Issues & Studies© 45, no. 2 (June 2009): 61-98.

The Spatial Organization of Elections and the Cube Law*

TSE-MIN LIN AND FENG-YU LEE

The single-member district plurality system for legislative elections, also known as first-past-the-post (FPTP), usually results in disproportional seat distributions among parties. The Cube Law, which stipulates that the ratio of seats won by the parties in a two-party system is a cubic

---

TSE-MIN LIN (林澤民) (Ph.D., University of Minnesota, 1990) is currently associate professor in the Department of Government at the University of Texas at Austin. His research interests include methodology, formal theory, and American and comparative political behavior. He can be reached at <tml@austin.utexas.edu>.

FENG-YU LEE (李鳳玉) (Ph.D., University of Texas at Austin, 2007) is currently assistant professor in the Department of Political Science at National Taiwan University. Her research interests include comparative political institutions, democratization, ethnic politics, and methodology and quantitative methods. She can be reached at <fylee323@ntu.edu.tw>.

*Earlier versions of this article were presented at the 67th Annual National Conference of the Midwest Political Science Association, Chicago, Illinois, April 2-5, 2009; the Conference on Electoral Politics in Taiwan: Origins and Consequences of the Electoral Reform, February 6-8, 2009, University of South Carolina, Columbia, South Carolina; and Taiwan's Election and Democratization Study International Conference, January 17-18, 2009, National Taiwan University, Taipei, Taiwan. Research for this article was partially funded by National Taiwan University's Special Project for the Advancement of Humanities and Social Science Research (提昇人文及社會科學研究能量專案). Some of the data analyzed in this article were compiled from Taiwan's Election and Democratization Studies, 2005-2008 (III): The Legislative Elections (TEDS 2008L). The coordinator of the multi-year project TEDS is Professor Chi Huang (黃紀) of National Chengchi University. The subject of TEDS2008L is the legislative elections of 2008. The principal investigator is Professor Yun-han Chu (朱雲漢) of National Taiwan University. We appreciate the assistance of the above-mentioned institutions and individuals in providing data. We alone are responsible for the views expressed herein. We also thank Professor Chih-sung Teng (鄧志松) of National Taiwan University for providing us with a part of the aggregate data analyzed in this article and Chih-cheng (Almond) Meng (蒙志成) and Liru Li (李立茹) for excellent research assistance.

function of the corresponding ratio of total votes, attests to the degree of disproportionality in representation under FPTP. This so-called law, however, is really just a benchmark, and the performance of FPTP can vary from country to country and from time to time.

It is well known that the validity of the Cube Law, and hence the proportionality of representation, depends on the distribution of vote share across constituencies. Scholars have pointed to contagion, heterogeneity, and the size of constituencies as factors that may affect the conditions under which the Cube Law can be sustained. In this article, we propose a spatial regression model which implies all these factors. Empirically, we investigate Taiwan's recent legislative elections to test our theory. Our findings show that, in this case, low spatial autocorrelation at the district level is associated with vastly disproportional election outcomes.

KEYWORDS: spatial organization of elections; Cube Law; Taiwan; FPTP system; proportionality.

* * *

論 One dominant issue in the study of electoral systems concerns disproportionality, i.e., overrepresentation of some parties at the expense of others in a parliament. The single-member district plurality system, which is also known as first-past-the-post (FPTP), usually results in disproportional seat distributions among parties. For example, in United Kingdom's (UK) parliamentary elections of 2005, New Labour gained 55.3 percent of the total seats with only 35.3 percent of the total votes. The Conservative Party won just slightly fewer votes (32.3 percent) than New Labour, but obtained only 30.7 percent of the seats.

The Cube Law, a benchmark for the FPTP system, attests to the degree of overrepresentation of the party that wins the popular vote in a two-party system. The Law was originally formulated on the basis of UK parliamentary elections. The results of elections in New Zealand around the mid-twentieth century also supported the Law.[1] More recently, Japan's elections in 2003 provided yet another example of how the Cube Law can predict how disproportional representation under the FPTP system can be.

[1]M. G. Kendall and A. Stuart, "The Law of the Cubic Proportion in Election Results," *British Journal of Sociology* 1, no. 3 (September 1950): 183-97.

However, the Cube Law has failed to hold true in other countries or even in the same aforementioned countries during different time periods.[2] Take Taiwan's seventh Legislative Yuan (立法院) elections in 2008 as an example. These elections produced a seat distribution that is much more disproportional than that prescribed by the Cube Law. With only 59 percent of the total votes, the pan-Blue coalition won 82 percent of the total seats, while about 41 percent of the total votes gave the pan-Green coalition only 18 percent of the seats. The United States' congressional elections in 2004 were even more disproportional.

Why does the performance of the FPTP system exhibit such variation? While the system produces disproportional election outcomes in general, some outcomes are more disproportional than others. In this article, we seek to explain the proportionality of representation under FPTP from the analytical framework provided in the Cube Law literature. Specifically, we propose a spatial regression model that encompasses three factors: contagion, heterogeneity, and the size of constituencies. Empirically, we investigate Taiwan's recent Legislative Yuan elections to test our theory.

The Cube Law

The Cube Law stipulates that, with two political parties competing for legislative seats under the FPTP system, the ratio of seats won by the parties is a cubic function of the corresponding ratio of total votes. More generally, if S and V represent, respectively, the proportions of seats and votes won by party 1, then the proportionality of the system can be summarized by a real number k such that

$$\frac{S}{1-S}=\left(\frac{V}{1-V}\right)^{k} \qquad (1)$$

---

[2]Edward R. Tufte, "The Relationship between Seats and Votes in Two-Party Systems," *American Political Science Review* 67, no. 2 (June 1973): 540-47.

The Cube Law stipulates that

$$k = \frac{\ln(S/(1-S))}{\ln(V/(1-V))} = 3\,.$$

From (1) S can be isolated as a function of V:

$$S = \frac{V^k}{V^k + (1-V)^k}$$

Kendall and Stuart (1950) show that, at $k = 3$, such a function can be approximated by the normal cumulative distribution function (cdf) with a specific standard deviation:

$$S = \frac{V^3}{V^3 + (1-V)^3} \approx \Phi(\frac{V-0.5}{s_0}) \quad \text{where } s_0 = 0.137 \qquad (2)$$

Therefore, the Cube Law can be rigorously stated as:

The Cube Law: With two political parties competing for legislative seats under FPTP, if Y, the district level vote share for party 1, is normally distributed with mean V and standard deviation $s_0 = 0.137$, then the two parties' seat ratio equals the cube of their vote ratio. Formally, if $Y \sim N(V, s_0^2)$, then $k = 3$.[3]

[Proof]

Let Y be a random variable representing party 1's vote share in a typical constituency, then Equation (2) implies that

[3] "The validity of the cubic proportion then depends on three things: (1) the empirical fact that the distribution of proportions p at an election is nearly normal, (2) the mathematical fact that the cubic-proportion law very closely approximates to a normal form with the same variance, and (3) the empirical fact that the variance of the cubic-proportion law is very closely approximated by the variance of the observed distributions." See Kendall and Stuart, "The Law of the Cubic Proportion in Election Results," 191.

$$\Pr(Y > 0.5) \approx \Phi(\frac{V - 0.5}{s_0}) \tag{3}$$

However, this suggests that Y is a random variable following the normal distribution with mean V and variance s_0^2, i.e.,

$$Y \sim N(V, s_0^2)$$

To see this, if $Y \sim N(V, s_0^2)$, then $Z = (Y - V)/s_0 \sim N(0,1)$, and hence

$$\Pr(Y > 0.5) = \Pr(Z > \frac{0.5 - V}{s_0}) = 1 - \Phi(\frac{0.5 - V}{s_0}) = \Phi(\frac{V - 0.5}{s_0})$$

The requirement that $E(Y) = V$ is automatically true by the Law of Large Numbers if there is a large number of districts all having the same size. Q.E.D.

Kendall and Stuart (1950) derive a simple formula to predict the bias of the Cube Law when the standard deviation of Y deviates from s_0. Suppose S_0 is the proportion of seats corresponding to $s_0 = 0.137$, and S is the proportion of seats corresponding to a standard deviation $s = s_0 + \delta s_0$ [i.e., $\delta = (s - s_0)/s_0$ is the change in proportion from $s_0 = 0.137$] and let $x = V - 0.5$, then

$$\frac{S}{1-S} - \frac{S_0}{1-S_0} = -\left(\frac{12x}{1-4x^2}\right)\delta \tag{4}$$

This implies that, if $\delta > 0$, then

$$\frac{S}{1-S} = \left(\frac{V}{1-V}\right)^k < \frac{S_0}{1-S_0} = \left(\frac{V}{1-V}\right)^3$$

or

$$k < 3$$

Conversely, if $\delta < 0$, then $k > 3$. Substantively, representation under FPTP will be more proportional than what is prescribed by the Cube Law if the standard deviation of Y is greater than $s_0 = 0.137$, and it is less proportional if the standard deviation of Y is smaller then $s_0 = 0.137$.[4]

Although the Cube Law literature has identified many systems for which $k \approx 3$, the Law is by no means universally true. And even when the Law is true, it is not at all clear why $s = s_0$. More likely, s, and hence k, varies from country to country and from time to time. In their efforts to explain why s or k would change, scholars have pointed to at least three types of causes: contagion, heterogeneity, and size of constituencies.

