The Chilean Plebiscite: Projections Without Historic Data

Eduardo Engel, Achilles Venetoulias

Journal of the American Statistical Association, Volume 87, Issue 420 (Dec., 1992), 933-941.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199212%2987%3A420%3C933%3ATCPPWH%3E2.0.CO%3B2-U

The Chilean Plebiscite: Projections Without Historic Data

EDUARDO ENGEL and ACHILLES VENETOULIAS*

On October 5, 1988, Chileans decided by plebiscite to oust General Pinochet from power and have free presidential elections in 1989. This article describes the projections that the authors made for the results of the plebiscite from early returns. From a statistical point of view, what made these projections different from those made in other countries was the complete lack of historic data. Furthermore, the Pinochet government carried out a campaign to discredit the projection effort. Uncertainty about both the data and the unpredictable political climate on the night of the plebiscite influenced the choice of the statistical methodology. The predictions, based on a 10% sample of the first one-third of the votes counted, were within one-half a percentage point of the true outcome. The described methodology could prove useful in projections of other elections that will take place under similar conditions (e.g., in Eastern Europe).

KEY WORDS: Election; Prediction; Projection; Sampling; Stratification.

On October 5, 1988, Chileans voted on whether they wanted to have General Pinochet remain in power for another eight years (YES alternative) or have free presidential elections in December 1989 (NO alternative). Voter registration for the YES/NO plebiscite began in February, 1987. At almost the same time, a group of public figures formed the Committee for Free Elections. Their original goal was to achieve an open presidential election instead of the YES/NO vote stipulated by Pinochet's constitution. They did not, however, succeed in this goal, and subsequently they decided to have a projection of the outcome of the plebiscite on the basis of early returns. The possibility of conducting exit polls was discarded, because it was believed that people would have been afraid to respond candidly.

There were at least two scenarios according to which such a projection could prove important. First, there was fear among Pinochet's opponents that the General would use some prefabricated incident to stop the vote count if he realized that he was losing the election. In this case a fast and reliable projection could prove a good deterrent and/or an effective proof for the opposition's victory. Second, there was the possibility of a close election in which both sides could claim victory. Their supporters would then take it to the streets to celebrate, and violence could erupt. In this case a fast and reliable projection could provide influential political actors who might be able to calm the situation (e.g., the Catholic Church) with accurate and objective information. Such information might have indicated that one side was the clear winner or that the election was too close to call and it was, therefore, necessary to wait for later returns (see Table 4).

* Eduardo Engel is Assistant Professor, Dept. of Industrial Engineering, University of Chile, Santiago, Chile. This work was done while he was working on his Ph.D. in economics at MIT. Achilles Venetoulias is Research Specialist, Capital Markets, Bank of America, 555 California Street, San Francisco, CA 94104. This work was done while he was teaching at the Sloan School of Management, MIT. For the duration of this project Achilles Venetoulias was partially supported by National Science Foundation Grant IST-8420614, by U.S. Army Grant DAAG-29-84-K-0207 and by the Center for Computational Research in Economics and Management Science. The authors thank Sergio Molina and Igor Saavedra for providing them with the opportunity to do this work; Persi Diaconis and John Tukey for their encouragement and advice; and Eduardo Bitrán, Guido Del Pino, Raúl Gormaz, Ross Ihaka, Warren Mitofsky, and Alan Zaslavsky for their help and suggestions.

This article is about the projections we made for the Committee for Free Elections. One of the authors, Eduardo Engel, was in Chile for the plebiscite and coordinated the implementation of the methodology. From a statistical point of view, what made these projections different from those made in other countries was the fact that methodologies used elsewhere relied strongly on data from recent elections. No such data were available for Chile, because the last presidential and congressional elections had taken place in 1970 and 1973. Furthermore, Pinochet's government carried out a campaign to discredit this effort. Consequently, the choice of the statistical methodology was influenced by (1) the uncertainty about the distributional form of the data and (2) the unpredictable political climate on election night. Our predictions, based on a 10% sample of the first one-third of the votes counted, were within one-half a percentage point of the final outcome of the plebiscite.

The methodology presented in this article is suitable for election projections in countries where elections have not taken place for a long time; for example, in Eastern Europe.

1. THE DATA

Voter registration for the plebiscite began in February 1987, and the opposition decided to participate in March 1988. Voters registered and voted in the county where they lived, the men separately from the women. Every 350 voters formed a "table," which was the basic voting unit; as soon as 350 voters registered at a table, that table was closed. Experience from previous elections (before 1973) and opinion surveys suggested that voter participation would be high (i.e., above 85%). There was a total of 22,131 tables throughout the country, and on election day each polling center housed approximately 50 tables. The results were reported on a "per table" basis.

The Committee for Free Elections decided to rely on its own data sources. To this end, it set up a nationwide organization for gathering firsthand information on the returns.

Journal of the American Statistical Association
December 1992, Vol. 87, No. 420, Applications and Case Studies

There were volunteers at 2,200 tables (i.e., 10% of the total number of tables) who watched the vote count and reported the results to a central computing center in Santiago, where projections were made. Security precautions were taken to guard the system from any potential interference by the government.

The decision to gather data independently from either the government or the opposition aimed at two goals: (1) to stress the independence of the Committee for Free Elections from the political parties of the opposition, and (2) to ensure the objectivity and the accuracy of the projections. On the night of the plebiscite, partial counts announced by the government underestimated the true NO vote, and those announced by the opposition overestimated it. (The first partial count by the government underestimated the NO vote by almost 10%; the corresponding count by the opposition overestimated it by approximately 7%.) The fear of biased partial counts made independent data gathering highly desirable.

