Electoral Studies 43 (2016) 124–132

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

Quick counts in the Mexican presidential elections: A Bayesian approach

Manuel Mendoza, Luis E. Nieto-Barajas*

Department of Statistics, ITAM, Mexico

ARTICLE INFO

Article history:
Received 9 September 2015
Received in revised form
4 May 2016
Accepted 23 June 2016
Available online 27 June 2016

Keywords:
Quick count
Bayesian approach
Noninformative prior
Normal model

ABSTRACT

In all democracies, anticipating the final results of a national election the same day the voters go to the polling stations is a matter of interest, for television stations and some civil rights organizations, for example. The most reliable option is a quick count, a statistical procedure that consists in selecting a random sample of polling stations and analysing their final counts to forecast the election results. In Mexico, a particularly important quick count is organized by the electoral authority. The importance of its results requires this exercise to be designed and executed with specially high standards far beyond those used in commercial studies of this type. In this paper, the model and the Bayesian analysis of the quick counts conducted by the Mexican authority, during the presidential elections in 2006 and 2012, are discussed.

1. Introduction

As in many other countries, in Mexico different organizations produce exit polls and quick counts the day of the election to get estimations of the final result. The largest and more sophisticated exercises are often those produced by the television networks although there are also studies generated by other media and, as usual in rather new democracies, a number of civil organizations, both foreign and local, report their results as well.

Predicting the result of an election has been of great interest for different authors. There are two main approaches. One based on opinion/exit polls before or during an election, and the other based on actual cast votes. When an opinion poll is conducted before the election day, a random sample of citizens who express their intention to vote, is selected. Once they reveal which party are willing to vote, this information is used to produce inferences. Alternatively, an exit poll is based on a random sample of polling stations. During the election day, a sample (usually of systematic type) of the citizens who cast their vote on each one of these stations is interviewed to know which party they voted for.

The use of opinion/exit polls has been severely criticised because they sometimes lead to misleading results. See for example, Brown et al. (1999) for the 1992 British general election, and Barreto et al. (2006) for the 2000 US presidential election. For the experience in Britain in 2005, Curtice and Firth (2008) not only discuss the difficulties that made impossible to conduct a traditional exit poll but describe a method which combines information from two consecutive elections. To this end, a panel of polling stations is defined where the information regarding each election is recorded by an exit poll. These data are used to estimate the change in each party's share of the vote. These estimates and the final results for the first election are used to forecast the results of the election in course. Inferences are successful although the method relies on specific characteristics of the British electoral system. Additionally, Anand and Jenkins (2004) raise some concerns about the fairness of opinion and exit polls in India since they claim can influence voting behaviour.

For the second approach based on final counts data, one possibility is to use the flow of results as they are produced when the polling stations close. This information is not the result of a random selection although a statistical model can be used to relate these final counts with those of previous elections in the same stations. In this regard, Bernardo and Girón (1992) proposed a Bayesian hierarchical model to predict the unobserved swings (difference between the votes from present and past elections). They work at polling station level and assume the swings to be exchangeable. On the other hand, Pavia-Miralles (2005) uses multivariate regression models to predict the current election data based on several past

* Corresponding author. Department of Statistics, ITAM, Rio Hondo 1, Col. Progreso Tizapan, 01080, Mexico D.F., Mexico.
E-mail addresses: mendoza@itam.mx (M. Mendoza), lnieto@itam.mx (L.E. Nieto-Barajas).

http://dx.doi.org/10.1016/j.electstud.2016.06.007
0261-3794/

elections at the poll level and carries out a frequentist analysis. In these two cases, the authors acknowledge that the final counts they use to fit their models, are not random. They treat the information as random and claim that their estimates improve as more polls become available.

Working with the first available poll-by-poll final results (not at random), has the risk to produce misleading results, specially in tight elections where the competing candidates votes are very close to each other. There are some strategies to make a non random sample to be representative. For instance, Sedransk and Clyde (1966) propose a post stratification technique based on population sizes and past elections. Interestingly, Brown et al. (1999) describes a method which combines nonrandom final counts data with prior information arising from an exit poll. Other recent example which uses non random final counts data is Fisher (2015).

Final counts data can also be obtained from a random sample of polling stations. Once these stations close and produce final counts, the information is analysed to obtain forecasts. This is the type of exercise usually known as quick count and is used by a number of organisations to anticipate the final results of an election. In particular, a quick count is considered a useful tool against fraud and other illegal practices (Estok et al., 2002). Statistical methods for the production of estimates in quick counts, usually rely on the well-known techniques for inference in survey sampling studies (e.g. Cochran, 2001). The basic problem is then to estimate both the total number of valid votes and the total number of votes in favour of each candidate. The proportion of interest is then estimated using the corresponding ratio. No explicit model is used to describe the raw data (nonparametric assumption) although a normal model is used to approximate the sampling distribution of the estimates based on asymptotic theory.

From a Bayesian point of view, one contribution that deals with actual cast votes is that of Bernardo (1984). He uses an information measure to define a set of polling stations whose final results were close to the national ones in a previous election (non random selection). For each of these stations, he collects the first 100 cast votes and uses this partial counts to produce inferences. Bernardo assumes a two stages model with a multinomial distribution for the votes in each polling station. The posterior mean of the proportion of votes in the station in favour of each party is obtained as an estimate. Finally, the distribution of the vector of log-odds of these estimated proportions is approximated with a multivariate normal and a hierarchical prior depending on the true national proportions is used to get the posterior distribution of interest. This type of study is not a quick count since the selected poll stations are not a random sample and only partial counts data are used from each station.

Basing the predictions on past election results might be technically correct and useful but is politically unacceptable for a quick count organized by the electoral authority in Mexico. Parties could incorrectly claim that the resulting forecasts are deliberately biased in favour of the *status quo*. Interestingly, political parties never questioned the sample design used in the Mexican quick counts which makes use of past electoral information.

In this article we propose a Bayesian parametric model which relies on final counts data collected from the polling stations in a random sample. This model was used by the National Electoral Institute (INE), formerly Federal Electoral Institute (IFE), during the 2006 and 2012 presidential elections in Mexico. We illustrate the performance of our model with the 2006 election which is the tightest election ever organised in Mexico. Results show that our model is reliable providing good forecasts with the right precision to call the winner.

The use of Bayesian models in political science is becoming a common practice. For instance, Darmofal (2009) compares the performance of Bayesian spatial frailty models versus non spatial and non frailty models using the U.S. House members' position announcements on NAFTA. Stegmueller (2013) implements a Monte Carlo experiment to compare frequentist and Bayesian approaches in the determination of the number of countries in a multilevel (hierarchical) model. Additionally, Hare et al. (2014) carries out Bayesian analysis of the Aldrich-McKelvey scaling to analyse American citizen's ideological preferences.

