Forecasts From Nonrandom Samples: The Election Night Case

Jose Manuel PAVÍA-MIRALLES

The 1990s was not the best of decades for electoral polls, with striking errors occurring in, among others, the British, French, and Spanish elections, including election night, when errors are more evident. This article proposes a model for predicting final election outcomes based on the consistency that polling stations show between elections. Using both past and incoming polling station vote proportions, the model produces continuously revised predictions. The method is validated predicting the 1995 Corts Valencianes (Valencia regional parliament) elections and displaying the real-time experience of the 1999 Corts Valencianes election night. The case study is completed by demonstrating the technique's efficacy in three additional elections. The results confirm that the procedure generates quick, highly reliable, and accurate forecasts. In fact, only a few minutes after starting the scrutiny, the proposal permits one to approximate the final results with great precision, even with only a small percentage of votes polled. The great flexibility of the procedure makes it possible to use the method under a wide variety of circumstances and electoral systems. Furthermore, this procedure has additional advantages, including robustness and lower cost, over other methods which can also be implemented during election night with the objective of forecasting final outcomes, like exit polls or quick counts of a meaningful sample of polling stations.

KEY WORDS: Generalized linear regression; Pseudodata augmentation; Sequence of estimates; Spanish elections; Vote predictions.

1. INTRODUCTION

Since George Gallup forecasted Roosevelt's triumph over Landon in the 1936 U.S. presidential elections, electoral polls and market surveys have become primary tools in the study of the attitudes and behavior of citizens. The media and campaign organizations have used scientific polling for decades, but the past several years have brought a real explosion in the number of public opinion polls, especially during electoral campaigns (Ladd and Benson 1992). This does not mean, however, that the final forecast results have been of a better quality. For instance, striking errors occurred in the 1992 British general election (Rallings and Thrasher 1999) and in the 1997 French legislative election (Jérôme, Jérôme, and Lewis-Beck 1999). In Spain, where the number of polls increased spectacularly in the 1990s, surveys have also failed in good measure. They did not correctly predict the general elections of 1993, 1996, and 2000, as well as in several regional elections (Sanz, Díaz, Rica, and Quesada 1996; Ibáñez et al. 2000). In 1993, although the polls forecasted a victory for the conservative party (PP), the socialists (PS) won the general election. In 1996 the surveys predicted an overwhelming majority for the PP; nevertheless, an absolute majority was 20 seats away. In 2000 the situation was the other way around. The polls, even those on the taken day of the election, claimed that the PP would not get an absolute majority but they did, easily. Moreover, not only have the election polls failed in Spain, but also the exit polls have systematically foreseen different results from the actual ones (Felip, Bellver, and Domenech 1996, p. 316).

The polls' inability to "get it right" over the last 10 years has seriously damaged their image, leaving people with very little or even no faith in them. Concerning elections in the U.K., Brown, Firth, and Payne (1999, p. 212) pointed out that "As a result of the experience in 1992, there was a considerable scepticism in the BBC, and the media generally, about the accuracy of both opinion polls and exit polls." Consequently, a search for new improved forecasting procedures was begun, especially for election night polls, when possible errors are more noticeable and influential on public opinion. This article proposes a procedure for predicting the final election outcome during election night, using information from past elections and current incoming results.

During election night, the activity is frantic. Mass media and political organizations make their exit poll forecasts public just after the polling stations close. The first results come out of the polling stations, and at the same time the forecasts obtained from a quick count of the first 100 scrutinized ballots from a group of polling stations previously chosen as representative are published. The agents in charge of communicating the results naturally want to check whether the results of the counted votes confirm or are compatible with the different estimates.

Unfortunately, especially at the beginning of the scrutiny, the available data cannot be considered a random or representative sample of the election results. In fact, outcomes from small rural polling stations typically come in early, and they usually have a different vote distribution than the overall vote distribution. For this reason, ordinary forecast procedures cannot be used.

Although the available votes cannot be treated as a random sample, certain consistency between elections can be expected (see Fig. 1). On the one hand, it is foreseeable that a polling station that usually has had a proportionally high, say, Republican or Democrat vote in previous elections will still have a proportionally high Republican or Democrat vote in the present elections. On the other hand, it is expected that the so-called *swings*, or vote changes between parties that happen between elections, are not concentrated in a few polling stations, but rather are spread throughout the territory, although not necessarily with the same intensity in all polling stations.

Therefore, because the results of past elections are available for each polling station, each new incoming result can be compared with the corresponding results of the previous elections to evaluate the direction and magnitude of the swing for each

Jose M. Pavía-Miralles is Lecturer in Statistics, Department of Applied Economics, University of Valencia, 46022 Valencia, Spain (E-mail: *pavia@uv.es*). This research was supported in part by DGI project SEJ2004-07924/ECON. The author thanks P. Muñiz and J. A. Alvarez for their help in establishing correspondence between the different polling stations, and two anonymous referees and an editor for their helpful comments and suggestions. The author also thanks Maria A. Ripoll for her assistance and for valuable discussions and ideas.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Applications and Case Studies
DOI 10.1198/016214504000001835

party or candidate. Thus the objective is to provide a sequence of forecasts on the final outcome as the votes reach the data center, combining the results that are already known with a prediction of results that have not yet arrived, solving the problem of nonrandom data with the aid of the tools provided by the two aforementioned considerations.

The idea of producing a series of election forecasts using past results has been used by, among others, Brown and Chappell (1999), Brown et al. (1999), Bernardo (1997), and Bernardo and Girón (1992). Brown and Chappell used both poll data and past relationships to update, in a Bayesian fashion, forecasts of the U.S. presidential elections once new polls are available. The proposal of Brown et al. is more complex. They estimated the change in the share of the vote for each major party in the U.K. in each constituency by means of a regression model using actual election outcomes, party shares of the vote in the previous election, dummy variables, and several socioeconomic variables. Bernardo and Girón combined previous election outcomes and campaign polls to sequentially forecast current elections from early incoming results, using either a hierarchical multivariate Bayesian regression model over the proportion of votes (Bernardo and Girón 1992) or a multivariate Bayesian regression model over the logit transformation of the proportion of votes (Bernardo 1997).

In this article, I construct a procedure that is in some sense inspired by the method proposed by Bernardo and Girón (1992), but that also has a number of important differences from the Bernardo and Girón procedure. Among other differences, the problem is approached from classical inference, it does not require prior information from surveys, and an a priori similarity in the size and direction of swings between polling stations is not presupposed. In particular, the proposed procedure is cheaper and more flexible in its application than the Bernardo and Girón method. Furthermore, as we discuss in Section 3, this procedure has shown a greater capacity than the Bernardo and Girón method to adapt in the face of drastic changes in citizens' political behavior.

The article is organized as follows. Section 2 describes the problem in mathematical terms and introduces the proposed forecasting model. Section 3 tests the model in predicting the 1995 regional election results. Section 4 presents results of the real-time 1999 Corts Valencianes election forecasts and suggests several necessary practicalities. Section 5 extends the case study to show the power of the technique in three new examples. Finally, Section 6 summarizes the results and presents some conclusions.

2. THE FORECASTING MODEL

Worldwide, there are as many different electoral systems as there are elections. However, candidate selection or seat distribution is always determined by counting in each electoral college or constituency into which the country or the region is divided, with the votes cast by each slate of electors (Nohlen 1981). Whatever the characteristics of the electoral system—with either single member or multiple members, with either a list system or specific candidates, using either proportional representation or a plurality method, and so on—in each election the proportion of votes that each candidate or political party obtains are the variables that ultimately determine (after applying the corresponding electoral system) the final election outcome. Thus any forecast procedure must be focused on properly estimating the proportion of votes that each party wins within each constituency.

Electoral data for a given electoral college are not always relevant to a different college. So a possible forecast strategy that simplifies the exposition could be to work college by college, combining the different electoral college forecasts into a final overall prediction in the final step. Then let there be a situation in which n voters of a constituency must choose c representatives or seats among the candidates of p competing parties. Let s be the number of polling stations into which the electorate is divided, and let n_j be the number of voters at polling station j.

