

Ecological Inference with Covariates: Voter Transition in Different Groups in South Korea*

WON-HO PARK[†]

July, 2008

Abstract

The building block of ecological inference strategies is to construct a two-by-two table that describes the individual-level relationship from aggregate information. Extensions to this baseline model, whichever particular technique is employed, have been developed in the context of constructing bivariate R-by-C tables. However, another important and substantively interesting extension is a model that would let the researcher include additional covariates into the model and is yet to be fully discussed and developed. In the paper, I propose a method of moment estimator that incorporates covariates into the ecological inference process by extending Thomsen (1987)'s voter transition model. I apply the developed model to estimate the impact of demographic variables on turnout in South Korean voters over time, especially around democratization, using precinct-level electoral returns and census records.

Introduction

Recent advances in the ecological inference techniques opened up new avenues of research in political science to delve into questions that survey-based studies could not reach before. As was demonstrated by Achen and Shively (1995) and King (1997), ecological inference is difficult but can be achieved with careful application of the underlying theory to the given problem and at hand. Yet at the same time, there are more theoretical work to be developed when it comes to applying the developed technique to tackle

*Revised and prepared for the 2008 Annual Political Methodology Conference. I thank Chris Achen, Ken Kollman, John Jackson, and Mike Traugott for their insights. Any errors are, of course, of my own.

[†]Assistant Professor, Department of Political Science, University of Florida, Gainesville, Florida, FL 32611. E-mail: wpark+comments@ufl.edu

more substantive and complex questions. Expanding the baseline 2×2 table into multinomial categories or choice structures (the “ $R \times C$ ” table) is an important front in this regard (for example, see Rosen et al. 2001), which will allow the practitioner to study more realistic models.

Another important missing link seems to be the fact that we do not have a clear idea on how to include additional covariates into the equation (and “control” for the variables). When our individual-level models address the *ceteris paribus* question, as is most often the case, ecological inference models should enable the researchers to grapple with the idea of a corresponding aggregate equation.

Consider, for example, the voter transition model, which is a typical ecological inference setup where the task is to estimate the loyalty (retention) and defection rates of voters across two elections. As I argued elsewhere (Park 2007), an ecological inference model introduced by Thomsen (1987) turns out to be extremely reliable and robust in estimating the movement of the voters across two elections, with a sound micro-foundation theory. Yet, the question the voter transition model asks can be characterized to be descriptive and somewhat coarse, even though the correct answer to the question is sufficiently hard to come across with aggregate data. The voter transition model typically studies the relative size of the movement of the voters, but it would usually be the case that the researcher will want to go a step further and ask *who* switched parties and *why* the transition occurred.

This paper considers this problem. I first discuss and develop an extension of the Thomsen model to accommodate additional covariates into the voter transition model. It is an attempt to equip the voter transition model with an analytical mechanism that will enable the researcher to study the correlates of voter movements across parties.

To this end, the Thomsen model is reinterpreted as a method of moment approach, and the model is extended to include covariates—simple fractions of people who belong to a certain demographic group, say, the proportion of workers, in a district—which will help the researcher to look into i) separate levels of party support in the demographic groups (different levels of support among workers and non-workers), and ii) separate sets of transition rates in the demographic groups (different loyalty and defection rates among workers and non-workers). This will enable the researcher to approach the more detailed mechanism that lies beneath the voter transitions: for example, it opens up a new way to study electoral realignments.

The example provided at the end of the paper is a detailed picture of how South Korean voters responded to the democratic reform of the country in 1987. To study the impact of the major political transformation of the country on the voters, I look into the change in the turnout pattern of South Korean

voters employing the developed model. More specifically, I examine whether formerly alienated voters started to come to the polling booth after the democratization; at the same time, I examine whether there would be any sign of decrease in turnout among formerly mobilized voters in post-democratization elections.

The Model

The Voter Transition Setup with Covariates

Essentially, we are interested in the difference in the voter transition rates among different demographic groups. The question would be equivalent to modeling the loyalty and defection rates as functions of the covariate. Suppose we are interested in the transition rates of workers, where a binary variable z_i would indicate whether an individual is a worker or not. If we had individual-level survey data, we would be able to estimate the parameters by a three-way cross tabulation, as is depicted by Figure 1, where x_i indicates the vote choice of individual i at time 1, and y_i indicates that at time 2.

	Worker ($z_i = 1$)		Non-Worker ($z_i = 0$)	
	$x_i = 1$	$x_i = 0$	$x_i = 1$	$x_i = 0$
$y_i = 1$	p_1	q_1	p_0	q_0
$y_i = 0$	$1 - p_1$	$1 - q_1$	$1 - p_0$	$1 - q_0$

Figure 1: Voter Transition with a Covariate

We should be able to write a regression relationship at the individual level as

$$\begin{aligned} \Pr(y_i = 1) &= p_i x_i + q_i (1 - x_i) \\ \text{where } p_i &= p_1 z_i + p_0 (1 - z_i) \\ q_i &= q_1 z_i + q_0 (1 - z_i) \end{aligned}$$

Or equivalently, we may write

$$y_i = z_i [p_1 x_i + q_1 (1 - x_i)] + (1 - z_i) [p_0 x_i + q_0 (1 - x_i)] + u_i \quad (1)$$

which is a fully saturated interactive model between the electoral choice and a covariate. Reorganizing the equation above yields

$$y_i = q_0 + (q_1 - q_0)z_i + (p_0 - q_0)x_i + [(p_1 - q_1) - (p_0 - q_0)]x_i z_i + u_i. \quad (2)$$

which is displayed in the usual OLS form of $y_i = \alpha + \gamma z_i + \beta x_i + \delta x_i z_i + u_i$. Loyalty and defection rate parameters in the two groups are all identified since

$$\begin{aligned} q_0 &= \alpha \\ q_1 &= \alpha + \gamma \\ p_0 &= \alpha + \beta \\ p_1 &= \alpha + \beta + \gamma + \delta. \end{aligned}$$

In other words, with individual-level data, OLS will exactly produce the parameters in Figure 1, and the estimate will have the usual desirable properties since $E(u_i) = 0$ and $Cov(u_i, x_i) = Cov(u_i, z_i) = 0$.

But what about the aggregate relationship? It is not straightforward to derive the aggregate equivalent of the equation analytically in the same fashion that justifies the Goodman regression (Achen and Shively 1995: 52–53). By averaging both sides of Equation (2), we may try to derive the relationship between the aggregate variables, Y_j , X_j , and Z_j , namely, party support in the two elections and the proportion of workers in district j , in the following manner. Averaging both sides of Equation (2) in district j , with n_j total voters, would yield

$$Y_j = q_0 + (q_1 - q_0)Z_j + (p_0 - q_0)X_j + [(p_1 - q_1) - (p_0 - q_0)] \frac{\sum_i x_i z_i}{n_j}. \quad (3)$$

If x_i and z_i are not correlated, then $Cov(x_i, z_i) = \frac{1}{n_j} \sum_i x_i z_i - X_j Z_j = 0$, in which case we should have an aggregate level regression relationship that corresponds to Equation (2).¹ This implies that even when the Goodman assumption of the constant parameter holds, a Goodman regression does not exist when it comes to estimating the impact of covariates that are correlated with the main independent variable.

¹Of course, in this case, “controlling” for z_i does not make any sense anyway.

The Baseline Thomsen Estimator

Extending the Thomsen model to have covariates is not straightforward. In this section, I will reinterpret the Thomsen estimator to be a method of moment estimator and develop an extension of it that allows the inclusion of covariates into the model.

The Thomsen model² is based upon the individual-level utility of voting for a given party in two elections which are assumed to distribute joint-bivariate normal in the population, with means μ_x and μ_y and correlation ρ between them:

$$f(x_i^*, y_i^*) = \mathcal{N}^2(\mu_x, \mu_y; \rho). \quad (4)$$

Voter decisions are simply a function of these random variables. If the utilities to vote for a given party in the two elections, x_i^* and y_i^* , exceed a threshold, which can be set at zero for identification, the voter will cast her ballot for the party; otherwise, she will vote for the opposing party. This is a usual probit setup where voting for a given party in elections can be denoted as indicator variables, say x_i for election 1 and y_i for election 2.

$$\begin{aligned} x_i &= 1 \text{ if } x_i^* > 0, \quad \text{otherwise, } x_i = 0 \\ y_i &= 1 \text{ if } y_i^* > 0, \quad \text{otherwise, } y_i = 0. \end{aligned} \quad (5)$$

Then, quite simply, we can see that our quantity of interest can be retrieved by evaluating

$$\begin{aligned} \Pr(x_i = y_i = 1) &= \Pr(x_i^* > 0 \text{ and } y_i^* > 0) \\ &= \int_0^\infty \int_0^\infty \mathcal{N}^2(x_i^*, y_i^* | \mu_x, \mu_y, \rho) dx_i^* dy_i^* \end{aligned} \quad (6)$$

where the double integration would estimate the probability that a voter supports a given party in both elections. The task then would be to estimate the three parameters of the bivariate-normal distribution, namely, the means of the two utility distributions, μ_x , μ_y , and their correlation, ρ .

First of all, since the μ_x is the mean of the normally distributed x_i^* , we may write from Equation (5) as

$$\begin{aligned} E [\Pr(x_i = 1)] &= \int_0^\infty \mathcal{N}(x_i^* | \mu_x, 1) dx_i^* \\ &= \int_{-\infty}^{\mu_x} \phi(z) dz \end{aligned}$$

²For more details, see Thomsen (1985) and Park (2007).

$$= \Phi(\mu_x)$$

where $\phi(z)$ denotes the standard normal density function and $\Phi(z)$ is the cumulative standard normal function. In a similar fashion, we may define $E [\Pr(y_i = 1)] = \Phi(\mu_y)$.