The outline of the paper is as follows. In Section 2 we review the political background that gave rise to the creation of INE. In Section 3, we describe the details of the INE quick count. In Section 4 we present the model used to process the data and show the results obtained in 2006 for the presidential election. Finally, Section 5 concludes with some final remarks.

2. Political background

For more than seventy years (1929–1994) the presidential elections in Mexico were won by the political party currently known as *Partido Revolucionario Institucional* (PRI). Only a few years after the revolution war (1910–1917) this party was created as an instrument for which the groups that emerged triumphant should organize themselves to share the power. For the general population, there was no distinction between government and party. In fact, nowadays that structure, which claimed to be a democratic system, is known as a "unique party regime". Any opposition was essentially annihilated by means of physical violence, all kind of threats, and many forms of bribery.

Only after a particularly well documented repression of students' political movement in 1968, when the army was used against demonstrations with the result of many deaths, in the 70's the government decided to allow the real existence of other political parties. As a consequence, a small number of organizations were able to participate in the electoral processes, although only in a marginal fashion since all aspects were under strict control of the government. For example, the approval of new parties and the regulation for the existing ones was entirely in hands of the Ministry of Interior. Even more, every election was also organized by the same Ministry.

At a very slow pace, Mexico evolved to a system which, nowadays, is close to its counterparts in many democratic countries. Now the elections are organized by INE, an autonomous body with a Board of Directors (*Consejeros*) appointed by the House of Representatives (*Cámara de Diputados*). Besides the role of organizer, INE also acts as a referee among the parties and can impose sanctions if anyone of them breaks an electoral rule. In addition, there is an special Court of Electoral Justice (*Tribunal Electoral*) where the parties can submit any complain regarding other parties, any government officer or even INE itself. It is interesting to know that for the presidential election, INE is, among other duties, in charge of providing public funds to the parties, to supervise the campaigns, to organize the logistics of the election, to hire all the required personnel, to train these staff as well as the citizens who collaborate as officers in the polling stations, to collect and count the votes and to announce the results. However, the final declaration of a winner, if there is one, does not come from INE. Since the parties might find reasons to complain, before or during the process, the winner is only announced by the Court of Electoral Justice after the analysis of every submitted complain.

Remarkably, as a result of the first presidential election organized by INE, in 2000, the office was won by an opposition (right wing) candidate. The same party won six years later, in 2006, when an amazingly closed election took place and a leftist opposition candidate ended in second place. In 2012 the old PRI party came back to the presidency.

One of the main concerns of INE has been not only to organize the elections but to increase the feeling of confidence and transparency on the whole electoral process. For these reasons, in addition to the usual activities related to the election organization, INE also invites a small group of specialists and creates a committee to design and operate the INE quick count.

3. The INE quick count

In Mexico, federal elections always take place on Sunday. Once a polling station is closed, votes are counted by the citizens in charge of the station, the results are recorded in a certificate and a copy of this document is posted on the wall outside the station. The following Wednesday, officials on each constituency (there are 300 of them over the country) accumulate the votes received from their polling stations and, by the following Sunday, they inform to INE about the election results. Finally, INE makes public the national results.

Thus, seven days must elapse after the election day to get official results of the process. In order to prevent unjustified victory claims during that period, INE organizes a quick count the same night of the election. This quick count faces most of the usual challenges for this kind of study. The role of the specialists is then to produce the sample design, to determine the sample size, to verify the systems that will be used to get the information from the polling stations and, of course, to propose and operate statistical methods to produce the desired inferences. An aspect that cannot be under emphasized is that while the sponsors of other quick counts can take the risk of announcing a false winner, this is simply unthinkable for INE. A wrong estimation by the INE quick count might destroy the confidence on the authorities organizing the election.

Taking all these elements under consideration, the committee explored in 2006 several alternatives for the sample design. The sampling unit was defined to be the polling station and the sampling frame consisted of 130,500 of them. The chosen design was stratified with allocation proportional to the size of the strata. Each one of the 300 constituencies was initially considered as a stratum and, in addition, the polling stations were classified according to its urban or non-urban condition to get a refined stratification. This choice has to do with the fact that one member of the House of Representatives is elected on each constituency and thus, after the campaigns, the citizens in a given constituency have been “treated” in an specific way, potentially different from those in other constituencies. In addition, electoral preferences in Mexico have been historically different in urban and rural areas, mostly because differences in the living conditions. In any case, since not all constituencies include both types of regions, the total number of strata is not $600 = 300 \times 2$ but 481 in 2006, and 483 in 2012.

The sample size was fixed in order to achieve an estimation error of 0.3% approximately. With a 95% of probability, and the proposed sampling scheme, this goal was reached with 7500 polling stations. The initial 7500 units were slightly increased in order to compensate for the delay on the arrival of information from the polling stations coming from the States in the western time zones. At the end, the sample sizes were, 7636 in 2006, and 7597 in 2012. Finally, the sample size for each stratum was allocated proportionally to its size.

To obtain the data, INE relies on a group of field employees (*capacitadores*) in charge of several activities which include reporting the finals counts of the polling stations in the sample. The data is collected in an information system with 300 nodes, one for each constituency.

The regulation allows INE to produce a quick count, but it explicitly states that the results must be announced no later than midnight. INE’s information system starts at 6 p.m. and, every 5 min, collects all the sample information sent. Thus, the system produces a sequence of accumulative files which are used to determine the available percentage of the sample, and are also examined to verify how this information is distributed over the country. In particular, these initial inspections allow the committee to identify regions where the sample is being collected at a slow rate.

In addition, this partial samples (full scrutiny information from an incomplete set of polling stations) might be analysed with the estimation methods to track the trend of the results. Obviously, the methods assume a complete probabilistic sample and these early analyses ignore the potential bias associated to the arrival pattern of the information. In any case, if, finally, the (almost) complete sample is obtained, it might be possible to get the final result and a report on the evolution of the inferences along the night. At this point it appears an advantage of the Bayesian analysis. If a stratum is not included in one available (partial) sample, other methods need to adjust the estimation procedure by assuming, for example, a stratification different from the originally planned. In contrast, the Bayesian analysis keeps the original stratification and uses the prior distribution for the parameters of the missing strata. We will discuss the selection of this prior in the following section.