It is assumed that past electoral information is available for each of the s polling stations. Let $\mathbf{X}(0)_j$ be the $p \times A$ matrix that contains the registered results in the last A elections in polling station j, where the (k, i)th element of the matrix, $x_k^i(0)_j$, represents the proportion of votes obtained by party k in the ith last election in polling station j; row k of the matrix, $\mathbf{x}_k(0)_j$, is a $1 \times A$ vector that captures the results of party k in the last A elections; and column i of the matrix, $\mathbf{x}^i(0)_j$, is a $p \times 1$ convex vector containing the proportion of votes registered for all of the parties in polling station j in the last election i.

Similarly, let $\mathbf{y}_j$ be the p-dimensional column convex vector where component k, y_{kj}, represents the proportion of valid votes that party k is obtaining in polling station j in the current election. So at a given moment t of the scrutiny, only the outcomes, $\mathbf{y}_j$, of $s(t)$ polling stations, with $0 \leq s(t) \leq s$, will be known. Thus the problem involves obtaining estimates, $\hat{\mathbf{y}}_j$, for the $s - s(t)$ unobserved polling stations using past information provided by the matrix $\mathbf{X}(0)_j$ and the incoming results, $\mathbf{y}_j$, available at the given time.

To estimate the unobserved proportions, it is necessary to take up the considerations pointed out in Section 1. On the one hand, the first assumption—that when a party has registered a high (low) proportion of the votes in a polling station in past elections, it will continue to show a high (low) proportion of the votes in the current election—leads us to assume that there is a linear relationship between the proportion of valid votes obtained in each polling station for each party with past elections. On the other hand, the second hypothesis—that related political behavior within the constituency is reasonable—suggests similar coefficients for all of the polling stations in the linear model. Nevertheless, because the second hypothesis also indicates that all of the polling stations of a constituency do not necessarily show the same intensity in their vote changes, we slightly modify this last assumption. Formally, it is admitted that

$$y_{kj} = \alpha_k + \mathbf{x}_k(0)_j \boldsymbol{\beta}_k + e_{kj}, \qquad k = 1, \dots, p, j = 1, 2, \dots, s, \tag{1}$$

where the subscript j refers to the jth polling stations, α_k and $\boldsymbol{\beta}_k$ are parameter vectors (specific for each party) of orders 1 and $A \times 1$, and the e_{kj} are mean-0 random disturbances, normally distributed with correlations to simplify the model, constant between parties and null between polling stations. That is, $E(e_{kj}e_{k^*j^*}) = \delta_{jj^*}\sigma_{kk^*}$, where δ is the delta of Kronecker.

The p linear relationships of (1) can be expressed in a compact manner. Let $\mathbf{e}_j = [e_{1j}, e_{2j}, \dots, e_{pj}]'$ be the $p \times 1$ disturbance vector, and let $\mathbf{C}$ be the $p \times p$ matrix where the (k, k^*)th element is equal to σ_{kk^*}, and thus $E(\mathbf{e}_j\mathbf{e}'_{j^*}) = \delta_{jj^*}\mathbf{C}$. Then (1) can be written as

$$\mathbf{y}_j = \boldsymbol{\alpha} + \mathbf{X}_j\boldsymbol{\beta} + \mathbf{e}_j, \qquad j = 1, 2, \ldots, s, \tag{2}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_p]'$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \ldots, \boldsymbol{\beta}'_p]'$ are coefficient vectors, $p \times 1$ and $(p \cdot A) \times 1$, and $\mathbf{X}_j$ is a $p \times (p \cdot A)$ block-diagonal matrix of past election proportions, defined by $\mathbf{X}_j = \text{block-diag}[\mathbf{x}_1(0)_j, \mathbf{x}_2(0)_j, \ldots, \mathbf{x}_p(0)_j]$.

The equation system (2) has linearly dependent disturbances, however. The sum of the proportion of votes obtained for the p parties in each polling station in any election is unity. We are looking at a generalized linear system with a singular covariance matrix, $\mathbf{W} = \mathbf{I}_s \otimes \mathbf{C}$, where $\mathbf{I}_s$ is the identity matrix of order s and $\otimes$ represents the Kronecker product. Therefore, for instance, following Theil (1971, pp. 274–280), we can obtain parameter estimates from available polling stations, and hence estimates for the nonavailable polling stations (see App. A).

Once we estimate the unobserved proportion of votes, we aggregate the proportion of votes from all polling stations—both observed and estimated—to estimate the share of the vote for each party in the constituency as a whole. It is assumed that observed outcomes correspond to the first $s(t)$ polling stations. Let v_j be the number of valid votes registered in polling station j, and let π be the proportion of voter participation. Then an estimate of the final distribution of the vote $\mathbf{z} = [z_1, z_2, \ldots, z_p]'$, where z_k represents the proportion of votes obtained by party k in the constituency, is given by (see App. B)

$$\hat{\mathbf{z}} = \sum_{j=1}^{s(t)} \omega_j \mathbf{y}_j + \sum_{j=s(t)+1}^{s} \omega_j \hat{\mathbf{y}}_j, \tag{3}$$

where $\omega_j = v_j/V$ if $j \in \{1, 2, \ldots, s(t)\}$, $\omega_j = \hat{\pi} n_j/V$ if $j \in \{s(t)+1, \ldots, s\}$, $\hat{\pi}$ is an estimate of π obtained from $\hat{\pi} = \sum_{j=1}^{s(t)} v_j / \sum_{j=1}^{s(t)} n_j$ and $V = \sum_{j=1}^{s(t)} v_j + \sum_{j=s(t)+1}^{s} n_j\hat{\pi}$.

The foregoing model supposes that the change parameters α_k and $\boldsymbol{\beta}_k$ and party correlations are constant for all polling stations. However, as noted earlier, the second hypothesis also indicates that all of the polling stations of a constituency do not necessarily show the same intensity in their vote changes. There are polling stations with staunch voters and polling stations whose voters are prone to changing their political opinion. In fact, as Figure 1 clearly shows—through the comparison of the proportion of valid votes obtained for socialists (PS) and conservatives (PP) in the 2,410 polling stations of the electoral college of the Spanish province of Valencia in the 1991 and 1995 Corts Valencianes elections—swings are not uniform neither by parties nor throughout polling stations. Therefore, it is necessary to identify and to group polling stations according to their change propensity for a proper model formulation to be possible. To do this, one can use a cluster analysis to construct homogeneous change propensity groups of polling stations.

Bernardo and Girón (1992) considered the geographical-administrative division as a relevant variable to properly cluster polling stations. They believed that election outcome similarity descends when polling stations are grouped through the sequence of section, district, county, and constituency. But a historical analysis does not support this hypothesis, however. It is possible to find in the same section polling stations with very different electoral change attitudes. The past election patterns available in the $\mathbf{X}(0)_j$ matrix should be used to properly cluster polling stations. It may be convenient to select a few clusters (or to use a hierarchical clustering procedure), to avoid the risk of having (especially at the beginning of scrutiny) a cluster with an insufficient number of polling stations as a basis for estimates. (For example, in the applications in the following sections, the electoral college of Valencia was divided into two clusters, whereas the electoral colleges of Alicante and Castello were not divided.) A graphic analysis (using figures such as Fig. 1) of the information available in the $\mathbf{X}(0)_j$ matrix can help determine the correct number of clusters into which a constituency should be divided (see App. C). Let g be the number of clusters in which the constituency is divided. Then, according to the second part of the second premise, specific coefficients in each cluster should be used. Therefore, for $m = 1, \ldots, g$, the new (1) relationships will be

$$y_{kjm} = \alpha_{km} + \mathbf{x}_{kjm}(0)\boldsymbol{\beta}_{km} + e_{kjm},$$
$$k = 1, \ldots, p, j = 1, 2, \ldots, s, \tag{4}$$

where m is the subscript for cluster and the rest of the elements must be interpreted as before, except that now the disturbance covariance matrix is different for each cluster, that is, $E(\mathbf{e}_{kjm}\mathbf{e}_{k^*jm}) = \sigma_{kk^*}(m)$. Therefore, in each cluster there is a different equation system with specific parameters—$\alpha(m)$'s, $\beta(m)$'s, and $\sigma(m)$'s—which can, similarly to (1), make estimates of the nonavailable polling stations from available polling stations.

