ISER Working Paper Series

The total survey error paradigm and pre-election polls: The case of the 2006 Italian general elections

Laura Fumagalli
Institute for Social and Economic Research
University of Essex

Emanuela Sala
University of Milano Bicocca

No. 2011-29
November 2011

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

www.iser.essex.ac.uk

Non technical summary

Pre-election polls sometimes fail to reach the purpose for which they are carried out: to provide accurate predictions of electoral outcomes. There are different reasons why this may occur: the list from which respondents are sampled may be inaccurate, respondents may decline participation in the study (non-response), respondents' voting behaviour may change between the time the polls are carried out and the Election Day.

This paper aims to assess the role that each of these factors plays in determining the accuracy of the pre-election polls carried out for the 2006 Italian General Elections. These polls overestimated the votes to the left-wing coalition. Our analysis shows that the first two factors may play a key role as respondents excluded from the list as well as non-respondents seem to be more likely to vote for the center-right coalition. The impact on the accuracy of the polls of respondents who change their mind remains, however, unclear. This paper also suggests two methods to overcome some of the limitations of the survey data. In some cases there is not a single survey which contains all the variables the researchers are interested in. Researchers often face the challenge of how to combine different data sets.

In this paper we use two data sets, the ITANES survey and the Multi Purpose survey. The former collects information on respondents' inclusion in the sampling list (but does not collect data on voting behavior) whereas the latter collects information on voting behavior (but does not gather data on inclusion in the sampling list). To fully exploit the information collected in both surveys, we develop two strategies. First, we use statistical matching to produce a single data set that contains the two pieces of information we are interested in, i.e. inclusion in the list and voting behavior. Second, we create a set of weights correcting for the distortion introduced by the inaccurateness of the list.

The Total Survey Error Paradigm and Pre-Election Polls: the case of the 2006 Italian General Elections

Laura Fumagalli*and Emanuela Sala†

November 8, 2011

Abstract

Pre-election polls sometimes fail to reach the purpose for which they are carried out: to provide accurate predictions of electoral outcomes. By looking at the 2006 Italian General Elections, this paper aims to assess the role that different factors play in determining the accuracy of the pre-election polls. We find strong evidence that the quality of the sampling frame and non-respondents may contribute to biasing the polls results. This paper also aims to show how to overcome some of the limitations of the survey data by using statistical matching techniques and weighing procedures.

Keywords: Total survey error, coverage error, statistical matching, weighting, pre electoral polls.

JEL Classification: C81, C83, D72

*ISER, University of Essex, UK, lfumag@essex.ac.uk (corresponding author).

†University of Milano-Bicocca, Italy.

[1]We are grateful to Erich Battistin, Francesco Figari, Romina Gambacorta, Olena Kaminska, Luke Keele, Peter Lynn, Stephanie McFall, Anja Neundorf, Rafael Novella, Riccardo Puglisi, and the participants of the XVII World Congress of Sociology (University of Gothenburg, Sweden), the 2010 Elections, Public Opinion and Parties Annual Conference (University of Essex, UK) for useful comments and suggestions.

1 Introduction

Pre-election polls sometimes fail to reach the purpose for which they are carried out: provide accurate predictions of electoral outcomes. In the last twenty years, for example, the polls failed to accurately predict the outcome of the 1992 British General Elections (Jowell, Hedges, Lynn, Farrant, and Heath, 1993; Smith, 1996), the 1998 Quebec Elections (Durand, Blais, and Vachon, 2001, 2002), the 2002 and 2007 French presidential elections (Durand, Blais, and Larochelle, 2004; Durand, 2008), the European elections in Portugal (Magalhães, 2005) and the 2008 Primary Elections in the States (AAPOR, 2009).

The case of Italian polls is not an exception: during the 2006 Italian General Elections, the majority of the pre-election polls carried out for the Elections of the Lower House put the center-left coalition (The Union) ahead by an average of 3 or 4 points. Contrary to what was predicted, the left-wing coalition won the elections with only a narrow margin of 0.1 percent of votes. The polls carried out from March, 1^{st} 2006 to the beginning of the embargo period (March, 24^{th})[1] overestimated the votes of the left-wing parties Communist Refoundation and Olive Tree by 0.7 and 1.5 points and underestimated the votes to the right-wing parties Go Italy and Northern League by 2.4 and 0.2 points.[2] Vote intentions for National Alliance were overestimated by 0.2 points.

The aim of the paper is to shed light on the role of different sources of error on the performance of the Italian polls carried out for the 2006 General Elections of the Chamber of the Deputies (or Lower House).[3] Following the approach of AAPOR (2009), our work is an attempt to extend the total survey error (TSE) paradigm to the case of pre-election polls.

The TSE paradigm is a theoretical framework aiming at developing best practices to optimize the quality of survey data under budgetary constraints (Biemer, 2010; Groves and Lyberg, 2010) and rests on a classification of error sources. In this paper we distinguish between two types of errors: sampling

[1] In Italy such an embargo period starts two weeks before the elections.

[2] We focussed on the 20 polls that recorded vote intentions at the party level.