It is straightforward to retrieve these two population parameters from the sample moments, since $E [\Pr(x_i = 1)]$ and $E [\Pr(y_i = 1)]$ are the respective first moments of the two binary variable, that is, the sample means. Note that this quantity can directly be retrieved by the weighted averages of the aggregate data:

$$\begin{aligned}\Phi(\mu_x) = E [\Pr(x_i = 1)] &= \frac{\sum_i^N x_i}{N} = \frac{\sum_j X_j n_j}{N} \\ \Phi(\mu_y) = E [\Pr(y_i = 1)] &= \frac{\sum_i^N y_i}{N} = \frac{\sum_j Y_j n_j}{N}\end{aligned}\tag{7}$$

where X_j and Y_j denote the proportion of party supporters in district j and n_j is the number of total voters in the district. N is the total number of national votes, so $\sum_j n_j = N$. The parameters of interest, μ_x and μ_y can be retrieved by inverting the probit function. Note that even though the actual estimates will be computed from the aggregate data, the quantities exactly are the first moments of individual level samples. In short, there is no cross-level inference, yet.

The third parameter, the correlation coefficient, can be written in terms of the sample moments as well, although inestimable because they are not observed directly:

$$\rho = \text{Corr} [x_i^*, y_i^*].\tag{8}$$

First, add a subscript j to the variables to indicate the district to which the voter belongs: then our task is to derive $\text{Corr} [x_{ij}^*, y_{ij}^*]$. Note that the aggregate-level information we have is in the form of

$$\begin{aligned}X_j = E(x_{ij}|j) &= \frac{\sum_i^{n_j} x_{ij}}{n_j} \\ Y_j = E(y_{ij}|j) &= \frac{\sum_i^{n_j} y_{ij}}{n_j}\end{aligned}$$

where each $\Phi^{-1}(X_j)$ and $\Phi^{-1}(Y_j)$ are the first sample moments of x_{ij}^* and y_{ij}^* within district j , respectively. According to Thomsen, these can be thought of as means of district-level distributions of individual utilities to vote for a given party.

Individual		Aggregate	
x_i and y_i	Binary variables for vote choices in time 1 and time 2	X_j and Y_j	Corresponding aggregate supports in District j
z_i	Binary variable indicating an individual attribute	Z_j	Corresponding aggregate fraction of such voters in District j
x_i^* and y_i^*	Underlying individual utilities to vote for a given party that determines x_i and y_i . $\Pr(x_i = 1) = \Phi(x_i^*)$ and $\Pr(y_i = 1) = \Phi(y_i^*)$.	X_j^* and Y_j^*	Corresponding aggregate utilities of voters in district j to vote for a given party. Estimated by $\Phi^{-1}(X_j)$ and $\Phi^{-1}(Y_j)$.

Table 1: Key Variables in the Extended Thomsen Model with a Covariate

The correlation between the two aggregate variables, the probit transformed $\text{Corr} [X_j^*, Y_j^*]$ or $\text{Corr} [\Phi^{-1}(X_j), \Phi^{-1}(Y_j)]$, can replace equation (8), when the ratio of the systematic versus the non-systematic parts of x_i^* and y_i^* equal the same ratio in X_j^* and Y_j^* respectively. A more detailed discussion on the assumption can be found in Appendix A. If we can establish that the probit transformed variables distribute normal, standard estimates from its sample moments and correlations will be consistent, converging in probability (Greene 2005).

Before I proceed, I will define key variables and parameters involved in the extended Thomsen model with covariates. Table 1 provides summary of the variables at both individual and aggregate levels: in general, individual manifest variables, such as x_i and y_i , are binary variables, while their counterparts are the aggregated fraction of them at district j and are conventionally written in upper case.

Adding a Covariate to the Thomsen Estimator

A. Two Joint-Bivariate Normal Distributions

As was shown in the previous section, the Thomsen estimator assumes that the two key latent variables that underlie the electoral choice is joint-bivariate normally distributed. To add a covariate, we can think about the variables, (x_i^*, y_i^*) , as draws from a mixture of two joint-bivariate distributions. More specifically, the probability density function of the two latent variables that capture voters' utility to support a given party in the two elections can be defined as:

$$f(x_i^*, y_i^* | z_i) = z_i \mathcal{N}_1^2 [\mu_{x,1}, \mu_{y,1}, \rho_1] + (1 - z_i) \mathcal{N}_0^2 [\mu_{x,0}, \mu_{y,0}, \rho_0]. \quad (9)$$

In other words, if a voter belongs to a particular group of interest—say, if she is a worker—, then her latent utilities to vote for the party in the two elections is drawn from the first bivariate normal distribution, \mathcal{N}_1^2 ; otherwise, they are drawn from the second distribution, \mathcal{N}_2^2 . There are six parameters to estimate to compute the transition rates in the two groups, three from each distribution.

One important point to make here is that such demographic variables, say, proportion of workers or the gender distribution, has a radically different theoretical interpretation from the vote shares within the context of the Thomsen model. In the context of the model, voting for a given party in the two elections is set up as an outcome of an underlying partisanship of a voter. In fact, neither y_i or x_i are exogenous to each other—Thomsen is only interested in studying the correlation between the two. On the contrary, many types of covariates, such as demographic variables, are truly exogenous variables that are measured directly. In the model above, they are accordingly set up as strictly exogenous factors that determine to which (joint-normal) distribution the sample belongs. Our task then is to study how different the two distributions are and use them to find out the transition rates in the two different demographic groups.

The key variables x_i^* and y_i^* are latent utility variables and if they were observed at the individual-level, we should be able to estimate directly the population parameters with the sample moments. However, since we do not have that information, it is necessary to take a detour and estimate auxiliary parameters that will help us study the two distributions. Since the voter preferences are distributed joint-bivariate normal in each group, it is possible to write the following regression relationships that accompany the two distribution functions:

$$\begin{aligned} y_i^* &= \alpha_1 + \beta_1 x_i^* + u_{i,1} & \text{if } z_i = 1 \\ y_i^* &= \alpha_0 + \beta_0 x_i^* + u_{i,0} & \text{if } z_i = 0. \end{aligned}$$

In a similar fashion that was shown in Equation (7), we may equate the first population moments to the sample means of probabilities to vote:

$$\Phi(\hat{\mu}_{y,1}) = E [\Pr(y_i = 1|z_i = 1)] = E [\Phi(\alpha_1 + \beta_1 x_i^* + u_{i,1})] \quad (10)$$

$$\Phi(\hat{\mu}_{y,0}) = E [\Pr(y_i = 1|z_i = 0)] = E [\Phi(\alpha_0 + \beta_0 x_i^* + u_{i,0})] \quad (11)$$

The significance of the above expression is that once we estimate the auxiliary parameters, β 's and α 's, we should be able to get estimates of $\mu_{y,1}$ and $\mu_{y,0}$ in terms of the function of x^* and x^* only. Most

importantly, as can be seen in the right hand side of the above equations, we do not have to worry about the covariate, z_i , any more once we collect the the auxiliary parameters.

In a symmetric fashion, write the reversed regression relationship with x_i^* as the dependent variable and y^* as the independent variable such as

$$\begin{aligned} x_i^* &= \delta_1 + \gamma_1 y_i^* + e_{i,1} & \text{if } z_i = 1 \\ x_i^* &= \delta_0 + \gamma_0 y_i^* + e_{i,0} & \text{if } z_i = 0 \end{aligned}$$

This may look counter-intuitive to have an earlier election as the dependent variable and a later election as the independent variable, but since the model does not claim any direct causal relationship between the two electoral variables—and they are only linked by a correlation—there is nothing that prohibits this. A different way to understand this is to say that it is the joint-distribution parameters of x_i^* and y_i^* that matter for our purpose, and regression coefficients are just auxiliary parameters to help us study the distributions. Thus, we write

$$\Phi(\hat{\mu}_{x,1}) = E [\Pr(x_i = 1|z_i = 1)] = E [\Phi(\delta_1 + \gamma_1 y_i^* + e_{i,1})] \quad (12)$$

$$\Phi(\hat{\mu}_{x,0}) = E [\Pr(x_i = 1|z_i = 0)] = E [\Phi(\delta_0 + \gamma_0 y_i^* + e_{i,0})] \quad (13)$$

The question now is how to estimate the four μ parameters with aggregate data. Take, the example of Equation (10). Once we have estimates of the auxiliary parameters, the expected value of the right-hand-side expression can be written as the following and can be estimated by aggregate data.

$$\begin{aligned} \Phi(\hat{\mu}_{y,1}) &= E [\Phi(\hat{\alpha}_1 + \hat{\beta}_1 x_i^* + \hat{u}_{i,1})] \\ &= E [\Phi(\hat{\alpha}_1 + \hat{\beta}_1 \Phi^{-1}[\overline{\Phi(x_i^*)}])] \\ &= \frac{\sum_j \Phi(\hat{\alpha}_1 + \hat{\beta}_1 X_j^*) Z_j n_j}{\sum_j Z_j n_j} \end{aligned} \quad (14)$$

The idea is to replace the individual utilities, x_i^* in district j , with X_j^* , the probit-transformed district vote shares of the party which are assumed to be the district means of aggregate utilities. Even though this is the solution I have at the current stage, it is not the perfect solution: the estimates will be slightly biased since the probit of means ($\Phi^{-1}[\overline{\Phi(x_i^*)}]$) is not always the mean of probit-transformed variables. ($\overline{x_i^*}$).