4. Bayesian inference in the INE quick count

4.1. Model

The objective of the quick count is to produce reliable inferences on the proportion of votes in favour of each one of the candidates. Now, this proportion can be calculated in several different ways. The upper part of the ratio is always the total number of votes in favour of the candidate. However, the denominator may be the total number of registered voters (many of which do not show up the election day); other possibility is to use the total number of ballots used by the voters (these include all invalid votes), another option is to use the total number of valid votes (defined as the total number of votes minus the null votes). The latter choice is the legally adopted in Mexico and defines the *effective* proportion of votes.

Remember that the sampling scheme was stratified and thus, our model needs to be placed accordingly. The intuitive reasoning behind the proposed model takes into account that on every polling station we observe the total number of votes in favour of each candidate. Here, it is worth noticing that, both in 2006 and 2012, at most 750 citizens were registered as potential voters on every station (on average, there was around 500 potential voters on each polling station). The file with the names and photographs of these citizens is called the nominal list of the station. Now as usual, given a candidate, a Bernoulli random variable can be defined for each potential voter such that it gets the value one if his/her vote is cast and it is in favour of this candidate, and zero otherwise. Thus, the total number of votes in favour of this candidate in the station might be viewed as the sum of a large number of Bernoulli variables. Moreover, since these voters belong to the same strata, we assume that the corresponding parameters are all equal across voters. If the votes were cast independently, we could think of the sum as a binomial variable and to use its usual normal approximation. However, there is no reason to assume independence and, hence, the variance of the sum will not be equal to the sum of the individual variances. A rather simple way to deal with this situation is to adopt a normal approximation where the variance is not related to the mean and must be estimated separately. In order to make this statement more precise, let us introduce the following notation.

If the total number of polling stations in the country are divided into N strata, each one with a nominal list of size n_i, for $i = 1,\ldots,N$, we can define X_{ij} as the number of people in favour of candidate j in stratum i and θ_{ij} as the proportion of people in the nominal list of stratum i whose preference is for candidate j. Here, $\theta_{ij} = X_{ij}/n_i$, for $i = 1,\ldots,N$ and $j = 1,\ldots,J$. For the sake of simplicity, in the sequel we will refer to the 2006 election parties (in 2012, the structure was the same, with somehow different parties). Thus, we are considering $J = 5$ *categories* which correspond to the three main contenders, PAN (rightist), APM (an aggregate of parties leaded by the PRI) and PBT (a coalition of leftist parties), a category that represents all other parties called OTHERS, and the null votes plus the abstentions called NULL, which represents the people who cancelled their vote or simply did not attend to the poll. Therefore, the proportion of people in the nominal list of the whole country whose preference is in favour of candidate j is θ_j, which according to the stratification scheme, can be obtained from the θ_{ij} as

$$\theta_j = \sum_{i=1}^{N} \frac{n_i}{n}\theta_{ij}, \tag{1}$$

with n the size of total nominal list in the country. It must be noticed that the parameters of interest are not the θ_j's, because we are looking for the *effective* proportions of votes, which excludes the NULL category. Our interest is then focussed on the λ_j's defined as

$$\lambda_j = \frac{\theta_j}{\sum_{l=1}^{4} \varphi_l}. \tag{2}$$

As input for the quick count, information is collected at a poll level, so we define X_{ij}^k the number of people in favour of candidate j in poll k of stratum i, $k = 1\ldots,K_i$, where K_i denotes the total number of polls in stratum i. Then,

$$X_{ij} = \sum_{k=1}^{K_i} X_{ij}^k$$

According to the sampling design, within each stratum, a random sample of c_i polls is taken, giving a total of $c = \sum_{i=1}^{N} c_i$ polls in the sample. Our model then assumes that

$$X_{ij}^k \Big| \theta_{ij}, \tau_{ij} \sim \mathrm{N}\left(n_i^k\theta_{ij}, \frac{\tau_{ij}}{n_i^k}\right), \tag{3}$$

for $k = 1,\ldots,c_i$, $i = 1,\ldots,N$ and $j = 1,\ldots,J$. Here, n_i^k stands for the size of the nominal list in poll k of stratum i and τ_{ij}/n_i^k is the precision where, for each candidate, τ_{ij} is assumed to be constant within the corresponding stratum (and unrelated to θ_{ij}). Moreover, we assume that X_{ij}^k is independent of $X_{ij'}^k$, for $j \neq j'$. This is perhaps a stronger assumption, however, previous analysis with a more complex model that assumed dependence between these variables showed that the dependence was too weak and could be disregarded. On the other hand, the dependence structure between the parameters of interest (the λ's), is recovered through the functional relation (2). Now, given a sample of size c_i from model (3), the likelihood for each stratum ($i = 1\ldots,N$) and each candidate ($j = 1,\ldots,J$) is given by

$$\mathrm{lik}(\theta_{ij}, \tau_{ij}|\mathbf{x}_{ij}) \propto \tau_{ij}^{c_i/2} \exp\left\{ -\frac{\tau_{ij}}{2}\sum_{k=1}^{c_i}\frac{1}{n_i^k}\left(x_{ij}^k - n_i^k\theta_{ij}\right)^2\right\}.$$

This likelihood must be combined with an appropriate prior distribution describing the initial knowledge regarding the parameters θ_{ij} and τ_{ij}. There are plenty of information sources available to elicit a prior probability regarding these parameters. However, it must be recalled that this model is to be used in a quick count organized by INE, who must be, and behave, as an absolutely impartial authority. As a consequence, no matter the previous information arising from past elections, experts, electoral observers or any type of surveys, the prior for the INE quick count must be completely neutral. This is accomplished by using a non informative prior for the parameters of the model to avoid any subjective influence in the estimation process. Thus, we use the prior

$$p(\theta_{ij}, \tau_{ij}) \propto \tau_{ij}^{-1} I(\tau_{ij} > 0) I(0 < \theta_{ij} < 1), \tag{4}$$

where $I(A)$ stands for the indicator function of the set A. This choice makes use of two facts. First, θ_{ij}, despite of being a proportion, plays the role of a location parameter and τ_{ij}, as usual, is a scale parameter. Thus, the proposed prior is widely accepted as noninformative for (θ_{ij},τ_{ij}). Second, the parameters are considered independent *a priori*. This is a quite common assumption. Finally, and highly relevant for practical purposes, the prior for θ_{ij} is *proper* and then it can be used to simulate draws of this parameter. The importance of this property will become clear later.