[3] Although the Italian Parliament is constituted of two Chambers, most of the polls are carried out to predict the outcome of the lower-House Chamber of Deputies. The two Chambers have different electoral systems, being the one for the election of the Senate regionally based and the one for the election of the Lower-House Chamber of Deputies nationally based.

error and non sampling error (Biemer, 2010). Sampling error is specific to the single polls and it depends on how the sampling has been implemented in terms of sampling scheme, sampling size and estimator choice. In order to study it, we would need to analyze the actual data coming from the polls. This is not the aim of this paper. Instead, we propose an alternative approach i.e. we use different survey data to analyze non sampling errors and we derive conclusions which can be generalized to any pre-election poll. Following Biemer's terminology, we are interested in three types of non sampling error: namely frame, specification and non-response error which, according to the most recent studies on the topic, are found to be the main drivers of the poor performance of the polls.

Frame error is the type of error arising when the frame used for the sampling does not exactly cover the population under study. This occurs when eligible units may be excluded from the sampling frame (non-coverage error), whereas ineligible units may be included. In the case of the pre-election polls, we are mainly concerned about non-coverage error, given that the frames commonly used usually fail to include the whole population of voters.

Polls are usually carried out by telephone on a sample of landline household numbers. This practice can be a source of non-coverage error, given that cell-only households are excluded. The expansion of mobile phones has been posing serious challenges to the quality of the sample frames used in CATI surveys worldwide. Two are the major sources of concern: the reduction of landline households and the differences between cell-only and landline households. With respect to the former, the recent AAPOR report on cell-phone surveys (AAPOR, 2010) states that the level of coverage of a sampling frame of landline telephone in the U.S. is about 80 percent and that the prevalence of cell-only household varies by State. The European case is in general very similar: in 2007, 73 percent of European households have a fixed telephone access. In Italy, it reaches a low 67 percent (Busse and Fuchs, 2011; Eurobarometer, 2010).

Compared to the U.S. where the sample is drawn from landline RDD (AAPOR, 2010, p. 30), some European countries face an additional problem, since they sample from directory-listed landline phone numbers (for a review, see Häder, Häder, and Kuhne, 2011). When the households have the option to opt out and choose whether to be included in the list, the problem of non-coverage error gets even bigger.

Respondents included in the sampling frame may be different from the

ones who are excluded and this may be a source of bias. Indeed, there is now a consistent body of evidence that shows that landline and cell-only households have different demographic and socio-economic characteristics (amongst others, Boyle, Bucuvalas, Piekarski, and Weiss (2009); Callegaro and Poggio (2004); Fuchs (2009); Keeter, Kennedy, Clark, Tompson, and Mokrzycki (2007); Vicente and Reis (2009)). If correlated with the outcome variables of interests, such differences may bias survey estimates as shown, for example, by Blumberg and Luke (2007) when looking at the impact of non-coverage error on key health estimates. Although very little is known on the impact of non-coverage error on polls accuracy, preliminary evidence shows that imperfect coverage is likely to lead to bias in favor of left-wing candidates (Durand, Blais, and Vachon, 2002; Mokrzycki, Keeter, and Kennedy, 2009). As Durand, Blais, and Vachon (2002) put it (p.38): 'excluding unlisted telephone numbers may have contributed to the bias against a more conservative party'. Very little is known about the characteristics of those households which have a landline phone, but they are not listed in the telephone directory. Therefore, nothing is known so far about how switching to RDD method could help European countries limit the effect of non-coverage error.

Italy shares with other European countries this additional problem, given that the sample frame used by pollsters is the Directory of Landline Phones which excludes households that decided not to be included in such a list. The share of the households which opt out is not negligible, for example in 2006, nearly 10 percent of the landline phone households opted out of the Directory of Landline Phone Numbers.[4] To sum up, the sampling frame used by Italian pollsters is from a directory of Landline Phones which excludes almost one third of the households (see table 1). These excluded households are cell-only households (about 20 per cent of the total) and landline households who did not want to be included in the telephone Directory.[5]

The second source of error which is relevant for our analysis is related to the impact of changes of mind. The voting behavior of the electorate may change between the time the polls are carried out and the election day. If the pools are not conducted late enough in the campaign, voters can change their mind: they can decide to stay at home or to vote for a different party. We

[4]Our calculations based on weighted data based on the 2006 'Multipurpose Survey'.

[5]Pre-election polls exclude also Italian voters who live abroad and who are entitled to vote. This form of non-coverage is very difficult to quantify as there are no data available.

interpret this as form of specification error, given that the construct that the pollsters would like to measure is the actual vote on the day of the elections, but the only construct which is possible to measure is vote intentions before the elections (see a discussion in Merkle, Langer, and Lambert, 2011).

Research on this topic is scant although there is evidence that suggests that changes in respondents' voting intentions do occur (Smith, 1996; Merkle, Langer, and Lambert, 2011; AAPOR, 2009). However, a few studies suggest that its impact on the accuracy of the polls is quite modest (Durand, Blais, and Vachon, 2002; Durand, Blais, and Larochelle, 2004; Jowell, Hedges, Lynn, Farrant, and Heath, 1993; Magalhães, 2005).

