objective function is differentiable. Finally, a smooth objective function permits the use of computationally faster derivative-based optimization methods in choosing the parameters. See Stern (1997) for additional information on smooth simulators.

[11] Given the interdependence in merger decisions (if A merges with B, then C cannot merge with A or

distance between the simulated probabilities of merger and the observed merger decisions. Additional details, including the GMM objective function, the optimal weighting matrix for the moment conditions, and expressions for the variance-covariance matrix, are provided in Appendix 3.

To summarize, estimation via simulation would proceed as follows:

> *Step 0: For each border, independently draw an unobserved match quality (ε_{ji}) from the type-I extreme value distribution. Do this R times and index the replications $r = 1, 2, ..., R$.*
>
> *Step 1: For each of the R replications, and given a set of initial parameter values, run the iterative algorithm described above in order to find a stable matching and the associated merger probabilities. The average of this probability across all simulations is the simulated merger probability.*
>
> *Step 2: Choose a new set of parameter values and return to step 1. Repeat until the GMM objective function is minimized.*

The estimation approach is also summarized graphically via a flow chart in Figure 3. As shown, the estimation involves both an inner loop, in which the simulated probabilities of merger are calculated given a set of parameters, and an outer loop, in which the set of parameters is chosen in order to minimize the GMM objective function.

Thus, we have developed an econometric model of discrete choice that overcomes the three key limitations of existing econometric models. In particular, by appealing to the economics of matching and the associated stability concept, this approach accounts for the two-sided nature of the merger protocol, allows each district to have an arbitrary number of potential merger partners, and accounts for the spatial interdependence of merger decisions. To illustrate the practical value of this approach, we next turn to an empirical application of school district mergers in the state of Iowa during the 1990s.

B) and our reliance on simulation in calculating the probability of mergers, maximum likelihood estimation is problematic. In particular, the likelihood function is defined over all potential *combinations* of merger decisions. Given that, in the empirical application, we have over 1,000 borders, the number of combinations in quite large, and even with a large number of simulation runs, we may not observe every combination of merger decisions. Thus, our simulation procedure would assign probability zero to combinations of mergers not observed in our simulation runs even though every combination of mergers occurs with positive probability in our empirical model (due to the fact that ε_{ji} is unbounded).

5 Empirical Application

We choose to look at the experience of school districts in the state of Iowa during the 1990s for several reasons. Most importantly, while the state did provide financial incentives and technical assistance for consolidation, the decision to integrate ultimately rested with the school districts themselves. That is, school districts decided both whether or not to consolidate (in contrast to some other state mandates for minimum district size), and if they did choose to consolidate, they chose with which of their neighbors to do so (conditional on the neighbor's agreement). Second, concentrating on more recent consolidation activity gives us access to better data on school district finances and the demographics of students and voters. Third, by looking at a period of consolidation beginning just after the 1989 Census was administered, we have access to the initial school district boundaries as geo-coded in the Census TIGER files.

Moreover, the consolidation environment in Iowa arguably satisfies two key assumptions underlying our methodology: symmetry in match quality and no side payments. Regarding the symmetry assumption, our conversations with Guy Ghan, who oversaw the consolidation process for many of the districts in our sample, suggest that this assumption is consistent with our particular historical setting. While the petition for reorganization was flexible enough to allow the two merging districts to write terms specific to their match, it appears that in practice, larger and more property-wealthy districts were universally attractive, rather than attractive in a match-specific way. Regarding side payments, we know of no explicit monetary transfers between merging districts. Two possible alternative forms of side payments regard existing debt and the closing of schools. Regarding debt, two initial districts could continue to maintain separate tax rates in order to pay off the debt they had before merging. Another potential side payment involved discretion over which schools would be closed (particularly politically important for high schools) following a merger. As pointed out in the theoretical literature, however, the promise of such side payments may suffer from credibility and enforcement problems (Alesina and Spolaore, 2003) in practice. We next describe the data before turning to exact measures and the empirical results.

5.1 Data Sources and Variable Definitions

We draw on a number of data sources to compile our district-year level data on Iowa school districts from 1989 to 2001. Our analysis requires data on the timing and composition of school district consolidations, a listing of potential merger partners, and pre-merger characteristics, including demographics, property values, revenues, and expenditures.[12] Demographic data on school districts come from the Census of Population and Housing for 1990 and 2000, and the Common Core of Data. The Census data from 1989 are tabulated at the school district level in the School District Data Book (SDDB), and we use the "Top 100" dataset from the SDDB. In order to use the 1989 Census data in analyzing mergers in each year of our panel, we created enrollment-weighted averages of those 1989 values for district-year observations that have experienced a merger between 1989 and the current year. These Census data include richer variables than found in the Common Core, including the distribution of adult educational attainment, self-reported home values, and area of the school district in school miles. Because the Census data are available only decennially, we use the Common Core of Data for less refined demographic variables on an annual basis. We use the number of total students enrolled in public school and enrollment by grade.[13] Data on school district finances are taken from the School District Finance Data (F33) file, available annually in our time period from the fall of 1989 to the fall of 2001. We use current instructional spending, converted into per-pupil measures using the corresponding enrollment variable. Finally, we have obtained administrative data from Iowa on property value assessments by year and school district; these data are available beginning in 1991.

In order to identify mergers, we have obtained administrative data on school district consolidations from the Iowa Department of Education dating to 1965. These data list the date on which each consolidation goes into effect, the names and Iowa state identification numbers of the districts merging, and the name and Iowa state identification number of the new district formed. In all cases except one, consolidations involved only two districts. One case did involve three districts; given the econometric complications involved with allowing for three-way mergers, we ignore the role of this single three-way merger in the empirical analysis to follow. Relatedly, in two cases, individual districts were involuntarily dissolved into surrounding districts. Because these cases were both infrequent and involuntarily, we disregard them in our estimation, which allows a given district to remain unmerged or to merge with any one of its neighbors.

In order to identify potential merger partners, we have obtained a map of school districts from 1989 as geo-coded in the Census TIGER files.[14] According to this map, there were 431

[12] School districts in Iowa are independent jurisdictions, meaning that they collect their own tax revenue rather than receiving revenue allocated to them by a parent government such as a town, city, or county.

[13] While race and ethnicity variables are available in the Census and the Common Core of Data, they provide limited variation in this instance and we exclude them from our analysis due to computational constraints.

[14] School district boundaries in Iowa bear little relationship to county boundaries; that is, existing district boundaries often cross county lines, and mergers also occur across county lines.

districts and 1,211 borders in 1989. Thus, districts had roughly 5 potential merger partners on average. Given the date of the map, our sample is defined over the period 1991 through 2001, the first and last years, respectively, for which we have complete data.

As mentioned above, our theoretical and econometric framework is purely static in nature. That is, we do not allow districts to consider how a merger today might alter the pool of potential merger partners in the future. Given our use of panel data, however, we must incorporate such changes in potential merger partners in the construction of our dataset. In particular, if two districts A and B merge in year t to form a new district AB, this new district AB now shares borders with all of A's original borders and B's original borders, and we allow for such subsequent mergers between AB and any of these potential merger partners. Empirically, subsequent mergers were rare; there were only two cases in which a school district, as it existed in 1989, went through two consolidations between 1989 and 2001.