After combining the prior distribution with the information from the data, we obtain that the posterior distribution for (θ_{ij},τ_{ij}), conditional on the data, is proportional to the product of a truncated normal for θ_{ij}, conditional on τ_{ij}, and a gamma distribution for τ_{ij}, that is,

$$p(\theta_{ij}, \tau_{ij}|X_{ij}) \propto N\left(\theta_{ij} \middle| \frac{\sum_{k=1}^{c_i} x_{ij}^k}{\sum_{k=1}^{c_i} n_i^k}, \tau_{ij}\sum_{k=1}^{c_i} n_i^k\right) I(0 < \theta_{ij} < 1)$$
$$\times\, Ga\left(\tau_{ij} \middle| \frac{c_i - 1}{2}, \frac{1}{2}\left\{\sum_{k=1}^{c_i}\frac{\left(x_{ij}^k\right)^2}{n_i^k} - \frac{\left(\sum_{k=1}^{c_i} x_{ij}^k\right)^2}{\sum_{k=1}^{c_i} n_i^k}\right\}\right) \tag{5}$$

for $i = 1,\ldots,N$ and $j = 1,\ldots,J$.

It is clear that the posterior distribution for τ_{ij} is proper only if the sample number of polls c_i, is bigger that 1. That is, we need to design the sample in such a way that each stratum has at least 2 polls, otherwise the posterior distribution of the corresponding parameter τ_{ij} will be improper. Note also that the truncation condition $0 < \theta_{ij} < 1$ in (5) affects both, θ_{ij} and τ_{ij}. So, in order to get draws from the posterior distribution we generate a τ_{ij} from the gamma distribution and then, conditionally on the value of τ_{ij}, generate a θ_{ij} from the normal distribution. If θ_{ij} belongs to the interval (0,1) we keep the generated pair (θ_{ij},τ_{ij}), otherwise we discard both values.

Posterior inference for the parameters of interest can be obtained via simulation. For each λ_j, a sample of size R can be easily generated from the marginal posterior distribution of θ_{ij} in all strata $i = 1,\ldots,N$ and for all categories $j = 1,\ldots,J$. Proportions θ_{ij} from all strata, are then used to compute the national proportions θ_j via the weighted average (1). Then, the θ_j's are used, via equation (2), to produce the desired samples of λ_j, $j = 1,\ldots,4$. Although the θ_j's are treated as independent, the λ_j's are not independent, due to the nonlinear transformation (2). So by computing this ratio (2), certain dependence within the λ_j's is introduced.

4.2. Results

To illustrate the results obtained from the model, we again concentrate on the 2006 election. During the night of the election day, the information was sent directly to a server which generated a file with the accumulated information up to that time. These files were generated every 5 min starting with the first at 18:00 h, just at the time when most of the polls were supposed to be closed. As expected, the first files did not have information from all strata. In fact, this was the situation for more than 3 h. The proposed model was able to cope with this issue since inferences on the parameters of interest (the λ_j's), are produced simulating draws from the corresponding distributions of the θ_{ij}'s and then using relations (1) and (2) to get the required parameters. If the available sample includes information from a given stratum, then draws are obtained from the joint posterior distribution defined by (5). However, in the case that no sample is available from a stratum, θ_{ij} is simulated from its uniform marginal prior, as in (4). Since this noninformative prior is the same for every candidate proportion in the stratum, we get that $P(\theta_{ij}>\theta_{ij'}) = P(\theta_{ij}<\theta_{ij'})$ for $j \neq j'$. This equilibrium could be broken only with information coming from the sample.

In any case, every 5 min we generated $R = 10{,}000$ draws from the joint posterior distribution of $\lambda=(\lambda_1,\lambda_2,\lambda_3)$, the effective proportion of votes in favour of PAN, APM and PBT, respectively. Using these draws we could approximate any characteristic of the joint and marginal distributions. In particular, we calculated posterior credible intervals to estimate the proportions of votes. More interesting, we were also able to report joint inferences for these parameters.

In order to facilitate the understanding of our reports, taking into account that among the final readers there would be many people not used to properly interpret statistical results (specially, if they come from a multivariate distribution), we devised a graphical report that proved to be clear and easy to understand. We used the multivariate draws from λ to produce the two-dimensional projections of the original distribution. The resulting figure was a four panel graph where the panels consisted of bivariate dispersion diagrams, that were forced to include the identity (45^o degrees) line, of the simulated draws from the posterior distribution of λ_1, λ_2 and λ_3. The colours chosen for each party are: blue for PAN, red for APM and yellow for PBT. On top of each diagram we included the estimation of the posterior probability of the proportion in favour of candidate j, λ_j, to be superior to the proportion in favour of candidate j', $\lambda_{j'}$. To do this we just divide the number of pairs in the simulated draws where $\lambda_j>\lambda_{j'}$ over 10,000. This probabilities allowed us to quantify, at any moment, the possibilities of one candidate winning the other. To facilitate the reading of these numbers, we printed them in the colour of the winning candidate at the time, that is, if $P(\lambda_j > \lambda_{j'}) > 0.5$ then we printed this probability in the colour of the party of candidate j, otherwise we used the