Third, we consider non-response error. Respondents and non-respondents can have different voting behaviors and this may lead to bias. Studies have found that non contacts, refusals, hard to reach respondents and non-disclosers are more likely to favor conservative parties. For example, while looking at the reasons for the poor performance of the polls for the 1998 Quebec Election, Durand, Blais, and Vachon (2002) find that (i) refusals are more likely to vote for conservative parties, (ii) hard to reach individuals do not vote differently from the others and (iii) non-disclosers are slightly more inclined to vote for conservative parties. The same conclusion is drawn by Smith (1996) when looking at the 1992 UK general elections. Therefore, pre-election polls may tend to underestimate right-wing votes.

2 Our contribution

In the context of pre-election polls, the accuracy of an estimate (A) is given by the difference between the population parameter measured as percentage of individuals who vote for a certain party and its estimate measured as percentage of respondents who declared to intend to vote for that party. We focus on frame error (here 'non-coverage error', NC), specification error (here 'changes of mind', CM) and non response error (NR). In brief, the determinants of the accuracy of the polls estimates we are interested in are the following:

$$A = NC + NR|NC + CM|NR, NC \tag{1}$$

Where NR is conditional on NC and CM is conditional on both NR and NC.

The paper makes a few important contributions. First, it applies the total survey error (TSE) paradigm to the analysis of pre-election polls and it analyzes different sources of error in an unitary perspective.

Second, it provides an original analysis of non-coverage error by applying a range of statistical techniques which have never been applied to this sort of problems: statistical matching via multiple imputation and inverse probability weighting. The evaluation of the impact of non-coverage error on bias lies on the comparison of voting behaviors of respondents excluded from and included in the sampling frame. When the two groups have different voting behaviors, non-coverage error may lead to non-coverage bias. Due to the limited availability of data on the political behaviors of respondents excluded from the sampling frame, such a comparison can hardy be performed and the effect of non-coverage on bias remains often unassessed. The Italian case is not an exception: to our best knowledge there are no Italian surveys that collect both information on vote intentions and telephone non-coverage. We overcome this problem in two different ways: by performing predictive mean matching between a dataset that collects data on inclusion in the sampling frame (but not on voting intentions) and a dataset with a wide range of information on political behaviors (but no information on inclusion on the sampling frame) and by weighting for a set of weights that correct for the probability of being included in the sampling frame. Moreover, as observed by O'Muircheartaigh and Lynn (1997) our analysis on inverse probability weighting can be seen as a validation of the use of procedures correcting for

non-coverage at the estimation stage.

Third, we make use of a rich dataset on vote intentions and actual vote to shed light on the role of specification error (here changes of minds) and non-response error.

Finally, the paper advances the current knowledge on the performance of Italian polls. Italy is an interesting case both because the share of cell-only households is bigger than in most of the other countries and because the decision of opting out is correlated with some indicators of trust and civic engagement (Sala and Fumagalli, 2011). Research into the causes of the poor performance of Italian polls is scant. Exception is work by Callegaro and Gasperoni (2008) and Gasperoni and Callegaro (2007, 2008) who, while assessing the accuracy of pre-election polls for the General Elections held in Italy since 2000, argue that non-coverage error may be one of the reasons for their poor performance. However, no evidence is provided to support this claim yet. This paper is meant to fill also this gap.

3 Background

3.1 The 2006 General elections

On April 9^{th} and 10^{th} 2006, Italians voted in a Parliamentary election by choosing between two major coalitions: the center-right coalition led by the incumbent prime minister Silvio Berlusconi and the center-left coalition led by the once president of the European Commission Romano Prodi. The center-right coalition named 'La Casa della Libertà' ('the House of Freedom') included the following parties: Berlusconi's party 'Forza Italia' ('Go Italy'), the biggest Italian right-wing party 'Alleanza Nazionale' ('National Alliance'), the centrist party 'Unione dei Cristiani Democratici' ('Union of Christian Democrats'), a joint list of regionalist movements: the northern 'Lega Nord' ('Northern League'), the southern 'Movimento per le Autonomie' ('Movement for the Autonomies') and the Sardinian 'Partito Sardo d'Azione' ('Sardinian Action Party'), the far right, neo-fascist 'Alternativa Sociale' ('Social Alternative') and 'Fiamma tricolore' ('Tricolor Flame') and other small parties including liberal reformers, eurosceptics, tiny groups of Christian Democrats and even a minor socialist party. The center-left coalition, 'L'Unione' ('the Union'), included 'l'Ulivo' (the 'Olive Tree'),[6] the small 'Movimento dei repubblicani Europei' ('European Republicans Movement'), the liberal/libertarian movement 'Rosa nel Pugno' ('Rose in the Fist'), the communists of the 'Partito della Rifondazione Comunista' ('Communist Refoundation Party'), the 'Partito dei Comunisti Italiani' ('Party of Italian Communists'), the 'Federazione dei Verdi' ('Federation of the Greens'), the 'Italia dei Valori' ('Italy of Values') led by Antonio di Pietro, the Christian party 'UDEUR', a small socialist party and the pensioners' party.