5.2 Financial Incentives

Financial incentives applied to school districts voting by November 30, 1990 to make their consolidations effective between July 1, 1991 and July 1, 1993. As Figure 2 shows, districts appear to have responded strongly to these time-specific incentives. Beginning in 1966, the start of our administrative data on consolidations, through 1990, there were zero to three consolidations per year (with 1966 the only year with more than two). In 1991, the first year for which districts received financial bonuses for consolidating, there were four consolidations. This rose to seven consolidations effective in 1992 and twenty in 1993. This was followed by three additional years of higher than average merger activity, even though districts whose consolidations first took effect in these years were not eligible for the financial incentives. We discuss two possible explanations for these post-1993 mergers below.

The financial incentives had two key components, which are summarized in Table 1. The largest incentive for districts to consolidate between 1991 and 1993 was a five-year reduction in their foundation tax rate. During our sample period, the foundation tax rate in Iowa was \$5.40 per \$1000 of assessed valuation (5.40 mills). By consolidating, taxpayers in districts with pre-merger enrollments of fewer than 600 students experienced a foundation tax rate of 4.40 mills in the first year post-consolidation, increasing by 0.20 mills per year until reaching 5.40 mills again in the sixth year after consolidation, where it would remain. Throughout this time, the district would receive supplemental state revenue equal to the decrease in local collections, so that the foundation tax reduction essentially transferred funds from state to local taxpayers with no reduction in total revenue available for local education expenditures.

To be clear, the enrollment limit is defined separately for each of the two potential merger partners; all property in the post-merger district will be eligible for the lower foundation rate if both partners had enrollment below 600 students. For mergers involving one district below 600 students and one district above 600 students, only the property in the district of the smaller partner is eligible for the lower foundation rate and thus property owners in the two districts effectively paid different tax rates for the five years following a merger. Thus, this merger incentive is measured separately for the two potential merger partners.

To compute the reduction in the foundation tax rate, we use enrollment figures in order to determine whether the district was above or below 600 students as well as annual administrative data on assessed property values. We then compute the present discounted value of the five-year stream of payments using an assumed discount rate of 3 percent, which is roughly the inflation rate during 1991, and, given the stagnant population in Iowa, an assumed nominal growth rate in housing values of zero.

As shown in Table 2a, mergers only occurred during this subsidy period 1991-1993 along borders in which at least one district had enrollments below 600 students, and the vast majority occurred along borders in which both districts had enrollments below 600 students. While these average merger rates of 2.4 percent appear relatively low at first glance, it is important to note that these merger rates are both along a given border and within a given year. Districts with enrollments below 600 merged with one of their multiple neighbors at an 8 percent rate in a given year during the 1991-1993 period. Even this district-level rate understates merger activity given that no subsequent mergers occurred during the 1991-1993 period. That is, almost one-third (31 percent) of districts with enrollments below 600 in 1991 merged with one of their neighbors at some point during the 1991-1993 period. Taken together with the spike in mergers during this incentive period, as demonstrated in Figure 2, this evidence suggests that districts strongly responded to the financial incentives in place during this period.

The second major incentive is related to the practice of whole grade sharing (WGS). Under WGS, two distinct districts do not merge their finances and thus maintain independent tax bases; instead, two districts divide responsibility over providing education services for particular grades. A common version of WGS involves both districts maintaining their own elementary schools, one district having a middle school serving students from both districts, and the other district having a high school serving students from both districts. Iowa had encouraged whole grade sharing by assigning an additional weight to students in whole grade sharing arrangements when making foundation payments to districts. Specifically, students in WGS arrangements counted as 1.1 "regular" students. The Iowa state legislature changed the school finance law to eliminate additional weights for students in WGS arrangements, but allowed school districts consolidating effective 1991-1993 to continue to weight their enrollments according to the proportion of students previously in WGS for five years after merging. This allowed consolidating districts to retain about $200 per pupil per year over a five-year period that they would have lost had they not merged.[15] Unlike the first incentive, which varied between the two potential partners, this second incentive provides extra funds directly to the new district, and both districts will thus share in the incentive post-merger. Many of the districts consolidating, both during the 1991-1993 eligibility window and afterwards, had been involved in WGS agreements.[16]

5.3 Heterogeneity Factors

We focus on three measures of heterogeneity: fiscal, demographic, and spatial. These latter two measures are emphasized in the work by Alesina and Spolaore (1997 and 2003) As a baseline measure of fiscal heterogeneity, we use the squared difference in per-pupil spending on education, adjusted for tax bases, between the two districts. That is, we estimate preferences for education by dividing per-pupil expenditures, using instructional spending and

[15] In order to estimate the monetary value of these whole grade sharing incentives, we first estimate the number of students involved in whole-grade sharing by school district. To generate this estimate, we make the simplifying assumption that a district's enrollment, as reported in the district-level files, is equally distributed across all thirteen (including kindergarten) grades. We then multiply this estimated grade-level enrollment by the number of grades in which there is no reported enrollment across all school-level files for the district. This whole-grade sharing enrollment estimate is thus an estimate of the district's gross exported students. We then multiply the number of students involved in whole-grade sharing by $247, which is 10 percent of the foundation payment in 1991, the first year in which the incentives were in place. Finally, we take the present discounted value of the 5-year stream of payments assuming a discount rate of 3 percent and a nominal growth rate in the foundation payment of 4.5 percent, which is roughly the growth rate realized during this period.

[16] Both the foundation tax rate reduction and continued use of supplemental WGS weights gave districts an incentive to consolidate effective 1991-1993. If we view the decision to consolidate as a choice between WGS and consolidation, districts may have chosen WGS over consolidation prior to 1991 because of the supplemental weights. This reason not to consolidate is not valid for mergers effective after 1993 (although they would still receive greater benefits from merging between 1991 and 1993), so may explain why more districts than average consolidated even after the greatest financial incentives were no longer applicable. Another possibility is that the school board had referred the merger to voters by November 30, 1990 but needed more time to build political consensus before voters ultimately approved the merger, albeit without the financial incentives, in subsequent years.

enrollments in the Census data, by housing values in the district, as self-reported by residents in Census data. To create a measure of heterogeneity, we then take the squared difference in the measures between the two adjacent districts.[17] For our demographic heterogeneity measures, we examine the squared difference across the two districts in two measures of adult educational attainment: the percent of adults with less than a high school degree, and the percent of adults with at least a four-year college degree. Finally, regarding spatial heterogeneity, we control for the population density of the district, as measured in total population per square mile, as well as the estimated distance between the two districts.[18]

5.4 Scale Factors

We are also interested in examining the role of economies and diseconomies of scale in these merger decisions. Let $c(N)$ denote the average cost of providing education services to N students. From the perspective of district i, the efficiency gains, or potentially losses, from a merger with district j can be expressed as:

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

For efficiency enhancing mergers $[c(N_i+N_j) < c(N_i)]$, our measure of efficiency gains will be positive. On the other hand, if $c(N_i+N_j) > c(N_i)$, our measure will be negative, suggesting efficiency losses. In terms of an empirical specification, we use the following average cost specification:

$$c(N) = N^{\beta+\gamma N}$$

[17] A related measure that would be interesting to examine is the fiscal transfer induced by mergers. That is, the relatively property-poor district may be subsidized by the property-rich district post-merger due to the differences in tax bases. Unfortunately, this fiscal transfer effect cannot be identified without district-specific voting data, which we have been unable to obtain for these mergers in Iowa. Given that an increase in the disparity between housing prices will increase the willingness to merge by the property-poor district but will decrease the willingness to merge by the property-rich district, the net effect on merger probabilities is thus ambiguous and cannot be identified without district-specific voting data. See Filer and Kenny (1980) for an analysis of the role of such fiscal transfers in city-county consoliation referenda during the 1949-1976 period.