³ However, it is not an unreasonable solution, as similar replacement was done in the simple Thomsen model.

The last expression shows that this can be computed by the weighted average of the aggregate sample. Note that expected turnout in district j is weighted by the number of the population group in the district, $Z_j n_j$. Following similar lines of logic, the rest of the necessary mean parameters of the two joint-bivariate normal distributions can be derived as:

$$\Phi(\hat{\mu}_{y,0}) = \frac{\sum_j \Phi(\hat{\alpha}_0 + \hat{\beta}_0 X_j^*) (1 - Z_j) n_j}{\sum_j Z_j n_j} \quad (15)$$

$$\Phi(\hat{\mu}_{x,1}) = \frac{\sum_j \Phi(\hat{\delta}_1 + \hat{\gamma}_1 Y_j^*) Z_j n_j}{\sum_j Z_j n_j} \quad (16)$$

$$\Phi(\hat{\mu}_{x,0}) = \frac{\sum_j \Phi(\hat{\delta}_0 + \hat{\gamma}_0 Y_j^*) (1 - Z_j) n_j}{\sum_j Z_j n_j} \quad (17)$$

Now to estimate the correlation parameter for the two joint distributions, we may write:

$$\hat{\rho}_1^2 = \hat{\beta}_1 \cdot \hat{\gamma}_1 \quad (18)$$

$$\hat{\rho}_0^2 = \hat{\beta}_0 \cdot \hat{\gamma}_0 \quad (19)$$

noting that the product of two reversed bivariate slope regression coefficients take the form of $\frac{Cov(x,y)}{Var(x)} \times \frac{Cov(x,y)}{Var(y)} = Corr^2(x,y)$. These complete the full two sets of parameters that are necessary to evaluate the two bivariate normal distributions. Quandt and Ramsey (1978) analyzed the problem of estimating parameters of such mixtures of normal distributions, and showed that the method of moment estimates are consistent. Estimates of the transition rates in the two different population groups are continuous functions of those consistent estimators, and hence consistent.

So far, I outlined how to estimate parameters of the two joint bivariate normal distributions from which the two groups of voters are modeled to be sampled. The six equations, (14)-(19), provide formulas to estimate the parameters, using auxiliary regression parameters. Now the task left is to show how to estimate them.

³There also is another source of bias in the equation. Since the expected value of the error term inside the cumulative normal function will not be zero in general, the estimates will slightly be biased.

B. Estimating the Auxiliary Parameters

The auxiliary regression models can be incorporated into the following two sets of nonlinear equations:

$$\Pr(y_i = 1|z_i) = z_i \Phi [\alpha_1 + \beta_1 x_i^* + u_{i,1}] + (1 - z_i) \Phi [\alpha_0 + \beta_0 x_i^* + u_{i,0}] \quad (20)$$

$$\Pr(x_i = 1|z_i) = z_i \Phi [\delta_1 + \gamma_1 y_i^* + e_{i,1}] + (1 - z_i) \Phi [\delta_0 + \gamma_0 y_i^* + e_{i,0}] \quad (21)$$

These equations are not estimable directly, since we do not have individual-level information, and the latent variables are unobservable. A good starting point is to sum the equations using the observed exogenous variable, z_i .

Now suppose within district j , the probability to draw a voter belonging to a demographic group of, say, the workers, will equal Z_j , then we may replace z_i with Z_j , and write

$$\Pr(y_i = 1|j) = Z_j \Phi [\alpha_1 + \beta_1 x_i^* + u_{i,1}] + (1 - Z_j) \Phi [\alpha_0 + \beta_0 x_i^* + u_{i,0}] \quad (22)$$

$$\Pr(x_i = 1|j) = Z_j \Phi [\delta_1 + \gamma_1 y_i^* + e_{i,1}] + (1 - Z_j) \Phi [\delta_0 + \gamma_0 y_i^* + e_{i,0}] \quad (23)$$

Taking the expectations within district j , we have

$$Y_j = Z_j E \left(\Phi [\alpha_1 + \beta_1 x_i^* + u_{i,1}] \right) + (1 - Z_j) E \left(\Phi [\alpha_0 + \beta_0 x_i^* + u_{i,0}] \right)$$

$$X_j = Z_j E \left(\Phi [\delta_1 + \gamma_1 y_i^* + e_{i,1}] \right) + (1 - Z_j) E \left(\Phi [\delta_0 + \gamma_0 y_i^* + e_{i,0}] \right)$$

Resorting to the same assumption that expectations of the cumulative normal functions can be approximated by using district means, we may write

$$Y_j = Z_j \Phi [\alpha_1 + \beta_1 X_j^*] + (1 - Z_j) \Phi [\alpha_0 + \beta_0 X_j^*] + \varepsilon_{Y,j} \quad (24)$$

$$X_j = Z_j \Phi [\delta_1 + \gamma_1 Y_j^*] + (1 - Z_j) \Phi [\delta_0 + \gamma_0 Y_j^*] + \varepsilon_{X,j}. \quad (25)$$

Both equations can be estimated by non-linear least squares or maximum likelihood with the identifying conditions: $\alpha_0 \neq \alpha_1$ or $\beta_0 \neq \beta_1$ for Equation (24) and $\delta_0 \neq \delta_1$ or $\gamma_0 \neq \gamma_1$ for Equation (25). In the trivially special case where Z_j is not continuous and is a binary variable indicating whether the unit is, say, urban ($Z_j = 1$) or rural ($Z_j = 0$), the estimation is equivalent to collecting two sets of parameters in the simple

transition problem without covariates, from two separate groups of geographic units.

With these results, it is now possible to estimate the voter transition model with covariates. In the following section, I demonstrate how the technique developed in this section can be applied to a real voter transition problem.

Recipe: The Thomsen Estimator with Covariates

In this section, I will provide a step-by-step example of the model developed in the previous section, using the example of the turnout rates in the South Korean elections in 1981-1985. The covariate here is the fraction of the more educated voters with college degrees or higher education. Using the model developed in the previous section, I will demonstrate that it is possible to estimate i) the turnout rates in the two different demographic groups and ii) the transition rates between the two elections in the two different groups.

1. **(DEFINITION OF VARIABLES)** First, definitions of the variables are as follows:

X_j : Turnout in the 1981 Election in District j ;

X_j^* : The inverse probit transformation of X_j , that is $\Phi^{-1}(X_j)$;

Y_j : Turnout in the 1985 Election in District j ;

Y_j^* : The inverse probit transformation of Y_j , that is $\Phi^{-1}(Y_j)$;

Z_j : Proportion of more educated voters, with in District j ; and

n_j : The total number of eligible voters in district j .

2. **(NLS ESTIMATION)** Estimate the nonlinear regression relationships for the auxiliary parameters:

$$Y_j = Z_j \Phi \left[\alpha_1 + \beta_1 X_j^* \right] + (1 - Z_j) \Phi \left[\alpha_0 + \beta_0 X_j^* \right] + \varepsilon_{Y,j}$$

The estimation can be carried out by maximum likelihood or nonlinear least squares. Here, I employ nonlinear least squares and the estimated equation is:

$$Y_j = Z_j \Phi \left[.671 + .281 X_j^* \right] + (1 - Z_j) \Phi \left[.564 + .496 X_j^* \right] + \varepsilon_{Y,j}$$

3. **(PROJECTED TURNOUTS OF THE TWO GROUPS IN THE SECOND ELECTION)** The next step is to compute

the predicted probabilities to turnout in the two groups.

$$E(Y_j|z_i = 1) = \frac{\sum_j \Phi \left[.671 + .281X_j^* \right] Z_j n_j}{\sum_j Z_j n_j} = .850$$

$$E(Y_j|z_i = 0) = \frac{\sum_j \Phi \left[.564 + .496X_j^* \right] (1 - Z_j) n_j}{\sum_j (1 - Z_j) n_j} = .810$$

Note that the predicted probabilities in the two population groups are weighted by the group sizes. The estimates indicate that around 85% of the more educated voters turned out in the 1985 election, while the turnout rate among other voters is estimated to be around 81%.

These two estimates constitute the first set of parameters that define the means of the two joint-bivariate normal distributions

$$\hat{\mu}_{y,1} = \Phi^{-1}(.850) = 1.036$$

$$\hat{\mu}_{y,0} = \Phi^{-1}(.810) = .878.$$

4. **(REVERSED NLS ESTIMATION)** The next parameters to estimate are the corresponding population means of the first election. Estimate the reversed nonlinear regression such as

$$X_j = Z_j \Phi \left[\delta_1 + \gamma_1 Y_j^* \right] + (1 - Z_j) \Phi \left[\delta_0 + \gamma_0 Y_j^* \right] + \varepsilon_{X,j}$$

which yields the following results:

$$X_j = Z_j \Phi \left[.055 + .043Y_j^* \right] + (1 - Z_j) \Phi \left[-.135 + 1.064Y_j^* \right] + \varepsilon_{X,j}$$

5. **(PROJECTED TURNOUTS OF TWO GROUPS IN THE FIRST ELECTION)** Again, compute the estimated projection of turnout in the two groups in the *first* election:

$$E(X_j|z_i = 1) = \frac{\sum_j \Phi \left[.055 + .043Y_j^* \right] Z_j n_j}{\sum_j Z_j n_j} = .611$$

$$E(X_j|z_i = 0) = \frac{\sum_j \Phi \left[-.135 + 1.064Y_j^* \right] (1 - Z_j) n_j}{\sum_j (1 - Z_j) n_j} = .793$$

Note that the turnout in the group that consists of more educated voters are significantly lower than that in the other group. With these results, we may estimate the second set of parameters,

$$\hat{\mu}_{x,1} = \Phi^{-1}(.611) = .282$$

$$\hat{\mu}_{x,0} = \Phi^{-1}(.793) = .817.$$

6. **(CORRELATION COEFFICIENTS)** The final parameters are the correlations in the two distributions, which can be retrieved by the multiplication of the two slope coefficients, β 's and γ 's as discussed before:

$$\hat{\rho}_1 = \sqrt{\hat{\beta}_1 \hat{\gamma}_1} = \sqrt{.281 \times .043} = .165$$

$$\hat{\rho}_0 = \sqrt{\hat{\beta}_0 \hat{\gamma}_0} = \sqrt{.496 \times 1.064} = .726.$$

As we can see, the correlation is much smaller for the group with higher education which indicates that more electoral change is happening in the group.