Contagion

If each voter votes independently with the same probability π for party 1, and there are n voters in a district, then

$$Y \sim N(\pi, \frac{\pi(1-\pi)}{n}) \text{ with } E(Y) = \pi \text{ and } Var(Y) = \frac{\pi(1-\pi)}{n}$$

For a constituency with n = 140,000, for instance, the standard deviation of Y is 0.0013 if $\pi = 0.5 \sim 0.6$, much smaller than required by the Cube Law. Thus, for the Cube Law to have any practical applicability, the assumption that voters vote independently must be abandoned.

Kendall and Stuart suggest that "voters cannot be regarded as scattered at random over the various constituencies" and that the observed variation of Y among constituencies "is due to the fact that voters of similar political views tend to occur in groups."[5] In elaboration, they provide a Markov scheme in which individual voting behavior is contagious. The scheme, as Kendall and Stuart demonstrate, can cause sufficient variation in Y to sustain the Cube Law.

---

[4]This is conditional on V not being too close to 0.5. If $V \approx 0.5$, $x = V - 0.5 \approx 0$, and the bias against the Cube Law would be negligible even if $s \neq s_0$.

[5]Kendall and Stuart, "The Law of the Cubic Proportion in Election Results," 188.

Kendall and Stuart's Markov scheme is simplistic because it models contagion among voters only in a sequential way. Coleman develops a contagious binomial model that allows for mutual dependence among voters.[6] Coleman, however, admits that his model alone cannot realistically explain the variability in Y required by the Cube Law. For the Cube Law to be sustained, the political processes must somehow generate the variability among constituencies.[7]

Heterogeneity

In addition to the Markov scheme, Kendall and Stuart also suggest a Lexian scheme which can generate large variation in vote share from constituency to constituency. The Lexian model assumes that there are *k* groups of voters. The voters in each constituency are all from one of the groups, but all groups contribute equally to the national electorate. An example of the Lexian scheme is the extreme form of regionalism in which constituencies are divided into spatially distributed groups with complete within-group homogeneity and between-group heterogeneity.[8] Although a Lexian scheme does not theoretically require constituencies from the same group to be geographically contiguous, spatial concentration is most likely to be the case. For example, ethnicity is often an important factor in voting behavior, and ethnic groups tend to concentrate geographically. Although contagion and heterogeneity are conceptually different, in practice they may not be distinguishable. This is because contagion can cause assimilation among voters spatially close to one another, which may be indistinguishable from the clustering of voters of similar political views without mutual influence.

Gudgin and Taylor point out that clusters that are much larger than legislative constituencies will lead to a non-normal distribution of Y

[6]James S. Coleman, *Introduction to Mathematical Sociology* (New York: Free Press, 1964).
[7]Ibid., 352.
[8]The *Social Science Encyclopedia* defines "region" as "an area of the earth's surface which is relatively homogeneous, and differs from its neighbors on certain criteria." See Adam Kuper and Jessica Kuper, eds., *Social Science Encyclopedia*, second edition (London: Routledge, 1996), 729.

with a large standard deviation.[9] Only when the clusters are units much smaller than constituencies, e.g., socially homogeneous districts within a township, will the spatial organization of elections be conducive to the conditions of the Cube Law (see figure 1).[10] Gudgin and Taylor's illustration reflects the common understanding that under FPTP, if the minority party's support is geographically concentrated (i.e., if it is regionalistic), then election results may not be as disproportional as prescribed by the Cube Law. It also underscores the critical conditions of the Cube Law, namely the normality of the district-level vote share and a certain degree of dispersion in the parties' electoral support. In other words, for the Cube Law to hold true, the parties' support must be dispersed across all districts in a certain way and to a certain degree. Attesting to the identification problem associated with heterogeneity and contagion, Gudgin and Taylor develop a Markov model that takes into account of the size of voter clusters.

Size of Constituencies

Taagepera develops a mathematical theory that the proportionality of representation

$$k' = \frac{\ln(V_{total})}{\ln(S_{total})}$$

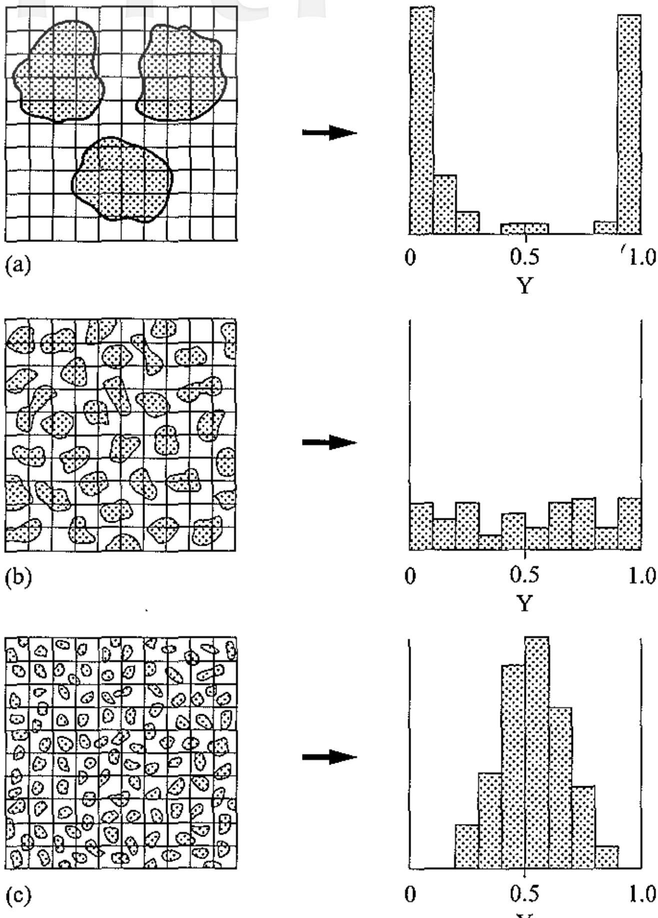
where V_{total} is the total number of popular votes, and S_{total} is the total number of seats.[11] Since V_{total}/S_{total} is the average size of a constituency, Taagepera's theory amounts to saying that proportionality is a function of constituency size in a sort of logarithmic scale. While Taagepera pro-

[9]Graham Gudgin and Peter J. Taylor, *Seats, Votes, and the Spatial Organization of Elections* (London: Pion, 1979).

[10]Adapted from ibid., 37.

[11]Rein Taagepera, "Seats and Votes: A Generalization of the Cube Law of Elections," *Social Science Research* 2, no. 3 (September 1973): 257-75; and Rein Taagepera, "Reformulating the Cube Law for Proportional Representation Elections," *American Political Science Review* 80, no. 2 (June 1986): 489-504.

Figure 1
Clustering of Voters and Variability of Election Outcomes



Source: Adapted from Graham Gudgin and Peter J. Taylor, *Seats, Votes, and the Spatial Organization of Elections* (London: Pion, 1979), 37, figure 3.1.

vides no substantive reason why size would affect proportionality, it is clear that smaller constituency size entails greater variability among constituencies. Empirically, Taagepera shows that for many legislative elections, the ratio of $\ln(V_{total})$ and $\ln(S_{total})$ is indeed close to 3.[12]

These theories essentially stipulate that the stronger the contagion in voting behavior, the more heterogeneous the constituencies, and the smaller the size (in logarithmic scale) of constituencies, the more proportional representation under FPTP will be. Although these theories do not explain why $s = s_0$ or why $k = 3$, they do provide clues as to why these quantities may vary from country to country and from time to time.

In this article, we argue that all these theories are implied by an econometric model of spatial interdependence. In the next section, we develop such a model. Our purpose is not so much to "explain" the Cube Law as to understand the reasons for its success *and* failure. Our model demonstrates that the standard deviation of Y is a function of spatial autocorrelation. The performance of the Cube Law is therefore a function of spatial autocorrelation.

Spatial Interdependence and the Cube Law

We began with noting that contagious voting behavior at the individual level may carry over to a hierarchy of spatial units (villages, townships, and legislative districts) and cause diminishing spatial interdependence as the level goes higher. Following Lin and Cohen,[13] we propose the following nonlinear spatial regression model for electoral outcomes at any level:

$$y_{i,t} = F\left(\rho(\sum_{j=1}^{m} w_{ij} y_{j,t-1} - 0.5) + \beta' x_{i,t} + \gamma\varepsilon_{i,t}\right) = \Phi\left(\frac{\rho(\sum_{j=1}^{m} w_{ij} y_{j,t-1} - 0.5) + \beta' x_{i,t} + \gamma\varepsilon_{i,t}}{\sigma}\right) \tag{5}$$

[12]Taagepera, "Reformulating the Cube Law," 490-92.

[13]Tse-min Lin and Matthew Cohen, "Spatial Regression as a Statistical Model of Regionalism" (Paper presented at the 66th Annual National Conference of the Midwest Political Science Association, Chicago, Illinois, April 2008), 1-41.

where

$y_{i,t}$ is party 1's share of the two-party vote in spatial unit i at time t;

$F(.)$ is the cumulative distribution function (cdf) of the normal distribution $N(0,\sigma^2)$;

$\Phi(.)$ is the cdf of the standardized normal distribution $N(0,1)$;

w_{ij} is the typical element of a standardized spatial weight matrix W defined as

$$w_{ij} = \begin{cases} \dfrac{1}{m_i} & \text{if } j \neq i \text{ and unit } j \text{ is a neighbor of unit } i,\ m_i \text{ being the number of } i\text{'s neighbors} \\ 0 & \text{otherwise} \end{cases}$$

m is the total number of units at a given level;

$x_{i,t}$ is a vector of exogenous independent variables;

$\varepsilon_{i,t}$ is a normally distributed error term. $\varepsilon_{i,t} \sim N(0,1)$;

ρ and γ are constant coefficients; and

β is a vector of coefficients.