[18] We estimate the distance between the two districts using data on each district's area in square miles, with the simplifying assumption that each district is square in shape. This allows us to calculate the distance between the two districts (from center to center) as the sum of one-half of the square root of each district's area.

As shown in figure 4, this specification allows for a wide range of shapes for the cost curve. That is, the parameter β describes the shape of the cost curve at low enrollment levels, while the parameter γ captures the shape of the cost curve at high enrollment levels. Thus, the former parameter can be interpreted as a measure of economies of scale and the latter can be interpreted as a measure of diseconomies of scale. As shown, if $\beta < 0$ but $\gamma > 0$, the cost-curve will be U-shaped, suggesting that mergers will be efficiency enhancing for smaller districts but potentially efficiency detracting for larger districts. Inserting this cost curve specification into our measure of efficiency gains, we have that:

$$\begin{aligned} \ln\left[\frac{c(N_i)}{c(N_i+N_j)}\right] &= \beta \underbrace{\left[\ln(N_i) - \ln(N_i+N_j)\right]}_{\text{economies of scale}} \\ &\quad +\gamma \underbrace{\left[N_i \ln(N_i) - (N_i+N_j)\ln(N_i+N_j)\right]}_{\text{diseconomies of scale}} \end{aligned}$$

Thus, as described above, our estimate of β can be considered an estimate of the role of economies of scale in merger decisions, while our estimate of γ can be considered a corresponding estimate of the role of diseconomies of scale.

Given that the merger incentives are targeted at small districts, it is important to be clear in describing how we distinguish between responsiveness to the incentives and the role of economies of scale. First, we have time-series variation in the merger incentives, which were available only during the 1991-1993 period, and, as shown above, merger rates were much higher during this period. A second source of identification is the 600-cutoff level for the merger incentives. As shown in Table 2b, merger rates were much higher along borders with two very small districts, defined as those with enrollments below 300, suggesting a role for economies of scale. Roughly speaking, we can use the variation in merger rates within the below-600 category to identify the role of economies of scale, while variation in merger rates between the below-600 and above-600 categories identifies the responsiveness of districts to the merger incentives.

5.5 Results

Table 3 provides summary statistics for our key variables at the level of an individual school district conditional on whether or not a merger occurred. As shown, those districts choosing to merge with one of their neighbors were significantly smaller than those that choose to remain unmerged. Moreover, these districts tended to be significantly smaller in a geographic sense, as captured by square miles, suggesting an important role for spatial heterogeneity and transportation costs. As shown in the final row, however, merging districts tend be less dense as the effect of low enrollment dominates the effect of small size in the geographic sense.

Similarly, Table 4 provides summary statistics at the level of a school district border. As shown, mergers were more likely to occur along borders that were eligible for the merger incentives, and this difference is statistically significant. Given the complexity of interpreting the economies of scale measures, we defer their discussion until the econometric analysis. Regarding heterogeneity measures, mergers were more likely to occur along borders with less heterogeneity in adult educational outcomes although this difference is only statistically significant for the college degree heterogeneity variable. Regarding the second measure of heterogeneity, differences in spending levels, no effect is detected in these summary statistics. Finally, mergers were more likely to occur among geographically close districts, and this difference is statistically significant. We next turn to a more formal econometric test of our hypotheses.

Table 5 provides the results from our simulated method of moments estimator. Column 1 presents our baseline results, which do not include any heterogeneity measures, while columns 2-4 introduce our three different heterogeneity measures. As shown in the baseline results, merger incentives have a positive effect on the decision to merge, providing evidence that is consistent with the suggestive evidence provided in figure 2 and table 2. While these results demonstrate a positive and statistically significant effect of financial incentives on merger decisions, Table 6 provides some evidence on the magnitude of the effect. In particular, we conduct a counterfactual experiment in which the financial incentives in place during 1991-1993 are eliminated. As shown, merger rates along borders in which both districts have enrollments below 600 and were thus eligible for the foundation rate incentives fall significantly from 2.89 percent to 0.46 percent, representing an 84 percent reduction in these merger rates.

Returning to the coefficients in column 1 of Table 5, the economies of scale variable has an expected negative coefficient, while the diseconomies of scale measure has an expected positive coefficient. In order to aid in the interpretation of these results, Figure 5 plots the log cost curve implied by the coefficients in column 1 against district enrollments. Recall that our assumed cost curve is given by $c(N) = N^{\beta+\gamma N}$, and thus we can write the log cost curve as follows:

$$\ln c(N) = (\beta + \gamma N)\ln(N) \tag{1}$$

As shown in Figure 5, these coefficients imply that average costs are minimized at enrollments of about 250 students. Thus, among equally sized districts, the most efficient mergers involve those with enrollments of about 125 each, and mergers involving larger districts may entail diseconomies of scale. It is important to note that these estimates of economies of scale and diseconomies of scale should be interpreted as those perceived by the voters when deciding whether or not to integrate. These revealed preference estimates may differ substantially from the economies of scale actually realized by districts through consolidation. Indeed, estimates of education cost functions, as summarized by Andrews, Duncombe, and Yinger (2002), imply that diseconomies of scale may not set in until enrollments reach 6,000 students, although, as the authors point out, this optimal size may be significantly lower in sparsely populated states, such as Iowa, due to transportation costs.

As shown in columns 2-4, the coefficients on the merger incentive measure and the size variables are similar after controlling for our various measures of heterogeneity. Turning to these measures, we find that all three sources of heterogeneity served as a repelling force in merger decisions, and all of these coefficients are statistically significant except for the adult high school degree heterogeneity measure. In particular, mergers are less likely to occur as the difference in the fraction of adults with a college degree increases, as shown in column 2. Similarly, as the difference in property-value adjusted spending levels increases, the propensity to merge falls, as shown in column 3. Finally, as shown in column 4, estimated distance between the two districts, conditional on population density, also serves as a repelling force, presumably due to the higher transportation costs. To provide a sense of the magnitude of these effects, Table 6 provides results from simulations in which the heterogeneity channel is shut down. For example, as shown in experiment 3, merger rates increase from 0.43 percent to 0.46 percent, an increase of 7 percent, if districts do not condition on such differences, or equivalently, if these disparities between districts are eliminated. Similarly, setting the distance between two districts equal to zero leads to an economically significant increase in merger rates.[19]

One implicit econometric assumption underlying the results in columns 1-4 of Table 5 is a lack of serial correlation. That is, we have assumed that the unobserved match qualities are independent over time for a given border, whereas it might seem more reasonable to assume that these match qualities are correlated over time. In order to test the robustness of this assumption, column 5 of Table 5 presents results from an alternative specification in which unobserved match qualities are drawn for a given border in 1991 and are then

[19] The large effect of estimated distance in experiment 4 may be due to two factors. First, this experiment holds constant population density, which would obviously tend to increase as the geographic size of the two districts is decreased. Second, there is likely multicollinearity in the underlying specification in Table 5 as it simultaneously controls for enrollment, district square miles, and population density.

held constant over the remaining sample period.[20] As shown, the coefficients and standard errors are qualitatively similar to those in our baseline results of column 1, suggesting that our assumption of independence is not driving the statistical significance of our baseline empirical results.