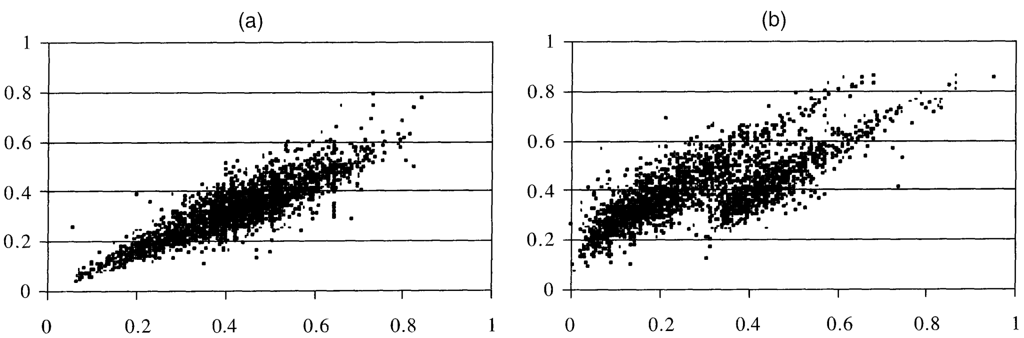

Figure 1. Proportion of Valid Votes Registered in Polling Stations of the Constituency of Valencia During 1991 (horizontal axes) and 1995 (vertical axes) Regional Elections for PS (a) and PP (b). Data from Abacus (1991, 1995).

3. METHOD VALIDATION: FORECASTING THE 1995 CORTS VALENCIANES ELECTIONS

Spain's geopolitical organization divides the country into 17 regions, called *Comunidades Autonomas*, which have a substantial level of self-government, greater than that of the German *länder*. In each of these regions, the citizens elect their own regional government. The Valencia region is one of the 17 autonomous regions in Spain. It is divided into three electoral colleges, called *provinces*, Alicante, Castello, and Valencia. The regional parliament, the *Corts Valencianes*, consists of a single House with 89 seats. The elected number of seats in each province is roughly proportional to its population; 30 seats are elected in Alicante, 22 in Castello, and 37 in Valencia. (In the 1995 regional elections, there were 997,625 voters in Alicante, 365,531 in Castello, and 1,736,161 in Valencia.) Voters cast their ballots for closed list parties. Only the parties that obtain at least 5% of the total regional vote can have seats in the Corts Valencianes. In each province the seats are distributed according to a corrected proportional system, usually known as the d'Hondt algorithm (see App. D or Bernardo 1984, sec. 2)—invented by Thomas Jefferson nearly a century before Victor d'Hondt popularized the system—and used, in various forms, in most parliamentary democracies with proportional representation systems (Nohlen 1981, pp. 129–131). The leader of the party or coalition that has a majority of the seats is appointed by the King to be president of the region.

In spring 1995, the citizens of Valencia were called to vote for the tenth time since democracy was reestablished in Spain. They were to elect the fourth regional parliament in their history. In previous regional elections, the socialists had obtained the majority of the seats. For this election, surveys predicted that only four parties would obtain seats in the parliament: the national conservative party (PP), the socialist party (PS), the communist party (EU), and the regional conservative party (UV). However, surveys disagreed about what combination of parties, either PP + UV or PS + EU, would reach a majority (Table 1). Almost all of the surveys indicated a conservative victory, but the same thing had happened in the 1993 Spanish national election, and the socialists won the election. Thus political analysts had some doubts about the conservative victory.

Moreover, the scrutiny process of the 1995 Corts Valencianes election night was characterized by an unusual lag in the convergence of the scrutinized votes to the final results (Table 2). Thus, for example, although the difference in the final number of seats between PP + UV and PS + EU was five, it was necessary to poll nearly three-quarters of the votes to find a difference greater than one seat between the two party blocks. Perhaps this lack of convergence may explain why Bernardo's article about this election contains only final result forecasts when 77% and 91% of votes were already polled (Bernardo 1997, p. 16, table 3), precisely when the forecasts are usually no longer useful. We apply the model proposed in the previous section to generate a sequence of estimates of the 1995 Corts Valencianes final results.

Equation (3) allows us to obtain point forecasts of the final proportions of votes that each party will reach in each polling station. However, single estimates do not tell us anything about estimate precision. Thus it is usual to use intervals to present uncertainty in the predictions. Consequently, confidence estimates are calculated. In particular, due to the hypothesis of normal distribution assumed for disturbances, estimation intervals based on the normal distribution with a .95 confidence coefficient are obtained in each constituency for the proportion of votes. Let $[\hat{z}_k^-, \hat{z}_k^+]$ be the .95 confidence interval for the final votes proportion of party k.

The relevant results for the Corts Valencianes elections, however, are not the vote proportions that each party will obtain, but rather are the number of seats that they will occupy in the parliament. So in this context, one question clearly emerges: Which of the values of each party interval will be used in the d'Hondt algorithm to approximate the seat distribution? To answer this question, we can refer to the meaning of the confidence intervals. At a confidence level, $\hat{z}_k^-$ and $\hat{z}_k^+$ represent the minimal and the maximum proportions that will be reached by party k. Thus estimates for the minimum and maximum number of seats of party k will be attained applying the d'Hondt algorithm (with a 5% barrier) over $(\hat{z}_1^+, \ldots, \hat{z}_{k-1}^+, \hat{z}_k^-, \hat{z}_{k+1}^+, \ldots, \hat{z}_p^+)$ and $(\hat{z}_1^-, \ldots, \hat{z}_{k-1}^-, \hat{z}_k^+, \hat{z}_{k+1}^-, \ldots, \hat{z}_p^-)$. That is, the number of seats that each party would have obtained at the moment t of scrutiny under the worst-case and best-case scenarios. Furthermore, to complete the estimates, an additional forecast must be made using a more realistic scenario. The most likely seat distribution is derived from $(\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_p)$.

Table 2, for 3%, 5%, 7.5%, 25%, 50%, and 75% of polled votes, presents the estimates that would be obtained if our procedure had been used during the 1995 Corts Valencianes election night. Despite the drastic political preference change demonstrated by the citizens of Valencia between the 1991 and 1995 regional elections (see Table 1), the proposed method exhibits great accuracy, even when the proportion of counted votes is very small. For instance, as can be observed in Table 2,

Table 1. Survey Forecasts of the 1995 Corts Valencianes Elections Published in the Main Newspapers

		Seats				*Percentages*			
Survey agency/newspaper	*Error*	*PS*	*PP*	*EU*	*UV*	*PS*	*PP*	*EU*	*UV*
1991 Results (May 26, 1991)		45	31	6	7	40.92	36.08	12.40	6.89
Demoscopia / El País[a]	±3.1	29	43	11	6	29.10	44.00	13.30	6.60
Sigma Dos / El Mundo[b]	±3.0	28–29	42–45	11–13	5	29.00	43.00	13.00	7.00
Gesfono / Las Provincias[b]	±2.8	28–31	43–46	11–12	3–4	30.70	46.00	13.20	6.00
EMERGfK / Levante[b]	±2.6	32–34	42–45	10–11	0–4	35.80	41.70	13.40	5.60
1995 Results (May 28, 1995)		32	42	10	5	34.25	43.36	11.65	7.09

Source: Felip et al. (1996).

[a]Published on May 20, 1995.

[b]Published on May 21, 1995.