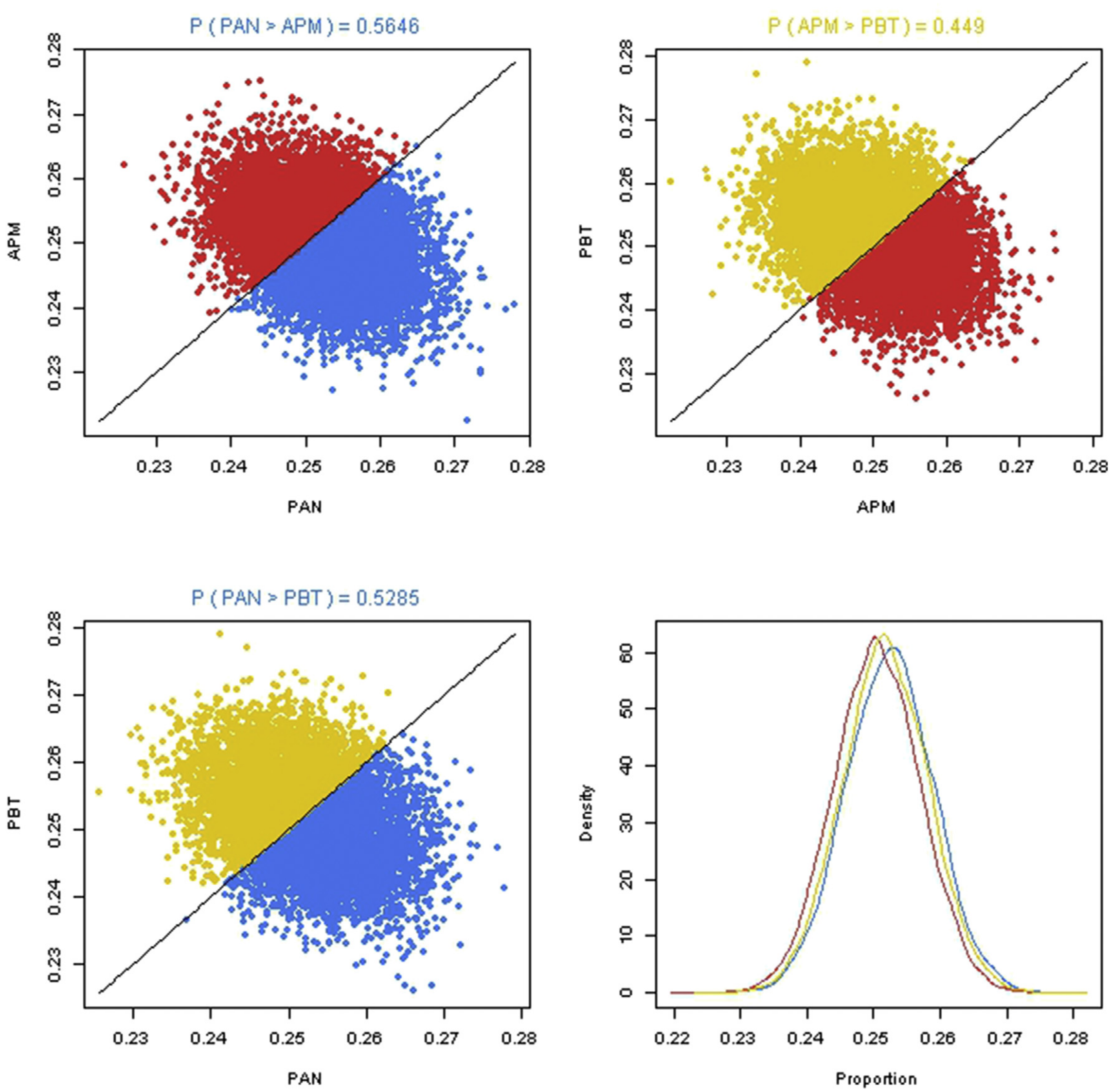

Fig. 1. Identity dispersion diagrams + posterior densities: 18:30 h.

colour of candidate j'. Finally, the fourth panel (right-bottom) contained a kernel smoothing estimator of the three posterior marginal densities for each λ_j, $j = 1,2,3$.

In 2006, the first file arrived at 18:00 h and contained only 3 polls (out of 7636), which corresponds to a 0.4% of the total sample. At 18:30 h we reached 2% of the sample. Fig. 1 shows how the proportions for the three main candidates were at this time. We can see that the three candidates were tied with the three posterior densities (fourth panel) placed on top of each other. This, as discussed, is a consequence of the lack of information and the proposed prior distribution.

As soon as 18:45 h, with 8.45% of the sample, the party APM had practically disappeared from scene. The first two panels in Fig. 2 mainly contain one colour, the red colour of party APM had almost vanished. In the fourth panel we can see a clear separation of the party APM with respect to the other two parties. It must be recalled that there was no guaranty that party APM was not coming back to the battle. These were partial results based on the available information at the time which not necessarily could be considered a (representative) random sample. It could have been perfectly possible that voters in favour of party APM were still not represented in the sample. At the end, party APM never came back, not even at 22:15 h when the committee decided to produce their final results with 95.12% of the planned sample. This decision was made after an examination of a series of maps where the distribution of the available sample was displayed over the country, showing that it was evenly distributed. It was also verified that all 300 constituencies appeared in the sample. All 481 but one strata were present as well.

To communicate the feeling that the entire sequence of dispersion diagrams caused among the members of the committee as well as among the officials of the INE, we present here some diagrams of the sequence. In order to save space and since the APM party was not in the competition anymore, from here onward we only include the diagrams of PBT versus PAN. Fig. 3 shows the diagrams at 19:00, 19:30, 20:00 and 20:30 h, respectively. These diagrams depict a rather thrilling tale; sometimes the probability of one party winning the other favoured PBT and some other times favoured PAN. From this graphs, the last one at 20:30 h with already 73% of the sample, it was unclear which party was going to win.

Fig. 4 includes the dispersion diagrams at 21:00, 21:30, 22:00 and 22:15 h. At 21:00 h with 83% of the sample, a clearer view was