In order to compare the results of our estimator relative to those in the existing literature, Table 7 presents the results from a bivariate logit model under the assumption of symmetry in match quality.[21] As shown in column 1, the merger incentives coefficient is positive and statistically significant. While this effect has the same sign as the corresponding coefficient in Table 5, it is difficult to compare the magnitude of the policy incentives effect in the two models given the independence assumption in the bivariate model. In particular, the results from policy experiments comparable to those in Table 6 would be difficult to interpret as there is no requirement that each district merge with only one partner in the bivariate probit model. While the size of these policy incentive effects are not directly comparable, the role of scale is more comparable: the bivariate logit coefficients suggest a minimum efficient size of 688, significantly larger than that suggested via our simulation approach. Regarding the role of heterogeneity, the bivariate probit finds statistically significant support for only one out of the three measures. As shown in column 4, geographic distance between districts creates a disincentive to merge, while columns 2 and 3 report no role for heterogeneity in the educational attainment of parents or differences in spending levels. One interpretation of these weaker heterogeneity results is that the simulation model, but not the bivariate logit model, accounts for the fact that districts choose one district from their many potential merger partners and thus more naturally allows for districts to choose the one district with which it is most appropriately matched.

[20] We treat new districts following mergers as having new borders. That is, if two districts 1 and 2 merge, creating district 12, then any borders of the new district 12 are treated as new borders and a new unobserved match quality is drawn from the type-I extreme value distribution.

[21] There are two differences between our bivariate logit model and the bivariate probit model estimated by Brasington. First, while Brasington's model estimates a parameter capturing the correlation between the unobserved utility of the two districts, we simply assume that this correlation equals 1 (we attempted to estimate a model with such a correlation but had difficulties with convergence as the correlation tended towards the extreme values of +1 or -1). Second, we use the logistic distribution rather than the normal distribution in order to get a closed form expression for the merger probability and to make our results more comparable with our simulation estimator, which assumes that match qualities are distributed type-I extreme value. As is well known, however, the normal and logistic distributions are both symmetric and are quite similar in shape, suggesting that our results are probably invariant to this distributional assumption.

6 Conclusion

In this paper, we develop an empirical approach to the study of school district mergers. This method is rooted in the economics of matching and thus overcomes several methodological problems with existing estimators. In particular, our approach allows for two-sided decision making, multiple potential merger partners for each district, and spatial interdependence in merger decisions. While the model does not generate an analytic expression for the probability of a merger, we show that the model can be estimated via simulation techniques. Applying this method to a spate of school district mergers in Iowa during the 1990s, our results demonstrate the importance of state subsidies, economies of scale as well as diseconomies of scale, and heterogeneity in explaining the patterns of mergers in Iowa during this time period. One caveat is that this analysis, which abstracts from racial heterogeneity, may not generalize to other states and time periods. Iowa has very little racial heterogeneity, and, as noted above, other studies, such as Alesina, Baqir, and Hoxby (2004), have found a strong role for such heterogeneity in terms of predicting the number of school districts within U.S. counties.

Appendix 1: Proof of Proposition

Chung (2000) has shown that no odd cycles implies the existence of a stable matching. The first part of our proof shows that, if there are two distinct stable matchings and strict preferences, then a cycle can be created. The second part of the proofs show that under the restriction of *symmetry in match quality* and strict preferences, there are no cycles. Thus, under the *symmetry in match quality* and strict preferences, a unique stable matching exists.

Claim: If there are two distinct stable matchings and strict preferences, then a cycle can be created.

Suppose there are $N > 2$ districts and two distinct stable matchings (A and B). In order for A and B to be distinct matchings, at least one district must be paired with different partners in A and B. Without loss of generality, denote this district as 1 and the partner in A as 2 and the partner in B as 4. Again, without loss of generality, assume that 1 prefers 2 over 4 ($2{\succ_1}4$). In order for matching B to be stable, it must be the case that district 2 is paired with another district, which we denote district 3, and further that district 2 prefers 3 over 1 ($3{\succ_2}1$).[22] In matching A, district 3 must either merge with 4 or a new district, say district 5. If 3 merges with 4, it must be that 3 prefers 4 over 2 in order for A to be stable ($4{\succ_3}2$). But, in order for matching B to be stable, it must be that 4 prefers 1 over 3 ($1{\succ_4}3$) and we thus have that ($2{\succ_1}4$), ($3{\succ_2}1$), ($4{\succ_3}2$), ($1{\succ_4}3$), which we refer to as the cycle 1234. On the other hand, if district 3 merges with district 5 in matching A, it must be the case that 3 prefers 5 over 2 ($5{\succ_3}2$) in order for A to be stable. Denote 6 as 5's partner in matching B. We thus know that 5 prefers 6 over 3 ($6{\succ_5}3$). Now, in matching A, 6 must merge with district 4 or a new district 7. If 6 merges with 4, it must be that 6 prefers 4 over 5 ($4{\succ_6}5$) in order for A to be stable. But, in order for B to be stable, 4 must prefer 1 over 6 ($1{\succ_4}6$) and we have the cycle 123564. On the other hand, if 6 merges with 7, etc. It is thus clear that, given a finite number of districts, this process will eventually lead to a cycle. Thus, if there are two distinct stable matchings and strict preferences, then a cycle can be created.

Claim: Under the restriction of *symmetry in match quality* and strict preferences, there are no cycles.

Suppose not and let the cycle of size C be given by $123...C$. Then, we know that the

[22] If district 2 is paired with itself in matching B, it must prefer district 1, its partner in matching A ($1{\succ_2}2$) in order for matching A to be stable. But it is then clear that matching B is unstable as district 1 prefers 2 over 4.

following preferences hold:

$$\begin{aligned}
U_{2,1} &> U_{C,1} \\
U_{3,2} &> U_{1,2} \\
U_{4,3} &> U_{2,3} \\
&\ldots \\
U_{C,C-1} &> U_{C-2,C-1} \\
U_{1,C} &> U_{C-1,C}
\end{aligned}$$

Inserting our specification and using the assumption that $Q_{i,j} = Q_{j,i}$, we have that:

$$\begin{aligned}
A_2 + Q_{1,2} &> A_C + Q_{1,C} \\
A_3 + Q_{2,3} &> A_1 + Q_{1,2} \\
A_4 + Q_{3,4} &> A_2 + Q_{2,3} \\
&\ldots \\
A_C + Q_{C,C-1} &> A_{C-2} + Q_{C-1,C-2} \\
A_1 + Q_{1,C} &> A_{C-1} + Q_{C,C-1}
\end{aligned}$$

Summing across these conditions, it is clear that the left hand side and right hand side are identical. Hence, a contradiction and no cycle.