2. STRATIFICATION

The goal of the Committee for Free Elections was to have a forecast of the final result of the plebiscite from the first available returns. Thus our objective was to make an accurate projection long before the sample of 2,200 monitored tables was complete. (A projection based on the complete sample would have taken too long and would have been of no interest.) The polls started closing at 6:00 P.M., and we expected to make a projection with approximately 600 tables by 9:00 P.M.

The statistical methodology behind most election projections relies on information available from previous elections. Projections are based on a preselected sample of polling places chosen so that it adequately represents the overall voter behavior (see, for example, Bernardo 1984). The selection of the sample is usually made on the basis of the data from recent elections. For Chile no such data were available, because the last presidential and congressional elections had taken place in 1970 and 1973.

The most basic problem in projecting the outcome of an election from early returns is that these returns may not be representative of the whole population. When the early returns do not represent a random sample from the entire population, the unweighted sample estimate is necessarily biased. This problem was addressed by stratifying the population; the tables were stratified according to variables that were expected to capture potential differences in the voting patterns. The use of predetermined weights (for the different strata) corrects for the fact that the early returns are not a random sample of the entire population and that an early unweighted sample estimate may, therefore, be biased. As long as there are no systematic trends in the returns *of any stratum,* stratification eliminates the bias of the estimates. Furthermore, stratification produces another important gain in that it reduces the standard error of the projection estimates. In addition to stratification, a simple test was implemented for detecting time trends in the returns of each stratum (Sec. 7); during the night of the plebiscite, this test showed no evidence of time trends in any of the strata.

As it turned out, the YES vote would have been overestimated if early returns were assumed to be representative of the whole population. The returns from low income neighborhoods came in much later than the rest, and, because the voters in those neighborhoods were mostly supporters of the NO alternative, treating the complete set of incoming tables as a random sample from the whole population would have resulted in a bias in favor of the YES alternative.

Both the possibility of overlooking additional sources of error (say, due to possible nonrandomness in the order of arrival of the returns) and the margin of error decreased as the number of strata increased. By the same token, a larger number of strata required a larger number of available tables at the moment of the projections. The reason was that a minimum number of tables had to be present in every stratum to assess the corresponding margin of error.

The variables used for stratification had to be observable at the level of tables, as these were the actual data. Thus, for example, age could not be used as a stratification criterion even though it was commonly believed that younger voters would vote mostly against Pinochet. The variables used for stratification were demographic: sex, city size, the closing date for the registration process of a table, and socioeconomic level.

Sex was chosen because men and women voted at different tables, and in Chile men have always exhibited more liberal voting patterns than women. City size was chosen because all the preplebiscite polls indicated that Pinochet had more support in rural areas than in urban areas (possibly because of the tighter political control exercised by the dictatorship in smaller cities). Tables were divided into three groups according to the size of the city to which they belonged: cities with more than 200,000 inhabitants, cities with more than 30,000 but less than 200,000 inhabitants, and rural areas (towns with less than 30,000 inhabitants).

The closing date of a table was chosen because informal evidence suggested that the opposition to Pinochet was stronger among those who registered to vote late than among those who registered early (due to the fact that the opposition did not decide to actively participate in the plebiscite until March 1988, a whole year after registration had begun). Tables in large- and medium-sized cities were divided into two groups, according to whether they were completed (roughly) during the first half or the second half of the registration process. Tables in rural areas were not classified according to this variable.

Socioeconomic level was chosen because all the opinion polls agreed that there was a larger proportion of Pinochet supporters in upper class neighborhoods. In Chile's larger cities (Santiago, Valparaíso-Viña del Mar, and Concepción-Talcahuano) it was possible to differentiate socially between different geographic neighborhoods, that is, the geographic units used to differentiate among tables corresponded to predominantly upper, middle, or lower class neighborhoods (as identified by market research institutions). Of course, a geographic unit could not be described solely in terms of a single socioeconomic level, and the true importance of socioeconomic level was underestimated to some degree. Nevertheless, this did not preclude a very useful stratification; the