7. **(SUMMING UP THE TWO JOINT DISTRIBUTIONS)** This finishes estimating necessary parameters from the two bivariate-normal distributions as:

$$\text{Educated Voters: } \mathcal{N}_1^2 [\hat{\mu}_{x,1}, \hat{\mu}_{y,1}; \hat{\rho}_1] = \mathcal{N}_1^2 [.282, 1.036; .165]$$

$$\text{The Rest: } \mathcal{N}_0^2 [\hat{\mu}_{x,0}, \hat{\mu}_{y,0}; \hat{\rho}_0] = \mathcal{N}_0^2 [.817, .878, .726]$$

8. **(FRACTION THAT VOTED IN BOTH ELECTIONS)** Now the task left is to find the fraction of voters in the two groups that voted in both elections. We can compute the double integrals of the estimated distributions to find the fraction of voters whose utilities exceed the threshold of zero in both elections.

$$\int_0^\infty \int_0^\infty \mathcal{N}_1^2 [x^*, y^* | .282, 1.036; .165] dx^* dy^* = .529$$

$$\int_0^\infty \int_0^\infty \mathcal{N}_0^2 [x^*, y^* | .817, .878, .726] dx^* dy^* = .718.$$

To reiterate, 53% of educated voters in South Korea voted in both elections, while 72% of the other voters turned out in both elections.

		Highly Educated The 1981 Election			The Rest The 1981 Election		
		Voted	Not Voted	Total	Voted	Not Voted	Total
The 1985 Election	Voted	0.529	0.321	0.850	0.718	0.092	0.809
	Not Voted	0.082	0.068	0.150	0.075	0.115	0.191
		0.611	0.389		0.793	0.207	

Table 2: Estimated Distribution of Voters across Elections in Different Education Groups: South Korean Elections 1981–1985

		Highly Educated The 1981 Election		The Rest The 1981 Election	
		Voted	Not Voted	Voted	Not Voted
The 1985 Election	Voted	0.866	0.825	0.905	0.444
	Not Voted	0.134	0.175	0.095	0.556

Table 3: Estimated Transition Rates in Different Education Groups: South Korean Elections 1981–1985

9. (FINISHING UP THE TABLE) Now fill in the fraction of the voters in the following three-way table, using the estimated total turnouts in each group. Note that the estimated quantities we already have are emphasized in the table. For example, we have estimated turnout rates in the two groups in the two election from Steps 3 and 4, which will enable us to define the marginal probabilities in the table. Also, since we have estimated the fraction of two-time voters in Step 8, the table is identified, and we may fill in the table in an obvious way.
10. (LOYALTY AND DEFECTION RATES: THE CONDITIONAL PROBABILITIES) Since loyalty rates and defection rates as defined in the literature are conditional probability terms, that is, the probability that a first time voter will return to the polling booth in the next election, the fractions can trivially be transformed into such rates. By dividing the fractions with the marginal turnout rates in the 1981 election, the voter transition rates can be retrieved and are shown in Table 3:

Among other things, Table 3 makes an interesting point on the estimated rate of more educated voters newly entering the election in 1985. It tells that around 83% of more educated non-voters in the previous election newly turned out in the 1985 election, perhaps because they were more responsive to the electoral

climate than the rest of the voters. The rest of the paper looks into the question of how the democratization movement in South Korea made an impact on the electoral participation of the voters by examining aggregate data and applying the model I have developed so far.

Application: The Impact of Democratization on Voter Turnout

Introduction

The year 1987 marks an important point in time in South Korean political history. After a series of authoritarian governments that ruled the country from the 1960s, the government conceded to the massive protest and wide-spread demand of the citizens for democratic changes, including restoration of civil rights, amendment of the constitution to ensure more democratic elections, and other political reforms (Han 1988; Oh 1999). The presidential election that was held at the end of the year was the first direct, competitive contest since the 1972 election.

If there ever was an election in South Korea that can be called a “critical” election in the same sense as V. O. Key (1955) defines, it is the 1987 election. Elections afterward showed voting patterns that were not existent before: including massive levels of regional voting pattern and high levels of partisan votes. In short, the democratization induced a critical electoral realignment.

Realignment, simply defined, is a systematic and enduring change in electoral preferences of the electorate over time. To study realignments, the researcher should naturally look into the change in partisanship or electoral support of voters. However, an equally, if not more, important aspect of electoral realignments is whether and why new voters are mobilized into the political arena and who they are. Change in the electoral environment or an issue that newly becomes salient may attract non-voters to the polling booth, and if such new voters are systematically favoring one party over other parties, they will constitute a central component of the electoral realignment.

In this vein, the dynamics of voter participation takes on an important meaning in understanding the electoral politics in South Korea. The democratic reform of the country—a system-level change with a profound impact on the political life of the voters and politicians—should entirely alter how people perceive politics and how political parties mobilize them: that is, who starts to vote and whom they vote for.

Authoritarian regimes will alienate certain groups of citizens, and as the literature on participation and

political efficacy suggests, we may expect a systematic influx of such voters into the electoral arena when such regimes reach their demise. Democratic reform however may also trigger the vanishing of traditional mobilization mechanisms. This will necessarily result in the decrease of turnout, and more importantly, the weakening of support for candidates and parties who have benefited from such mobilization networks and connections of the authoritarian government.

This section focuses on how the democratic reform in South Korea attracted voters that were formerly uninterested in the authoritarian regime, and how such voters are aligning into the new party system. Of course, it is not to say that such influx of new voters is the one and only cause of the emergence of the new party system, but I argue that it constitutes a major part of the electoral realignment. In this section, I simply try to answer the following questions: first, did the democratization of the country attract new voters into the political arena? If so, who are they, and what are the consequences of such new voters? And finally, what caused such systematic movement in the electorate?

The following section sets up the background on the dynamics of turnout in South Korea. A simple adaptation of the voter transition setup will be applied to model the entrance and exit rates of the voters.

Background: The Dynamics of Voter Turnout in South Korea

The dynamics of voter turnout in South Korea, especially focusing upon the elections around 1987, shows that there are several complex story lines. Figure 2 shows the overall trend of national voter turnout in presidential, legislative, and local elections.

First, it is quite obvious that there is a short-term effect of democratization on turnout rates: the turnout rate in the 1987 presidential election, which is around 89%, is higher than any previous election since the 1970s; turnout in the National Assembly elections peaks in 1985—a widely contested election that triggered the democratization movement. However, it should be noted that turnout decreased by about 10 percentage points in the 1988 National Assembly election that took place a year after democratization. There is also a long-term trend of decline in turnout after the democratization of the country.

As the literature on political participation finds a positive relationship between political efficacy and participation, the democratization of a political system should facilitate higher turnout. The literature is quite clear on the relationship between participation and political efficacy: alienated voters are less likely to vote. For example, Luttbeg and Gant (1995: 134–136) distinguish between internal and external

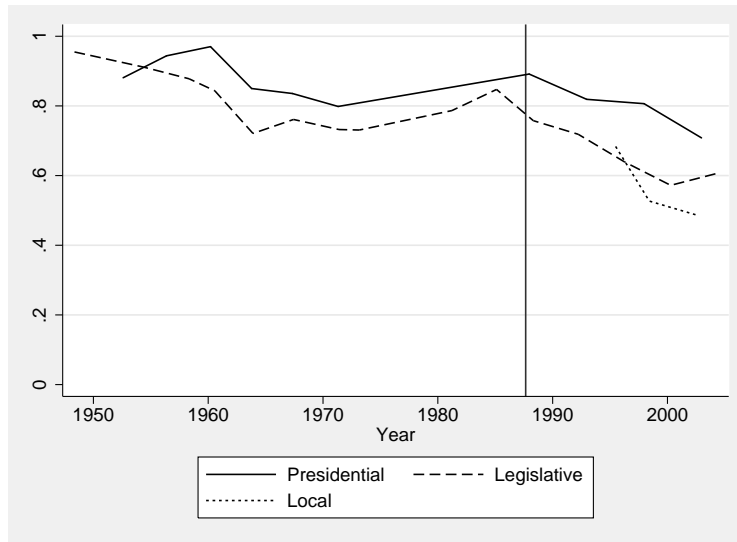


Figure 2: Turnout in South Korean Elections

efficacies, and highlight that both strongly contribute to the voter turnout. The case in point here is that we may expect an increase in turnout in a more democratic regime (external efficacy), and especially from internally efficacious voters with more political resources.