3.2 The methodology of the 2006 General Election pre-election polls.

The methodology of the 2006 General Election pre-election polls is described in details in Callegaro and Gasperoni (2008). Here we focus on three aspects, namely the mode of data collection, sampling methods and questionnaire content. In Italy polls are usually CATI (Computer Assisted Telephone

[6]The Olive Tree is a confederation of the social-democratic 'Democratici di Sinistra' ('Democrats of the Left') and the center-left party 'La Margherita' ('the Daisy').

Interviewing) surveys, whereas web-based polls are carried out only in very few cases. When the polls are CATI, the sampling frame used is drawn from the telephone numbers listed in the Directory of Landline Phones which is available on cd (Callegaro and Poggio, 2004). Calls to unlisted telephone numbers (e.g. mobile phones, respondents who opted out from the Directory of Landline Phone etc.) are not made. The details on the sampling design used by pollsters are unclear,[7] whereas the sample size is usually quite small and consisting of about 1,000 productive interviews. Pre-election polls collect different kinds of information including vote intentions where political preferences are often recorded at a single party level. However, procedures for screening for likely voters are not in place. In Italy, the results of the polls can be made public till two weeks before the election.

[7]To describe the methodology of the polls, Callegaro and Gasperoni (2008) analyzed the documentation deposited with the relevant Italian Government body, the Presidenza del Consiglio dei Ministri (www.sondaggipoliticoelettorali.it). With respect to sampling methods, this documentation proved to be poor and inaccurate (see also Pisati, 2008).

4 Telephone coverage and political behavior: evidence from the Italian Multipurpose Survey

4.1 Data

We use the data from the 2006 Multipurpose survey to assess differences in political behaviors of respondents included in/excluded from the sampling frame. The Italian 'Multipurpose Survey' ('Indagine Multiscopo sulle Famiglie') is a nationally representative survey run every year by ISTAT, the Italian National Statistics Institute. It is a repeated cross-section survey that is carried out face to face at respondents' home. The sample, which includes about 49,000 individuals living in almost 19,000 households, is a two stages clustered sample where municipalities are used as primary sampling units, while households are at the secondary level. Strata are constructed by taking into account macro regions (north, center and south), administrative regions (e.g., Sicily, Lombardy etc.), provinces (e.g., province of Milan, Rome etc.) as well as characteristics of the municipalities defined in terms of urban/rural classification and dimension. The Multipurpose survey collects a wide range of information, including data on household composition, housing, education, training, employment, leisure, health and use of social services. It also collects different measures of political behaviors (e.g., frequency of participation in party meetings etc.) as well as key information on telephone coverage (ownership of a landline phone and registration with the Phone Directory). To our best knowledge, no other surveys collect such detailed information on telephone coverage. The Multipurpose survey, however, does not gather information on voting intentions nor political preferences.

4.2 Non-coverage error

We consider non-coverage rates as estimates for non-coverage error. To estimate the magnitude of non-coverage error, we computed an indicator for respondents' inclusion in the sampling frame which is equal to one when respondents own a landline phone and are registered with the Directory of Landline Phones. Table 1 shows the distribution of non-coverage rates by region. In 2006, 31 percent of households, 28 percent of individuals and 27 per cent of individual aged 18 or older (potential voters) are excluded from

Table 1: Telephone non coverage. Households and individuals excluded from the sampling frame (Percentages by region)

Region	Households	Individuals	Ind 18+
North West	*29.3*	*26.0*	*25.4*
Piemonte	31.3	29.0	27.6
Lombardia	29.0	25.2	25
Liguria	25.6	22.6	21.6
North East	*23.4*	*20.8*	*19.6*
Trentino	23.6	18.7	18.4
Veneto	20.8	18.1	17.3
Friuli Venezia Giulia	23.4	20.7	20
Emilia Romagna	26.1	24.3	22.27
Center	*29.2*	*24.8*	*24.5*
Toscana	22.3	18.6	18
Umbria	25.1	20.9	20.1
Marche	25.5	21.9	20.8
Lazio	35.5	30.5	31
South	*36.2*	*33.6*	*31.8*
Abruzzo	31.1	25.6	25.2
Molise	29.4	24.8	23.7
Campania	35.6	33.1	31.2
Puglia	38.4	36.4	34.6
Basilicata	29.6	25.7	24.8
Calabria	35.7	38.4	36
Islands	*42.2*	*40.1*	*37.7*
Sicilia	44.2	42.4	40
Sardegna	32.1	32.8	31
Italy	*31.0*	*28.2*	*26.9*
N	*18,803*	*47,611*	*39,561*

Note: Weighted data from the Multipurpose Survey

the sampling frame. The level of non-coverage is not homogeneous across the different regions of Italy. When looking at regional differences, a clear pattern emerges: telephone coverage is lower in the Southern regions of Italy and in the main islands. For example, 31 percent of the potential voters living in Campania are excluded from the sampling frame compared to 25 percent of those living in Lombardia. Such a regional imbalance in sample coverage can potentially become a source of bias, given that in Italy the voting behavior differs across regions (see Diamanti, 2005).