Equation (5) may also be written as

$$\Phi^{-1}(y_{i,t}) = \frac{\rho(\sum_{j=1}^{m} w_{ij} y_{j,t-1} - 0.5) + \beta' x_{i,t} + \gamma \varepsilon_{i,t}}{\sigma} \tag{6}$$

where $\Phi^{-1}(.)$ is the inverse normal cdf.

Note that the definition of w_{ij} entails that $\sum_{j=1}^{m} w_{ij} y_{j,t-1}$, the spatial lag of the dependent variable, is party 1's average vote share among i's neighboring units at time $t-1$. Equation (5) thus stipulates that party 1's vote share in unit i is a nonlinear function of, among other things, the party's majority margin (i.e., $\sum_{j=1}^{m} w_{ij} y_{j,t-1} - 0.5$) among i's neighboring units at time $t-1$. The nonlinearity of the functional form is necessary because $0 \leq y_{i,t} \leq 1$. The extent of the nonlinearity is captured by the functional form of $F(.)$ or, equivalently, the parameter σ.

Equation (5) is essentially a spatial regression model with a limited dependent variable. Lin and Cohen (2008) have shown that the model can exhibit increasing returns for certain parameter values.[14] For example, when $\rho = 1$, $\beta = 0$, and $\gamma = 0$, the model will exhibit increasing returns if $\sigma < \phi(0) \approx 0.4$. In this regime, the function

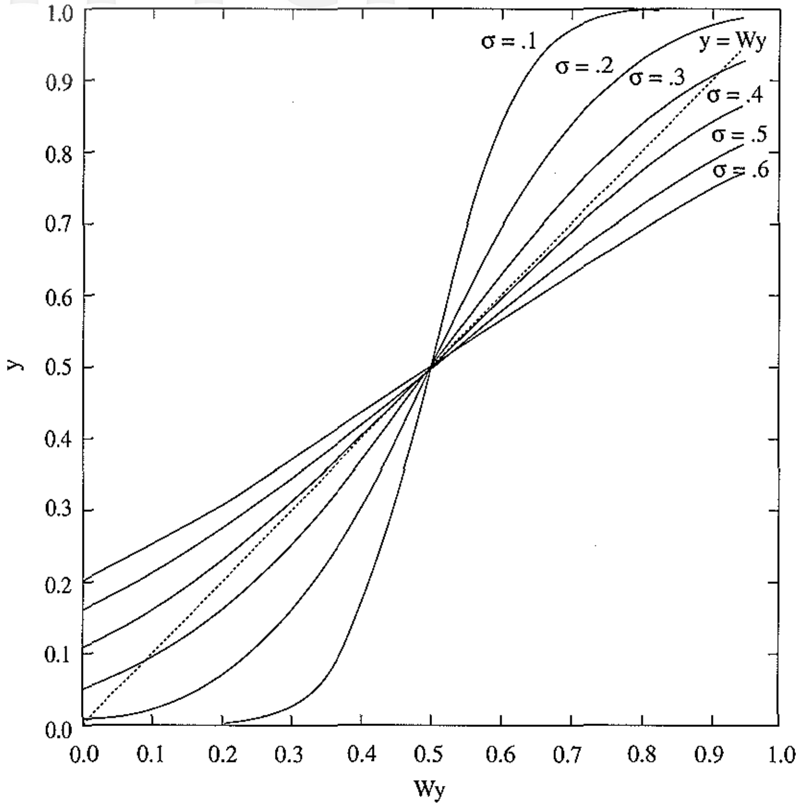
$$y_{i,t} = \Phi\left(\frac{\sum_{j=1}^{m} w_{ij} y_{j,t-1} - 0.5}{\sigma}\right)$$

considered as a function of $\sum_{j=1}^{m} w_{ij} y_{j,t-1}$ intersects with $y = Wy$ at three fixed points (see figure 2). Substantively, this means that unit i's electoral support for party 1 is subject to positive feedback from its neighbors. To the extent that a majority/minority of the voters in i's neighboring units support party 1, then a majority/minority of i's own voters will also support party 1. Since the neighboring relationship is mutual, the electoral influence is also mutual. As mutual influence goes *ad infinitum*, an equilibrium state may ultimately emerge. Lin and Cohen (2008) show that a pattern of regionalism can emerge out of increasing returns. When an equilibrium is reached, $y_{i,t} = y_{i,t-1}$ and the temporal lag in Equations (5) and (6) is no longer necessary.

In figure 3, we show that Equation (5) can generate the relationships between clustering and the variability of election outcomes observed by Gudgin and Taylor (1979). The figures in figure 3 are the quasi-equilibrium states of a simulated "national constituency" consisting of a grid of 100 × 100 spatial units (i.e., cells). The simulation was based on Equation (5) with $\rho = 1$, $\beta = 0$, $\gamma = 0.1$, $t = 250$, and various values of σ (or ρ/σ, a more appropriate measure of spatial interdependence) as indicated. As expected, increasing spatial interdependence among the units, i.e., decreasing σ or increasing ρ/σ, leads to clustering of units

[14]For the notion of increasing returns, see W. Brian Arthur, "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *The Economic Journal* 99, no. 394 (March 1989): 116-31; and W. Brian Arthur, *Increasing Returns and Path Dependence in the Economy* (Ann Arbor: University of Michigan Press, 1994).

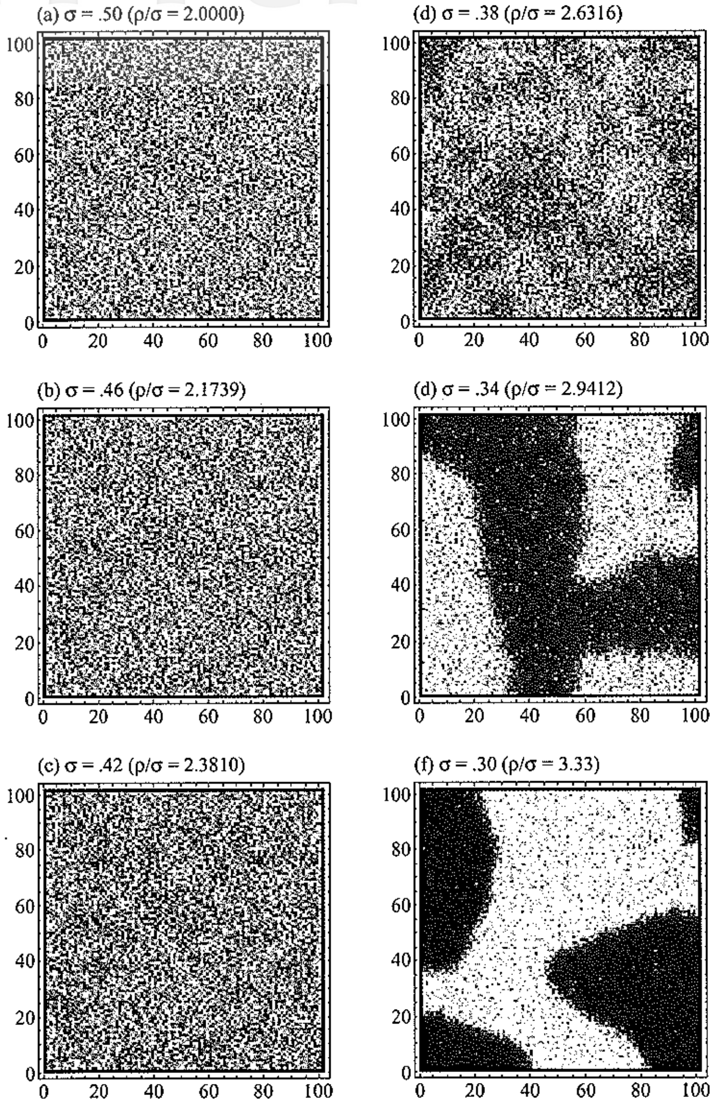
Figure 2
The Neighborhood Response Function



Source: Adapted from figure 1 of Tse-min Lin and Matthew Cohen, "Spatial Regression as a Statistical Model of Regionalism" (Paper presented at the 66th Annual National Conference of the Midwest Political Science Association, Chicago, Illinois, April 2008).