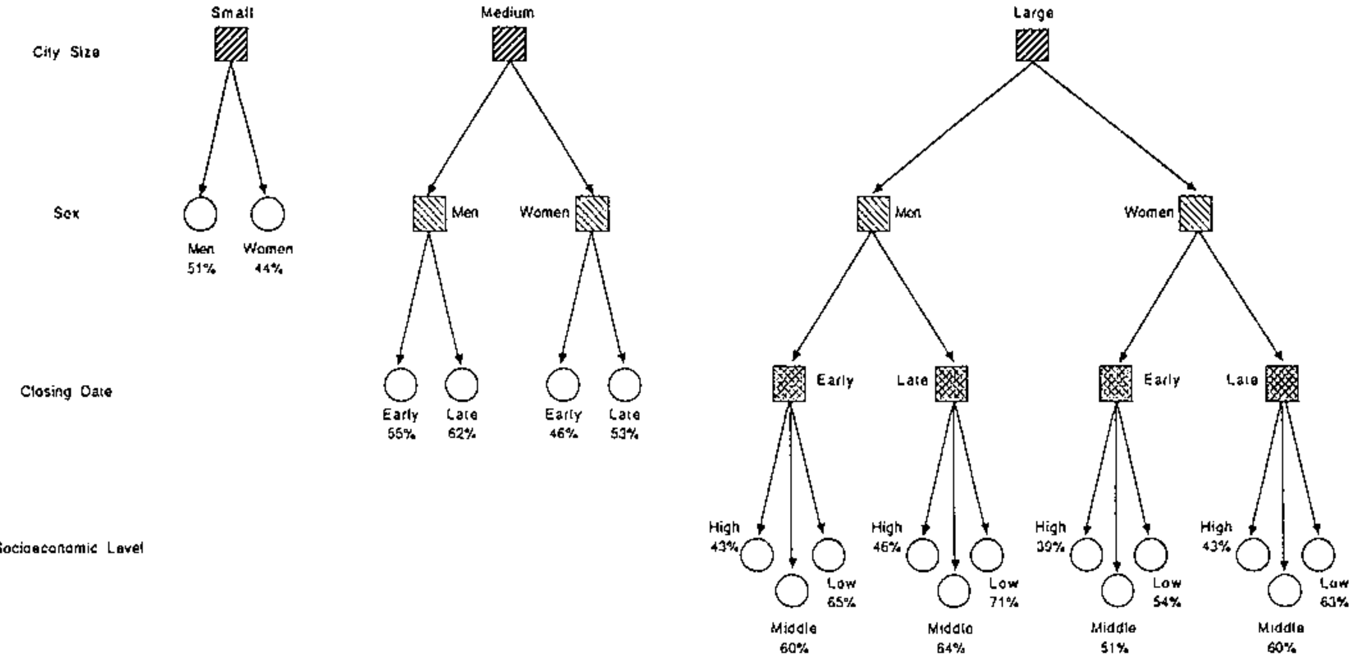

Figure 1. The 18 Strata and Their Respective Percentages of NO Votes.

tables in the largest cities were stratified according to whether they were located in neighborhoods of lower, middle, or upper socioeconomic level.

The sample of 2,200 tables was selected by proportional sampling: For a given stratum a random sample was drawn that consisted of 10% of the total number of tables in that stratum. Any more sophisticated sampling criterion would have required prior knowledge of some measure of dispersion within the strata; such information was not available.

The adopted stratification resulted in 18 strata, which are shown schematically in Figure 1. Table 1 shows the percentages of tables for the different levels of the stratifying variables and the results for the plebiscite within these subpopulations. It should be kept in mind that closing date is nested only within medium and large cities and that socioeconomic level is nested only within large cities. (Hence the relative sizes of the socioeconomic and the closing date strata do not add up to 100%.) Also, the percentages of YES and NO votes add up to less than 100% because of blank and invalid votes. Table 1 shows that the defined strata were more homogeneous than the population of the entire country. The percentages of YES votes for the 18 strata ranged from 32% to 62%.

The significance of the individual variables and the degree of interaction between them was evaluated, after the plebiscite, with an analysis of variance. This analysis confirmed, as expected, the significance of all four stratification variables; it also showed that two pairwise interactions (between sex and socioeconomic level and between sex and closing date) were significant.

3. THE PROJECTION: THE DIFFERENCE METHOD

The estimation method used to carry out the projections used the difference between the YES and the NO votes as its basic unit; for this reason, it is referred to as the *difference* method. The method assumed that the available returns (at any given instant in time) *for any given stratum* could be treated as a random sample. A simple test of this assumption is presented in Section 7. The method is described in this section, with the following notation: Let M_k be the total number of tables in stratum k ($k = 1, 2, \ldots, 18$), let m_k be the number of tables in stratum k whose returns have been received when the projection is made, let y_{ik} be the number of YES votes in the ith received table in stratum k, let n_{ik} be the number of NO votes in the ith received table in stratum k, let b_{ik} be the number of blank and null votes in the ith received table in stratum k, and let d_{ik} be the difference between YES and NO votes in the ith received table in stratum k.

The difference method made a projection at a given instant of time as follows. The difference between the number of YES and NO votes in every stratum at that moment was computed. These differences were weighted according to the true number of tables in the corresponding stratum (including those whose result had not yet been received and those that were not included in the sample). The weighted differences were added to yield a projection for the nationwide

Table 1. Percentages of YES and NO Votes According to the Stratifying Variables

Stratifying variable	*Relative size*	*Yes*	*No*
Large cities	45%	39%	59%
Medium cities	38%	43%	54%
Rural and small towns	17%	49%	48%
Male	52%	39%	59%
Female	48%	46%	51%
Registered early	37%	45%	53%
Registered late	46%	38%	60%
High socioeconomic level	5%	56%	42%
Middle socioeconomic level	28%	39%	59%
Low socioeconomic level	12%	34%	63%

difference between YES and NO votes. This was the quantity of main interest; it was positive if and only if the projection predicted the YES alternative as the winner. A related question was whether the available returns at that point allowed a determination as to which side would win. This was equivalent to whether zero belonged to the confidence interval for the estimated difference between the YES and the NO votes. The difference method estimates for the numbers of voters were

$$\bar{y}_k = \frac{1}{m_k} \sum_i y_{ik}$$

= average YES vote in received tables in stratum k,

$$\bar{n}_k = \frac{1}{m_k} \sum_i n_{ik}$$

= average NO vote in received tables in stratum k,

$$\bar{b}_k = \frac{1}{m_k} \sum_i b_{ik} = \text{average blank and null vote in received tables in stratum } k,$$

$$\bar{d}_k = \frac{1}{m_k} \sum_i d_{ik} = \text{average difference between YES and NO votes in received tables in stratum } k,$$

$$\hat{y} = \sum_k M_k \bar{y}_k = \text{projected number of YES votes},$$

$$\hat{n} = \sum_k M_k \bar{n}_k = \text{projected number of NO votes},$$

$$\hat{b} = \sum_k M_k \bar{b}_k$$

= projected number of blank and invalid votes,

and

$$\hat{d} = \sum_k M_k \bar{d}_k$$

= projected difference between YES and NO votes.