Appendix 2: Smooth simulator

For each simulation r, the probability of a merger between two districts i and j can be expressed as the probability of deviations from the stable matching. In particular, denote U_i^* and U_j^* as the equilibrium utility, or value of the game, for districts i and j under the stable matching. After calculating these utilities, we provide each district a small amount of additional information $(\tau\eta_{ij})$,which is also distributed type-I extreme value and is assumed symmetric between the two partners, regarding each of their options. The parameter τ is referred to as the smoothing parameter. For two bordering districts i and j that are not merged together under the stable matching, we can then calculate the probability of a deviation as follows:

$$\begin{aligned}\Pr(\text{deviation}_{ij}) &= \Pr(U_{ij} + \tau\eta_{ij} > U_j^* + \tau\eta_j, U_{ji} + \tau\eta_{ij} > U_i^* + \tau\eta_i \\ &= \frac{1}{1 + \exp[(U_j^* - U_{ij})/\tau] + \exp[(U_i^* - U_{ji})/\tau]}\end{aligned}$$

where τ is the smoothing parameter and is chosen to be small. As this smoothing parameter converges to zero, the probability of a merger approaches 0 given that at least one district must prefer its equilibrium utility to this deviation option in order for this matching to be stable ($U_j^* > U_{ij}$ or $U_i^* > U_{ji}$). Thus, in the limit, the smooth simulator approaches the frequency simulator. But, for any positive τ, the probabilities are bounded between zero and one. This simulated probability can thus be interpreted as the probability of a deviation, allowing districts to make mistakes, where the magnitude of the mistakes depends upon the smoothing parameter τ.

For merged borders ($y_{ij}^r = 1$), create a set of willing deviation partners for i ($U_{ik} > U_k^*$); to guarantee that this set is not empty, include district i itself in this set where $U_{ii} = 0$. Denote this set as B_i with elements indexed by k. Create a similar set for j; denote this set B_j with elements indexed by l.

$$\begin{aligned}\Pr(\text{no deviation}_{ij}) &= \\ \Pr(U_{ji} + \tau\eta_{ij} &> U_{ki} + \tau\eta_{ki} \text{ all } k,\ U_{ij} + \tau\eta_{ij} > U_{lj} + \tau\eta_{lj} \text{ all } l) \\ &= \frac{1}{1 + \sum_k \exp[(U_{ki} - U_{ji})/\tau] + \sum_l \exp[(U_{lj} - U_{ij})/\tau]}\end{aligned}$$

Again, as the smoothing parameter converges to zero, the probability of no deviation, and thus a merger, approaches 1 as each term in the summation (for example, $\exp[(U_{ki} - U_{ji})/\tau]$) approaches zero given that district i must prefer j over k in order for the matching to be stable. Thus, the smooth simulator again approaches the frequency simulator.

Appendix 3: GMM Estimator and Inference

For estimation purposes, we use a simulated method of moments approach, where the objective function is defined below:

$$[I'(y-p)]'W[I'(y-p)]$$

where y is an $N \times 1$ vector of observed merger indicator variables, p is an $N \times 1$ vector of simulated merger probabilities, and I is a $N \times k$ matrix of instruments, or exogenous variables. Finally, W is a $k \times k$ weighting matrix. The optimal weighting matrix is given by the inverse of the variance-covariance matrix:

$$\begin{aligned} W &= var[I'(y-p)]^{-1} \\ &= [I'var(y-p)I]^{-1} \\ &= [I'(E(yy')-pp')I]^{-1} \end{aligned}$$

Note that $E(yy')$ is not necessarily diagonal due to the interdependence in merger decisions. However, we can estimate this matrix via our simulation approach as follows:

$$E(yy') = (1/R)\sum_{r=1}^{R} y^r y^{r\prime}$$

Let $m \geq k$ denote the number of parameters in the vector θ. Then, we calculate the standard errors according to the following variance-covariance matrix:

$$Var(\theta) = (1/N)I'[dp/d\theta]W[dp/d\theta]'I.$$

References

[1] Alesina, Alberto and Enrico Spolaore, 2003, The Size of Nations, MIT Press: Cambridge.

[2] Alesina, Alberto and Enrico Spolaore, 1997, "On the Size and Number of Nations," *Quarterly Journal of Economics*, 112: 1027-1056.

[3] Alesina, Alberto and Enrico Spolaore and Romain Wacziarg, 2000, "Economic Integration and Political Disintegration," *American Economic Review,* 90: 1276-1296.

[4] Alesina, Alberto and Reza Baqir and Caroline Hoxby, 2004, "Political Jurisdictions in Heterogenous Communities," *Journal of Political Economy*, 112: 348-396.

[5] Andrews, Matthew and William Duncombe, and John Yinger, 2002, "Revisiting Economies of Size in American Education: Are We Any Closer to a Consensus?," *Economics of Education Review*, 21: 245-262.

[6] Becker, Gary, 1973, "A Theory of Marriage: Part I" *Journal of Political Economy,* 81(4): 813-846.

[7] Bolton, Patrick, and Gerald Roland, 1997, "The Breakup of Nations: A Political Economy Analysis." *Quarterly Journal of Economics*, 112: 1057-1090.

[8] Brasington, David. 1999. "Joint Provision of Public Goods: the Consolidation of School Districts." *Journal of Public Economics*, 73: 373-393.

[9] Brasington, David. 2003a. "Snobbery, Racism, or Mutual Distaste: What Promotes and Hinders Cooperation in Local Public-Good Provision?" *Review of Economics and Statistics*, 85: 874-883.

[10] Brasington, David. 2003b. "Size and School District Consolidation: Do Opposites Attract?" *Economica*, 70: 673-690.

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

[12] Filer, John and Lawrence Kenny. 1980. "Voter Reaction to City-County Consolidation Referenda," *Journal of Law and Economics,* 23(1): 179-190.

[13] Fox, Jeremy. 2005. "Nonparametric Estimation of Matching Games with Endogenous Prices", working paper, University of Chicago.

[14] Gordon, Nora and Brian Knight, 2005. "The Fiscal Impact of Political Integration: Evidence from School District Consolidations" *mimeo.*

[15] Hooker, Clifford and Van Mueller, 1970, The Relationship of School District Reorganization to State Aid Distribution Systems, National Education Finance Project Special Study No. 11.

[16] Kenny, Larry and Amy Schmidt, 1994. "The Decline in the Number of School Districts in the U.S.: 1950-1980" *Public Choice*, 79: 1-18.

[17] Lerman, R. and C. Manski, 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities" in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, Cambridge: MIT Press, 1981.

[18] McFadden, Daniel, 1989, "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration" *Econometrica*, 57: 995-1026.

[19] Persson, Torsten and Guido Tabellini, 2000, Political Economics: Explaining Economic Policy, MIT Press: Cambridge.

[20] Poirer, Dale. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics*, 12: 209-217.

[21] Rodrigues-Neto, Jose Alvaro. 2005. "Representing Roommates Preferences with Symmetric Utilities." forthcoming, *Journal of Economic Theory.*

[22] Roth, Alvin, 1990. Two-sided matching : a study in game-theoretic modeling and analysis, Cambridge University Press: Cambridge.

[23] Sorensen, Morten. 2005. "How Smart is Smart Money: An Empirical Two-sided Matching Model of Venture Capital", working paper, University of Chicago.