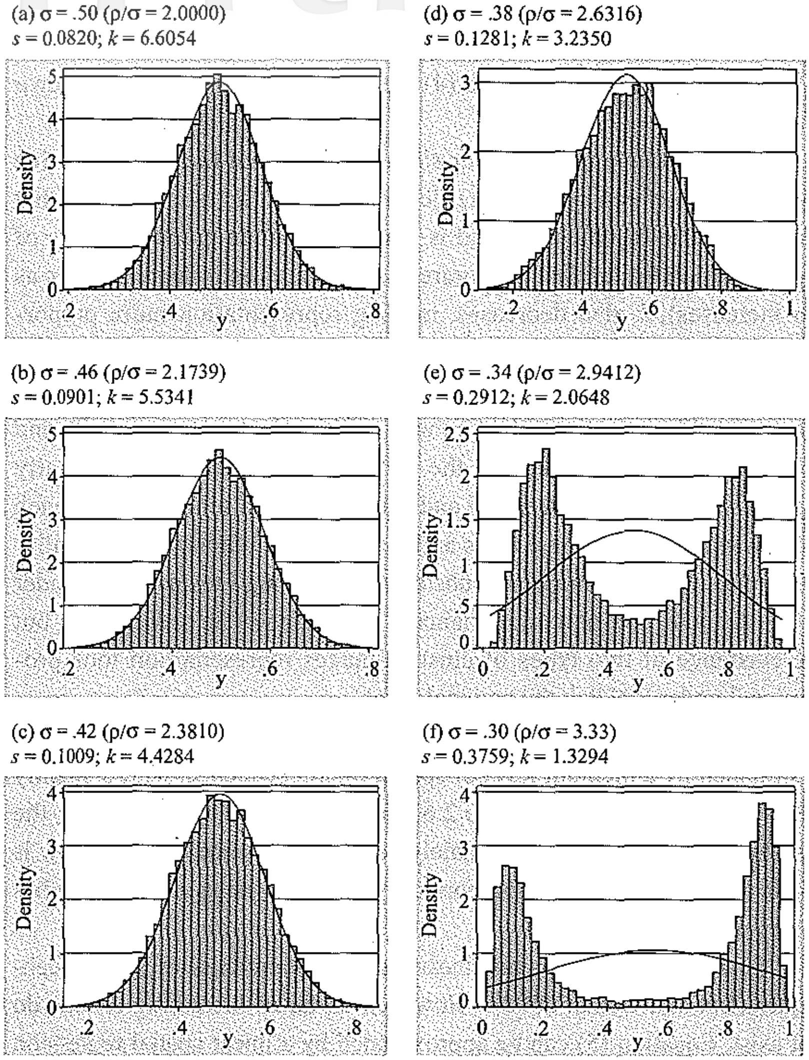
and ultimately to the emergence of a regionalistic pattern that exhibits clear within-region homogeneity and between-region heterogeneity. For each simulation, the distribution of $y_{i,250}$ is shown in figure 4 together with its standard deviation, s, across all districts. Clearly, as σ decreases (or as ρ/σ increases), s, the variability of election outcomes, increases. Furthermore, as σ dips below a certain point, clustering becomes in-

Figure 3
Spatial Autocorrelation, Clustering, and the Emergence of Regionalism



Source: Based on simulations of Equation (5) with $\rho = 1$, $\beta = 0$, $\gamma = 0.1$, $t = 250$, and various values of σ (or ρ/σ). The simulation was carried out on a grid of 100 × 100 cells. A uniform distribution was imposed as the initial condition of each cell at $t = 0$.

Figure 4
Spatial Autocorrelation and Variability of Election Outcomes



Source: Based on simulations of Equation (5) with $\rho = 1$, $\beta = 0$, $\gamma = 0.1$, $t = 250$, and various values of σ (or ρ/σ). The simulation was carried out on a grid of 100 × 100 cells. A uniform distribution was imposed as the initial condition of each cell at $t = 0$.

creasingly obvious, and the normality of the distribution collapses, causing s to increase at a much higher pace. Although our model, which is essentially a Markov model of mutual influence based on spatial contiguity, depicts aggregate rather than individual voting behavior, the patterns shown in figures (3) and (4) are strikingly similar to those shown in figure 1.

An immediate result of our simulations is that there is a positive correlation between spatial interdependence and k, the exponent that links seat ratio to vote ratio and is a measure of the proportionality of representation under FPTP. Figure 4 shows that as σ decreases (or as ρ/σ increases), k also decreases, making representation more proportional. The reason for this is that increasing clustering leads to increasing spatial concentration of the parties' loyal supporters. When the districts are grouped into partisan clusters with within-cluster homogeneity and between-cluster heterogeneity, the national constituency is essentially organized according to Kendall and Stuart's Lexian scheme. With each party winning big in some units and losing big in others, the variability of election outcomes increases substantially, causing k to decrease. Under such a scheme, one-party dominance is the norm within each cluster, but the existence of two types of clusters makes representation look proportional under FPTP at the national level.

The simulations represented in figure 4 also show that the Cube Law is sustained by a medium level of spatial interdependence, with σ slightly below 0.38 or ρ/σ slightly above 2.6316. Note that at this value of σ, the distribution of y is still approximately normal, and hence the Cube Law is applicable. Beyond this point, normality start collapsing as clustering becomes prevalent.

Since Equation (5) is nonlinear, its estimation requires special techniques (Lin and Cohen 2008). However, as far as increasing returns do not exist, Equation (5) can be approximated by a linear spatial regression model

$$y = \rho W y + X\beta + \varepsilon \tag{7}$$

for $0 \leq y_i \leq 1$, which is always true.[15] A linear spatial regression model can be easily estimated with software such as Anselin's Geoda or LeSage's Matlab toolbox. In Appendix 1, we show that, within the framework of the linear regression model (7), it is also true that the variance (and hence standard deviation) of y increases as ρ increases.

Note that if spatial autocorrelation is the carryover of contagion at the individual level, it should be stronger at a lower level, such as villages or townships, than at a higher level, such as legislative districts. Since a lower level unit necessarily has a smaller voting population, our theory, like Taagepera's (1973, 1986), also predicts a smaller k or higher proportionality at a lower level. In a sense, our theory provides a behavioral interpretation for Taagepera's mathematical theory.

We now turn to our empirical analysis of the Taiwan case. Our purpose is to show that the disproportional election outcomes in the country's recent legislative elections are associated with the low degree of spatial interdependence at the district level.

The Taiwan Case: Research Design

The seventh Legislative Yuan elections in January 2008 were the first in Taiwan after the country adopted a single-member district plurality, two-vote system. Under the system, voters cast two votes, one for a district seat and the other for a party list. Overall there are 73 single-member FPTP districts, 2 three-member SNTV districts with a total of 6 seats reserved for Taiwanese aborigines, and one national constituency with 34 at-large seats. The major parties competing for the elections were the Kuomintang (KMT,

---

[15]For the empirical cases that were analyzed and presented below, we estimated Equation (5) using the gprobit (i.e., probit with grouped data) procedure. See William H. Greene, *Econometric Analysis*, fifth edition (Upper Saddle River, N.J.: Prentice Hall, 2003), 686-89; Damodar N. Gujarati, *Basic Econometrics*, fourth edition (Boston: McGraw-Hill and Irwin, 2003), 629; and Lin and Cohen, "Spatial Regression as a Statistical Model of Regionalism," 14-16. We did not find significant increasing return effects. Hence, we resorted to the linear spatial regression model and presented only its estimates in this article.

國民黨) and the Democratic Progressive Party (DPP, 民主進步黨). There were also minor parties and nonpartisan alliances, but all political forces could be fairly accurately divided into the pan-Blue and the pan-Green camps, with the KMT leading the former and the DPP the latter. As is well known, the single dominant political cleavage in Taiwan is national identity.[16] The pan-Blue camp supports a more conciliatory approach toward China, which considers Taiwan as a renegade province. The pan-Green camp advocates ultimate Taiwan independence.[17] The election outcomes are given in table 1. The seat ratio between the two camps is vastly disproportional to the vote ratio.

Since our interest is in investigating the Cube Law, we focus on the single-member district part of the elections. Also, because our theory relates the Cube Law to spatial interdependence among the districts, we exclude the three island counties (Jinmen 金門縣, Lianjiang 連江縣, and Penghu 澎湖縣) which are isolated districts. This leaves us with 70 districts on Taiwan proper. The races in two of these districts (Taipei County's [台北縣] ninth district and the Hsinchu County [新竹縣] district), however, were uncontested by the pan-Green camp. Because the pan-Blue candidates in these two districts won 100 percent vote share as a result, we do not include the two districts in the calculations involving the Cube Law.

---

[16]Tse-min Lin and Yun-han Chu, "The Structure of Taiwan's Political Cleavages toward the 2004 Presidential Election: A Spatial Analysis," *Taiwan Journal of Democracy* 4, no. 2 (December 2008): 133-54.

[17]Before the seventh Legislative Yuan elections, the pan-Blue camp included, in addition to the KMT, the People First Party (PFP, 親民黨), the New Party (NP, 新黨), the Non-Partisan Solidarity Union (無黨團結聯盟), and independents who were former KMT members, and the pan-Green camp included, in addition to the DPP, the Taiwan Solidarity Union (TSU, 台灣團結聯盟) and the Taiwan Independence Party (建國黨). After the change of the electoral system, however, the KMT and the DPP, in anticipation of a two-party system, dropped their alliances with the minor parties within their respective camp. For this reason, in our analysis of the seventh Legislative Yuan elections, our definition of pan-Blue/pan-Green vote share in a district relates only to votes won by the KMT/DPP. In districts where the KMT/DPP did not nominate their own candidates, but instead reached pre-election agreements with other parties to support their candidates, pan-Blue/pan-Green vote share refers to votes won by those minor parties. In our analysis of the sixth Legislative Yuan elections, we include votes won by minor parties in the calculation of vote share for each camp.

Table 1
Taiwan's Seventh Legislative Yuan Election Outcomes

	District seats (%, % popular vote)	At-large seats (% popular vote)	Total seats (% total seats)
KMT	61 (77.22, 53.49)	20 (51.23)	81 (71.68)
DPP	13 (16.46, 38.17)	14 (36.91)	27 (23.89)
Others*	5 (6.33, 8.34)	0 (11.86)	5 (4.42)
Total	79 (100, 100)	34 (100)	113 (100)
Turnout	58.50%	58.28%	

*All the minor parties that won district seats are pan-Blue parties. None of the minor parties exceeded the 5 percent threshold required to win at-large seats.

While our primary objective is to investigate Taiwan's seventh Legislative Yuan elections, we also include the electoral results of the sixth Legislative Yuan elections in our analyses. These elections were held under the old SNTV system. We followed Wu and Lee in using their district-level returns to re-create simulated electoral races between the pan-Blue and pan-Green camps under the single-member district FPTP system of the seventh Legislative Yuan elections.[18]

In addition to district-level election outcomes, we also use the actual outcomes of the seventh and sixth Legislative Yuan elections *at the township level* to simulate elections under FPTP at that level and investigate the relationship between spatial interdependence and the Cube Law for these simulated elections.