Table 2. Forecasts of the 1995 Corts Valencianes Elections

	Polled[c]	Seats[a] PP	PS	EU	UV	Percentages[b] PP	PS	EU	UV
Results	3%	39	34	11	5	40.10	37.09	11.91	6.50
Forecast		$42_{(39–43)}$	$33_{(31–35)}$	$9_{(9–11)}$	$5_{(3–5)}$	41.36–43.08	34.08–35.70	11.15–12.23	6.50–7.39
Results	5%	39	34	11	5	39.58	37.23	12.54	6.59
Forecast		$41_{(39–42)}$	$33_{(32–35)}$	$10_{(8–11)}$	$5_{(4–5)}$	41.23–42.57	34.48–35.74	11.40–12.23	6.62–7.30
Results	7.5%	38	35	11	5	40.21	36.53	12.47	6.88
Forecast		$41_{(41–43)}$	$32_{(32–35)}$	$11_{(9–11)}$	$5_{(4–5)}$	41.75–42.84	34.27–35.30	11.60–12.28	6.63–7.17
Results	25%	40	33	11	5	41.95	35.19	12.49	6.81
Forecast		$42_{(41–42)}$	$32_{(31–33)}$	$10_{(10–11)}$	5	42.93–43.63	33.66–34.33	11.98–12.39	6.91–7.22
Results	50%	40	33	11	5	41.95	35.19	12.49	6.81
Forecast		42	32	10	5	42.93–43.63	33.66–34.33	11.98–12.39	6.91–7.22
Results	75%	42	32	10	5	43.14	34.03	12.18	6.98
Forecast		42	32	10	5	43.06–43.46	33.99–34.29	11.62–12.01	6.94–7.12
Results	100%	42	32	10	5	43.36	34.25	11.65	7.09

[a]The result lines display the provisional number of seats corresponding to each political option at the moment of scrutiny. The forecast lines show the most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained at the moment of scrutiny under its worst-case and best-case scenarios.

[b]The result lines show the proportions of valid votes that each political party was receiving at the moment of scrutiny. The forecast lines portray the estimated .95 confidence interval for the final votes proportion.

[c]The percentage of census polled at each moment of scrutiny.

when the tallied votes were a mere 3% of total votes and outcomes indicated a PS + EU victory with a one seat difference, the forecast procedure already advanced, toward the more likely scenario, a clear PP + UV majority with a difference of five seats. The method proves to be very robust. It provides highly accurate estimates in an environment of political change and lag in vote convergence.

4. A REAL–TIME APPLICATION: PREDICTING THE 1999 VALENCIA REGIONAL ELECTIONS

Once the model had been tested for forecasting the 1995 Corts Valencianes elections, the procedure was used in real time during the night of the 1999 Corts Valencianes elections to predict its final seat distribution. On June 13, 1999 the municipal elections and the European Parliament elections were celebrated in Spain. On the same day, citizens from 12 of the 17 Spanish regions were also called to elect their regional parliaments. In addition to electing mayors and European parliamentarians, the Valencian voters had to elect their fifth regional Parliament. More than 3 million Valencians (1,096,759 voters in Alicante, 384,195 in Castello, and 1,857,306 in Valencia), divided into 5,125 polling stations (1,864 in Alicante, 652 in Castello, and 2,609 in Valencia), voted to distribute 89 seats among 16 parties. Nevertheless, only five parties had real possibilities of surpassing the barrier of 5% of valid votes required by the Valencian electoral system (DOGV 1987) (Table 3). Those parties were the four parties (PP, PS, EU, and UV) that already occupied seats in Corts Valencianes in 1995 alongside the coalition of the left nationalist party and the ecologist party (BV). Thus the prediction effort was focused on the five parties with real possibilities of being elected to the regional parliament. So the hypothesis was admitted that only six organizations were involved in the elections—PP, PS, EU, UV, BV, and an additional party that could not obtain any seats though its vote proportion surpassed 5%—to guarantee the congruence of the system.

However, between elections there are always changes in the composition and the number of polling stations. Indeed, in each new election there are creations of new polling stations and fusions and divisions among the existing ones. Moreover, in polling stations that have apparently not changed there are new incoming voters (due to young people reaching voting age and new residents) and leaving voters (due to deaths and changes of residence). So it is necessary to establish a proper correspondence between polling stations in different elections to collect accurate past information on them. The correspondence criteria are as follows. It is assumed that incoming and leaving voters in existing polling stations are random, so a direct correspondence is established between polling stations that apparently have not changed. When two or more stations are joined to create a new one, the aggregate results of the original polling

Table 3. Survey Forecasts of the 1999 Corts Valencianes Elections Published Just Before the Elections

Agency	Date	Seats* PP	PS	EU	UV	Percentages PP	PS	EU	UV	BV
Results	May 28, 1995	42	32	10	5	43.4	34.3	11.7	7.1	2.8
Gesfono	May 99	$48_{(45–48)}$	$29_{(29–32)}$	$8_{(6–9)}$	$4_{(3–5)}$	47.7	31.5	10.0	6.7	
Gesfono	May 99	$47_{(46–48)}$	$30_{(29–32)}$	8	$4_{(3–5)}$	47.3	31.6	10.3	6.4	
Sigma Dos	May 99	$48_{(47–50)}$	$32_{(31–34)}$	$7_{(6–7)}$	$0_{(0–3)}$	48.7	33.2	8.5	4.9	2.7
EMER/GfK	May 99	$46_{(46–48)}$	$32_{(30–33)}$	$6_{(6–7)}$	$3_{(3–5)}$	47.2	32.6	8.7	6.3	4.3
Demoscopia	May 99	$49_{(47–49)}$	$29_{(29–32)}$	8	3	49.0	32.0	9.9	5.2	3.6
CIS	June 99	$48_{(47–50)}$	$31_{(29–32)}$	7	3	48.1	31.9	9.5	5.1	3.2
Results	June 13, 1999	49	35	5	0	48.4	34.3	6.1	4.7	4.6

Source: ARGOS (2000).

*The most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained under its worst and its best scenarios.

stations are considered historical data for the new polling station. For stations arising as a consequence of the division of an existing polling station, the vote proportions of the original station are chosen to represent the historical vote proportions of these new stations. Finally, either section, district, or city average vote proportions are assigned as historical data for newly created polling stations, because they are usually located in the expansion areas of the cities.

The polling stations closed at exactly 8:00 PM, at which time the public count of the votes started. First, the municipal election votes were counted, then the scrutiny continued with the Corts Valencianes election votes. So the lag in communicating results by big polling stations was cumulative. For instance, the first results were received in the analysis center almost 1/2 hour after the polling stations closed, and these were from a polling station with only 20 voters. In fact, the mean of valid votes of the first 5% received from polling stations was 301.75, which contrasts with the value of 583.57 of the last 5% of scrutinized stations. Throughout the night, 126 predictions about the Corts Valencianes distribution were made. The first prediction was made at 9:20 PM with only .18% of the census polled, and the last prediction was performed at 2:07 AM with 96.08% of the census polled, the scrutiny ended at 5:00 AM. These predictions were automatically transferred to a webpage of restricted use, which kept the agents in charge of communicating polled results properly informed. Table 4 gives some of the predictions made throughout that night.

As we can see in Table 4, the early estimates were highly reliable. The forecasts presented with 1.23% of the census polled (1.20% in Alicante, 2.25% in Castello, and 1.03% in Valencia), however, have much less narrow seat intervals than posterior predictions. In that moment, the predictions indicated that BV would perhaps surpass the electoral system's 5% barrier. The fact that results from small polling stations and from the

Table 4. Samples of the 1999 Corts Valencianes Forecasts From Election Night Outcomes