[24] Stern, Steven, 1997. "Simulation-Assisted Estimation", *Journal of Economic Literature,* 35(4)*:*2006-2039.

Table 1: Summary of Merger Incentives

	WGS pre-91	WGS post-91	Reorganize pre-91 or post 93	Reorganize 91-93 (for 5 yrs after reorganization)
Foundation tax rate (paid to state on district assessed property valuation)	=$5.40 per $1000	=$5.40 per $1000	=$5.40 per $1000	year 1 = 4.4 year 2 = 4.6 year 3 = 4.8 year 4 = 5.0 year 5 = 5.2
Per-pupil foundation payment from state to district (F)	F per non-WGS student 1.1F per WGS student	F per non-WGS student F per WGS student	F per non-WGS student F per previous WGS	F per non-WGS student 1.1F per previous WGS

Table 2a: Merger Activity by District Enrollments

600 <= smaller district < 1200		0.00%
smaller district < 600	2.36%	0.36%
	larger district < 600	600 <= larger district < 1200

Table 2b: Merger Activity by District Enrollments, further details

300 <= smaller < 600		1.79%
smaller district < 300	3.90%	2.49%
	larger district < 300	300 <= larger district < 600

Table 3: Summary Statistics for Key Variables

Observation = school district / year

mean (standard deviation)

	merger n=108	no merger n=4579
enrollment	452 (311)	1288 (2499)
percent of adults with less than HS degree	0.2091 (0.0459)	0.2062 (0.0529)
percent of adults with college degree	0.1066 (0.0280)	0.1178 (0.0529)
per-pupil instructional spending (scaled by housing values)	0.0801 (0.0295)	0.0703 (0.0288)
area (square miles)	98.7517 (43.0814)	134.9075 (69.1751)
population density	4.5581 (2.2789)	17.3308 (54.3544)

Table 4: Summary Statistics for Key Variables

Observation = border / year

mean (standard deviation)

	merger n=50	no merger n=12184
merger incentive	2.4172	0.4737
	(2.4516)	(1.1630)
economies of scale	-0.7772	-0.8420
	(0.4422)	(0.6197)
diseconomies of scale	-3140.61	-11969.15
	(2067.28)	(28582.14)
heterogeneity in percent adults	0.0030	0.0032
with less than HS degree	(0.0036)	(0.0049)
heterogeneity in percent adults	0.0015	0.0037
with college degree	(0.0020)	(0.0098)
heterogeneity in spending	0.0010	0.0009
(property-value adjusted)	(0.0028)	(0.0033)
estimated distance	9.6427	11.4768
between districts	(1.1579)	(2.1751)
population density	4.4837	14.9497
	(1.9605)	(46.2279)

Table 5: Determinants of School District Consolidations

	column 1	column 2	column 3	column 4	column 5
merger incentive	0.5139**	0.5177**	0.5179**	0.3235**	0.5234**
	(0.1347)	(0.1244)	(0.1139)	(0.1354)	(0.0945)
economies of scale	-0.3184**	-0.2992**	-0.3066**	-0.2106**	-1.0722**
	(0.0249)	(0.0211)	(0.0189)	(0.0347)	(0.0382)
diseconomies of scale	0.0002**	0.0002**	0.0002**	0.0002**	0.0002**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
heterogeneity in percent adults		-6.8629			
with less than HS degree		(8.9018)			
heterogeneity in percent adults		-11.0788**			
with college degree		(2.2978)			
heterogeneity in spending			-68.6160**		
(property-value adjusted)			(10.4858)		
estimated distance				-0.5045**	
between districts				(0.0360)	
population density				-0.1184**	
				(0.0004)	
constant	-6.7812**	-6.7925**	-6.7383**	-8.1039**	-6.7316**
	(0.0717)	(0.0654)	(0.0626)	(0.0748)	(0.0428)
sample size	12234	12234	12234	12234	12234
serial correlation assumption	independent	independent	independent	independent	dependent

Standard errors below coefficients

* significant at 10%; ** significant at 5%

Table 6: Counterfactual experiments

	merger rates	
	baseline	**counterfactual**
Experiment 1: remove merger incentives, 1991-1993 (merger rates for borders with both districts below 600 enrollment)	2.89%	0.46%
Experiment 2: eliminate adult educational heterogeneity	0.43%	0.43%
Experiment 3: eliminate heterogeneity in spending levels	0.43%	0.46%
Experiment 4: set estimated distance to zero	0.0041	11.54%

Table 7: Results from Bivariate Logit Model

	column 1	column 2	column 3	column 4
merger incentive	0.4226**	0.4219**	0.4146**	0.2982**
	(0.0762)	(0.0762)	(0.0754)	(0.0792)
economies of scale	-2.5942**	-2.5964**	-2.4376**	-1.0391
	(0.9728)	(0.9735)	(0.9090)	(1.5017)
diseconomies of scale	0.0005**	0.0005**	0.0005**	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0003)
heterogeneity in percent adults		-7.9141		
with less than HS degree		(37.0851)		
heterogeneity in percent adults		-1.9801		
with college degree		(60.5672)		
heterogeneity in spending			-117.6041	
(property-value adjusted)			(78.9324)	
estimated distance				-0.5644**
between districts				(0.2149)
population density				-0.1560
				(0.1425)
constant	-4.9715**	-4.9401	-4.6454**	1.5591
	0.4492	(0.4719)	(0.4596)	(2.8054)
sample size	12234	12234	12234	12234

Standard errors below coefficients

* significant at 10%; ** significant at 5%

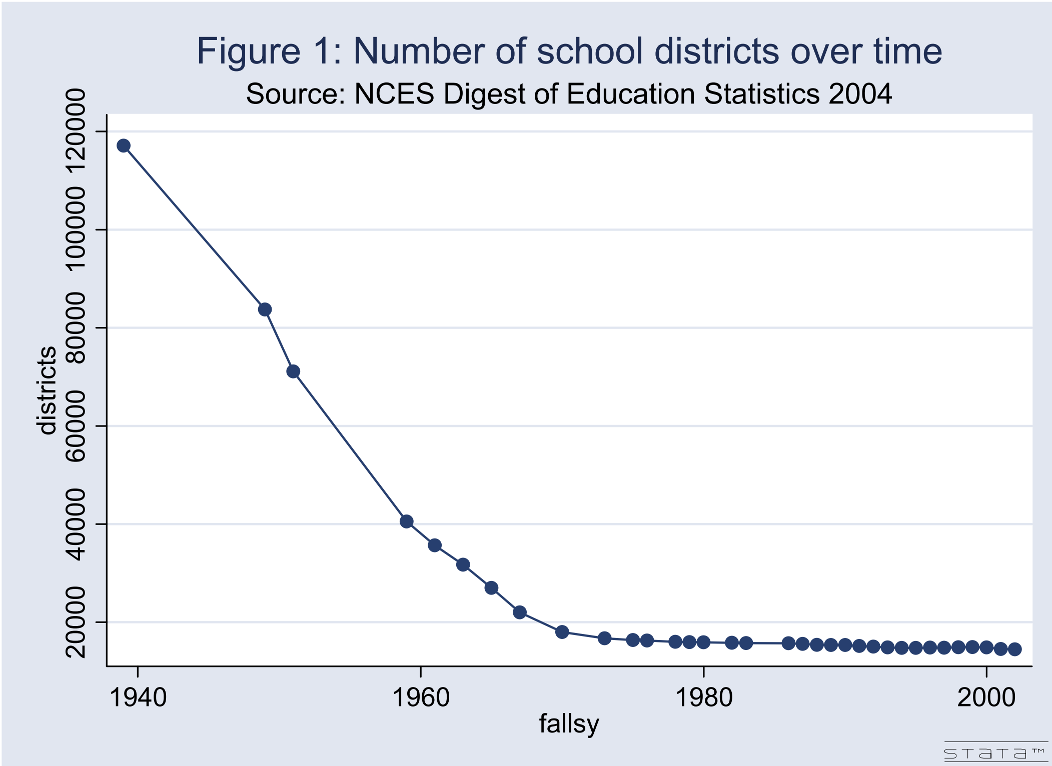


Figure 1: Number of school districts over time
Source: NCES Digest of Education Statistics 2004
districts
20000
40000
60000
80000
100000
120000
1940
1960
1980
2000
fallsy