The same method yielded a projection of YES and NO percentages. The total numbers of YES and NO votes were projected separately (with the same approach) and were converted into percentages. This approach required that the total number of voters also be estimated. These estimates were

$$\hat{V} = \hat{y} + \hat{n} + \hat{b} = \text{projected number of actual votes},$$

$$\widehat{py} = \hat{y}/\hat{V} = \text{projected percentage of YES vote},$$

and

$$\widehat{pm} = \hat{n}/\hat{V} = \text{projected percentage of NO vote}.$$

The accuracy of the difference estimator was derived from standard results in sampling theory (see Cochran 1977, p. 95). First, the variance of the estimator was obtained for each stratum separately; then, the individual variances were added to compute an estimate for the overall variance. An assessment of accuracy at this stage would have been significantly harder if the total number of YES and NO votes had been treated separately (rather than together as a difference). The estimate for the variance of the projected difference between YES and NO votes was

$$\hat{s}^2 = \sum_k \frac{M_k^2}{m_k}\left(1 - \frac{m_k}{M_k}\right)\frac{1}{m_k - 1}\sum_{i=1}^{m_k}(d_{ik} - \bar{d}_k)^2, \quad (1)$$

and the estimate for the standard deviation of the estimates of the percentages was

$$\hat{s}_p = \hat{s}/2\hat{V} = \text{estimate of the standard error for the percentage of both YES and NO votes.}$$

The expression for $\hat{s}^2$ is derived in the Appendix and assumes (1) that the correlation between the numerator and the denominator is negligible, and (2) that the coefficient of variation of the denominator (compared to that of the numerator) is small. It is important to mention that the standard error estimate $\hat{s}_p$ was not really needed; what was needed was a p value for the statement "this side has won." Such a p value could have been obtained from the standard deviation estimate, $\hat{s}$, of the difference between YES and NO votes.

One of the main characteristics of the difference method is its simplicity. If necessary, it could have been implemented with 19 pocket calculators (one for each stratum and one for combining the results). This was an important consideration, because the possibility of government intervention could not be ruled out. Had we been forced to leave the computing center, we still could have carried out the calculations somewhere else by hand.

Figure 2, taken from the official results of the plebiscite, shows the projections of the difference method for the percentage of NO votes. These projections are based on the actual returns of the plebiscite and are shown as a function of the available tables. For each projection, the actual estimate and the 99% confidence interval are shown (with the confidence bands formed by the connected points). For comparative purposes, the true percentage of NO votes (54.7%) is also shown as a horizontal line.

4. SIMULATIONS

The performance of the difference estimator was evaluated before the plebiscite with extensive simulations. This was a necessary measure for assessing the sensitivity of the method to various assumptions. For example, we (wrongly) expected that data from rural areas would arrive later than data from urban areas. This combination of events would have underestimated the true importance of rural strata and overestimated the NO vote.

The estimation procedure could not have been tested on historic data from recent elections, because such data did not exist. Even worse, it was difficult to guess the characteristics of the data without having seen any other similar data. As a result, it was basically impossible to validate the truth of any assumptions (e.g., normality of the estimators, normality of the NO votes in each stratum, uniform rate of arrival for the returns across strata). At the same time, precautions had to be taken to avoid embarrassing errors (i.e., projecting a false victory for one of the two options). For

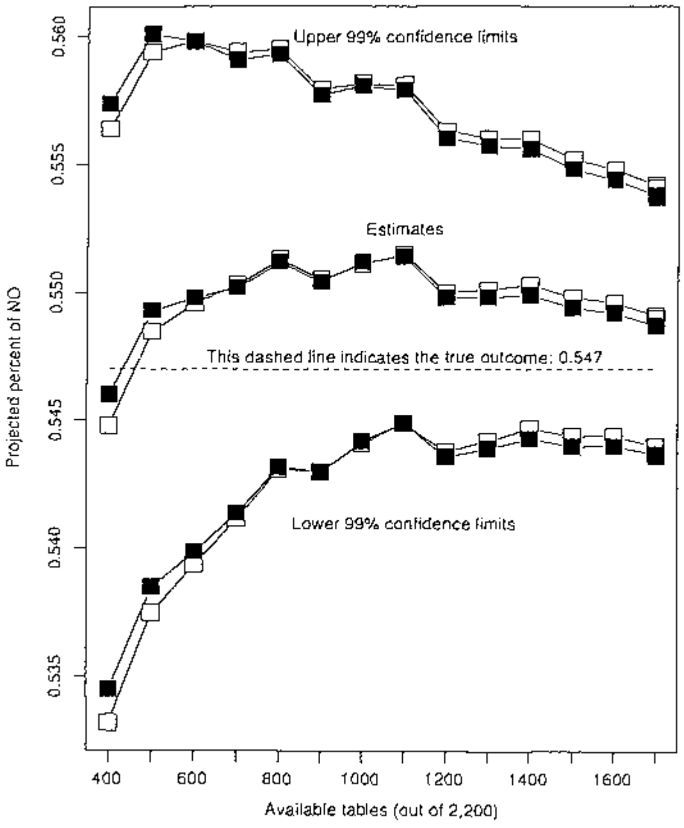

Figure 2. Projections of the Percentage of NO at 99% Confidence Level. □ indicates the difference method. ■ indicates the percentage method (see Sec. 6).

these reasons, the sensitivity of the projection method was tested with simulations.