Table 2 shows the distribution of non-coverage rates by a set of other variables which could be correlated with voting behavior. Unsurprisingly,

Table 2: Telephone non coverage. Households and individuals excluded from the sampling frame (Percentages by gender, education, age)

Variable	Individuals	Ind 18+
Gender		
Men	28.67	27.47
Women	27.73	26.46
Age		
0-17	34.15	-
18-29	32.65	32.65
30-59	30.13	30.13
>= 60	18.35	18.35
N	47,611	39,561
Education		
No education	31.19	29.33
Primary school	23.14	22.36
Lower secondary school	31.57	32.2
2 or 3 years high school	26.77	26.94
4 or 5 years high school	25.9	25.89
University degree	22.69	22.69
N	45,107	39,561

Note: Weighted data from the Multipurpose Survey

the non coverage error is higher among young people and it decreases quite significantly when older respondents are considered. Again, the highest non-coverage rate is estimated for respondents having no education, while the lowest is estimated for those holding a university degree. However, the relation between non-coverage and education is not clear and no pattern emerges. Finally, non coverage error does not seem to depend on gender.

4.3 Differences in political behaviors

We compare the distribution of a set of indicators of political behaviors for respondents who are included in/excluded from the sampling frame. We first look at frequency of speaking and getting informed about politics. These can be considered as indicators of political awareness, in line with Barnes, Kaase, Allerback, Farah, Heunks, Inglehart, Jennings, Klingemann, Marsh, and Rosenmay (1979); Luskin (1990). Table 3 shows that individuals included in the sampling frame are characterized by a higher level of political awareness, given that they speak and get informed about politics more often than respondents excluded from the sampling frame. The differences between the two groups are statistically significant ($p < 0.01$) and big in magnitude.

Table 3: Political awareness of respondents included in and excluded from the sampling frame

	How often the respondent speaks about politics ***		How often the respondent get informed about politics ***	
	Included (%)	Excluded (%)	Included (%)	Excluded (%)
Every day	10.4	8.3	36.4	26.9
Some times a week	23.6	19.2	21.4	20.2
Once a week	5.7	5.0	3.8	4.0
Once a month	15.3	13.2	8.1	8.5
Some times a year	11.4	12.0	6.4	7.3
Never	33.7	42.3	24.1	33.1
N	30,144	10,452	30,165	10,436

Pearson chi-square test: *** $p < 0.01$

Note: Weighted data from the Multipurpose Survey.

Then we look at a set of indicators of political participation. Table 4 shows the percentage of respondents included in and excluded from the sampling frame who reported having participated in six different forms of political activities in the twelve months preceding the interview. The comparison between the two groups confirms the previous conclusions: with the exception of the participation in rallies, respondents who are included in the sampling frame are more likely to be politically active than those excluded. The table also shows that the differences between the two groups are larger and more significant when 'institutionalized' forms of political participation are considered.[8]

[8]For a discussion of 'institutionalized' versus 'mobilized' and 'formal' versus 'informal'

Table 4: Political behaviors of respondents included and excluded in the sampling frame

	Included (%)	Excluded (%)	n
Having listened to political debates***	26.4	20.3	41,269
Participation in rallies	5.3	5.1	41,311
Participation in political demonstrations**	5.2	4.5	41,242
Participation in party meetings***	4.0	3.1	41,179
Donated to parties***	3.4	2.2	41,259
Volunteered for a party***	1.6	1.1	41,253

Mean comparison t-tests: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Note: Weighted data from the Multipurpose Survey.

forms of participation see Seligson (1980).

5 Telephone coverage and vote intentions. Evidence from the Italian National Election Studies

As we have demonstrated in the previous section, respondents who are included in the sampling frame are different with respect to their political behavior from those excluded from such a frame. We conclude that non-coverage error may lead to biased estimates for some indicators of political participation.

In this section we address the following question: Do respondents included in and excluded from the sampling frame also differ with respect to their vote intentions? If they do, non-coverage error may also bias the predictions of the polls.

Following Groves, Fowler, Couper, Singer, and Tourangeau (2009), we can express non-coverage bias as:

$$\overline{Y}_C - \overline{Y} = \frac{U}{N}(\overline{Y}_C - \overline{Y}_U) \tag{2}$$

where:

- $\overline{Y}$ is the estimate of the share of a given party in the target population.
- $\overline{Y}_C$ is the estimate of the share of a given party among those included the sampling frame
- $\overline{Y}_U$ is the estimate of the share of a given party among those excluded from the sampling frame
- $\frac{U}{N}$ is the non-coverage rate which we estimated in paragraph 4.2

Our aim, therefore, is to measure $(\overline{Y}_C - \overline{Y})$. Non-coverage bias, however, can not be measured directly, given that the vote intentions of the population are unknown. We use survey data to overcome this problem.

Ideally, to estimate the impact of non-coverage error on bias, one would need a survey that collects at the same time information on coverage (i.e. registration with the Directory of Landline Phones) and vote intentions. As previously mentioned, such a dataset is not available. To address this problem, we follow two strategies.

First, we use statistical matching techniques to impute to the ITANES survey, a rich dataset on political behaviors, which also includes vote intentions, the inclusion indicator that we previously computed for the Multipurpose survey. This makes it possible to assess directly the role of non-coverage error on bias by comparing the political preferences of respondents included in and excluded from the sample frame (paragraph 5.2).