			Seats[a]					*Percentages*[b]				
College		*Polled*[c]	*PP*	*PS*	*EU*	*UV*	*BV*	*PP*	*PS*	*EU*	*UV*	*BV*
Alicante	Result	1.20	15	13	1	0	1	46.52	39.81	5.91	1.65	3.55
	Forecast		$16_{(15-17)}$	$13_{(12-13)}$	$1_{(1-2)}$	0	$0_{(0-1)}$	47.2 ±1.0	37.6 ±.9	5.7 ±.4	2.2 ±.3	4.3 ±.4
Castello	Result	2.25	11	8	2	0	1	46.71	32.75	9.75	3.53	4.97
	Forecast		$13_{(12-13)}$	$8_{(8-9)}$	1	0	$0_{(0-1)}$	50.4 ±1.2	33.2 ±1.1	6.3 ±.6	4.4 ±.5	5.1 ±.5
Valencia	Result	1.03	20	13	2	0	2	47.67	32.23	6.61	5.14	6.17
	Forecast		$21_{(20-21)}$	$13_{(12-14)}$	$3_{(2-3)}$	0	$0_{(0-2)}$	48.7 ±1.2	31.9 ±1.1	7.1 ±.6	4.9 ±.5	5.0 ±.5
Region	Result	1.23	46	34	5	0	4	47.22	34.67	7.04	4.03	5.03
21h, 47m	Forecast		$50_{(47-52)}$	$34_{(32-36)}$	$5_{(4-6)}$	0	$0_{(0-4)}$	48.4 ±1.1	33.9 ±1.0	6.5 ±.5	3.9 ±.4	4.8 ±.5
Alicante	Result	2.49	15	13	2	0	0	47.12	39.93	5.99	1.74	2.57
	Forecast		$15_{(15-16)}$	$13_{(12-13)}$	$2_{(1-2)}$	0	0	47.1 ±.7	38.8 ±.6	5.7 ±.3	2.1 ±.2	4.0 ±.3
Castello	Result	7.01	13	8	1	0	0	48.29	32.85	6.90	4.46	5.21
	Forecast		13	8	1	0	0	51.0 ±.7	33.3 ±.8	5.4 ±.3	4.3 ±.3	5.4 ±.3
Valencia	Result	2.80	22	13	2	0	0	49.47	30.71	6.52	5.42	5.93
	Forecast		$21_{(20-22)}$	$14_{(13-14)}$	$2_{(2-3)}$	0	0	49.7 ±.7	32.0 ±.6	6.7 ±.3	5.7 ±.3	4.8 ±.3
Region	Result	3.18	50	34	5	0	0	48.59	33.54	6.49	4.26	4.91
22h, 02m	Forecast		$49_{(48-51)}$	$35_{(33-35)}$	$5_{(4-6)}$	0	0	49.0 ±.8	34.4 ±.6	6.2 ±.3	4.4 ±.3	4.6 ±.3
Alicante	Result	9.30	15	13	2	0	0	45.52	41.78	6.38	1.50	2.70
	Forecast		$16_{(16-17)}$	12	$2_{(1-2)}$	0	0	49.6 ±.3	37.2 ±.3	5.9 ±.2	1.7 ±.1	3.9 ±.1
Castello	Result	15.93	13	8	1	0	0	49.09	33.21	5.59	4.40	5.48
	Forecast		$13_{(12-13)}$	$8_{(8-9)}$	1	0	0	51.1 ±.6	34.3 ±.7	4.9 ±.2	4.3 ±.3	5.2 ±.2
Valencia	Result	9.28	21	14	2	0	0	48.34	31.52	6.59	5.70	5.82
	Forecast		21	14	2	0	0	48.1 ±.5	32.6 ±.4	6.5 ±.2	6.1 ±.2	5.1 ±.2
Region	Result	10.05	49	35	5	0	0	47.66	34.81	6.34	4.24	4.85
22h, 24m	Forecast		$50_{(49-51)}$	$34_{(34-35)}$	$5_{(4-5)}$	0	0	48.9 ±.4	34.3 ±.4	6.2 ±.2	4.4 ±.2	4.7 ±.2
Alicante	Result	49.66	16	12	2	0	0	49.05	39.03	6.76	1.63	3.53
	Forecast		16	12	2	0	0	49.6 ±.2	36.8 ±.3	6.1 ±.1	1.7 ±.1	3.7 ±.1
Castello	Result	59.59	13	8	1	0	0	51.49	34.20	4.68	4.05	5.58
	Forecast		$12_{(12-13)}$	$9_{(8-9)}$	1	0	0	50.1 ±.4	34.9 ±.3	4.3 ±.2	4.5 ±.2	5.3 ±.2
Valencia	Result	48.54	21	14	2	0	0	48.26	33.50	6.86	6.28	5.09
	Forecast		21	14	2	0	0	47.7 ±.3	33.2 ±.2	6.5 ±.1	6.4 ±.1	4.9 ±.1
Region	Result	50.18	50	34	5	0	0	48.97	35.36	6.52	4.49	4.67
23h, 35m	Forecast		$49_{(49-50)}$	$35_{(34-35)}$	5	0	0	48.6 ±.3	34.6 ±.3	6.1 ±.1	4.6 ±.1	4.6 ±.1
Alicante	*Result*	100	16	12	2	0	0	49.53	36.57	6.15	1.76	3.74
Castello	*Result*	100	12	9	1	0	0	49.36	34.51	4.16	4.34	5.32
Valencia	*Result*	100	21	14	2	0	0	47.52	32.91	6.50	6.53	4.92
Region	*Result*	100	49	35	5	0	0	48.39	34.28	6.11	4.73	4.59

[a]The result lines display the provisional number of seats corresponding to each political option at the moment of scrutiny. The forecast lines show the most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained at the moment of scrutiny under its worst-case and its best-case scenarios.

[b]The result lines show the proportions of votes that each political party was receiving at the moment of scrutiny. The forecast lines portray the estimated .95 confidence interval for the final votes proportion.

[c]The percentage of census polled at each moment of scrutiny.

constituency of Castello—where BV had a relatively higher increase in votes—were received quickly temporarily improved the BV perspectives. Nevertheless, this situation soon changed. Only 15 minutes later and with a mere 3.18% of the census polled (2.49% in Alicante, 7.01% in Castello, and 2.80% in Valencia), predictions indicated that both UV and BV would not reach the 5% of valid votes in the region necessary for parliamentary representation. Notwithstanding, we must emphasize the enormous difficulties that predictions would have encountered in reducing the size of the seat intervals has the proportion forecasts of UV and/or BV been very near 5% for a long time.

Finally, we need to mention two additional issues arising when analyzing Table 4. On the one hand, we observe that the seat intervals are opened, but by only one seat, even when the scrutiny is very advanced. This happened because the last seat in Castello was assigned at the end of the scrutiny for only 1,332 votes, barely .35% of the census of the constituency. On the other hand, it is perceived that first forecasts tended to slightly underestimate UV final proportions. This results from the lag that greater polling stations registered in supplying outcomes. In fact, UV harvests most of its votes in the province of Valencia, especially in the city of Valencia, where stations are larger. Indeed, the average census of Valencia polling stations is 712, in contrast to the average census of 589 of the Alicante and Castello polling stations.

5. SOME OTHER EXAMPLES

In the foregoing sections, the method was discussed for the 1995 and 1999 Valencia regional elections. The model was also tested in three other elections. On March 12, 2000, the procedure was run to forecast the seat distribution in the three constituencies of the Valencia region in the 2000 Spanish general election. On May 25, 2003, it was used to anticipate the results of the 2003 Corts Valencianes elections and also to predict the number of councillors that each party would return in the four biggest cities of the Valencia region (Valencia, Alicante, Elche, and Castello) in the 2003 Spanish local elections.

In the 2000 Spanish general election, the voters had to elect representatives to (divided into 52 provinces) the 350 seats of the Congreso de los Diputados, who in turn elected the President of the Spanish Government. (For an overview of the Spanish electoral system, see Bernardo 1984, sec. 2.) Thirty-two of these seats were elected in the provinces of the Valencia region (16 in Valencia, 11 in Alicante, and 5 in Castello). As an example, Table 5 gives some of the predictions made throughout that night for the electoral college of Alicante, where, despite the fact that the provisional results did not converge with the final outcome until the scrutiny process was rather advanced, the method was accurate.

The method was again used in the 2003 Corts Valencianes elections, and (as shown in Table 6) the obtained forecasts were quick and accurate. In fact, despite the enormous bias that the results from early outcomes showed—especially in percentages—the first estimates already evinced great precision. Moreover, in the local elections, where the number of polling stations in each college was is sensitively smaller (895 in Valencia, 402 in Alicante, 271 in Elche, and 190 in Castello), the method also produced highly reliable estimates, as shown in Table 7, which gives a sample of forecasts for the 2003 local elections for the city of Alicante.

6. SUMMARY AND CONCLUDING REMARKS

This article has proposed a procedure for predicting on election night the final shares of votes of a group of parties or candidates competing in the election. The model is based on the consistency demonstrated in polling stations between elections. In particular, forecasts of the final vote proportions of each party are obtained using both past polling station vote proportions and current polling station incoming results. The method can produce continuously revised predictions as soon as a new station is polled. The results demonstrate that the method produces highly reliable and accurate forecasts. Moreover, in a context where election polls have evidenced a serious lack of accuracy, such as the 1990s European polls, the proposed model clearly outperformes those similar strategies that also include poll data.