Meanwhile, as the literature on voter mobilization and participation claims, there are reasons to expect a decrease in participation after the democratization of the country, since the ruling party lost its edge in mobilizing voters. (For example, see Rosenstone and Hansen 2002.) More specifically, under the authoritarian regime, there was a strong connection between local wards of governing parties and local governmental offices, which enjoyed uncontested and effective voter mobilization, usually in the form of strong turnout drives. Abolition of such practices, or guarantee of the neutrality of local governmental offices, was one of the main focuses in the discussion of democratic reform, and by the 1987 presidential election, old mechanisms of voter mobilization would not work for the ruling Democratic Justice Party.⁴

In short, there is no simple trend in turnout rates after democratization that we may expect at the national level. It is fair to summarize that the change in the mobilization patterns and the level of political efficacy affects turnout in opposite directions and cannot easily be separated although we may postulate that there is a short term increase and a long-term decline. And of course, there is no individual-level survey data from this period to look into this question.

⁴See *Donga Ilbo* [*Donga Daily*] (Feb. 1985) on the turnout drive lead by the government. Due to the accuracy of the citizen census register that the government holds, it was possible to identify exactly who voted and who did not. Also, see the discussion on the change in the organization of local party wards.

Examining Entrances and Exits

Figure 3 depicts the possible consequences of changes in voter turnout. In times of changes we would observe higher rates of voter replacement, which could be defined by the proportion of voters who newly enter or do not anymore come to the polling booth. The apparent questions are: who exactly are “entering” and “exiting” the polling booth? ; and what impact does this have on the ensuing party system?

		BEFORE (Time 1)	
		Vote	Not Vote
AFTER (Time 2)	Vote	<i>“Enter”</i>	
	Note Vote	<i>“Exit”</i>	

Figure 3: Replacement of Voters

A direct adaptation of the voter transition model will allow us to investigate the turnout rates in the two elections before and after 1987. We may rewrite the Thomsen ecological estimation function as

$$T_{1987,j} = \Phi \left[\alpha + \beta \Phi^{-1}(T_{1985,j}) \right]$$

where T_{1987} and T_{1985} indicate turnouts in the two consecutive elections. The “entrance” rate will be the usual defection rate and the “exit” rate will be one minus the usual loyalty rate.

		1985 Election	
		Vote	Not Vote
1987 Election	Vote	0.878 (0.875, 0.881)	0.782 (0.768, 0.795)
	Not Vote	0.122 (0.119, 0.125)	0.218 (0.205, 0.232)

$N = 3922$

Note: Aggregate turnout rates in the two elections were 81% (1985) and 86% (1987).

Table 4: Voter Transition Rates Around Democratization

Table 4 shows the estimated entrance and exit rates retrieved by the Thomsen estimator. Entries in the parentheses represent the 95% confidence interval of the estimates, as was defined in Achen (2000). We see that about 12% of the previous voters did not vote in the 1987 election, while 78% of the previous

non-voters came to the polling booth in the election. If the causes that systematically determine who votes and who does not are working in both elections, we should see a clear pattern of consistently high coefficients on the diagonal entries: voters will continue to vote, and non-voters will stay that way. Of course, since this is a time when a major electoral earthquake is taking place, we see a sizable proportion of the previous non-voters coming to vote in the first election after democratization.

Table 5 reports such entrance and exit rates in multiple elections in South Korea over time. The table provides an overlook of the composition of the voter turnout in the elections and will be a baseline estimation for the analysis in the next section.

First of all, consider the entrance rates for legislative elections over time. It becomes apparent that there is a significant number of new voters coming into the electoral system in the mid-1980s and into the early 1990s. Also, we note that the exit rates are kept quite small until the 1988 election, meaning that once the voter turns out, she mostly stayed a voter. This trend changes somewhat in the later elections where we see a significant number of voters consistently exiting the elections.

Another important point we observe is that the initial transformation in the pattern of who votes and who does not first started in the 1985 election, not after 1987. This is consistent with what was shown in Figure 2 and with what we know about the 1985 election, which featured a strong opposition party fielding competitive candidates that was a rarity in South Korean elections for many years. In a way, the results show that the democratization movement in 1987 was not a political episode that suddenly erupted, but there already was an underlying movement in the electorate that showed a surge in turnout.

Legislative			Presidential		
Year	Enter	Exit	Year	Enter	Exit
1978	0.439	0.131			
1981	0.341	0.102			
1985	0.504	0.077	1987	0.782	0.121
1988	0.321	0.171			
1992	0.385	0.167	1992	0.617	0.143
1996	0.219	0.210	1997	0.514	0.134
2000	0.192	0.206			

Note: Estimates for the 1987 Election was retrieved in comparison to the 1985 Election.

Table 5: Entrances and Exits from the Polling Booth

Transition rates between presidential elections are less subtle. In the first two elections right after 1987, the majority of previous non-voters turned out in the elections. As it reaches down to more recent

	Year	Enter			Exit		
		Urban	Rural	Diff	Urban	Rural	Diff
Legislative Elections	1978	0.448	0.614	-0.167	0.170	0.143	0.027
	1981	0.366	0.576	-0.210	0.144	0.139	0.005
	1985	0.580	0.501	0.078	0.101	0.084	0.017
	1988	0.321	0.323	-0.001	0.187	0.159	0.028
	1992	0.496	0.401	0.095	0.237	0.112	0.125
	1996	0.535	0.309	0.226	0.379	0.188	0.190
	2000	0.372	0.236	0.136	0.342	0.176	0.166
Presidential Elections	1987	0.793	0.710	0.083	0.124	0.116	0.008
	1993	0.557	0.645	-0.088	0.140	0.145	-0.005
	1997	0.485	0.515	-0.030	0.123	0.141	-0.018

Table 6: Entrance and Exit Rates in Urban and Rural Districts

years, the entrance rate dropped down while the exit rate climbed up. In any case, the presidential elections picture a more volatile electorate where moving in and out of the polling booth seems more frequent.

Unpacking the Entrances and Exits

A. Mobilization and the Turnout of Urban and Rural Voters

Based on our setup, we are ready to examine the entrance and exit rates in different demographic groups. Since we are interested in “identifying” the voters who are newly coming into and leaving from the electoral arena, it requires us to employ the technique developed earlier.

Here I first start out by comparing the entrance and exit rates of the rural and urban voters. It should be noted that this round of estimation was carried out on two separate geographical samples—thus we may infer the rates as those of the urban voters and those of the rural voters. The important point to consider before examining the results in the table is the possible impact of the change in mobilization patterns. It can be assumed that former mobilization mechanisms were more concentrated in the rural areas, and we may see how the turnout patterns change over time there. Most importantly, we should observe lower exit rates and higher entrance rates in rural areas before 1987. After democratization, we can expect abrupt changes in such patterns—where we would start to observe higher exit rates and possibly lower entrance rates of the rural voter.

Table 6 presents estimated entrance and exit rates of the rural and urban voters in several legislative presidential elections. Among many things in the table, we first observe larger entrance rates of the rural voters than their urban counterparts before the 1985 election—something that is consistent with

the strong mobilization in the rural areas. This changes rapidly while we move down to later legislative elections where we see the entrance rates are almost always (except the case of the 1988 election) larger in urban areas. We observe that the rural entrance rate goes down slowly and never recovers to the level above the urban rates in legislative elections, after which it stabilizes. In other words, the recruitment of new voters is now far less effective in the rural areas than it once was.

The exit rates in legislative elections tell a slightly different story. Arguably, the exit rates in the two areas were comparable until the 1992 election, where the urban exit rates took off. Contrary to what we speculated—that the lack of mobilization would result in an immediate release of rural voters—it stays well below the level of urban exit rates. Still, we see a gradual increase in exit rates in rural areas, which may be the long-run effect of the changes in the mobilization pattern. In any case, the extraordinarily large exit rates in the more recent elections are directly related to the rapid decline of overall turnout in recent elections that was shown in Figure 2. The results above indicate that the decline in turnout is heavily driven by the urban voters and less by rural voters.

To summarize, it seems to be the case that the impact of democratic reform—here, essentially the cut-off of previous mobilization structures—had an immediate impact on how new voters were recruited in the rural areas. Democratization did not immediately drive away the voters: but gradually and in a couple of elections voters were moving away from the polling booth at an astonishing rate, especially in urban areas. An overall comparison between urban and rural voters in legislative elections reveal that the urban turnout is always much more volatile than that of the rural voters. Especially in post-democratization elections, both the entrance rates and exit rates are larger in urban areas, indicating more urban citizens are coming in and out of the polling booth, making them less consistent voters.

Presidential elections tell a different story. In both areas, the movement of the voters are volatile with larger entrance rates, and there are not many distinguishable patterns in the three elections, perhaps except the fact that the urban-rural difference diminishes in these elections. Again, this could be due to the fact that we only have three observation points, all of which being “high-profile” elections. But we may learn from a non-finding here: when stakes are high, urban and rural voters behave similarly. Now we look into other demographic attributes.

B. Estimated Turnout Among Different Groups

While the previous section was able to address the question of political mobilization by dividing the observations into different types of geographic units, the question of political efficacy will be more complicated.

For example, it is possible to argue that the particular type of mobilization under the authoritarian regimes in South Korea was targeted at specific types of geographic areas, while attributes such as voter efficacy are truly individual-level characteristics. It would require a full-fledged individual level election study that investigates deep into the psyche of the voters to get a good sense of the interaction between voter efficacy and turnout in South Korea. Additionally, since we are interested in the dynamic aspect of the relationship around the democratization period, a perfect data set to address the question would consist of at least several panels around 1987. In this section, we approach the problem with aggregate demographic information and try to depict the impact of democratization on turnout through voter efficacy.