Each simulation created a population of 22,000 tables for the whole country, selected an appropriate subset as the observed sample, and used the estimation method to make a projection on the basis of that sample. These projections were compared to the "true" outcome, which was obtained from the whole population. The factors that were likely to affect the behavior of the estimator were modeled as input parameters to the simulations. The values of these parameters were allowed to vary, and the performance of the estimator was evaluated for different combinations of parameter values. The input parameters were the size of the observed sample (i.e., the number of reported tables), the percentages of voter turnout and their standard deviations (per stratum), the percentages of NO (or YES) votes and their standard deviations (per stratum), and the distribution of the NO votes and the presence (or absence) of significant numbers of outliers in the observed samples.

The percentages of voter turnout and the percentages of NO votes for each stratum were selected to reflect prior beliefs about the behavior of the particular stratum. Voter turnout was expected to be high: The means of the voter turnouts (for the 18 strata) ranged from 80% to 90%, and their standard deviations were set to either 5% or 10%. A normal distribution was used to generate the percents of voter turnout for the 18 strata.

The percentages of NO votes were expected to fluctuate more: The means of the NO percentages ranged from 35% to 65%, and their standard deviations were set to either 10% or 25%. These parameters were chosen, differently for each stratum, so that they would capture the expected pro-government constituency of the strata; they were chosen in consultation with the political scientists and the sociologists involved in the projection effort. Three different distributions were used to generate the simulated samples of NO votes: a normal, a mixture of normals with fat tails, and a slash distribution. These corresponded to three degrees of departure from the desired normality. Also, the populations from certain strata were selectively contaminated with outliers. When a particular stratum was selected, some of its tables were arbitrarily set to consist of 80% YES votes and 20% NO votes. This choice reflected the fact that some tables in districts near military bases consisted mostly of members of the armed forces, who were expected to strongly support General Pinochet.

For each combination of parameter values, the simulation was repeated a number of times. This number was determined sequentially. A group of 25 simulations were performed; (only) if the particular combination of parameter values appeared to adversely affect the performance of the estimator, another 100–200 simulations were performed. The simulations were implemented in the New S (Becker, Chambers, and Wilks 1988) and were run on a Sun 3/160 (with a Sun floating point accelerator and 8 megabytes of memory). The run time performance of the simulation module was essentially determined by the size of the observed sample (i.e., by the number of available tables). A projection took on average 3.07 seconds with 386 tables, 3.70 seconds with 735 tables, 4.33 seconds with 1,127 tables, and 5.14 seconds with 1,603 tables. These run times were sufficiently small to enable adequate experimentation with different combinations of parameter values.

The conclusions from the simulations were encouraging. The estimation method was fairly robust to the adversity of the hypothetical scenarios and performed adequately well even if the conditions were not ideal. As an example, consider a scenario where the populations from certain strata were contaminated with outliers. The contamination was introduced by forcing 10% of the tables for some strata to be outliers. Eight strata were contaminated: They were chosen to be the eight most favorable for the NO vote. This scenario was created so that it would tend to introduce biases against the NO vote. Table 2 shows the simulated accuracy of the projection method under these conditions. For this table the percentages of NO votes were generated by either a normal

Table 2. Simulated Coverage Probabilties of Normal Confidence Intervals

	Normal distribution		*Slash distribution*	
Tables received	*95% Confidence interval*	*99% Confidence interval*	*95% Confidence interval*	*99% Confidence interval*
386	.94	.98	.91	.98
735	.96	.99	.95	.99
1,127	.95	.98	.97	.99
1,603	.96	.97	.93	.97

or a slash distribution. These coverage probabilities were obtained from 200 simulations for each case.

The results of Table 2 were representative of the results for a wide range of simulated scenarios. It became obvious from the simulations that neither the 95% nor the 99% confidence intervals could be trusted to be real 95% or 99% intervals. This, of course, was anticipated. The more informative conclusion was that the loss in accuracy could be tolerated, especially if some conservative precautions were taken (e.g., report 99.9% intervals as 99% intervals). This conclusion concurred with John Tukey's advice when, before the plebiscite, we described the statistical methodology to him: "My experience has shown that a theoretical 99% confidence interval corresponds to what really is only a 95% confidence interval." (Tukey 1987, personal communication). The robustness of the method to departures from ideal conditions permitted considerable optimism about the success of the projections.

5. THE NORMALITY ASSUMPTION

Once estimates and their standard deviations had been calculated, the next step was to obtain confidence intervals. To do so, assumptions were needed for the distributions of the estimator. Ideally, it would have been desirable to have computed bootstrap confidence intervals (see, for example, Efron 1982, 1987; Efron and Tibshirani 1986; and Hall 1988), because they generally require weaker distributional assumptions than their classical (i.e., normal theory) counterparts. This was an appealing approach, because the complete lack of historic data precluded the validation of any distributional assumptions. Unfortunately, it was also an infeasible approach, because the computation involved would have taken much too long. Therefore, it became necessary to assess on a priori grounds the expected degree of normality for the distribution of the difference estimator.