Second, we use the information contained in the Multi Purpose Survey to create a set of weights that correct for non-coverage bias and we then use these weights to weight the distribution of the vote intentions of the ITANES survey (paragraph 5.2). The comparison of the weighted and unweighted distribution constitutes an indirect way of assessing the role of non-coverage error on bias.

5.1 Data

The 'Italian National Election Studies' (ITANES) consists of a series of face-to-face surveys (which include also a panel component) on political behaviors run by the Cattaneo Institute in Bologna (Italy) since 1990. Its sample is a nationally representative sample of respondents aged 18 and older drawn from the electoral registry which report the name of anyone in the country who has the right to vote (i.e. the registration is not on a voluntary base). ITANES has a multistage sample with municipalities, poll sections and individuals as primary, secondary and tertiary sampling units. Municipalities are stratified by region, dimension and main political orientation, while individuals are stratified by gender. By design, this sampling strategy offers full coverage of the voting population. ITANES collects a diverse set of information including respondents' attitudes to a wide range of social, economic and political issues (from abortion to tax evasion), interest in politics, different forms of political participation, intended and actual vote choice (recorded at the party level) etc. We use two components of the ITANES carried out in 2006 when the General Election was held. More specifically, we use the 2006 pre and post-electoral panel and the 2006 cross-sectional post electoral survey. The former has a sample size of 2005 respondents, 69 percent of which were successfully interviewed after the elections. The latter has a sample size of 2,011 respondents. There is high variation in the level of missing data for the variables that collect respondents' vote intentions and actual vote. For the pre-electoral wave of the panel component, item non response on vote intentions is 42.3 percent; for the pre and post-electoral (balanced) panel, item

non response on actual vote is 46.2 percent (of which 31.3 percent is due to attrition); for the cross-sectional post electoral survey, item non response on actual vote is 26.4 percent.

5.2 Statistical Matching. A direct way to assess the role of coverage error on bias

Statistical matching is a set of statistical methods aimed at integrating two datasets when the variables of interest are not jointly observed.[9] Let us suppose we need to investigate the relationship between variable x and variable y. No dataset contains simultaneously the two variables of interests. These, however, are present in two separate datasets (X and Y) which also contain a set of common variables **z**. Statistical matching is a method that permits to match the sample members of dataset X with the sample members of dataset Y on the basis of the jointly observed characteristics **z**. This permits to create a new dataset W containing **x**, **y** and **z**. In that, statistical matching can be considered as an extreme form of imputation.

While performing statistical matching, decisions regarding (i) the role of the datasets, (ii) the statistical method used, (iii) the harmonization of the datasets and the choice of the **z** variables have to be taken.

The role of the datasets We used the Multipurpose survey (dataset Y) to create an indicator of inclusion in the sampling frame (**y**) and we imputed such an indicator to the ITANES survey (dataset X) through statistical matching. Therefore, the 'donor' data file is the Multipurpose Survey, the 'recipient' data file is the ITANES survey and the new dataset W is the ITANES survey with the additional indicator **y**. Other strategies could be feasible including the possibility of using the two datasets both as donor and recipient and imputing the inclusion indicator to the ITANES survey and the variables on voting behavior to the Multipurpose survey. This approach, which resembles closely the approach that was first introduced by Rubin (1986), has the advantage of exploiting all the available information. In our case, however, this was not a viable solution for at least two reasons. First, the differences in sample sizes between the two datasets. The sample of the Multi Purpose Survey is about 14 times bigger than the one of the ITANES survey. In cases like this, Singh, Mantel, Kinack, and Rowe (1993)

[9]For a good overview, see D'Orazio, Di Zio, and Scanu (2006).

warn that 'if the smaller file was the donor, some records in the donor file would be imputed more than once in the recipient, artificially modifying the variability of the distribution of the imputed variables in the final synthetic file'. Second, the nature of the voting behavior variables. Such variables are categorical, they have a great number of categories (the Italian political spectrum used to be very fragmented) and the categories can not be easily ordered along the classical left-right continuum. This may therefore magnify the effect of matching noise.

The statistical method In our application, we use predictive mean matching (pmm) which was first introduced by Little (1988) on the basis of the method proposed by Rubin (1986). Pmm is convenient since it is a semi parametric method derived from a combination of parametric and non parametric techniques (see also Schenker and Taylor, 1996). As such, it offers the advantages of combining the parsimony of parametric models and the robustness of non parametric models against model misspecification (D'Orazio, Di Zio, and Scanu, 2006).

Predictive mean matching is particularly attractive when one needs to impute categorical variables with an underlying continuous distribution. Let us define the variable y (the individual is in the list) as follows:

$$\begin{aligned} y = 1 \; if \; y^* > 0 \\ = 0 \; otherwise \end{aligned} \tag{3}$$

where y^* is a latent variable indicating the propensity of being in the list.

The matching is carried out by using a two step procedure which first estimates the parameters of a simple linear regression for y^* and then imputes live y values through a hot-deck model.[10]

In particular, pmm:

- fits

$$y^* = \beta \mathbf{z} + \epsilon \tag{4}$$

  using dataset Y

- estimates $\widehat{\beta}$ and $\widehat{\sigma^2}$

[10]For an application of hot deck methods for statistical matching, see Singh, Mantel, Kinack, and Rowe (1993).