For each (actual or simulated) election we investigate, we use the Cube Law as a benchmark to analyze the extent of proportional representation for Taiwan proper, the North, and the South. Taiwanese politics is known to exhibit some degree of regionalism, with the North relatively pro-Blue and the South pro-Green. For each region, we calculate a battery of statistics to assess the applicability of the Cube Law and the degree of

[18]Chin-En Wu and Feng-yu Lee, "Electoral Systems and the Moderation of Party Positions on Ethnicity," *Zhengzhi xuebao* (Chinese Political Science Review) (Taipei), no. 43 (June 2007): 71-99.

spatial interdependence. These include Blue vote share, Blue seat share, predicted Blue seat share under the Cube Law, seats exceeding the prediction of the Cube Law, k, k', the mean and standard deviation of Blue vote share across districts, Kolmogorov-Smirnov snd skewness-kurtosis test statistics for normality tests, and the spatial autocorrelation statistic Moran's I.[19]

For each (actual or simulated) election we investigate, we also run a full linear spatial regression model for Taiwan proper. The independent variables of these models include a spatial lag term as well as male population, average age, median income, college-educated population, ethnic population, region, and public opinion on unification/independence as control variables. The purpose of this analysis is to see if the spatial interdependence, if any, indicated by Moran's I can be explained by the control variables.

Following our theoretical results, and the fact that the election outcomes are vastly disproportional, our expectation is that Taiwan's legislative districts should exhibit relatively low spatial autocorrelation. However, we do expect that spatial autocorrelation will be higher at the township level. The reason is that spatial autocorrelation originates from contagious individual behavior, and hence the carried over effect should be higher at a lower level of spatial units. Furthermore, most neighboring townships belong to the same district and are, therefore, subject to the same campaign dynamics, which entail higher contagion and spatial autocorrelation.

---

[19]In spatial econometrics, Moran's I is a measure of spatial autocorrelation. Suppose y is a vector of observations y_i in deviation form and W is the row-standardized spatial weight matrix, then Wy is a vector of $\sum_j W_{ij} y_j$, which is the average y value of all neighbors j of spatial unit i. Moran's I is formally defined as

$$I = \frac{y'Wy}{y'y}$$

Since the elements of y are deviations, I is equivalent to the slope coefficient in a regression of Wy on y. See Luc Anselin, "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association" (Paper presented at the GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, The Netherlands, December 1-5, 1993). West Virginia University, Regional Research Institute, Research Paper 9330.

Demographic data were compiled from the 2000 Population and Housing Census conducted by the Directorate-General of Budget, Accounting, and Statistics (DGBAS) of the Executive Yuan (行政院主計處). Income data were compiled from the *Special Volume of the 2002 General Income Tax Return Filing and Appraising Statistics* published by the Financial Data Center of the Ministry of Finance. Ethnic population data were compiled from the 2004 National Hakka Population Survey conducted by the Council for Hakka Affairs of the Executive Yuan. Election returns were compiled from data available on the website of Taiwan's Central Election Commission.

District-level public opinion on unification/independence was estimated with a Bayesian multilevel (or hierarchical) model based on both individual-level survey data and aggregate-level demographic data. The variable is a 5-point scale measuring opinion on unification and independence.[20] The survey we used to compile individual-level opinion data was Taiwan's Election and Democratization Studies, 2005-2008 (III): The Legislative Elections (TEDS 2008L). TEDS 2008L includes a pre-election telephone interview component (TEDS 2008L-T) and a post-election personal interview component (TEDS 2008L-C). Since the question about unification/independence was asked in both components, we combine them to get more observations. As a result of the relatively large total N, all the districts included in our analysis are represented in our estimation sample.[21] The Bayesian approach provides better estimates for public opinion at the subnational level because it compensates for the small subsample within each subnational unit by incorporating more information, including population information. For technical details on how we implement the BHM methodology, see Appendix 2.

[20]As we recoded it, the scale is 1 = Seek unification as soon as possible; 2 = Maintain the status quo now and seek unification later; 3 = Maintain the status quo forever or maintain the status quo now but decide on unification or independence later; 4 = Maintain the status quo now and declare independence later; and 5 = Declare independence as soon as possible.

[21]For TEDS 2008L-T, N = 3,843. For TEDS 2008L-C, N = 1,238 for the independent sample and N=1,381 for the panel. Combining all these components, our valid N = 3,595, included 2,035 from telephone interviews and 1,560 from personal interviews.

Empirical Results and Discussion

We start our analysis at the township level to investigate the extent of clustering below legislative districts. Lin, Wu, and Lee have shown that Taiwanese township residents are subject to mutual influence in the formation of their national identity, which was a dominant factor in the 2008 elections.[22] Lay, Yap, and Chang and Lay, Chen, and Yap have shown the existence of spatial autocorrelation in Taiwan's presidential elections at, respectively, the village and township levels.[23] We seek to find out whether interdependence at the individual level spread to the township level in Legislative Yuan elections and, if it did, to what extent.

We analyze election returns at the township level not only for the seventh Legislative Yuan elections but also for the sixth Legislative Yuan elections. We compile a pan-Blue vote share for each township and treat the townships as if they were single-member districts under the FPTP system. For Taiwan proper, the North, and the South, we investigate the relationship between "seats" and votes using the Cube Law as a benchmark. We also compute Moran's I's as a measure of spatial autocorrelation. The results are shown in tables 2 and 3.

As mentioned earlier, we expect spatial autocorrelation at the township level to be significant because townships are relatively small and neighboring townships belonging to the same districts are subject to the same campaign dynamics. Our results demonstrate that this is indeed the case. The Moran's I's are of moderate values for all cases. For the seventh Legislative Yuan elections, they are 0.6321 for the whole of Taiwan proper, 0.6377 for the North, and 0.3678 for the South. For the sixth Legislative

---

[22]Tse-min Lin, Chin-En Wu, and Feng-yu Lee, "Neighborhood Influence on the Formation of National Identity in Taiwan: Spatial Regression with Disjoint Neighborhoods," *Political Research Quarterly* 59, no. 1 (March 2006): 35-46.

[23]Jinn-guey Lay, Ko-hua Yap, and Chy-chang Chang, "Spatial Perspectives and Analysis on Voting Behavior: A Case Study of the 2004 Taiwan Presidential Election," *Xuanju yanjiu* (Journal of Electoral Studies) 14, no. 1 (May 2007): 33-60; and Jinn-guey Lay, Yu-wen Chen, and Ko-hua Yap, "Spatial Variation of the DPP's Expansion: Between Taiwan's Presidential Elections," *Issues & Studies* 42, no. 4 (December 2006): 1-22.

Table 2
Taiwan's Seventh Legislative Yuan Elections and the Cube Law, Simulated Township-Level Elections (2008)

	Taiwan proper (N = 335)	North (N = 161)	South (N = 147)
Blue vote share#	0.5772	0.6104	0.5086
Blue seat share [seats]	0.7463 [250]	0.8944 [144]	0.5442 [80]
Predicted Blue seat share under cube law [seats]	0.7180 [240.5201]	0.7937 [127.7859]	0.5257 [77.2836]
Seats exceeding cube law	9.4799	16.2141	2.7164
$k = \ln(\frac{S}{1-S}) \Big/ \ln(\frac{V}{1-V})$	3.4636	4.7572	5.1627
$k' = \ln(V_{total})/\ln(S_{total})$	2.7610	3.0750	2.9955
Mean of Blue vote share	0.5775	0.6136	0.5193
Standard deviation of Blue vote share	0.1198	0.1016	0.1141
Kolmogorov-Smirnov test for normality	p = 0.5620	p = 0.5160	p = 0.1860
Skewness and kurtosis test for normality	p = 0.0376*	p = 0.0183*	p = 0.0001***
Moran's I	0.6321	0.6377	0.3678

*p<0.05; **p<0.01; ***p<0.001

#Based on actual township-level election returns, this simulation considers each township as a "district" that elects a candidate under the FPTP system. Fourteen northern townships were not included in the calculations because there were no pan-Green candidates. The only exception is Moran's I for which the calculation cannot have missing values.

Yuan elections, they are 0.5395 for the whole of Taiwan proper, 0.4168 for the North, and 0.4377 for the South.

Concerning the relationship between seats and votes, it appears that representation at the township level is less proportional than prescribed by the Cube Law, although the benchmark is more farfetched for the two regions than for the whole of Taiwan proper. For the seventh Legislative Yuan elections, *k* is 4.7572 for the North, 5.1627 for the South, but 3.4636

Table 3
Taiwan's Sixth Legislative Yuan Elections and the Cube Law, Simulated Township-Level Elections (2004)

	Taiwan proper (N = 349)	North (N = 175)	South (N = 147)
Blue vote share[#]	0.5363	0.5670	0.4682
Blue seat share [seats]	0.6390 [223]	0.7943 [139]	0.3946 [58]
Predicted Blue seat share under cube law [seats]	0.6074 [211.9826]	0.6920 [121.1000]	0.4055 [59.6085]
Seats exceeding cube law	11.0174	17.9000	-1.6085
$k = \ln(\frac{S}{1-S}) \Big/ \ln(\frac{V}{1-V})$	3.9230	5.0078	3.3591
$k' = \ln(V_{total})/\ln(S_{total})$	2.7441	3.0297	2.9950
Mean of Blue vote share	0.5507	0.5746	0.4990
Standard deviation of Blue vote share	0.1144	0.0931	0.1132
Kolmogorov-Smirnov test for normality	p = 0.5640	p = 0.7810	p = 0.0430*
Skewness and kurtosis test for normality	p = 0.0065**	p = 0.3629	p = 0.0005***
Moran's I	0.5395	0.4168	0.4377

*p<0.05; **p<0.01; ***p<0.001

[#]Based on actual township-level election returns, this simulation considers each township as a "district" that elects a candidate under the FPTP system.

for the whole of Taiwan proper. For the sixth elections, it is 5.0078 for the North, 3.3591 for the South, and 3.9230 for Taiwan. Not surprisingly, the standard deviations are all smaller than $s_0 = 0.137$. In fact, there are indications that the pan-Blue vote share may not be normally distributed in some of the areas or regions. Although the Kolmogorov-Smirnov test cannot reject normality for all but one region (the South in the sixth elections), the more powerful skewness-kurtosis test rejects normality for all but one region (the North in the sixth elections). The distributions of vote share clearly cannot sustain the Cube Law. Interestingly, though, Taagepera's generalized Cube Law holds well at this level. In fact, the k's for both regions in both elections are almost exactly 3.