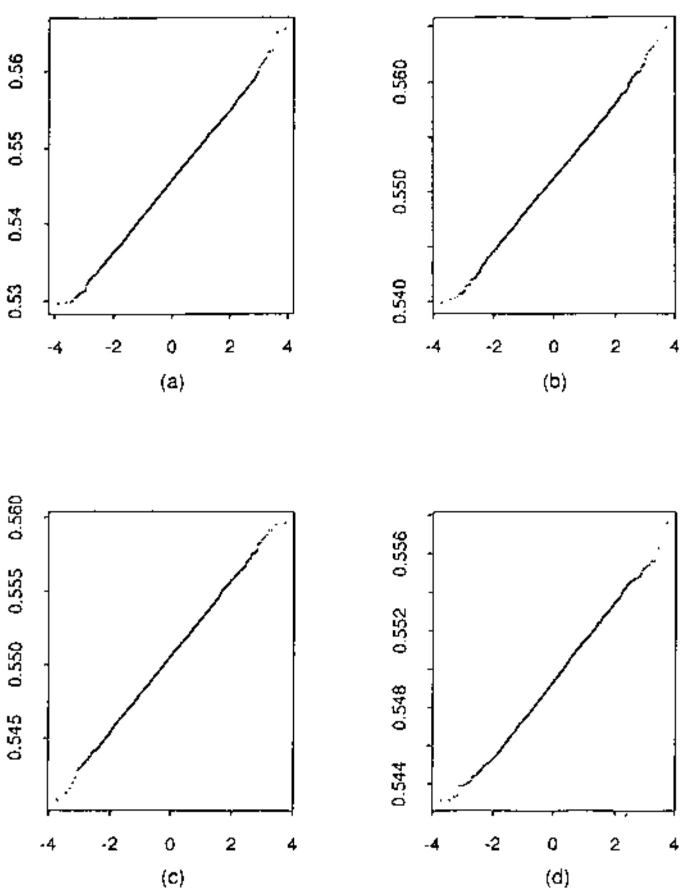

Figure 3. Probability Plots for the Bootstrap Distributions of the Difference Method. The quantiles of the standard normal distribution are on the x axis and the quantiles of the empirical distribution of projections are on the y axis. (a) 386 available tables; (b) 735 available tables; (c) 1,127 available tables; (d) 1,603 available tables.

Table 3. 99% Confidence Intervals for the Percentage of NO

Tables received	Normal intervals			Bootstrap intervals		
	L99	$\widehat{NO\%}$	U99	L99	$\widehat{NO\%}$	U99
386	53.30	54.47	55.64	53.32	54.47	55.61
735	54.28	55.14	56.00	54.26	55.14	56.03
1,127	54.42	55.08	56.73	54.43	55.08	55.73
1,603	54.46	54.98	55.50	54.48	54.98	55.50

NOTE: L99 and U99 denote the lower and the upper endpoints of the 99% confidence intervals and $\widehat{NO\%}$ denotes the corresponding estimate (i.e., the center of the interval).

A priori, there were strong arguments in favor of normality for the difference estimator, because the central limit theorem entered twice in its computation: first, when the difference between YES and NO votes for each stratum was estimated, and second, when these estimates were added for the final projection. Thus normality for the values of the difference estimator seemed justifiable, even if the distributions generating those values were not normal at the level of tables.

The actual data from the plebiscite confirmed this assertion: Some strata populations differed significantly from normality, but the values of the difference estimator were indeed normal. This fact was verified, after the plebiscite, by extensive bootstrapping of the estimation process. The bootstrap samples were constructed from the actual returns of the plebiscite as follows. First, a point in time was chosen and the corresponding observed sample (from the 2,200 monitored tables) was determined. Second, the tables in a given stratum of the observed sample were sampled with repetition to create a bootstrap sample of the same size as the observed one. Once such a bootstrap sample was generated for each stratum, the estimation process was carried out. A total of 5,000 bootstrap samples were generated, and the bootstrap distribution of the estimator was obtained. This analysis was repeated for four different points in time, namely for observed samples of 386, 735, 1,127, and 1,603 tables.

Probability plots (qq plots) of the bootstrap distributions clearly supported the normality assumption for the difference estimator (see Fig. 3). Furthermore, the 99% bootstrap confidence intervals were identical to those suggested by normal theory. (The bootstrap confidence intervals were obtained with the percentile method; see Efron 1982 for details.) Table 3 presents a comparison between normal theory and bootstrap confidence intervals. Discrepancies appeared only at the extreme tails (at four or more standard deviations from the mean), and even then they were probably due to the inadequacy of the number of bootstrap samples (i.e., 5,000 bootstrap replications may just not have been enough to detect differences so far out in the tails of the distribution).

6. AN ALTERNATIVE ESTIMATION METHOD

The difference estimator was unbiased but would have resulted in a relatively large error if the number of voters per table varied substantially within any given stratum. If this phenomenon had occurred, and if there were no correlation between the number of voters and the percentage of YES votes per table in any given stratum, then an estimator based on percentages instead of on actual results could be expected to perform more accurately than the difference estimator. This motivated the development of a second estimation method, the *percentage* method, which used the percentage of YES (and NO) votes at every table as its basic unit.

The percentage method is sketched out below for the percentage of YES votes. (If blank and null votes did not exist, this would be equivalent to considering the percentage of NO votes; this subtle difference was, for all practical purposes, negligible.) Approximate formulas to assess the accuracy of the percentage method were developed but are not presented here.

The percentage method estimated the percentage of YES votes in a given stratum by the average of the percentages of the YES votes in the tables of that stratum. The nationwide percentage of YES votes was estimated by a weighted average of the individual stratum estimates. Ideally, the weights would have been equal to the percentages of voters who belonged to the different strata. These quantities, however, required knowledge of the true voter turnout in every stratum and were not known before the vote count was completed. Instead, the weights were estimated from the observed voter turnout in the available sample.

The percentage method requires additional assumptions to produce an asymptotically unbiased and consistent estimator. One such assumption is that the actual number of voters and the percentage of YES votes at tables within any given stratum are independent. This assumption seemed realistic; a test implemented to detect any departures from it showed no such evidence during the night of the plebiscite (Sec. 7). Estimating percentages at the stratum level by pooling available returns would have provided an unbiased estimate equivalent to the stratum estimates of the difference method. This idea was abandoned because its estimates could not have been more precise than those derived from the difference method.