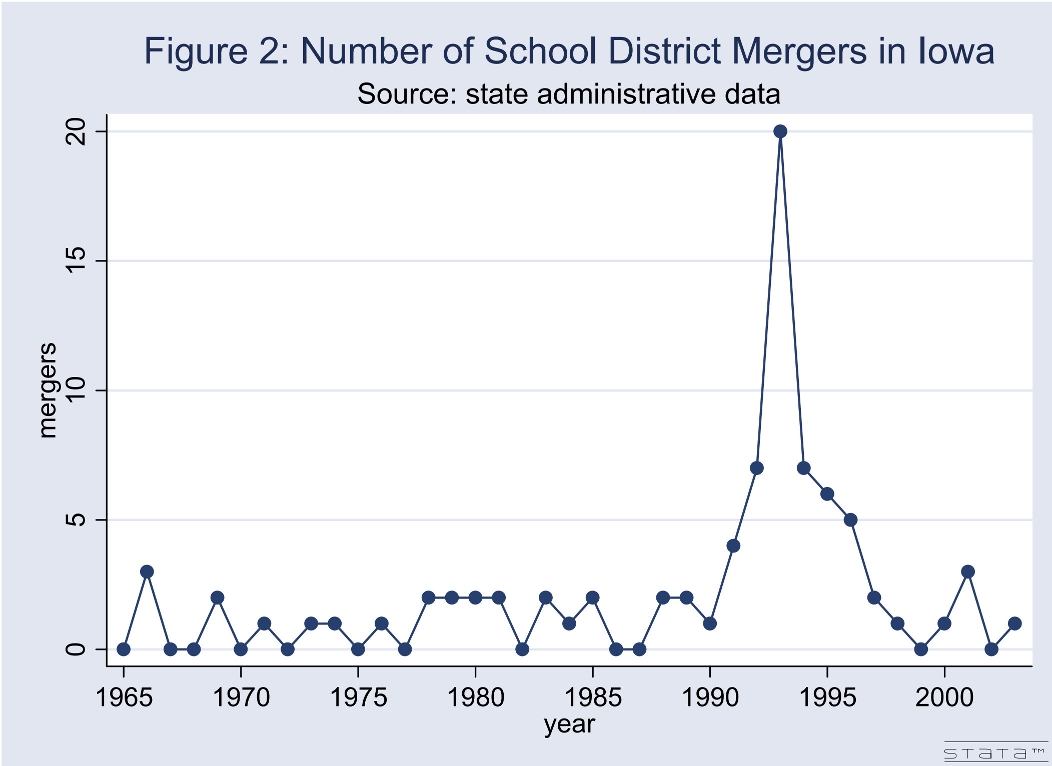


Figure 2: Number of School District Mergers in Iowa
Source: state administrative data
mergers
0
5
10
15
20
1965
1970
1975
1980
1985
1990
1995
2000
year