Ecological estimates shown in this section require even stronger assumptions to hold true, and sometimes, the estimations are less than stable. As was sketched and discussed in previous sections, there is still more work to be added for the improvement of the estimator. That said, the following analysis will show that the ecological inference strategy developed here can be used here to address the question of the impacts of covariates (demographic information) on turnout, and more specifically, on entrance and exit rates.

Figure 4 is a good place to start. Entries are estimated turnout rates of selected demographic groups, which was retrieved by the sample predictions shown in equation (14) after estimating equation (24) with nonlinear least squares. The exact figures are available in the appendix at the end of this chapter. The covariates (Z_j) here would indicate the demographic composition of the unit, say, proportion of young voters. Essentially, the estimates are predicted turnouts of homogeneous groups—for example, young voters—and will provide insights into how one can look at the dynamics of turnout in South Korean legislative elections over time.

The overall turnout rate is similar to what was shown in Figure 2, even though the turnouts here are computed from the data directly—after the matching of the units over time and joining them to census units, the data had to go through a process of dropping some problematic units⁵. The rate hovers around mid 70% before the 1980s, and peaks at the 1985 election. After 1985, we see a continuous and rapid decline in turnout rates. The task here is to decompose the turnout in different demographic groups.

The demographic variables I investigate here are: 1) the percentage of young voters in their 20s and 30s at the time of the election; 2) the percentage of older voters who are in their 50s and above; 3) the

⁵Joining geographic units from different sources of data sets will always create problems. For example, the boundaries can change over time or maybe are just defined differently. Merging adjacent units and aggregating them sometimes will help—cases that I was not able to salvage through this process were dropped. Also, sometimes it is the case that certain geographic units go through drastic population changes. Units that more than doubled in population and that lost more than half of their population were also dropped in the cleaning process.

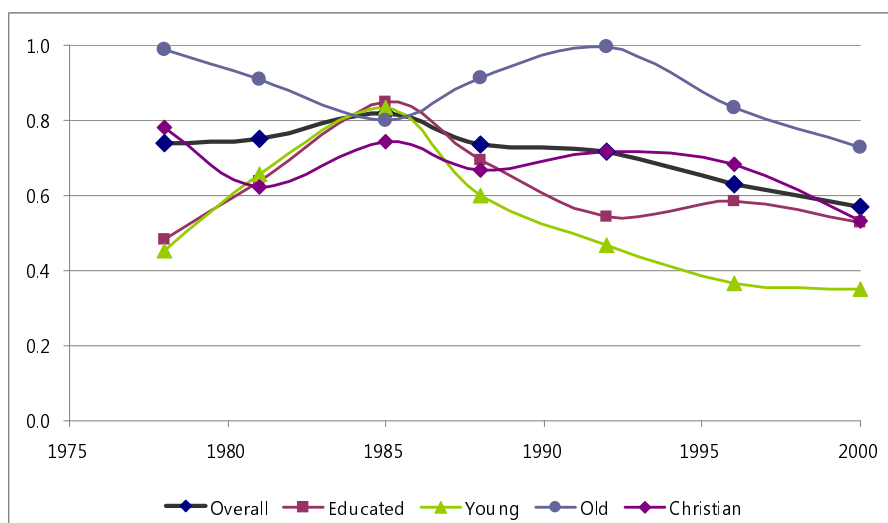


Figure 4: Estimated Turnouts in Legislative Elections, Selected Demographic Groups

percentage of voters with higher education defined by those with college or above education; and 4) the percentage of Christians.

The estimated turnout of high-education voters starts out with a fairly low level of estimated turnout—at around 48% which is about 25% lower than the national turnout. At the time of the important 1985 election, the estimated turnout of voters with high education reaches its peak at around 85%: in fact, the estimated turnout is higher than that of any other demographic groups we have in the analysis. The turnout afterward declines and stays in the mid-50%.

A similar pattern can be found in the young voter category: it starts at the mid 40% level; peaks at the 1985 election; and rapidly dies down. The difference from the previous group of people is that the projected turnouts drop down even more radically where it hits the mid-30% mark by the 2000 election.

As was argued previously, these two groups of voters would have been the most alienated and/or uninterested voters under the authoritarian regime, and we can understand why they display the lowest estimated level of turnout before 1985. It is interesting to note both groups reach their peaks in the 1985 election: not just compared with their turnout level in other elections, but it marks the only time when the estimated turnout among young and highly educated voters record a higher turnout than the national average.

The fast decline—faster than other groups—in the estimated turnout also reveals an interesting point. In the short term, the 1988 legislative election which took place after the 1987 presidential would display

a sharp decline of estimated turnouts among the young and highly educated voters. This could be due to the after-effect of the 1987 presidential election where the candidate from the traditional ruling bloc, Roh Tae Woo, would finally win the electoral contest flattening down all the heightened expectations. In any case, we see that the “critical” legislative election was the 1985 election, not the 1988 election.

In the long run, we see the continuous decline of turnout over time, especially among the young voters, since 1985. We note that the overall decline in turnout is mainly driven by the low turnout of young voters which has consistently been dropping. The estimated turnout for the highly educated holds at around the mid-50%, which distinguishes it from that of the young voters. We will see the difference between the two demographic groups when it comes to looking at their respective entrance and exit rates.

An interesting pattern emerges when we look at the estimated turnout of the “old” voters. It essentially starts at around 98% in the 1973 election and remains to be the highest estimated turnout group except in the 1985 election. It reaches another peak in the 1992 election before showing some decline in later elections.

C. Voter Efficacy and The Dynamics of Turnout

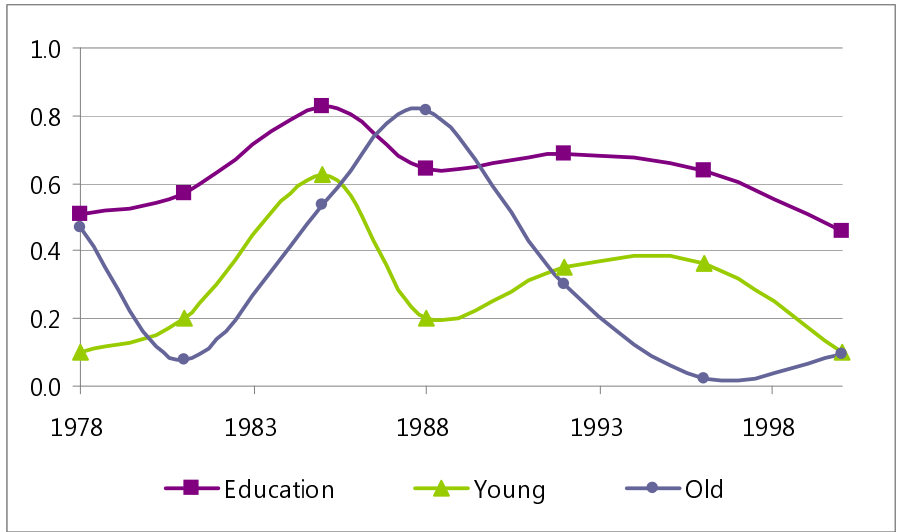
So far, we have described the projected turnout rates among different groups of voters. In this section, we conclude our investigation into the impact of democratization on the turnout of South Korean voters by looking at the estimated exit and entrances rates of voters from different demographic groups.

As was constructed earlier, we want to study the entrance and exit rates among different demographic groups, and see how differently they reacted to democratization. As theory would suggest, we expect to observe larger entrance rates around democratization among those who who would have been politically alienated and who would feel politically more efficacious in the post-democratization era. Particularly, those with higher education (political resources) and the young voters (democratic values) would be a good example.⁶

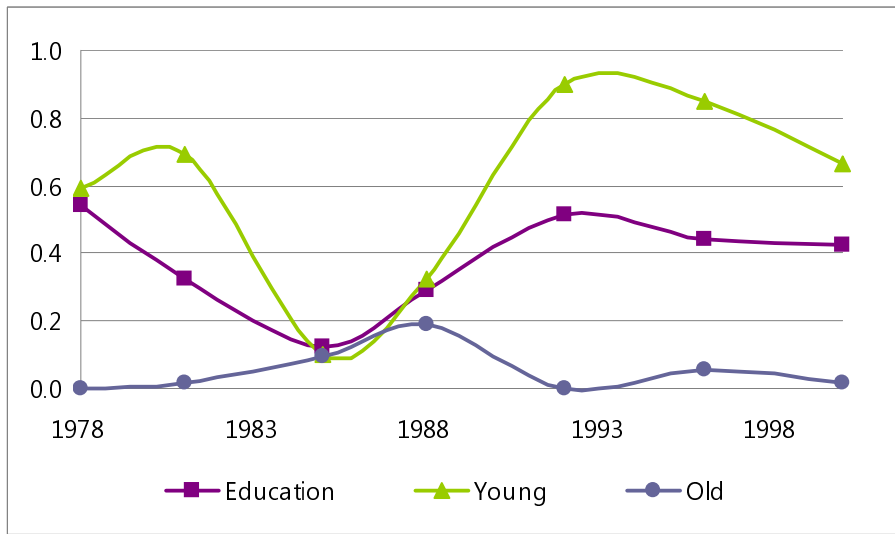
Figure 5 plots the estimated entrance and exit rates of the three different demographic groups, namely, those with higher education, those in their 20s and 30s, and those who are over 50. The plots partially confirm what we expect, and at the same time provide some new insights. More detailed figures with other demographic variables are in Appendix B.