- simulates new parameters $\widehat{\beta_{\mathbf{s}}}$ and $\widehat{\sigma^{\mathbf{2}}_{\mathbf{s}}}$ from their joint posterior distribution
- predicts $\widehat{y_i^*} = \widehat{\beta_{\mathbf{s}}}\mathbf{z_i}$ on dataset X and $\widehat{y_j^*} = \widehat{\beta_{\mathbf{s}}}\mathbf{z_j}$ on dataset Y
- matches the entries $i \in X$ with a set of nearest neighbors (potential donors) $j \in Y$ by using the absolute distance $|\widehat{y_i^*}\text{-}\widehat{y_j^*}|$ as distance function
- randomly draws an imputed value from the set of neighbors.

The advantages of the non parametric model are now clear: the linear regression is meant to estimate $\widehat{y_i^*}$, which is a variable summarizing the probability of being included in the list on the basis of the information provided by variables $\mathbf{z}$, while the hot-deck component permits to preserve the shape of the distribution of the imputed data, which would not otherwise be possible by using the parametric regression only.

Obviously, the method relies on the 'Conditional Independence Assumption (CIA)' i.e. that $\mathbf{y}$ and $\mathbf{x}$ are statistically independent conditionally on $\mathbf{z}$ or, more formally:

$$P(y, z|\mathbf{z}) = P(y, |\mathbf{z})P(x, |\mathbf{z}) \tag{5}$$

and

$$f(\mathbf{x}, \mathbf{y}, \mathbf{x}) = f_{\mathbf{y}|\mathbf{z}}(\mathbf{y}|\mathbf{z}) f_{\mathbf{x}|\mathbf{z}}(\mathbf{x}|\mathbf{z}) f(\mathbf{z}). \tag{6}$$

This means that variables $\mathbf{z}$ are jointly able to explain both $\mathbf{x}$ and $\mathbf{y}$ without leaving any unobserved endogenous effect in the error term. The CIA assumption, however, is a standard practice in statistical matching. Unfortunately little can be done to test it empirically (see Renssen, 1998; Rodgers, 1984) and almost nothing can be done do relax it.

The harmonization of the datasets and the choice of the z variables The harmonization of the datasets is crucial for the performance of statistical matching. While discussing these issues, Van der Laan (2000) lists four forms of harmonization: i) harmonization of the definition of units, ii) harmonization of reference periods, iii) harmonization of variables and iv) harmonization of classification.

Our units of analysis are respondents eligible to vote (i.e. respondents who are 18 years old and older). In case of the ITANES survey, no harmonizations was needed as its sample is composed of individuals who are aged 18 and over; for the Multipurpose survey, we limited the analysis to respondents who are 18 years old and over. No harmonization of the reference period was needed, as both surveys were carried out in 2006.

The decisions concerning the choice of the $\mathbf{z}$ variables are interlinked with issues related to their harmonization. In the absence of widely shared criteria to guide the choice of the matching variables, we adopted the following:

- presence of the variables in the different datasets,
- correlation of the $\mathbf{z}$ variables to either $\mathbf{x}$ or $\mathbf{y}$ (see Van der Laan, 2000),
- similarity in question wording and comparability of response categories.

Six variables met our criteria, namely: age, sex, region of residence, employment status, sector of employment and education. Before performing the statistical matching, some variables were recoded to ensure that the response categories were harmonized. Descriptive statistics for the harmonized variables can be found in Appendix 1. Notice the similarities in the variable distributions across datasets.

Another final issue concerns the number of imputations. We combined statistical matching with multiple imputation by performing ten different sets of imputations (see also Kamakura and Wedel, 1997; Rassler, 2004). The number of imputations performed is in line with what discussed in Rubin (1987) and Rubin (1996). Notice that the imputations provide quite stable results[11] and the between imputation variance is limited.

[11] This is likely to be due both to the availability of a relatively large number of the variables $\mathbf{z}$ and to the big sample size of the Multipurpose Survey, as discussed above.

Results of the statistical matching To investigate the effect of coverage error on the accuracy of the predictions of the polls, we compare vote choice of respondents who are included in and excluded from the sampling frame. We use two of the datasets of the ITANES study: the pre-electoral component of the panel and the cross-sectional post-electoral survey. Each of these datasets has pros and cons with respect to the variables collected to measure respondents' voting behavior and the level of missingness on these variables. The former collects respondents's vote intentions, the same variable that is used by pollsters when predicting the outcome of the elections. However, as mentioned previously, this variable has the disadvantage of having a high level of missing data. Conclusions derived from this part of the analysis need therefore to be interpreted with caution as we can not exclude *a priori* that missingness is systematic. The latter gathers respondents' actual vote which is a different variable from the one collected by pollsters. However, this variable has a lower level of missingness.

Table 5 shows the results of a set of 10 imputations for the pre-electoral component of the ITANES survey. We focus on five main parties, which, altogether, obtained 78 percent of the votes (see table 8, column 1). For each imputation we report (i) the vote intentions for respondents who are included in or excluded from the sampling frame, (ii) the differences in vote intentions between the two groups and (iii) the level of significance of the t-tests of equality of the means between groups.