Furthermore, compared with other methods, such as exit polls or quick counts of a meaningful sample of polling stations, that also can be implemented with the same objective, this procedure has the important advantages of being low cost, very robust, and accurate. Indeed, both exit polls and quick counts are very costly procedures. They require a vast effort in logistics (including technological infrastructures) and in personnel, which make their application very expensive. The proposed method, in contrast, can be applied by means of a connection

Table 5. Samples of the 2000 Spanish General Election Forecasts for the Province of Alicante

			Seats[a]					Percentages[b]				
	Time	*Polled*[c]	*PP*	*PS*	*EU*	*UV*	*BV*	*PP*	*PS*	*EU*	*UV*	*BV*
Result	20h, 46m	3.19	6	5	0	0	0	50.39	38.34	5.14	.26	1.09
Forecast		79	7	4	0	0	0	53.95 ±.5	34.80 ±.5	5.13 ±.3	.29 ±.1	1.27 ±.2
Result	21h, 00m	11.14	6	5	0	0	0	51.61	37.64	5.55	.25	1.20
Forecast		248	7	4	0	0	0	54.38 ±.3	34.71 ±.2	5.30 ±.1	.27 ±.0	1.46 ±.1
Result	21h, 16m	22.07	6	5	0	0	0	52.36	37.49	5.61	.29	1.36
Forecast		470	7	4	0	0	0	54.37 ±.3	34.82 ±.2	5.36 ±.1	.30 ±.0	1.48 ±.1
Result	21h, 40m	48.25	7	4	0	0	0	53.41	35.97	5.41	.30	1.57
Forecast		943	7	4	0	0	0	54.36 ±.2	34.81 ±.2	5.34 ±.1	.29 ±0	1.52 ±.1
Result	0h, 55m	100	7	4	0	0	0	54.36	34.76	5.34	.29	1.53

[a]The result lines display the provisional number of seats corresponding to each political option at the moment of scrutiny. The forecast lines show the most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained at the moment of scrutiny under its worst-case and its best-case scenarios.

[b]The result lines show the proportions of valid votes that each political party was receiving at the moment of scrutiny. The forecast lines portray the estimated .95 confidence interval for the final votes proportion.

[c]The percentage of census polled at each moment of scrutiny appears in the result rows and the number of stations polled in the forecast rows.

Table 6. Samples of the 2003 Corts Valencianes Forecasts From Election Night Outcomes

	Time	Polled[c]	Seats[a] PP	PS	EU	UV	BV	Percentages[b] PP	PS	EU	UV	BV
Result	21h, 00m	.23	48	41	0	0	0	41.22	47.75	4.61	.81	3.55
Forecast		38	49 (48–50)	35 (35–36)	5 (4–5)	0	0	48.69 ±.5	35.98 ±.4	5.85 ±.3	2.23 ±.4	4.91 ±.1
Result	21h, 30m	2.31	44	40	5	0	0	44.25	42.30	5.97	1.98	4.17
Forecast		174	48	36 (35–36)	5 (4–5)	0	0	47.71 ±.3	35.86 ±.3	6.04 ±.3	2.90 ±.1	4.89 ±.1
Result	21h, 50m	10.41	45	39	5	0	0	45.37	39.97	6.35	2.25	4.72
Forecast		685	48 (47–48)	36 (36–37)	5 (4–5)	0	0	47.67 ±.3	36.62 ±.3	6.35 ±.1	2.91 ±.1	4.87 ±.1
Result	22h, 15m	25.24	47	37	5	0	0	46.36	38.60	6.70	2.44	4.56
Forecast		1,544	48 (47–48)	36 (36–37)	5 (4–5)	0	0	47.69 ±.3	36.70 ±.3	6.44 ±.1	3.00 ±.1	4.81 ±.1
Result	22h, 50m	52.23	47	37	5	0	0	47.09	37.72	6.57	2.68	4.61
Forecast		3,075	48	36	5	0	0	47.78 ±.2	36.58 ±.2	6.43 ±.1	3.01 ±.1	4.75 ±.1
Result	3h, 25m	100	48	36	5	0	0	47.96	36.50	6.44	3.04	4.77

[a]The result lines display the provisional number of seats corresponding to each political option at the moment of scrutiny. The forecast lines show the most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained at the moment of scrutiny under its worst and its best scenarios.

[b]The result lines show the proportions of valid votes that each political party was receiving at the moment of scrutiny. The forecast lines portray the estimated .95 confidence interval for the final votes proportion.

[c]The percentage of census polled at each moment of scrutiny appears in the result rows, and the number of stations polled appears in the forecast rows.

with the computer that centralizes the scrutiny and that previously organized the polling stations' historical data. So, a small group of people (even one analyst alone) could carry out all of the work. In addition, in the event that monetary incentives were not sufficient, there are other reasons for using the suggested procedure. On the one hand, exit polls are strongly influenced by *response errors*, which can generate estimates of a very low quality, as happened recently in Spanish exit polls. In contrast, the proposed model is based on real votes and so does not have response error problems, and, as Tables 2 and 4 show, it produces very high-quality estimates. On the other hand, a comparison with the quick-counts strategy offers additional arguments in favor of the proposed model. First, it is doubtful that quick-count produce earlier predictions. Second, quick-counts forecast errors are not necessarily smaller. Third, quick counts usually require data from all of the selected polling stations to compose accurate forecasts. So a lag or a problem in a station can seriously damage quick-count estimates, whereas our proposed model is more flexible and is not affected by delays in specific stations because it does not depend on particular polling stations.

The proposed model's flexibility is further manifested when changes in electoral colleges occur. Sometimes administrative or population variations lead to a redistribution in the boundaries of the constituencies, as happened, for example, in the 1997 Canadian general election. In these cases, some forecasting methods, such as the procedure of Brown et al. (1999), must undergo painstaking analysis to impute the votes that would have been cast for each party at previous elections had they been held within the new boundaries. The proposed model, however, need not appeal to complicated solutions. It is sufficient to consider the polling stations in its actual constituency and to use their past values in the model. In fact, this was the solution implemented for a number of polling stations that moved from the constituency of Castello to that of Valencia in the 1999 Valencian regional elections.

Finally, I would like to make one last reflection. In real inference problems, it is not usual to have the possibility of comparing estimates and true values. Therefore, election night forecasts are exceptions. Consequently, in the light of the obtained results, the proposed methodology could be used to solve other problems, where disposing of random samples is difficult

Table 7. Samples of the 2003 Local Election Forecasts for the City of Alicante

	Time	Polled[c]	Seats[a] PP	PS	EU	BV	Percentages[b] PP	PS	EU	BV
Result	20h, 50m	1.34	10	16	1	0	35.84	52.91	6.24	2.74
Forecast		6	13 (13–14)	13 (12–13)	1 (1–2)	0	45.63 ±.4	42.37 ±.4	6.51 ±.3	3.10 ±.2
Result	21h, 00m	7.19	13	13	1	0	43.02	45.59	6.35	2.54
Forecast		33	14	12	1	0	48.74 ±.3	39.86 ±.3	6.34 ±.2	2.8 ±.2
Result	21h, 20m	20.53	14	12	1	0	46.84	41.33	6.43	2.86
Forecast		92	14 (14–15)	12 (12–11)	1	0	48.97 ±.3	39.62 ±.3	6.44 ±.2	2.89 ±.2
Result	21h, 30m	34.72	14	12	1	0	45.96	40.07	6.22	2.84
Forecast		150	14	12	1	0	49.00 ±.2	39.46±.2	6.48 ±.1	2.94 ±.1
Result	21h, 40m	50.01	14	12	1	0	47.20	40.90	6.40	2.97
Forecast		214	14	12	1	0	49.02 ±.2	39.48 ±.2	6.49 ±.1	2.96 ±.1
Result	1h, 00m	100	14	12	1	0	49.11	39.37	6.53	2.98

[a]The result lines display the provisional number of seats corresponding to each political option at the moment of scrutiny. The forecast lines show the most likely seat distribution and, in subscripts and brackets, the number of seats that each party would have obtained at the moment of scrutiny under its worst-case and its best-case scenarios.