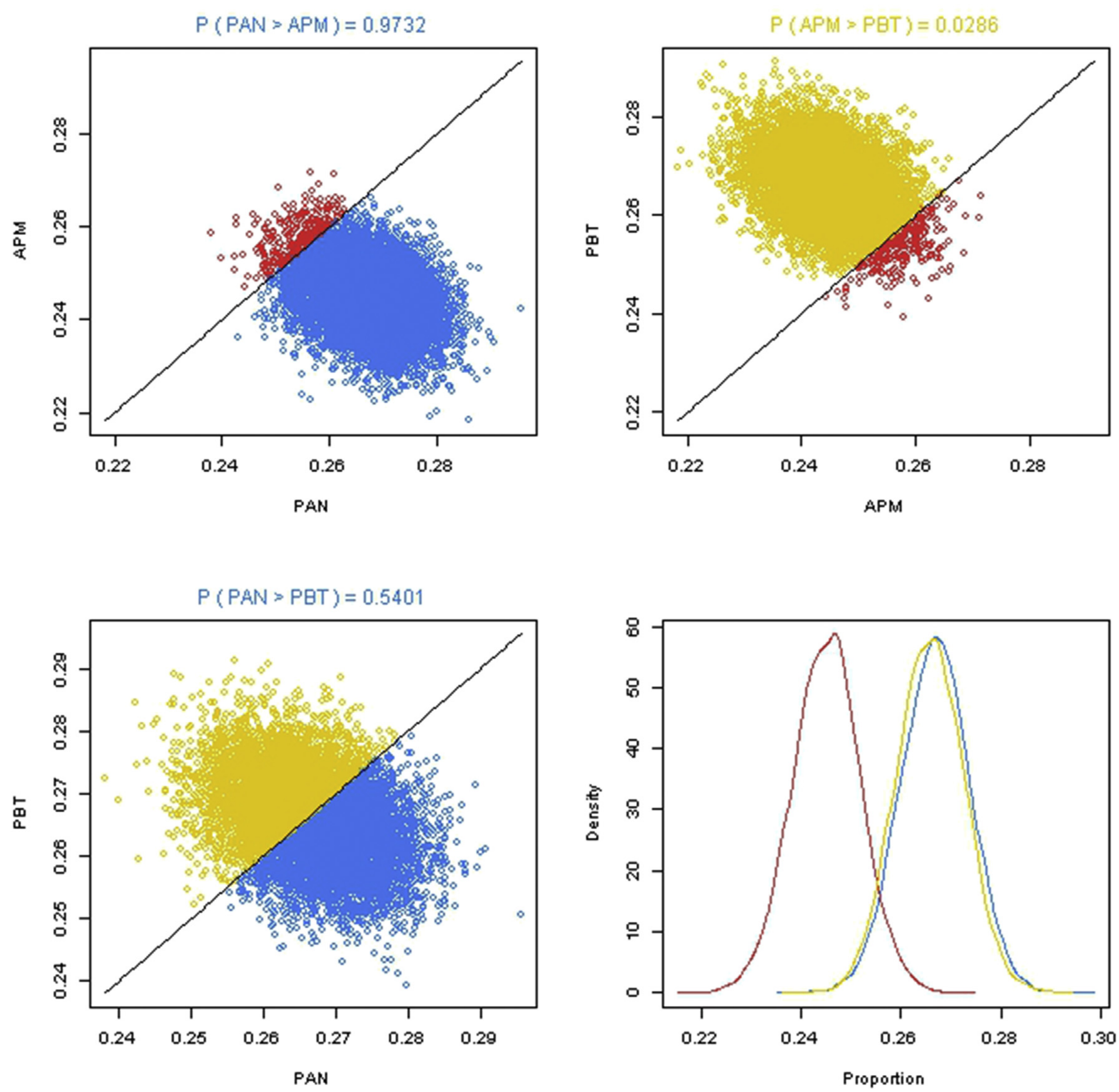

Fig. 2. Identity dispersion diagrams + posterior densities: 18:45 h.

appreciated where, if this partial sample could be considered itself a random sample, the probability of PAN overtaking PBT would be 0.8462. However there is no way to guarantee the randomness of the partial sample and this probability must be read with extreme caution. Nevertheless, with more than 80% of the planned sample collected, this number suggested that it was going to be difficult to see another reversal of the result. As already mentioned, at 22:15 h with 95% of the sample, the final decision was made. There was a smashing probability of 0.9994 of PAN overtaking PBT, the cloud of points in the last panel of Fig. 4 was practically painted in blue located in one side of the identity line. The rest of the sample continued to arrive after 22:15 h presenting no change with respect to the results shown here.

The final Bayesian interval estimates, based on 99% posterior probability intervals, for the three main parties, together with the reported official results before complaints and refutations (PREP, 2006) are presented in Table 1. For comparison purposes, we also computed the frequentist 99% confidence intervals given by $\widehat{\lambda}_j \pm Z_{\alpha/2}\sqrt{\widehat{V}(\widehat{\lambda}_j)}$, with $\widehat{\lambda}_j = \widehat{T}_j/\widehat{T}$, a ratio estimate, where $\widehat{T}_j = \sum_{i=1}^{N} \frac{N_i}{c_i} \sum_{k=1}^{c_i} X_{ij}^k$, $j = 1,\ldots,4$, $\widehat{T} = \sum_{j=1}^{4} \widehat{T}_j$ and $Z_{\alpha/2}$ is the corresponding normal quantile with α=0.005. As for the approximated variance, we used (Särndal et al., 2003),

$$\widehat{V}\left(\widehat{\lambda}_j\right) = \sum_{i=1}^{N} N_i^2 \left(\frac{1}{c_i} - \frac{1}{N_i}\right) \widehat{V}_i(G_j),$$

with

$$\widehat{V}_i(G_j) = \frac{1}{c_i - 1} \sum_{k=1}^{c_i} \left(G_{ij}^k - \overline{G}_{ij}\right)^2, \text{ and } G_{ij}^k = \frac{X_{ij}^k - \widehat{\lambda}_j T_i^k}{\widehat{T}},$$

where, $T_i^k = \sum_{j=1}^{4} X_{ij}^k$ and $\overline{G}_{ij} = \frac{1}{c_i} \sum_{k=1}^{c_i} G_{ij}^k$.

It is worth emphasising that since one of the strata was not observed (stratum 388 specifically), to calculate the frequentist estimates, the missing stratum had to be collapsed with its neighbour stratum. However, for the Bayesian analysis no collapsing is needed since inference relies on the (noninformative) prior distribution for that stratum.

From Table 1 it can be verified that the Bayesian estimates were accurate for all proportions in the sense that they all include the