Compared to the difference method, the percentage method presents various drawbacks. It requires additional assumptions to yield asymptotically unbiased estimates, it provides a priori less support for the normality of the resulting estimator (see Sec. 5), and it is less amenable to calculations under an emergency (see Sec. 3). The only advantage of the percentage method is that it may be expected to be more accurate when voter turnout varies greatly across tables. Yet even this potential advantage is not compelling, because part of the additional accuracy that is gained at the stratum level is lost in the estimation of voter participation in every stratum. By comparison, in the difference method the appropriate measure of stratum size is given by the total number of tables in the stratum, which was a known quantity.

The difference method and the percentage method would have led to identical estimates if all the tables in any given stratum had exactly the same number of voters. The nature of the registration process ensured that initially there were 350 registered voters at every table, except for the last table in every county. (There was a significant number of such tables only in rural areas.) In reality, of course, the actual number of voters per table was smaller than 350. The average number of voters was 327 (with a standard deviation of 29.84). Deaths, relocations, registration cancellations, and abstentions accounted for this discrepancy. Abstention rates in Chile have always been much lower than those in the U.S., usually around 10%. Opinion surveys held before the plebiscite showed that this pattern was likely to prevail.

On the night of the election, the percentage method was used only once, as an additional check just before the first public announcement was made. Its projection coincided with that of the difference method (see Fig. 2). Furthermore, the returns showed that voter participation was even higher than was expected; the actual abstention rate was only 2.7%. For this reason it was deemed unnecessary to continue using the percentage method, and all additional projections were made using the difference method.

7. ADDITIONAL PRECAUTIONS

The difference (and the percentage) method assumed that the available returns (at any given instant in time) *for any given stratum* could be treated as a random sample. This assumption could have been wrong if there were "forgotten variables"; that is, overlooked variables that should have been used in the stratification. A forgotten variable would have introduced bias in the estimates if there was some dependence between that variable and the arrival times of the returns. This observation motivated a check for time trends in the data.

The following simple test, from Alan Zaslavsky, was built into the software that made the projections (Zaslavsky 1988, personal communication). The data in each stratum were divided into two halves, according to the chronological order of arrival of the tables in that stratum. Every time the program ran, it computed the difference and the percentage estimators from both halves of the data for each stratum. Any trend (say, in the order in which the data were coming in) could have been detected by identifying significant discrepancies between the estimators from the first half and the second half. This test also would have detected any correlation between table size and election results within a given stratum (because returns from tables with fewer voters were expected to arrive first). The test did not show any sign of time trends for either method throughout the night of the plebiscite, however.

The software also counted the number of monitored tables in every stratum each time it made a projection. This step, from Raúl Gormaz, was not necessary (because these numbers remained constant), but it provided an extra check on the consistency of the data base effectively at no extra cost (i.e., computing time) (Gormaz 1988, personal communication). Only a few hours before the plebiscite, this seemingly naive precaution led to the detection and elimination of a substantial data entry error and averted a debacle. (When some final changes were made to the data base, more than 200 tables had been assigned to the wrong stratum.)

8. ELECTION NIGHT

During election night, 99.99% confidence intervals (plus or minus four standard deviations) were reported as 99% confidence intervals. There were three reasons for this conservative choice. First, the cost of making a potential error (e.g, predicting a false victory for either option) was enormous. Second, conversations with politicians and journalists in Chile during the days before the plebiscite led to the conclusion that public opinion perceived 99% confidence statements as almost certain but 95% confidence statements as unreliable. Furthermore, the public did not much care about the difference between a 99% and a 99.99% confidence interval. Third, there were John Tukey's advice and the results from the simulations, which strongly suggested such a precaution (see Sec. 4).

Table 4 shows the projections at four instants in time, including the two projections made public (735 and 1,603 tables). The table presents the projections for the percentage of the NO votes and the corresponding 99.99% confidence intervals. In Table 4, the first column shows the actual time of the projection, the second column shows how many tables had been received by that time, and the third column shows the percentage of the total number of tables that the received tables represented. The fourth and sixth columns show the lower and upper endpoints of a 99.99% confidence interval for the percentage of NO votes, and the fifth column shows the corresponding estimate. The last column shows the smallest difference (in percentages) between the NO and the YES vote for which at that time it would have been possible to call the winner. For example, it was possible to project the winner in a 51.5%–48.5% race by 9:30 P.M., in a 51.1%–48.9% race by 10:00 P.M., and in a 50.9%–49.1% race by 11:30 P.M.

The first projection, based on 735 tables, was made at 9:30 P.M. The president of the Committee for Free Elections, Sergio Molina, announced the projection in a press conference at 10 P.M. The projection predicted the true result with an error of (slightly less than) one-half a percentage point. (For the YES vote, the error was .4%.) This projection was based on 10% of the first one-third of the votes cast that night. Even more important was the fact that the projection asserted beyond any reasonable doubt that the NO vote had won; a confidence interval 30 standard deviations wide still would not have included an even race. This fact was known even earlier, because the first computer output, available before 9 P.M., had already shown that Pinochet had been defeated.

By midnight the results from three quarters of the monitored sample (7.3% of the total vote) had arrived. At that point the difference between the projection and the actual result was less than .3%. The standard error had decreased from .34% (for the first public projection) to .20% (for the second one). Naturally, the width of the confidence intervals shrank as the number of observed tables increased (inversely proportional to the square root of the number of available tables).

9. CONCLUSIONS

From a scientific point of view, the projections of the Committee for Free Elections were both accurate and expeditious. The estimation method used to carry out the projections (the difference method) can be highly recommended for its simplicity and reliability. It provides a tool for making electoral projections from early returns when data from previous elections cannot be obtained (or used) and exit polls cannot be trusted.