Figure 3

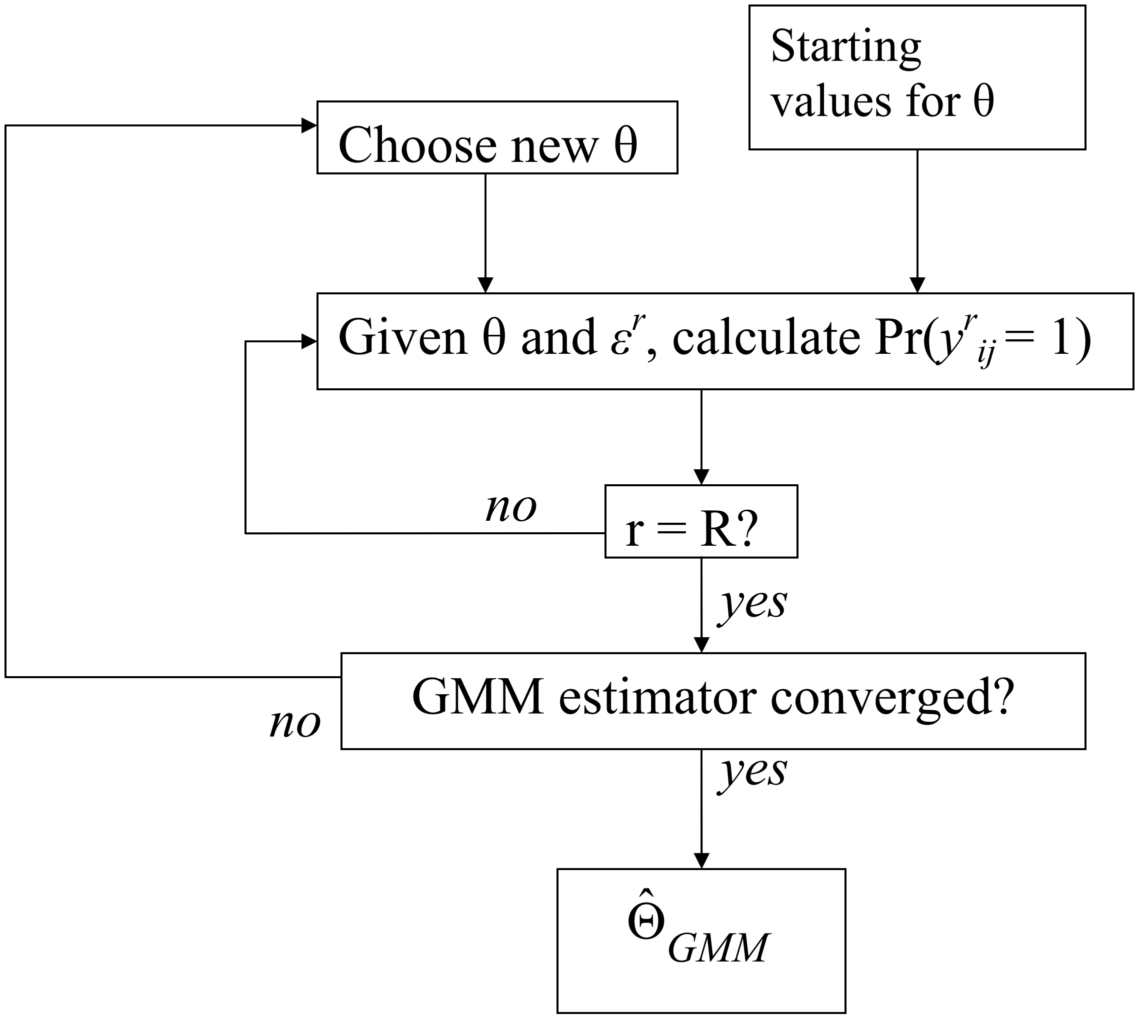
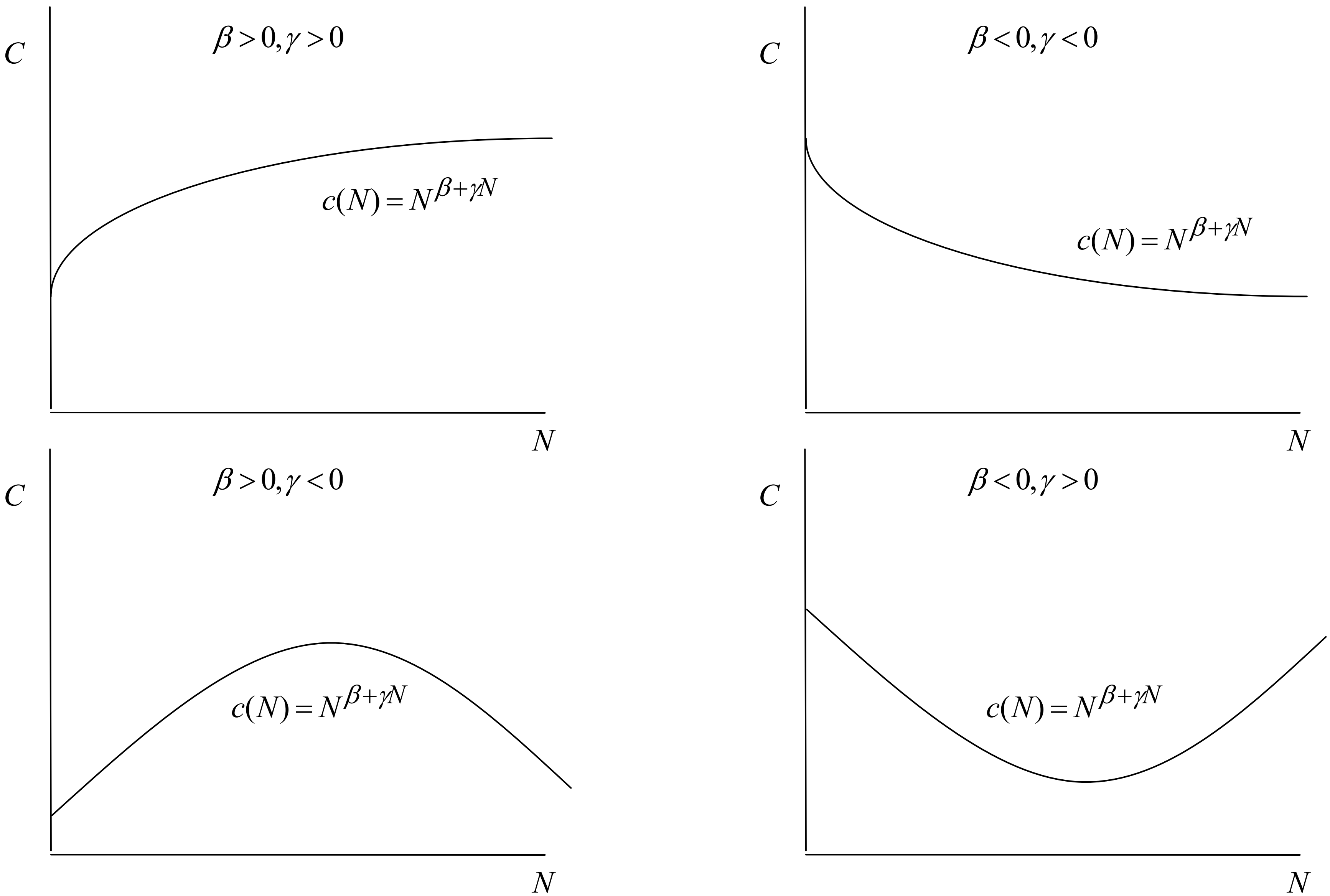

Starting values for θ
Choose new θ
Given θ and ε^r, calculate $\Pr(y^r_{ij} = 1)$
no
r = R?
yes
GMM estimator converged?
no
yes
$\hat{\Theta}_{GMM}$

Figure 4: Cost curve specification



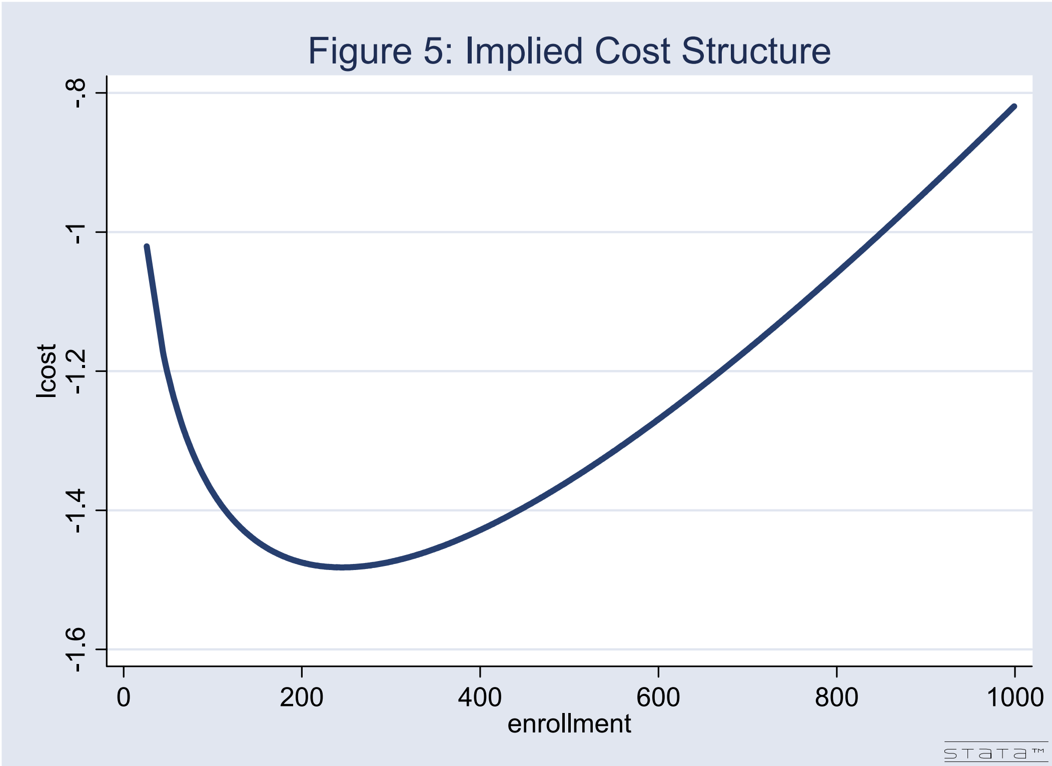


Figure 5: Implied Cost Structure
-.8
-1
-1.2
-1.4
-1.6
lcost
0
200
400
600
800
1000
enrollment