[b]The result lines show the proportions of valid votes that each political party was receiving at the moment of scrutiny. The forecast lines portray the estimated .95 confidence interval for the final votes proportion.

[c]The percentage of census polled at each moment of scrutiny appears in the result rows and the number of stations polled in the forecast rows.

or to find alternative solutions to problems treated almost exclusively from a random-sample perspective. For instance, the method could be applied to election polls by collecting a few detailed samples of a few small areas, which would be cheaper than a large sample covering the whole constituency.

APPENDIX A: FORECASTING NONAVAILABLE POLLING STATIONS

The reorganization of (1) offered in (2) is very graphic, but is not very operative. Let $\boldsymbol{\gamma}_k = [y_{k1}, y_{k2}, \ldots, y_{ks}]'$ be an $s \times 1$ vector of the proportion of valid votes obtained in all the polling stations for party k, let $\mathbf{X}_k(0) = [\mathbf{x}_k(0)'_1, \mathbf{x}_k(0)'_2, \ldots, \mathbf{x}_k(0)'_s]'$ be an $s \times A$ matrix of the past proportion of votes obtained in all polling stations for party k in A last elections, and let $\boldsymbol{\varepsilon}_k = [e_{k1}, e_{k2}, \ldots, e_{ks}]'$ be an $s \times 1$ vector of disturbances. Then the linear relationship (1) can be written as

$$\boldsymbol{\gamma}_k = \mathbf{X}_k{}^{\boldsymbol{\alpha}}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k \quad \text{for } k = 1, 2, \ldots, p, \tag{A.1}$$

where ${}^{\boldsymbol{\alpha}}\boldsymbol{\beta}_k = [\alpha_k, \boldsymbol{\beta}'_k]'$ is an $(A+1) \times 1$ vector of party k parameters and $\mathbf{X}_k$ is an $s \times (A+1)$ matrix defined by $\mathbf{X}_k = [\mathbf{i}_s, \mathbf{X}_k(0)']'$, with $\mathbf{i}_s$ an $s \times 1$ vector of 1's.

Furthermore, to express (A.1) more compactly, let $\boldsymbol{\gamma} = [\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2, \ldots, \boldsymbol{\gamma}'_p]'$ be an $s \cdot p \times 1$ vector of current vote proportions, let $\mathbf{X} = \text{block-diag}[\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$ be a matrix $s \cdot p \times (A+1) \cdot p$ of past proportions, and let ${}^{\boldsymbol{\alpha}}\boldsymbol{\beta} = [{}^{\boldsymbol{\alpha}}\boldsymbol{\beta}'_1, {}^{\boldsymbol{\alpha}}\boldsymbol{\beta}'_2, \ldots, {}^{\boldsymbol{\alpha}}\boldsymbol{\beta}'_p]'$ be a $p \cdot (A+1) \times 1$ coefficient vector. Then the p linear relationships of (4) can be denoted compactly by

$$\boldsymbol{\gamma} = \mathbf{X}^{\boldsymbol{\alpha}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{A.2}$$

where $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \ldots, \boldsymbol{\varepsilon}'_p]'$ is an $s \cdot p \times 1$ normal mean-0 random vector of disturbances with covariance matrix $\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{C} \otimes \mathbf{I}_s$, where the (k, k^*)th block is equal to $E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}'_{k^*}) = \sigma_{kk^*}\mathbf{I}_s$.

Note, however, that, (A.2) has linearly dependent disturbances in the same way as (2); the sum of the proportion of votes obtained for the p parties in each polling station is unity $(\mathbf{i}'_p \otimes \mathbf{I}_s)\boldsymbol{\gamma} = \mathbf{i}_s$, where $\mathbf{i}_p$ is a $p \times 1$ vector of 1's, so $\boldsymbol{\Omega}$ is singular. Consequently, the best linear unbiased estimator of the parameters is obtained from

$${}^{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^+\boldsymbol{\gamma}, \tag{A.3}$$

where $\boldsymbol{\Omega}^+$ is the Moore–Penrose generalized inverse of $\boldsymbol{\Omega}$ and ${}^{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}} \sim \mathrm{N}[{}^{\boldsymbol{\alpha}}\boldsymbol{\beta}, (\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X})^{-1}]$.

A difficulty in computating this Aitken estimator is that the matrix $\boldsymbol{\Omega}$ is unknown. So we must replace $\boldsymbol{\Omega}$ by an estimate. Because $\boldsymbol{\Omega} = \mathbf{C} \otimes \mathbf{I}_s$, we need only estimate $\mathbf{C}$. The estimator of $\mathbf{C}$ that we use is the matrix of the mean squares and products of the p sets of the OLS residuals obtained from estimating (A.1). In particular, to estimate the proportions of the as-yet uncounted votes at a given moment t of the scrutiny, the proposed procedure is as follows:

1. Construct the vectors and matrix $\boldsymbol{\gamma}_k$, $\boldsymbol{\gamma}$, $\mathbf{X}_k$, and $\mathbf{X}$ using the data from the $s(t)$ polling stations for which current outcomes are available.
2. Estimate by OLS each of the p equations of (A.1), and calculate an estimation for $\boldsymbol{\Omega}$ from the obtained residuals.
3. From the estimation of $\boldsymbol{\Omega}$ obtained in (2), $\hat{\boldsymbol{\Omega}}$, estimate the parameter vector of (A.3), ${}^{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^+\boldsymbol{\gamma}$.
4. Finally, using the hypothesis of noncorrelation between polling stations, approximate the nonavailable y_{kj} current results by $\hat{\boldsymbol{\gamma}}_* = \mathbf{X}_*{}^{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}$, where $\boldsymbol{\gamma}_*$ and $\mathbf{X}_*$ correspond to the $s - s(t)$ polling stations for which actual results are not yet available.

APPENDIX B: AGGREGATION OF POLLING STATIONS PROPORTION OF VOTES

Let V_k be the total number of votes of party k. It is easy to prove that

$$V_k = z_k \pi n = z_k \pi \sum_{j=1}^{s} n_j = z_k \sum_{j=1}^{s} v_j = \sum_{j=1}^{s} v_j y_{kj}, \tag{B.1}$$

and hence

$$z_k = \frac{\sum_{j=1}^{s} v_j y_{kj}}{\sum_{j=1}^{s} v_j} = \frac{\sum_{j=1}^{s(t)} v_j y_{kj}}{\sum_{j=1}^{s(t)} v_j + \sum_{j=s(t)+1}^{s} v_j} + \frac{\sum_{j=s(t)+1}^{s} v_j y_{kj}}{\sum_{j=1}^{s(t)} v_j + \sum_{j=s(t)+1}^{s} v_j}. \tag{B.2}$$

Therefore, at a given moment t of the scrutiny, an estimate for z_k can be obtained from equation

$$\hat{z}_k = \frac{\sum_{j=1}^{s(t)} v_j y_{kj}}{\sum_{j=1}^{s(t)} v_j + \sum_{j=s(t)+1}^{s} v_j} + \frac{\sum_{j=s(t)+1}^{s} v_j \hat{y}_{kj}}{\sum_{j=1}^{s(t)} v_j + \sum_{j=s(t)+1}^{s} v_j}. \tag{B.3}$$

However, at the moment t, the values v_j—the number of votes registered in polling station j—are unknown for $j > s(t)$. So to compute this equation, it is necessary to replace the v'_js by estimates. So, given that $v_j = \pi_j n_j$, an estimate for v_j can be obtained using π to approximate π_j. Therefore, the estimator of v_j used is $\hat{v}_j = \hat{\pi} n_j$, where $\hat{\pi}$ is the proportion of voters' participation being registered at moment t. Consequently, (B.3) is

$$\hat{z}_k = \frac{\sum_{j=1}^{s(t)} v_j y_{kj}}{\sum_{j=1}^{s} v_j + \sum_{j=s(t)+1}^{s} \hat{\pi} n_j} + \frac{\sum_{j=s(t)+1}^{s} \hat{\pi} n_j \hat{y}_{kj}}{\sum_{j=1}^{s} v_j + \sum_{j=s(t)+1}^{s} \hat{\pi} n_j}. \tag{B.4}$$

Note, however, that it would not be correct to use the relative number of voters, $\eta_j = n_j/n$, to aggregate the proportion of votes from all polling stations, because the resulting estimate would be biased.