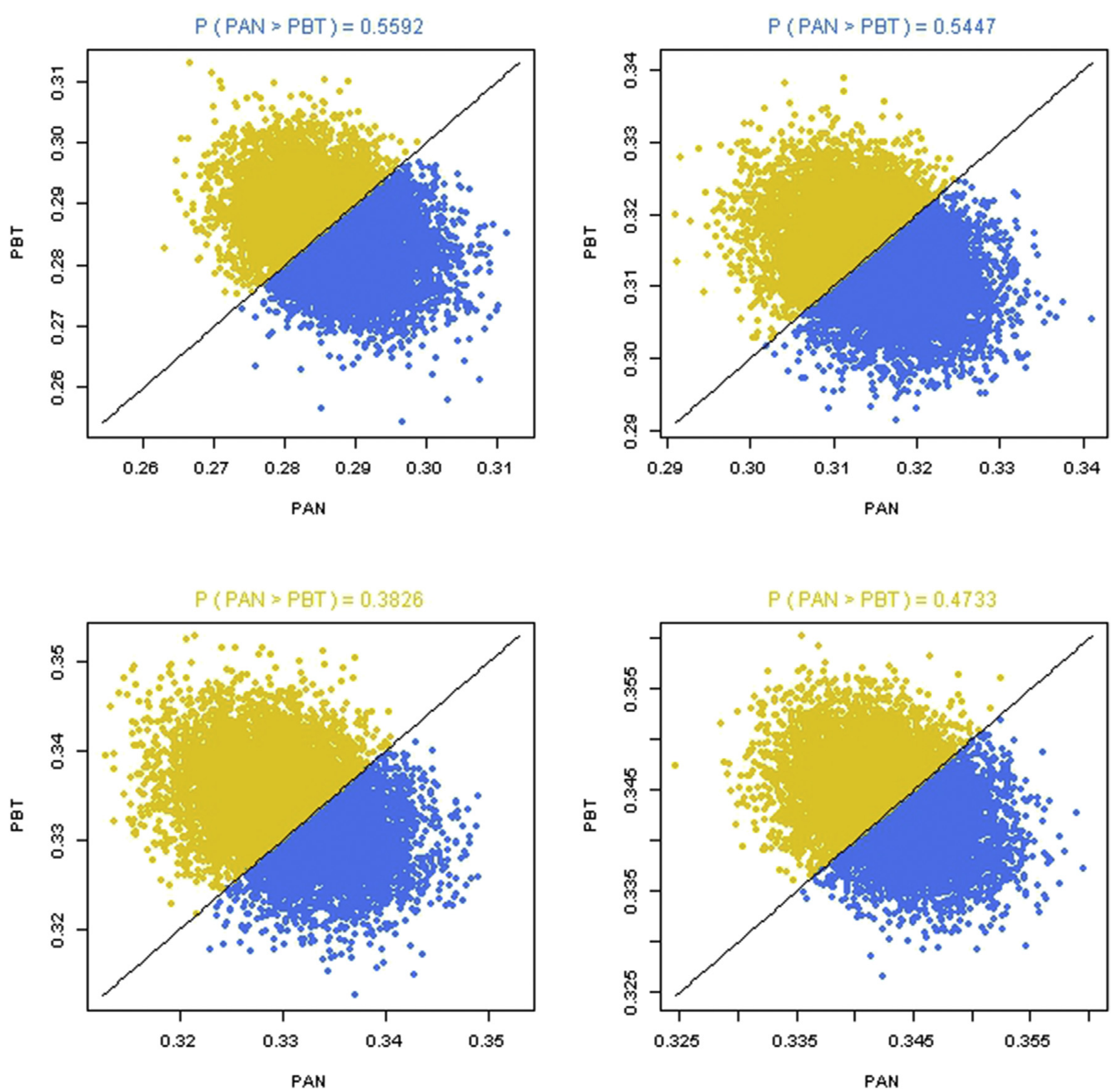

Fig. 3. Identity dispersion diagrams: From left to right and from top to bottom, 19:00, 19:30, 20:00 and 20:30 h.

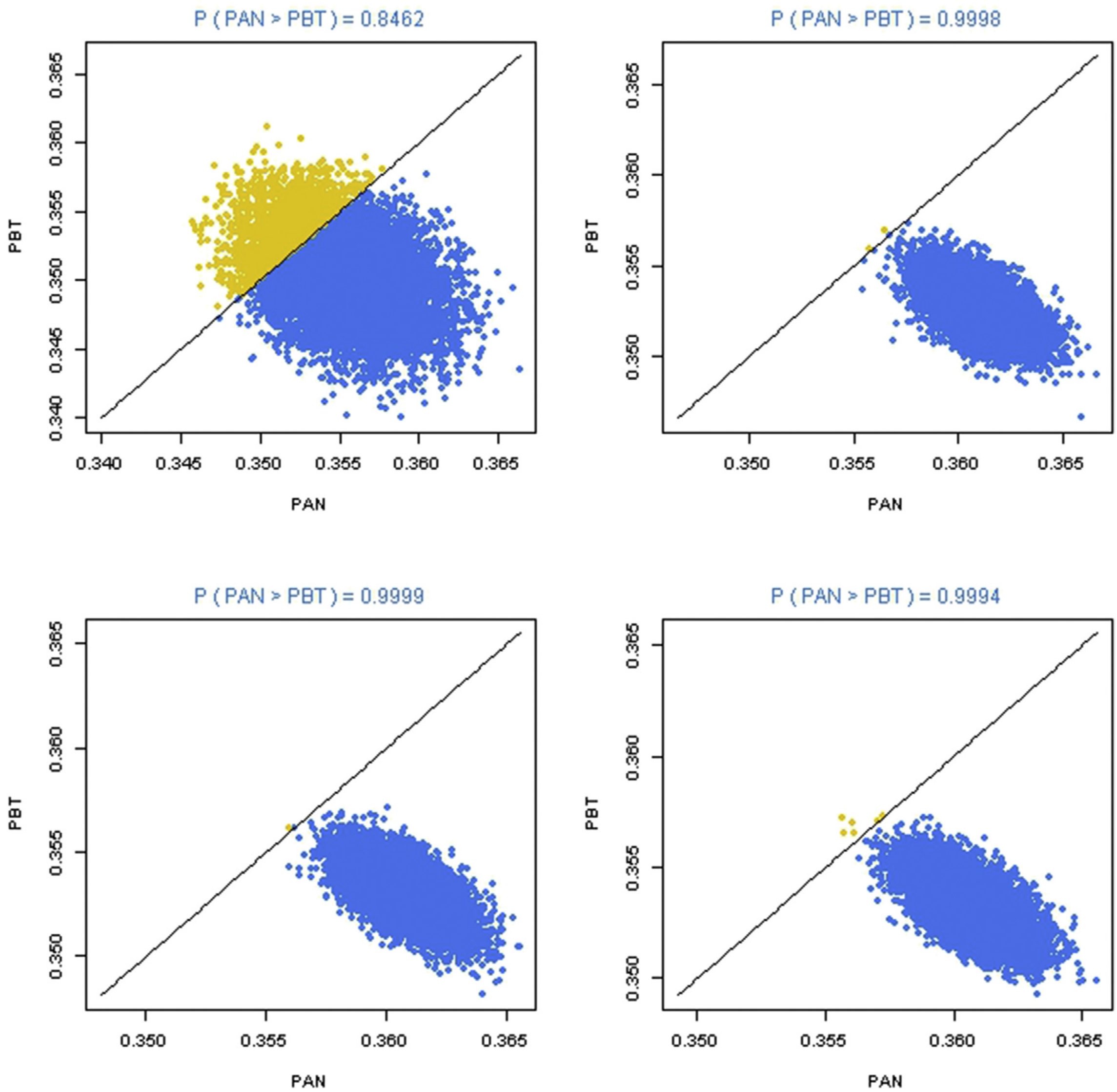

Fig. 4. Identity dispersion diagrams: From left to right and from top to bottom, 21:00, 21:30, 22:00 and 22:15 h.

final proportions computed after counting all votes in the country. In contrast, the frequentist estimates underestimate the uncertainty producing very narrow intervals. In consequence, two of the three proportions (PAN and APM) lie outside their corresponding intervals. It must be recognised however, that both intervals (frequentist and Bayesian) would fail to capture the true proportions if the level of confidence (probability) were reduced. The fact that the frequentist intervals are narrower than the Bayesians is not surprising. The uncertainty regarding the unknown variance of the estimate is ignored when an approximated estimation is used to produce the intervals whereas the posterior distribution takes that lack of information into account. Moreover, the frequentist estimates are nonparametric whereas the Bayesian estimates are based on the parametric assumption (3).

Table 1
Estimates of λ_j, j=1,2,3 and final results before refutations. Intervals are computed with 99% probability (confidence).