Despite the fact that moderate spatial autocorrelation fails to generate enough variability in vote share, it does generate a certain degree of clustering in the support for the two political camps. The pan-Blue completely dominated in the North in both elections ($V = 0.6104$ with $S = 0.8944$ in the seventh and $V = 0.5670$ with $S = 0.7943$ in the sixth). The pan-Green dominated the South in the sixth elections ($V = 0.5318$ with $S = 0.6054$), although it lost the dominance there in the seventh ($V = 0.4914$ with $S = 0.4558$). This clustering of electoral support is consistent with the spatial distribution of ethnic groups in Taiwan, with the pro-Blue mainlander and Hakka populations relatively concentrated in the North.

If clustering at the subdistrict level is generated by spatial autocorrelation, however, there appears to be more dynamics than ethnic politics. The results of a fully specified spatial regression of pan-Blue vote share at the township level for the seventh elections reveal that, controlling for demographic variables, ethnic populations, regions, and public opinion on unification/independence, the spatial lag terms remain statistically highly significant (see table 4). This means that the control variables cannot completely explain the spatial autocorrelation indicated by Moran's I's. Mutual influence at the township level goes beyond ethnic politics and national identity at the township level. The story, however, is different at the district level to which we now turn.

The seventh Legislative Yuan elections in Taiwan proper and especially in the North did indeed lead to vastly disproportional representation, with k as large as 4.6954 for the whole of Taiwan proper and 6.8854 for the North (see table 5). Using the Cube Law as a benchmark, these translate to 6-7 seats above what would be considered "normal" under the FPTP system. Surprisingly, representation in the South is quite proportional, with a 0.5087 pan-Blue vote share resulting in a 50 percent seat share. As explained in a footnote of table 2, the calculated value $k = 0$ is a mathematical artifact and not very meaningful.

Table 4
Spatial Regression Models of "Blue" Vote at the Township Level, the Seventh Legislative Yuan Elections (2008)

Dependent variable: Blue parties vote share

Independent Variables	Model 1	Model 2	Model 3
Constant	0.5875***	0.6586**	0.6535**
	(0.1554)	(0.2282)	(0.2138)
Male	-0.3957+	-0.2709	-0.3942+
	(0.2233)	(0.2369)	(0.2234)
Age	-0.0027	-0.0043*	-0.0026
	(0.0018)	(0.0019)	(0.0018)
Income	0.0000	0.0003*	0.0000
	(0.0001)	(0.0001)	(0.0001)
College education	-0.1460	-0.2927*	-0.1500
	(0.1159)	(0.1221)	(0.1163)
Taiwanese	-0.1354***		-0.1324***
	(0.0236)		(0.0250)
Mainlander	0.3083***		0.3096***
	(0.0644)		(0.0645)
District-level unification/independence		-0.1113*	-0.0218
		(0.0495)	(0.0496)
South	-0.0125	-0.0028	-0.0104
	(0.0105)	(0.0125)	(0.0118)
East	-0.0363+	0.0176	-0.0374+
	(0.0199)	(0.0194)	(0.0201)
Spatial lag ($\hat{\rho}$)	0.6602***	0.7809***	0.6565***
	(0.0409)	(0.0359)	(0.0413)
N	349	349	349
R-squared	0.6817	0.6315	0.6815
SER	0.0810	0.0871	0.0810

+p<0.10; *p<0.05; **p<0.01; ***p<0.001; two-tailed tests.

Table 5
Taiwan's Seventh Legislative Yuan Elections and the Cube Law (2008)

	Taiwan proper (N = 68)	North (N = 44)	South (N = 22)
Blue vote share[1]	0.5762	0.6088	0.5087
Blue seat share [seats]	0.8088 [55]	0.9545 [42]	0.5000 [11]
Predicted Blue seat share under cube law [seats]	0.7154 [48.6447]	0.7903 [34.7716]	0.5262 [11.5764]
Seats exceeding cube law	6.3553	7.2284	-0.5764
$k = \ln(\frac{S}{1-S}) \Big/ \ln(\frac{V}{1-V})$	4.6954	6.8854	0.0000[2]
$k' = \ln(V_{total})/\ln(S_{total})$	3.8037	4.1276	4.8367
Mean of Blue vote share	0.5782	0.6092	0.5089
Standard deviation of Blue vote share	0.0797	0.0693	0.0511
Kolmogorov-Smirnov test for normality	p = 0.9960	p = 0.8630	p = 0.8450
Skewness and kurtosis test for normality	p = 0.0981	p = 0.0134*	p = 0.4514
Moran's I	0.3877	0.1782	0.0337

*p<0.05; **p<0.01; ***p<0.001

1. Two northern districts, District 9 of Taipei County and the Hsinchu County District, were not included in the calculations. The only exception is Moran's I for which the calculation cannot have missing values.

2. For the South, the exponent k in $\left(\frac{S}{1-S}\right)=\left(\frac{V}{1-V}\right)^k$ is artificially low because S = 0.5, and, therefore, $\frac{S}{1-S}=1$, rendering $k = 0$.

Simulated as under FPTP, the results of the sixth Legislative Yuan elections are only slightly worse than that indicated by the Cube Law in the whole of Taiwan proper ($k = 3.5419$), but k is fairly large in the North (k = 6.3437) and shoots to an incredibly large magnitude in the South (k = 12.4212 . These results are reported in table 6.

Table 6
Taiwan's Sixth Legislative Yuan Elections and the Cube Law (2004)

	Taiwan proper (N = 70)	North (N = 46)	South (N = 22)
Blue vote share[1]	0.5328	0.5673	0.4538
Blue seat share [seats]	0.6143 [43]	0.8478 [39]	0.0909 [2]
Predicted Blue seat share under cube law [seats]	0.5973 [41.8110]	0.6926 [31.8596]	0.3644 [8.0168]
Seats exceeding cube law	1.1890	7.1404	-6.0168
$k = \ln(\frac{S}{1-S}) \Big/ \ln(\frac{V}{1-V})$	3.5419	6.3437	12.4212
$k' = \ln(V_{total}) / \ln(S_{total})$	3.7841	4.0906	4.8359
Mean of Blue vote share	0.5327	0.5664	0.4513
Standard deviation of Blue vote share	0.0813	0.0639	0.0450
Kolmogorov-Smirnov test for normality	p = 0.7800	p = 0.6300	p = 0.6230
Skewness and kurtosis test for normality	p = 0.5061	p = 0.5633	p = 0.2936
Moran's I	0.5206	0.2936	0.2728

*p<0.05; **p<0.01; ***p<0.001

1. The sixth Legislative Yuan elections, held in 2004, were the last elections under SNTV. We follow Wu and Lee in simulating the elections as under FPTP based on actual election returns. For detail see Chin-En Wu and Feng-yu Lee, "Electoral Systems and the Moderation of Party Positions on Ethnicity," *Zhengzhi xuebao* (Chinese Political Science Review) (Taipei), no. 43 (June 2007): 71-99.

The standard deviations of pan-Blue vote share across districts are all very low compared with the required value, $s_0 = 0.137$, for the Cube Law. In fact, they are all significantly smaller than their counterparts at the township level. It is thus not surprising that the Cube Law was not sustained, even though the Kolmogorov-Smirnov test and the skewness-kurtosis test cannot reject normality in all cases but one (the North, seventh elections). We can in general conclude that under-dispersion in vote share has caused worse disproportion in representation than the already disproportional representation marked by the Cube Law.[24] Note that, as predicted by Taagepera's theory, his k's are all significantly greater at the district level than at the township level.

The spatial concentration that we observed at the township level not only carries over but is pronounced at the district level. In the seventh elections, the pan-Blue completely dominated in the North (a vote share of 0.6088 with a seat share of 0.9545), and there was a balance of electoral powers in the South (evenly divided vote shares and seat shares). Furthermore, the fact that the standard deviation for the whole of Taiwan proper (0.0797) is greater than those for the two regions (0.0693 and 0.0511) reflects a slight within-region homogeneity and between-region heterogeneity. The phenomenon is even more conspicuous in our analysis of the simulated sixth elections. In these elections, the pan-Blue and the pan-Green both showed a clear spatial concentration, with the former commanding a vote share of 0.5673 (with a seat share of 0.8478) in the North and the latter a vote share of 0.5462 (with a seat share of 0.9091) in the South. Between-region heterogeneity and within-region homogeneity was also clear in terms of the relative magnitude of the standard deviations: 0.0813 for the whole of Taiwan proper and 0.0639 and 0.0450, respectively, for the North and the South. Overall there was the kind of large-size "clustering" that Gudgin and Taylor (1979) refer to, the kind that is supposed to produce a standard deviation so large that would destroy the normality of the distribution in the national constituency (see figure 1a). However, the standard deviations are far smaller than what is required by the Cube Law.