APPENDIX C: CHOOSING THE NUMBER OF CLUSTERS, A SENSITIVITY ANALYSIS

It seems that the number of clusters chosen is up to the analyst. Several different values for g could be selected. Hence some doubts about the power of the method may arise, depending on the impact of this choice. So a study of the sensitivity of predictions to the number of clusters is merited. Final outcomes alternative estimates for the electoral college of Valencia in the 1995 Corts Valencianes elections have been obtained using four different values for g. Table C.1 gives some of the predictions reached. Although differences between forecasts for each party are not large, it is observed that as a rule, predictions are jointly more accurate as g increases. Nevertheless, it is worth noting that an excessive number of clusters could result in an unjustified extrawork and even decrease the estimate's quality. For instance, with 2.5 % of polled votes, forecasts with $g = 3$ are better than those with $g = 4$, because in this case only eight polling stations were polled in one of the clusters.

APPENDIX D: THE D'HONDT ALGORITHM

Let z_k be, for $k = 1, 2, \ldots, p$, the proportion of valid votes obtained by the kth party in the electoral college. To distribute the c seats of the college, the d'Hondt rule proceeds as follows: (1) Calculate the $p \times c$ matrix of quotients where the (k, r)th element is $q_{k,r} = z_k/r$; (2) Select the c largest elements of the matrix obtained in (1); and (3) assign to party k a number of seats equal to the number of these c largest elements found in its corresponding row. Thus if, for example,

Table C.1. Forecasts for the College of Valencia Using Different Numbers of Clusters in the 1995 Elections

No. of clusters	Polled[b]	Percentages[a] PP	PS	EU	UV	Error[c]	No. of clusters	Polled[b]	Percentages[a] PP	PS	EU	UV	Error[c]
Results	2.5%	37.27	34.76	12.66	10.65	5.52	Results	15%	38.15	33.79	13.27	10.99	4.56
$g=1$		39.69	33.50	12.34	10.49	2.01	$g=1$		40.35	32.57	12.88	10.66	.79
$g=2$		39.73	33.36	12.42	10.52	1.77	$g=2$		40.22	32.74	12.72	10.67	.60
$g=3$		40.03	33.18	12.37	10.50	1.32	$g=3$		40.26	32.74	12.67	10.64	.54
$g=4$		39.95	33.41	12.26	10.48	1.72	$g=4$		40.34	32.77	12.69	10.58	.37
Results	7.5%	37.43	34.85	13.08	10.66	5.81	Results	25%	38.61	33.26	13.59	10.83	3.74
$g=1$		40.05	33.15	12.64	10.53	1.04	$g=1$		40.52	32.43	12.91	10.62	.75
$g=2$		40.02	33.14	12.69	10.53	1.02	$g=2$		40.45	32.60	12.72	10.75	.59
$g=3$		40.14	33.15	12.67	10.50	.89	$g=3$		40.49	32.56	12.65	10.72	.58
$g=4$		40.22	33.17	12.66	10.42	.85	$g=4$		40.61	32.61	12.68	10.63	.37
Results	100%	40.57	32.76	12.69	10.46		*Results*	100%	40.57	32.76	12.69	10.46	

[a] The result lines show the proportions of valid votes that each political party was receiving at the moment of scrutiny. The g-lines portray the final votes proportion forecasts obtained after dividing the electoral college into g clusters.

[b] The percentage of census polled at each moment of scrutiny.

[c] The error column displays the sum of the differences in absolute values between final outcomes and provisional data (either results or predictions): $\sum_i |z_i - \hat{z}_i|$.

17 seats must be distributed among four parties A, B, C, and D, which have achieved 45%, 32%, 15%, and 8%, of the valid vote, we must construct the matrix below, from which A, B, C, and D would obtain 8, 6, 2, and 1 seats.

Party	Seats 1	2	3	4	5	6	7	8	9	⋯	17
A	45.0	22.5	15.0	11.3	9.0	7.5	6.4	**5.6**	5.0	⋯	2.6
B	32.0	16.0	10.7	8.0	6.4	**5.3**	4.6	⋯	⋯	⋯	1.9
C	14.0	**7.5**	4.7	⋯	⋯	⋯	⋯	⋯	⋯	⋯	.8
D	**8.0**	4.0	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	.5

As can be observed, the d'Hondt rule provides a proportional system that enhances the representation of the larger parties to the detriment of smaller ones. Indeed, if only one seat is allocated, then the d'Hondt algorithm reduces to a majority rule. Nevertheless, the correction becomes smaller as the number of seats increases, so almost perfect proportional representation can be achieved if the number of seats is sufficiently large. For instance, in the foregoing example an exact proportional distribution would lead to 7.65, 5.44, 2.28, and 1.36 seats for each party—figures not too different from the distribution provided.

[Received June 2003. Revised September 2004.]

REFERENCES

Abacus (1991), *Base de Datos de Resultados de Eleccions a Corts Valencianes de 1991*, Valencia, Spain: Author.

——— (1995), *Base de Datos de Resultados de Eleccions a Corts Valencianes de 1995*, Valencia, Spain: Author.

ARGOS (2000), *Estimaciones de Voto para Elecciones a Cortes Valencianas*, Valencia, Spain: Presidencia de la Generalitat Valenciana. Subsecretaría de Planificación y Estudios, *http://www.pre.gva.es/argos/demoscopia/index2.html*, data acceded October 26, 2000.

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

——— (1997), "Probing Public Opinion: The State of Valencia Experience," in *Bayesian Case Studies 3*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 3–21.

Bernardo, J. M., and Girón, F. J. (1992), "Robust Sequential Prediction From Nonrandom Samples: The Election Night Forecasting Case," *Bayesian Statistics*, 4, 61–77.

Brown, L. B., and Chappell, H. W. (1999), "Forecasting Presidential Elections Using History and Polls," *International Journal of Forecasting*, 15, 127–135.

Brown, P., Firth, D., and Payne, C. D. (1999), "Forecasting on British Election Night 1997," *Journal of the Royal Statistical Society*, Ser. A, 162, 211–226.

DOGV (1987), "Ley 1/87, Ley Electoral Valenciana," *Diario Oficial de la Generalitat Valenciana*, 561, 31 Marzo, 1987.

Felip, J. M., Bellver, R., and Domenech, E. (1996), "Sondeos Y Resultados Electorales: Las Elecciones a Cortes Valencianas de 1995," *Revista Valenciana d'Estudis Autonòmics*, 15, 311–328.

Ibáñez, J., Martínez, A., Wert, J. I., Prat, J. C., Ortí, A., Corral, A., and Santamaría, J. (2000), "Debate Abierto: Elecciones Generales 12M/2000," *Investigación y Marketing*, 66, 55–72.

Jérôme, B., Jérôme, V., and Lewis-Beck, M. S. (1999), "Polls Fail in France: Forecasts of the 1997 Legislative Election," *International Journal of Forecasting*, 15, 163–174.

Ladd, E. C., and Benson, J. (1992), "The Growth of New Polls in American Politics," in *Media Polls in American Politics*, eds. T. E. Mann and G. R. Orren, Washington, DC: Brookings, pp. 19–31.

Nohlen, D. (1981), *Sistemas Electorales del Mundo*, Madrid, Spain: Centro de Estudios Constitucionales.

Rallings, C., and Thrasher, M. (1999), "Local Voters, National Forecasts Using Local Government by Elections in Britain to Estimate Party Support," *International Journal of Forecasting*, 15, 153–162.

Sanz, S., Díaz, B., Rica, C., and Quesada, A. (1996), "Edición Especial Elecciones Generales," *España Hoy: Noticias*, 93, 4 Marzo, 1996.

Theil, H. (1971), *Principles of Econometrics*, London: Wiley.