	PAN	APM	PBT
Final results	35.91	22.19	35.29
Bayesian	(35.73, 36.38)	(21.72, 22.22)	(35.05, 35.62)
Frequentist	(36.04, 36.19)	(21.91, 22.01)	(35.27, 35.39)

Our Bayesian model produced disjoint 99% intervals, for the two leading parties, whose limits where really close to each other but separated. This fact is relevant because, non expert readers might think that the risk of a false winner call was large. This was not true and the evidence is provided precisely by the 22:15 dispersion diagram (bottom right panel in Fig. 4), where the probability of PAN being the winner is 0.9994. In fact, the marginal intervals could intersect and nonetheless the relevant information to call a winner is provided by the corresponding joint bivariate posterior distribution.

Another interesting aspect of predicting the results of an election is the evolution of the estimates along time when more information was available. In Fig. 5 we include the evolution of the 99% posterior intervals for the two leading parties over time. This graph allows us to assess how tough the competition was between parties PAN and PBT. It is worth noting that this figure needs to be read with care, no matter how the evolution of the estimates were, the valid estimates are the last ones in the final cut at 22:15 h. Moreover, the dynamic described by this evolution graph, only

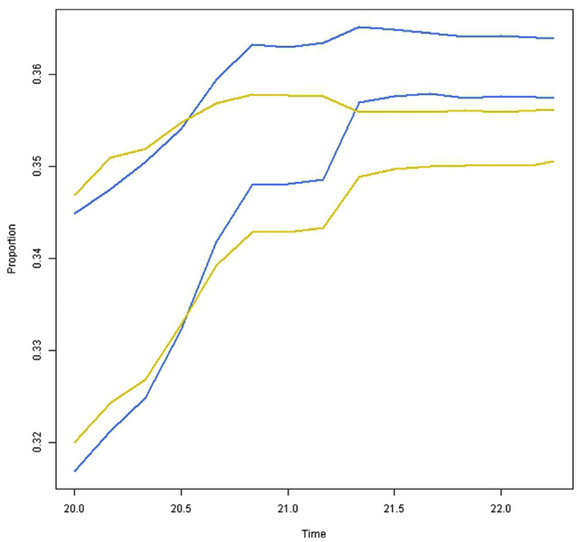

Fig. 5. 99% posterior credible intervals along time for the proportions of the two leading parties, PAN and PBT.

reflects the way the sample arrived to the server. If a party, which was in a winning position at some point of time, is later in a second or third place it does not mean a change in the voters behaviour; it only reflects the fact that the information from some polls where the party won, arrived first.

5. Discussion and concluding remarks

In Mexico, the authority in charge of the national elections is the National Electoral Institute. As an unusual practice if we review most of the electoral authorities in other countries, INE has decided to conduct a quick count the night of the election. This exercise faces a number of specific challenges. For example, it must reassure the confidence on the authority by eliminating the possibility of a false call and, it must deliver results before midnight.

The Bayesian model proposed here was used by the authors on behalf of INE during the presidential elections in 2006 and 2012. It was shown to be reliable producing better results than the classical methods for the particular dataset analysed in this paper, where the difference between the two leading candidates was only 0.62 percentage points. The proposed method also provides an easy way to communicate the results to a wider audience, in particular with respect to the possible ties among candidates via the dispersion diagrams of the joint posterior samples.

The model is simple and does not require of collapsing strata when no data is available, as is the case for the classical approach. However, implementation of our model requires simulation to produce inferences. This can be easily performed in few seconds in any personal computer so that the sequential files arriving every 5 min can be analysed without any difficulty.

It must be stressed that this model uses final counts from a random sample of polling stations. If the polling stations are not randomly selected the probabilistic analysis used to measure the credibility of the forecast does not longer hold.

In 2012 INE decided to implement again this Bayesian approach in the respective quick count analysis; the final results were not so close and the Bayesian model accurately estimated the proportions and detected the winner without any difficulty. We believe that this method will be used in several other future elections in Mexico.

Acknowledgements

This research was supported by *Asociación Mexicana de Cultura, A.C.* and by *Sistema Nacional de Investigadores* (CONACYT), Mexico.

References

Anand, S.A., Jenkins, C.J., 2004. Exit polls: debating freedom or fairness. Econ. Political Wkly. 39, 4971–4973.
Barreto, M.A., Guerra, F., Marks, M., Nuño, S.A., Woods, N.D., 2006. Controversies in exit polling: implementing a racially stratified homogenous precinct approach. Political Sci. Polit. 39, 477–483.
Bernardo, J.M., 1984. Monitoring the 1982 Spanish socialist victory: a Bayesian analysis. J. Am. Stat. Assoc. 79, 510–515.
Bernardo, J.M., Girón, F.J., 1992. Robust sequential prediction from nonrandom samples: the election night forecasting case. Bayesian Stat. 4, 61–77.
Brown, P.J., Firth, D., Payne, C.D., 1999. Forecasting on British election night 1997. J. R. Stat. Soc. Ser. A 162, 211–226.
Cochran, W.G., 2001. Sampling Techniques. Wiley, Chichester.
Curtice, J., Firth, D., 2008. Exit polling in a cold climate: the BBC-ITV experience in Britain in 2005. J. R. Stat. Soc. Ser. A 162, 211–226.
Darmofal, D., 2009. Bayesian spatial survival models for political event processes. Am. J. Political Sci. 53, 241–257.
Estok, M., Nevitte, N., Cowan, G., 2002. The Quick Count and Election Observation. A NDI Handbook for Civil Organizations and Political Parties. National Democratic Institute. Available at: https://www.ndi.org/files/1417_elect_quickcounthdbk_0.pdf.
Fisher, S.D., 2015. On-the-Night Prediction of the Scottish Independence Referendum 2014. Wiley StatsRef: Statistics Reference Online, pp. 1–3.
Hare, C., Armstrong II, D.A., Bakker, R., Carroll, R., Poole, K.T., 2014. Using Bayesian Aldrich-McKelvey scaling to study citizen's ideological preferences and perceptions. Am. J. Political Sci. 59, 759–774.
Pavia-Miralles, J.M., 2005. Forecasts from nonrandom samples: the election night case. J. Am. Stat. Assoc. 100, 1113–1122.
PREP, 2006. Memorias del Programa de Resultados Electorales Preliminares. Available at: http://www.ine.mx/docs/INE-v2/UNICOM/UNICOM-PREP/2006/Memorias_PREP_2006.pdf.
Särndal, C.-E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer, New York.
Sedransk, J.H., Clyde, R.W., 1966. A generalized technique for predicting election results from early returns. Public Opin. Q. 30, 416–422.
Stegmueller, D., 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. Am. J. Political Sci. 57, 748–761.