First of all, there is no question about whether there was a “shock” to the electoral system in 1985

⁶See J. H. Rhee (2001) for the political efficacy arguments; Jung (2000) has an interesting discussion on the “Democratization Generation.”



(a) Estimated Entrance Rates



(b) Estimated Exit Rates

Figure 5: Estimated Entrance and Exit Rates by Education and Age

that had a profound impact—at least a short-term effect. We observe surges of entrance rates in both the young voter group and the high-education group at the time; exit rates are the lowest for both groups as well.

Secondly, we are left to wonder whether this “shock” that was created by democratization had any lasting impact. Both groups of voters immediately go back to their pre-democratization levels of entrance and exit rates. One noticeable long-run trend is the drop in entrances and surge of the exit rates, especially among the young voters, but we cannot be sure that this is the direct impact of democratization. In any case, we note that the overall decline of turnout among the young voters that we have seen in Figure 4 in fact is a function of both the drop of the entrance rate and the surge of the exit rate in elections after the 1985 election.

Most interestingly, the impact of democratization can be seen in the older voter group as well. Generally, this is a group of people with very stationary turnout patterns, as can be seen by its minuscule exit rates in the second sub-figure of Figure 4: elderly voters consistently turn out to vote. The only time this is perturbed is in 1985 and 1988 which indicates that a significant amount of these voters stopped coming into the polling booth. Combine this with the sudden surge in the entrance rate of the “old” group around 1988. This indicates that a new batch of older voters was absorbed into the system while a significant chunk of them left in the two elections around democratization.

Discussion

The above analysis on South Korean elections highlights several important points around the democratization period that have not been empirically investigated before. First, it was the 1985 National Assembly election, not the 1987 presidential election, that was the critical election to define the post-democratization realignment. In fact, it was centered around a certain group of voters that initiated the start of such immense political transformation through an election that took place two years before democratization.

More specifically, the group of voters with more political resources to participate—the more educated, the younger, those who live in urban areas, and most importantly, those who chose not to participate in previous elections under the authoritarian regime—suddenly decided to participate and were able to create a solid opposition against the government. An apparent explanation is that there were underlying political demands for more democracy all along, and it somehow solidified as a political force in the 1985

election and was instrumental in the democratization movement.

Second, consistent with the literature on electoral mobilization, this movement was later sustained by the institutional reforms in 1987. As seen in the analysis above, a long-term drop of turnout in rural areas with high exit rates is likely to be due to the demise of the local wards of the governing party and is a continuing trend.

Third, the initial impact around democratization wears off among the voters with more political resources, but they continue to show higher turnout rates and voice their opinions in presidential elections more effectively than their counterparts.

Finally, the observations are consistent with the comparative democratization literature on the relevance of the civil society in its role in democratic transformation. As seen in the section above, we see a clear pattern where an underlying demand for political reform is first manifested by a group of middle class voters in an election. As the social structure went through a gradual change in the 1970s, the middle class gained political weight in South Korea. Their political demand was expressed, among other channels, in the 1985 election, which later would become a precursor to the democratic reform of the country.

References

- [Achen(2000)] Christopher Achen. The thomsen estimator for ecological inference. Unpublished Manuscript, 2000.
- [Achen and Shively(1995)] Christopher Achen and Phillips Shively. *Cross-Level Inference*. The University of Chicago Press, Chicago, IL, 1995.
- [Anselin and Tam Cho(2002)] Luc Anselin and Wendy K. Tam Cho. Spatial effects and ecological inference. *Political Analysis*, 10:276–297, 2002.
- [Berglund and Thomsen(1990)] Sten Berglund and Søren Thomsen. *Modern Political Ecological Analysis*. Abo Academy Press, Abo, Finland, 1990.
- [Cramer(1998)] Erhard Cramer. Conditional iterative proportional fitting for gaussian distributions. *Journal of Multivariate Analysis*, 65(2):261–276, 1998.
- [Duff et al.(2007)Duff, Hanmer, Park, and White] Brian Duff, Michael J. Hanmer, Won-Ho Park, and Ismail K. White. Good excuses: Understanding who votes with an improved turnout question. *Public Opinion Quarterly*, 71:67–90, March 2007. doi: 10.1093/poq/nfl045.
- [Goodman(1959)] Leo Goodman. Some alternatives to ecological correlation. *American Journal of Sociology*, 64:610–625, 1959.
- [Gosnell(1942)] Harold F. Gosnell. *Grass Roots Politics*. American Council on Public Affairs, Washington, D.C., 1942.
- [Greene(2003)] William Greene. *Econometric analysis*. Prentice Hall, Upper Saddle River N.J., 5th ed. edition, 2003. ISBN 9780130661890.
- [Han(1988)] Sung-Joo Han. South korea in 1987: The politics of democratization. *Asian Survey*, 28(1): 52–61, jan 1988. ISSN 0004-4687.
- [Hanushek and Kain(1974)] John E. Jackson Hanushek, Eric and John F. Kain. Model specification, use of aggregate data, and the ecological correlation fallacy. *Political Methodology*, 1:89–107, 1974.
- [Jackson(2002)] John E. Jackson. A seemingly unrelated regression model for analyzing multiparty elections. *Political Analysis*, 10(1):49–65, 2002.

- [Key(1955)] V. O. Jr. Key. A theory of critical elections. *The Journal of Politics*, 17(1):3–18, feb 1955. ISSN 0022-3816.
- [Kil and Moon(2001)] Soong Hoom Kil and Chung-In Moon. *Understanding Korean Politics: An Introduction (Suny Series in Korean Studies)*. State University of New York Press, 2001. ISBN 0791448908.
- [Kim(1980)] Chong Lim Kim. *Political Participation in Korea: Democracy, Mobilization, and Stability*. Studies in international and comparative politics. Clio Books, Santa Barbara, CA, 1980. ISBN 0874362962.
- [King(1997)] Gary King. *A Solution to the Ecological Inference Problem*. Princeton University Press, Princeton, NJ, 1997.
- [King et al.(1999)King, Rosen, and Tanner] Gary King, Ori Rosen, and Martin Tanner. Binomial-beta hierarchical models for ecological inference. *Sociological Methods & Research*, 28:61–90, 1999.
- [King et al.(2004)King, Rosen, and Tanner] Gary King, Ori Rosen, and Martin Abba Tanner. *Ecological Inference: New Methodological Strategies*. Analytical methods for social research. Cambridge University Press, Cambridge, UK, 2004. ISBN 0521835135.
- [Kousser(2001)] J Morgan Kousser. Ecological inference from goodman to king. *Historical Methods*, 34: 101–127, 2001. ISSN 0161-5440.
- [Langbein and Lichtman(1978)] Laura Irwin Langbein and Allan J Lichtman. *Ecological Inference*. Sage Publications, Beverly Hills, Calif, 1978. ISBN 0803909411.
- [Lee(2002)] Junhan Lee. Economic voting in south korea. Paper presented at the Annual Meeting of the Midwest Political Science Association, 2002.
- [Lewis(2001)] Jeffrey B Lewis. Understanding king’s ecological inference model: A method-of-moments approach. *Historical Methods*, 34:170–88, 2001.
- [Luttbeg and Gant(1995)] Norman R Luttbeg and Michael M Gant. *American Electoral Behavior, 1952-1992*. F.E. Peacock Publishers, Itasca, Ill, 2nd edition, 1995. ISBN 0875813860.
- [Oh(1999)] John Kie-Chiang Oh. *Korean Politics: The Quest for Democratization and Economic Development*. Cornell University Press, 1999. ISBN 0801484588.

- [Palmquist(1993)] Bradley Palmquist. *Ecological Inference, Aggregate Data Analysis of U.S. Elections, and the Socialist Party of America*. PhD thesis, University of California, Berkeley, 1993.
- [Quandt and Ramsey(1978)] R. Quandt and J Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of American Statistical Association*, 73:730–738, 1978.
- [Robinson(1950)] W. S. Robinson. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15:351–357, 1950.
- [Rosen et al.(2001)Rosen, Jiang, King, and Tanner] Ori Rosen, Wenxin Jiang, Gary King, and Martin Tanner. Bayesian and frequentist inference for ecological inference: The $r \times c$ case. *Statistical Neerlandica*, 55:134–156, 2001.
- [Rosenstone and Hansen(1993)] Steven J Rosenstone and John Mark Hansen. *Mobilization, Participation, and Democracy in America*. New topics in politics. Macmillan Pub. Co, New York, 1993. ISBN 0024036609.
- [Tam Cho(1998)] Wendy K Tam Cho. Iff the assumption fits ... a comment on the king ecological inference solution. *Political Analysis*, 7:143–163, 1998.
- [Thomsen(1987)] Søren Thomsen. *Danish Elections 1920–79: A Logit Approach to Ecological Analysis and Inference*. Politica, Aarhus, Denmark, 1987.
- [Wellhofer(2001)] E Spencer Wellhofer. Party realignment and voter transition in italy, 1987-1996. *Comparative Political Studies*, 34:156–186, March 2001. ISSN 0010-4140.
- [Zellner(1962)] Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368, jun 1962. ISSN 0162-1459.

Appendix A: Identification of the Thomsen Model

Thomsen’s estimator achieves identification by equating individual (utility) correlation with aggregate (transformed vote fractions) correlation. To see when this postulate works, let us formulate the implied correlations at individual and aggregate levels and derive the bias.