We argue that the spatial concentration exhibited in the sixth and seventh elections is a result of Taiwan's ethnic demographics and politics, not of contagious political behavior. The Moran's I's, like the standard deviations, do reflect spatial heterogeneity: the I's for Taiwan proper

[24]The reason that the Cube Law works in the South in the seventh elections despite a small standard deviation is because , the pan-Blue vote share, is so close to 0.5 that $x = V - 0.5 \approx 0$. According to Equation (4), the bias against the Cube Law due to under- or over-dispersion would be negligible.

are greater than those for the regions (see tables 5 and 6). However, these figures are small compared with their moderate counterparts at the township level. This result is consistent with our conjecture that mutual influence should be greater at the township level than at the district level. It is also consistent with our theoretical argument that under-dispersion can be explained by the lack of spatial autocorrelation.

To demonstrate the validity of our argument, we estimate the fully specified spatial regression model for pan-Blue vote share for the seventh elections. The results, shown in table 7, are strikingly different from those pertaining to the township level. After controlling for demographic variables, ethnic population, region, and public opinion on unification/independence, the spatial lag term is now statistically insignificant. This means that the control variables completely explain the small spatial autocorrelation indicated by Moran's I, and no other form of contagion can be detected at this level.

In sum, our empirical analysis of the Taiwan case has the following findings. First, spatial autocorrelation indeed affects the variation in vote share. Specifically, higher standard deviations at the township level are associated with higher spatial autocorrelations, while lower standard deviations at the district level are associated with lower spatial autocorrelations. Second, spatial autocorrelation, presumably carried over from individual-level contagion, is correlated with the size of constituencies. Specifically, spatial autocorrelation is stronger at the township level where the size of constituencies is smaller, and it is weaker—to the extent of being nonexistent—at the district level where the size is larger. Third, at the township level, spatial clustering exists independent of the effects of ethnicity, national identity, and regional heterogeneity, but this is not true at the district level. The fact that spatial interdependence did not carry over from lower to higher spatial units suggests that mutual influence in voting behavior is not strong; it is certainly weaker than ethnic and identity preferences at the district level. Given the relationship between spatial autocorrelation and standard deviation, weak mutual influence may explain the disproportionality of representation in Taiwan's first legislative elections under FPTP. Finally, even though Taiwanese politics exhibits a certain degree of elec-

Table 7
Spatial Regression Models of "Blue" Vote at the District Level, the Seventh Legislative Yuan Elections (2008)

Dependent variable: Blue parties vote share

Independent Variables	Model 1	Model 2	Model 3
Constant	1.0851*	1.2907*	1.4888*
	(0.5434)	(0.6159)	(0.5847)
Male	-0.8695	0.0754	-0.7418
	(0.9466)	(0.9590)	(0.9305)
Age	0.0015	0.0029	0.0028
	(0.0036)	(0.0040)	(0.0036)
Income	0.0002+	0.0004**	0.0003*
	(0.0001)	(0.0001)	(0.0001)
College education	-0.8766**	-0.7223*	-0.8819**
	(0.3006)	(0.3174)	(0.2950)
Taiwanese	-0.2329***		-0.2047**
	(0.0623)		(0.0643)
Mainlander	0.5888*		0.4605+
	(0.2465)		(0.2539)
Unification/independence		-0.3351***	-0.1679+
		(0.0974)	(0.1002)
South	-0.0755**	-0.0581*	-0.0653**
	(0.0246)	(0.0274)	(0.0253)
East	-0.0333	0.0135	-0.0356
	(0.0579)	(0.0595)	(0.0572)
Spatial lag ($\hat{\rho}$)	0.0558	0.1742	0.0198
	(0.1534)	(0.1558)	(0.1535)
N	70	70	70
R-squared	0.5416	0.4638	0.5588
SER	0.0711	0.0769	0.0697

+p<0.10; *p<0.05; **p<0.01; ***p<0.001; two-tailed tests.

toral regionalism, the North versus South is more a result of spatial heterogeneity associated with the spatial concentration of ethnic populations and national identity rather than a result of spatial interdependence.

Conclusion

It is commonly known that barring the geographic concentration of political parties, FPTP leads to a two-party system in which the minority party is underrepresented. The Cube Law provides a benchmark for the proportionality of representation under FPTP. The traditional Cube Law literature addresses the condition concerning geographic concentration by associating proportionality with the dispersion of the distribution of vote share across districts. In this article, we have related the proportionality of representation to a behavioral factor, namely, the spatial interdependence of constituencies which may have been carried over from mutual influence at the individual level. Since spatial interdependence is an empirical condition that can change from country to country and from time to time, our work provides an approach in which the applicability of the Cube Law can be more meaningfully assessed. Our study of Taiwan's recent legislative elections illustrates the usefulness of our generalization. Even the Cube Law fails to predict the vastly disproportional election outcomes, which we associate with the low spatial autocorrelation at the district level. The Cube Law does not have the status of a scientific law, but to understand why it does or does not work helps us understand the proportionality of representation as an important issue in democratic elections.

Appendix 1
Spatial Autocorrelation and Variance

$$y = \rho Wy + X\beta + \varepsilon \qquad |\rho| < 1$$

$$Var(\varepsilon) = \Sigma = [\sigma_{hk}]_{NxN} = diag(\sigma_1^2, \sigma_2^2, \ldots \sigma_N^2)$$

$$(I - \rho W)y = X\beta + \varepsilon$$

$$y = (I - \rho W)^{-1}(X\beta + \varepsilon)$$

$$y = (I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots)(X\beta + \varepsilon)$$

$$y = X\beta + \rho(WX)\beta + \rho^2(W^2X)\beta + \rho^3(W^3X)\beta + \ldots + \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + \rho^3 W^3 \varepsilon + \ldots$$

$$\begin{aligned} &Var(y \mid X) \\ &= Var(\varepsilon + \rho W\varepsilon + \rho^2 W^2 \varepsilon + \rho^3 W^3 \varepsilon + \ldots) \\ &= Var(\varepsilon) + Var(\rho W \varepsilon) + Var(\rho^2 W^2 \varepsilon) + \ldots \\ &= Var(\varepsilon) + \rho^2 Var(W\varepsilon) + \rho^4 Var(W^2 \varepsilon) + \ldots \\ &= \Sigma + \rho^2 W \Sigma W' + \rho^4 W^2 \Sigma (W^2)' + \ldots \end{aligned}$$

(Note: $Var(W\varepsilon) = E[W\varepsilon(W\varepsilon)'] = E(W\varepsilon\varepsilon' W') = WE(\varepsilon\varepsilon')W' = W\Sigma W'$)

Let $S = W\Sigma W'$ and let n_i and n_j, respectively, be the number of neighbors of spatial units i and j. Also, let $N(i)$ denote the set of unit i's neighbors. For simplicity here we assume that $N(i) \neq \phi$ (i.e., $n_i \neq 0$) and $N(j) \neq \phi$ (i.e., $n_j \neq 0$). Recall that $w_{ih} = \frac{1}{n_i}$ if and only if $h \in N(i)$; otherwise, $w_{ih} = 0$. Similarly, $w_{jh} = \frac{1}{n_j}$ if and only if $h \in N(j)$; otherwise, $w_{jh} = 0$.

$$S_{ij} = \sum_{h=1}^{N}\sum_{k=1}^{N} w_{ih}\sigma_{hk}w_{jk} = \sum_{h=1}^{N} w_{ih}\sigma_h^2 w_{jh} = \sum_{h=1}^{N} w_{ih}w_{jh}\sigma_h^2 = \frac{\sum_{h \in N(i) \cap N(j)} \sigma_h^2}{n_i n_j},$$

where the last summation is run through all common neighbors of i and j.

Therefore, the i^{th} diagonal element of $S = W\Sigma W'$ is

$$S_{ii} = \frac{\sum_{h \in N(i)} \sigma_h^2}{n_i n_i} = \frac{\overline{\sigma}_{N(i)}^2}{n_i}$$

where $\overline{\sigma}_{N(i)}^2$ is the average variance of i's neighbors.

Therefore, if we ignore higher-order terms, the variance of the i^{th} spatial unit is

$$Var(y_i \mid X) \approx \sigma_i^2 + \rho^2 (\frac{\overline{\sigma}_{N(i)}^2}{n_i})$$

If $\sigma_i^2 = \sigma^2$, a constant, for all i, this can be further reduced to

$$Var(y_i \mid X) \approx (1 + \frac{\rho^2}{n_i})\sigma^2 > \sigma^2$$

Thus, as far as N(i) $\neq \phi$ (i.e., $n_i \neq 0$) and $\rho > 0$, $Var(y_i \mid X)$ is greater than σ^2.

Now consider the matrix $Z = W^2\Sigma(W^2)' = WW\Sigma W'W' = WSW'$.

$$Z_{ij} = \sum_{p=1}^{N}\sum_{q=1}^{N} w_{ip} S_{pq} w_{jq} = \frac{1}{n_i n_j} \sum_{p \in N(i)} \sum_{q \in N(j)} S_{pq}$$

where the summation $\sum_{p \in N(i)}$ is over unit i's neighbors (indexed as p), and the summation $\sum_{q \in N(j)}$ is over unit j's neighbors (indexed as q). Therefore,

$$Z_{ii} = \frac{1}{n_i^2} \sum_{p \in N(i)} \sum_{q \in N(i)} S_{pq} .$$

The double summation involves n_i^2 terms. These terms can be divided into two parts: the first part consists of terms associated with $p = q$ (i.e., the same neighbor of i), while the second part consists of $p \neq q$.

$$Z_{ii} = \frac{1}{n_i^2} [\sum_{p \in N(i)} S_{pp} + \sum_{p \in N(i)} \sum_{q \in N(i) - \{p\}} S_{pq}]$$

$$= \frac{1}{n_i^2} [\sum_{p \in N(i)} \frac{\overline{\sigma}_{N(p)}^2}{n_p} + \sum_{p \in N(i)} \sum_{q \in N(i) - \{p\}} \frac{\sum_{k \in N(p) \cap N(q)} \sigma_k^2}{n_p n_q}]$$

where k indexes common neighbors of p and q.