Three factors that contributed to the success of the projections were the small variability in the number of actual voters per table, the fairly large sample size, and the successful stratification. The small variability in the number of voters reduced the uncertainty associated with the estimation procedures. The fairly large sample size reduced the margin of error. The stratification dealt with the handicap of having to make a projection on the basis of a sample in which late returns could not have been included. The stratification also produced impressive gains in the precision of the projection estimates.

This work was not a scientific exercise, however. Its main purpose was to act as a deterrent against any possible attempt by the Pinochet government to interfere with the electoral process. It is, therefore, difficult to assess to what extent it achieved its purpose.

One measure of the impact of this work is the extent to which the Pinochet government tried to discredit it before the plebiscite. The government's efforts included a media campaign (with radio, television, and newspaper commentaries), which reached its peak on October 2, 1988, with a Sunday editorial in Chile's most influential newspaper. The

Table 4. Projections of the Night of the Plebiscite

Time	Tables received	Percentage of votes	Lower bound	Percentage of NO	Upper bound	Range of nondetectability
9:00 P.M.	386	1.6	52.3	54.2	56.1	52.0–48.0
9:30 P.M.	735	3.3	53.8	55.2	56.6	51.5–48.5
10:30 P.M.	1,127	5.1	54.1	55.1	56.1	51.1–48.9
11:30 P.M.	1,603	7.3	54.2	55.0	55.8	50.9–49.1

editorial echoed the concerns of a minister in the Pinochet government and described the projection effort as a communist plot. The purported plot consisted of a false projection of a victory for the NO, which was to be followed by a takeover of the streets by the Communist Party.

On October 5, Pinochet's government initially presented partial vote counts that were biased in its favor. But by 2:00 A.M. on October 6 the goverment finally conceded that it had lost the election. This turn of events might give the impression that this work was not necessary after all. However, as the pro-Pinochet weekly magazine *Que Pasa* later revealed, there was a government plan for interfering with the election. By 10 P.M. police and soldiers would leave the streets. The government media would then announce a projection (based on a selective sample of 1 million votes) that predicted a victory for Pinochet, and the supporters of the General would be called to the streets to celebrate. This invitation would provoke the supporters of the opposition to also take to the streets, and violence would ensue. The military would intervene to restore order, and Pinochet would have the opportunity to declare a state of siege and ultimately claim victory.

There are many possible explanations for why the government finally decided to accept its defeat at the polls; for more details see Cavallo, Salazar, and Sepúlveda (1988) and Drake and Valenzuela (1989). Two possibilities are lack of agreement within Pinochet's own power base and foreign pressure; three days before the plebiscite the U.S. State Department issued a strong warning against any attempt by Pinochet to interfere with the election. Yet it seems fair to conclude that having an accurate and fast projection of the final outcome of the plebiscite in the hands of various political actors (e.g., the Catholic Church, foreign embassies, political parties) acted as an additional deterrent.

APPENDIX: THE ESTIMATE OF THE STANDARD ERROR

Consider two positive, uncorrelated random variables, X and Y. Let μ_X, μ_Y be their expected values, let σ_X^2, σ_Y^2 be their variances, and let CV_X, CV_Y be their coefficients of variation. When CV_X is small relative to μ_X,

$$\mathrm{var}\left(\frac{Y}{X}\right) \simeq \frac{\sigma_Y^2}{\mu_X^2}\left(1 + (1 + \mathrm{CV}_Y^2)\frac{\mathrm{CV}_X^2}{\mathrm{CV}_Y^2}\right).$$

Consequently, if $\mathrm{CV}_X \ll \mathrm{CV}_Y$ and $\mathrm{CV}_Y < 1$, then $\mathrm{var}(Y/X) \simeq \sigma_Y^2/\mu_X^2$.

Proof. From the first-order Taylor expansion of $f(X) = 1/X$ around μ_X,

$$\mathrm{var}\left(\frac{Y}{X}\right) \simeq \mathrm{var}\left[Y\left(\frac{2}{\mu_X} - \frac{X}{\mu_X^2}\right)\right].$$

The term being neglected is small by assumption, because it is of the order of CV_X^2/μ_X. Because X and Y are independent, a standard calculation yields the desired result.

[*Received July 1989. Revised December 1991.*]

REFERENCES

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language*, Pacific Grove, CA: Wadsworth & Brooks/Cole.

Bernardo, J. M. (1984), "Monitoring the 1982 Spanish Socialist Victory: A Bayesian Analysis," *Journal of the American Statistical Association*, 79, 510–515.

Cavallo, A., Salazar, M., and Sepúlveda, O. (1988), *The Hidden History of the Military Regime* (in Spanish), Santiago, Chile: Ediciones La Epoca.

Cochran, W.G. (1977), *Sampling Techniques*, New York: John Wiley.

Drake, P., and Valenzuela, A. (1989), "The Chilean Plebiscite: A First Step Toward Redemocratization," *LASA Forum*, 19, 18–36.

Efron, B. (1982), *The Jackknife, The Bootstrap, and Other Resampling Plans*, Philadelphia: SIAM Publications.

——— (1987), "Better Bootstrap Confidence Intervals" (with discussion), *Journal of the American Statistical Association*, 82, 171–200.

Efron, B., and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy" (with discussion), *Statistical Science*, 1, 54–77.

Hall, P. (1988), "Theoretical Comparison of Bootstrap Confidence Intervals" (with discussion), *The Annals of Statistics*, 16, 927–953.