Variances	Within-Unit	Across-Unit
Dimensional	σ^2	ω^2
Non-Dimensional	1	τ_t^2

Figure 6: Decomposition of Variances

First, decompose the variance structure of individual utilities. There can be two sources of variances: the dimensional and the non-dimensional that pertain to the latent partisanship dimension. For the entire population, the variation occurs both within and across the observed units, while the aggregate data would only reveal the across-unit variances in ecological data. Thus, the total individual variance can be decomposed into:

The individual correlation would include all four of the variances, while the ecological estimation of the correlation would only involve the across-unit variances, the second column. To find the individual correlation, invert equation (??), and define the left-hand-side variable as voter i 's utility (difference) to vote for a given party over the other party. Denote d_{ij}^* as the underlying individual partisanship for i , D_j^* as its mean, u_{ij} as d_{ij}^* 's deviation from D_j^* . Then we may write within district j ,

$$U_{tij} = \Phi^{-1} [\Pr(d_{tij} = 1)] = \alpha_t + \beta_t d_{ij}^* + e_{tij}$$

where $\text{Var}(d_{ij}^*) = \text{Var}(D_j^* + u_{ij}) = 0 + \sigma^2$ and $\text{Var}(e_{tij}) = 1$

The within-district correlation can then be written as

$$\rho_{within}(U_{1ij}, U_{2ij}) = \frac{\text{Cov}(U_{1ij}, U_{2ij})}{\sqrt{\text{Var}(U_{1ij})}\sqrt{\text{Var}(U_{2ij})}} = \frac{\beta_1 \beta_2 \sigma^2}{\sqrt{\beta_1 \sigma^2 + 1} \sqrt{\beta_2 \sigma^2 + 1}} \quad (26)$$

For the entire nation, variances include the second column, the across-unit variances:

$$U_{tij} = \alpha_t + \beta_t d_{ij}^* + \xi_{tij}$$

where $\text{Var}(d_{ij}^*) = \text{Var}(D_j^* + u_{ij}) = \omega^2 + \sigma^2$ and $\text{Var}(\xi_{tij}) = \tau_t^2 + 1$

Then the individual correlation for the entire nation is

$$\begin{aligned}\rho_{ind}(U_{1ij}, U_{2ij}) &= \frac{Cov(U_{1ij}, U_{2ij})}{\sqrt{Var(U_{1ij})}\sqrt{Var(U_{2ij})}} \\ &= \frac{\beta_1\beta_2(\omega^2 + \sigma^2)}{\sqrt{\beta_1(\omega^2 + \sigma^2) + (\tau_1^2 + 1)}\sqrt{\beta_2(\omega^2 + \sigma^2) + (\tau_2^2 + 1)}}\end{aligned}\quad (27)$$

However, (27) cannot be computed directly since we do not observe individual utilities. Thus, we have to estimate it using the ecological observations. To derive the ecological correlation, invert equation (??) and write

$$\begin{aligned}\Phi^{-1}(D_{tj}) &= \alpha'_t + \beta'_t D_{tj}^* + \varepsilon'_{tj} \\ \text{where } \alpha'_t &= \frac{\alpha_t}{\sqrt{1 + \beta_t^2 \sigma^2}}, \beta'_t = \frac{\beta_t}{\sqrt{1 + \beta_t^2 \sigma^2}} \text{ and } \varepsilon'_{tj} = \frac{\varepsilon_{tj}}{\sqrt{1 + \beta_t^2 \sigma^2}}\end{aligned}$$

with across-unit variances only: $Var(D_j^*) = \omega^2$ and $Var(\varepsilon_{tij}) = \tau_t^2$.

Then the aggregate correlation we obtain is

$$\begin{aligned}\rho_{agg}[\Phi^{-1}(D_{1j}), \Phi^{-1}(D_{2j})] &= \frac{Cov[\Phi^{-1}(D_{1j}), \Phi^{-1}(D_{2j})]}{\sqrt{Var[\Phi^{-1}(D_{1j})]}\sqrt{Var[\Phi^{-1}(D_{2j})]}} \\ &= \frac{\beta_1\beta_2\omega^2}{\sqrt{\beta_1\omega^2 + \tau_1^2}\sqrt{\beta_2\omega^2 + \tau_2^2}}\end{aligned}\quad (28)$$

Now the task is to find under what conditions (28) approximates (27). By assuming $\tau_1 = \tau_2$, the difference between the squared aggregate and individual correlation can be factored as

$$\begin{aligned}\rho_{agg}^2 - \rho_{ind}^2 &= [\text{Equation(28)}]^2 - [\text{Equation(27)}]^2 \\ &= G(\omega^2 - \tau^2\sigma^2) \\ \text{where } G &= \frac{\beta_1^2\beta_2^2\omega^4[\omega^2(\omega^2 + \sigma^2)(\beta_1 + \beta_2) + \tau^2\sigma^2 + 2\omega^2\tau^2 + \omega^2]}{(\beta_1\sigma^2 + 1)(\beta_2\sigma^2 + 1)[\beta_1(\omega^2 + \sigma^2) + (\tau^2 + 1)][\beta_2(\omega^2 + \sigma^2) + (\tau^2 + 1)]}\end{aligned}\quad (29)$$

Equation (29) implies that the aggregate correlation equals the individual correlation when the ratios of the variances shown in Figure 6 are constant — $\sigma^2 : \omega^2 = 1 : \tau^2$. In other words, if the proportion

of the aggregate variance to total variance is the same in both systematic and non-systematic variances, the ecological correlation is an unbiased estimate of the individual correlation, and thus, the Thomsen estimator is identified (Achen 2000).

$$\text{Identification Condition: } \frac{\omega^2}{\omega^2 + \sigma^2} = \frac{\tau_t^2}{\tau_t^2 + 1} \quad (30)$$

A couple of substantive interpretations are implied by the identification condition. First, without spatial heterogeneity, the ecological correlation is unbiased. This is equivalent to Thomsen’s original “homogeneous district assumption” that districts are (almost) perfectly similar dimensionally. When there are no cross-district variances, $\tau_t^2 = \omega^2 = 0$, and the condition above is trivially satisfied. Moreover, in such cases, the expression G in (29) becomes zero, eliminating the bias. However, the assumption is still troublesome in the sense that it assumes away any cross-unit variances, which is, in fact, assuming away of the aggregation problem. In this case, equation (28) can take any value between zero and 1 based on the sampling error. The only possible implication of Thomsen’s district homogeneity assumption is that the total individual correlation (28) is equivalent to estimating a within-district correlation (26), the entire nation being one large district. Then, for (28) to approximate this quantity, our observations should be something like a randomly sampled individual probability to vote for a party, rather than the expectation of it in each district.

Second, with spatial heterogeneity, the aggregation bias of the ecological correlation can be reduced by fine-grained data. Note that the quantities in (30) will increase and get closer to unity as the between-unit variances increase and the within-unit variances decrease. Instead of assuming away the aggregation problem, we may allow that voters of similar preferences are more likely to live in the same unit, which is a more realistic assumption, especially in the case of a voter transition setup. In such cases, the smaller the unit size is, ω^2 and τ_t^2 will become larger relative to their counter parts, making the quantities in (30) closer to unity, and satisfying the identification condition. Additionally, a smaller aggregate unit size will more likely satisfy the constant within-unit variance assumption. In other words, the estimator is disaggregation-consistent.

Appendix B: Estimation Group Turnouts, Entrance and Exit Rates

Table 7: (Appendix) Estimated Turnout Rates by Different Demographic Groups in South Korean Elections

	Year	Overall	Educated	Old	Young	Christian
Legislative Elections	1978	0.741	0.482	0.988	0.455	0.781
	1981	0.753	0.639	0.908	0.656	0.624
	1985	0.819	0.850	0.798	0.833	0.743
	1988	0.737	0.695	0.914	0.599	0.668
	1992	0.716	0.543	0.998	0.470	0.719
	1996	0.628	0.584	0.835	0.367	0.681
	2000	0.569	0.529	0.727	0.351	0.534
Presidential Elections	1987	0.861	0.917	0.790	0.933	0.873
	1992	0.823	0.794	0.909	0.719	0.760
	1997	0.797	0.848	0.781	0.814	0.909

Table 8: (Appendix) Estimated Entrance Rates of Different Demographic Groups in South Korean Elections

	Year	Overall	Education	Young	Old	Christian
Legislative Elections	1978	0.439	0.511	0.100	0.468	0.414
	1981	0.341	0.568	0.200	0.080	0.580
	1985	0.504	0.825	0.624	0.534	0.664
	1988	0.321	0.644	0.200	0.817	0.360
	1992	0.385	0.686	0.350	0.300	0.785
	1996	0.219	0.637	0.363	0.020	0.560
	2000	0.192	0.457	0.100	0.092	0.252
Presidential Elections	1987	0.782	0.074	0.946	0.799	0.109
	1992	0.617	0.201	0.339	0.797	0.230
	1997	0.514	0.143	0.342	0.754	0.058

Table 9: (Appendix) Estimated Exit Rates of Different Demographic Groups in South Korean Elections

	Year	Education	Young	Old	Christian	Overall
Legislative Elections	1978	0.541	0.592	0.000	0.000	0.131
	1981	0.322	0.694	0.019	0.365	0.102
	1985	0.125	0.100	0.097	0.220	0.077
	1988	0.292	0.322	0.191	0.228	0.171
	1992	0.515	0.900	0.001	0.323	0.167
	1996	0.442	0.850	0.056	0.245	0.210
	2000	0.427	0.662	0.018	0.306	0.206
Presidential Elections	1987	0.894	0.000	0.160	0.841	0.121
	1992	0.756	0.379	0.111	0.698	0.143
	1997	0.811	0.123	0.175	0.832	0.134