This expression can be simplified by some assumptions:

(1) Suppose all i's eighbors have the same number of neighbors, n_i, as i does, and that they don't have any common neighbors except i. Also suppose that $\sigma_i^2 = \sigma^2$, a constant, for all i. In this case,

$$Z_{ii} = \frac{1}{n_i^2}[\frac{1}{n_i}\sum_{p\in N(i)}\sigma^2 + \frac{1}{n_i^2}\sum_{p\in N(i)}\sum_{q\in N(i)-\{p\}}\sigma^2] = \frac{1}{n_i^2}[\sigma^2 + \frac{n_i(n_i-1)\sigma^2}{n_i^2}] = \frac{(2n_i-1)\sigma^2}{n_i^3}$$

(2) Suppose that i and all its neighbors form a disjoint neighborhood. That is, they are all mutual neighbors and they do not have other neighbors outside this group. Thus, each unit has n_i neighbors, and any two units have $n_i - 1$ common neighbors. Also suppose that $\sigma_i^2 = \sigma^2$, a constant, for all i. In this case,

$$Z_{ii} = \frac{1}{n_i^2}[\frac{1}{n_i}\sum_{p\in N(i)}\sigma^2 + \frac{1}{n_i^2}\sum_{p\in N(i)}\sum_{q\in N(i)-\{p\}}(n_i-1)\sigma^2]$$
$$= \frac{1}{n_i^2}[\sigma^2 + \frac{n_i(n_i-1)^2\sigma^2}{n_i^2}] = \frac{(n_i^2-n_i+1)\sigma^2}{n_i^3}$$

Under assumption (1) and ignoring higher-order terms,

$$Var(y_i\,|\,X) \approx (1+\frac{\rho^2}{n_i})\sigma^2 + \frac{\rho^4(2n_i-1)\sigma^2}{n_i^3} = [1+\frac{\rho^2}{n_i}+\frac{(2n_i-1)\rho^4}{n_i^3}]\sigma^2$$

For example, if $n_i = 4$ and $\rho = .8$,

$$Var(y_i\,|\,X) \approx [1+\frac{\rho^2}{n_i}+\frac{(2n_i-1)\rho^4}{n_i^3}]\sigma^2 = [1+.1600+.0448]\sigma^2 = 1.2048\sigma^2$$

Under assumption (2) and ignoring higher-order terms,

$$Var(y_i\,|\,X) \approx (1+\frac{\rho^2}{n_i})\sigma^2 + \frac{\rho^4(n_i^2-n_i+1)\sigma^2}{n_i^3} = [1+\frac{\rho^2}{n_i}+\frac{(n_i^2-n_i+1)\rho^4}{n_i^3}]\sigma^2$$

For example, if $n_i = 4$ and $\rho = .8$,

$$Var(y_i\,|\,X) \approx [1+\frac{\rho^2}{n_i}+\frac{(n_i^2-n_i+1)\rho^4}{n_i^3}]\sigma^2 = [1+.1600+.0832]\sigma^2 = 1.2432\sigma^2$$

Appendix 2
Bayesian Multilevel Estimation with Poststratification: District-Level Estimates of Public Opinions on Unification/Independence

An opinion survey is often targeted at a national population. The sample of a national survey is a representative sample of the national population but the subsample associated with each subnational unit may not be representative of that unit's population. Even if it is, it is often too small to be of practical use. To estimate public opinions at a subnational level such as counties or legislative districts, subsamples from a national survey are therefore often inadequate or insufficient to provide good estimates. The Bayesian multilevel or hierarchical model overcomes the difficulty by supplementing subsamples with information from other units or other levels, including the population. The procedure first requires the specification of a multilevel model for the individual response variable, y, often by relating it to some demographic variables. The model is then estimated with a Bayesian approach which treats the coefficients of the model as random variables following certain prior distributions. Once estimated, population-level demographic data are plugged in to the model to produce estimates of y for all units at the desired level.*

Our model for the 5-point scale unification/independence variable is

$$y_i = \beta_0 + \beta_{district(i)} + \beta_{male} male_i + \beta_{age} age_i + \beta_{college} college_i + \beta_{Taiwanese} Taiwanese_i \\ + \beta_{Hakka} Hakka_i + \beta_{Mainlander} Mainlander_i + \beta_{previous\_vote} previous\_vote_i + \varepsilon_i$$

with the district-specific intercept, $\beta_{district(i)}$, specified as a function of the county in which the district is located:

$$\beta_{district(i)} = \beta_j \sim N(\beta_{county(j)}, \sigma_j^2)$$

We assigned normal distributions to all other coefficients.

Thus specified, the model was fit to TEDS 2008L survey data with WinBUGS as called from R using Gelman's Bugs.R, which implemented the Markov Chain Monte Carlo (MCMC) method to estimate the model. Once estimated, district-level demographic and electoral data were plugged in to derive $\hat{y}$ for each district.

---

*For more information about this approach, see Bruce Western, "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach," *American Journal of Political Science* 42, no. 4 (October 1998): 1233-59; David K. Park, Andrew Gelman, and Joseph Bafumi, "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls," *Political Analysis* 12, no. 4 (Autumn 2004): 375-85; Andrew Gelman, "Multilevel (Hierarchical) Modeling: What It Can and Can't Do," *Technometrics* 48, no. 3 (August 2006): 432-35; and Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge and New York: Cambridge University Press, 2007).

BIBLIOGRAPHY

Anselin, Luc. 1993. "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association." Paper presented at the GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, The Netherlands, December 1-5, 1993. West Virginia University, Regional Research Institute, Research Paper 9330.

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal* 99, no. 394 (March): 116-31.

______. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.

Coleman, James S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.

Gelman, Andrew. 2006. "Multilevel (Hierarchical) Modeling: What It Can and Can't Do." *Technometrics* 48, no. 3 (August): 432-35.

______, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge and New York: Cambridge University Press.

Greene, William H. 2003. *Econometric Analysis*, fifth edition. Upper Saddle River, N.J.: Prentice Hall.

Gudgin, Graham, and Peter J. Taylor. 1979. *Seats, Votes, and the Spatial Organization of Elections*. London: Pion.

Gujarati, Damodar N. 2003. *Basic Econometrics*, fourth edition. Boston: McGraw-Hill and Irwin.

Kendall, M. G., and A. Stuart. 1950. "The Law of the Cubic Proportion in Election Results." *British Journal of Sociology* 1, no. 3 (September): 183-97.

Kuper, Adam, and Jessica Kuper, eds. 1996. *Social Science Encyclopedia*, second edition London: Routledge.

Lay, Jinn-guey (賴進貴), Yu-wen Chen (陳玉文), and Ko-hua Yap (葉高華). 2006. "Spatial Variation of the DPP's Expansion: Between Taiwan's Presidential Elections." *Issues & Studies* 42, no. 4 (December): 1-22.

Lay, Jinn-guey (賴進貴), Ko-hua Yap (葉高華), and Chy-chang Chang (張智昌). 2007. "Spatial Perspectives and Analysis on Voting Behavior: A Case Study of the 2004 Taiwan Presidential Election" (投票行為之空間觀點與空間分

析方法: 以臺灣 2004 年總統選舉爲例). *Xuanju yanjiu* (選舉研究, Journal of Electoral Studies) (Taipei) 14, no. 1 (May): 33-60.

Lin, Tse-min (林澤民), Chin-En Wu (吳親恩), and Ferng-yu Lee (李鳳玉). 2006. "Neighborhood Influence on the Formation of National Identity in Taiwan: Spatial Regression with Disjoint Neighborhoods." *Political Research Quarterly* 59, no. 1 (March): 35-46.

Lin, Tse-min (林澤民), and Yun-han Chu (朱雲漢). 2008. "The Structure of Taiwan's Polit-ical Cleavages toward the 2004 Presidential Election: A Spatial Analysis." *Taiwan Journal of Democracy* 4, no. 2 (December): 133-54.

Lin, Tse-min (林澤民), and Matthew Cohen. 2008. "Spatial Regression as a Statistical Model of Regionalism." Paper presented at the 66th Annual National Conference of the Midwest Political Science Association, Chicago, Illinois.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12, no. 4 (Autumn): 375-85.

Taagepera, Rein. 1973. "Seats and Votes: A Generalization of the Cube Law of Elections." *Social Science Research* 2, no. 3 (September): 257-75.

______. 1986. "Reformulating the Cube Law for Proportional Representation Elections." *American Political Science Review* 80, no. 2 (June): 489-504.

Tufte, Edward R. 1973. "The Relationship between Seats and Votes in Two-Party Systems." *American Political Science Review* 67, no. 2 (June): 540-47.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach." *American Journal of Political Science* 42, no. 4 (October): 1233-59.

Wu, Chin-En (吳親恩), and Feng-yu Lee (李鳳玉). 2007. "Electoral Systems and the Moderation of Party Positions on Ethnicity" (選舉制度與台灣政黨族群議題立場的和緩). *Zhengzhi xuebao* (政治學報, Chinese Political Science Review) (Taipei), no. 43 (June): 71-99.