Let us now look at the mean difference in vote intentions for the 5 parties across the two groups of interest (last row). The mean difference is the multiple imputation estimate of the difference in vote intention between the two groups (Rubin, 1987). We find evidence that respondents included in the sampling frame may vote differently from respondents who are excluded from this frame. The former are more likely to vote for left-wing or center-left parties (the sign of the mean difference is positive for Communist Refoundation and the Olive Tree). Respondents who are included in the sampling frame are also more likely to vote for the regional movement the Northern League. On the contrary, respondents who are excluded are more inclined to vote for center-right or right-wing parties (the sign of the mean difference for Go Italy and National Alliance is negative).

We look in more details at the results of the 10 imputations for the Olive Tree, Go Italy and National Alliance.These are parties with the largest absolute values in the mean differences.

When looking at the differences in the shares estimated for the Olive

Tree two findings stand out: (i) five of the 10 mean differences are positive (imputations 2, 5, 7, 8, 10), (ii) the only three differences that are statistically significant are the differences with a positive sign (imputations 5, 7, 8). We interpret these findings as an indication that respondents included in the sampling frame appear to be more likely to vote for the Olive Tree.

We find similar, but opposite, findings when looking at the differences in the shares estimated for National Alliance: eight of the 10 differences are negative (imputations 2, 3, 4, 5, 6, 7, 8, 10) and 4 of the five statistically significant differences are the differences with a negative sign (imputations 2, 3, 5, 7, 9). We conclude that respondents included in the sampling frame seem to be less likely to vote for National Alliance.

If we consider the differences in the shares estimated for Go Italy, we find mixed results. On the one hand, six of the ten differences are negative (imputations 1, 2, 6, 7, 8, 9), on the other, none of the 10 differences reaches the conventional level of statistical significance. Further evidence is therefore needed to interpret the differences in vote intentions for Go Italy between respondents included in or excluded from the sampling frame.

We explore further the link between voting behavior and coverage error by looking at the differences in actual vote between respondents who are included in or excluded from the sampling frame. Table 6 shows the results of this analysis. With few exceptions, they confirm the previous findings; respondents included in the sampling frame are more likely to have voted for the Olive Tree and less likely to have voted for Go Italy. Interestingly, the relationship between actual vote for National Alliance and respondents'inclusion in or exclusion from the sampling frame is now weaker.

Taken together, the results shown in table 5 and 6 suggest that (i) respondents included in the sampling frame have a different voting behavior from respondents who are excluded from the sampling frame, (ii) the former are more likely to vote for the Olive Tree and (iii) the latter are more likely to vote for center-right parties such as Go Italy and National Alliance. It is worth noting that that statistical matching is likely to weaken the association between x and y. Consequently, the association between being on the frame and party is likely to be under-estimated. We conclude that coverage error may bias the estimate of the polls and, in particular, may overestimate the votes for left-wing parties. This is consistent with the direction of the bias of the 2006 General Election pre-electoral polls.

Table 5: Vote intentions of respondents included in and excluded from the sampling frame (ITANES data, pre-electoral component of the panel)

		Vote intentions									
		1		2		3		4		5	
		Communist		Olive		Go		National		Northern	
imputation		Refoundation		Tree		Italy		Alliance		League	
number	Sample	Share	Diff.	Share	Diff.	Share	Diff.	Share	Diff.	Share	Diff.
1	Included	9.33	-0.46	36.64	-1.12	22.24	-3.63	13.59	1.35	4.03	0.18
1	Excluded	9.79	ns	37.76	ns	25.87	ns	12.24	ns	3.85	ns
2	Included	9.78	1.26	37.46	2.05	22.38	-2.87	12.01	-4.71	4	0.07
2	Excluded	8.52	ns	35.41	ns	25.25	ns	16.72	**	3.93	ns
3	Included	9.85	1.49	36.89	-0.09	24.08	3.5	11.86	-5.18	4.27	1.05
3	Excluded	8.36	ns	36.98	ns	20.58	ns	17.04	**	3.22	ns
4	Included	9.36	-0.32	36.73	-0.69	23.34	0.76	12.8	-1.72	4.15	0.6
4	Excluded	9.68	ns	37.42	ns	22.58	ns	14.52	ns	3.55	ns
5	Included	9.49	0.18	38.43	6.02	23.26	0.5	12.27	-3.94	4.28	1.18
5	Excluded	9.31	ns	32.41	**	22.76	ns	16.21	**	3.1	ns
6	Included	9.47	0.11	36.84	-0.28	22.69	-1.72	12.87	-1.51	4.21	0.87
6	Excluded	9.36	ns	37.12	ns	24.41	ns	14.38	ns	3.34	ns
7	Included	9.56	0.44	38.34	5.57	22.96	-0.69	11.54	-6.7	3.26	-2.82
7	Excluded	9.12	ns	32.77	**	23.65	ns	18.24	***	6.08	**
8	Included	8.51	-3.5	38.18	4.74	22.93	-0.77	13	-0.96	3.66	-1.21
8	Excluded	12.01	**	33.44	*	23.7	ns	13.96	ns	4.87	ns
9	Included	10.06	2.37	36.84	-0.28	22.22	-3.53	16.05	3.77	4.33	1.32
9	Excluded	7.69	ns	37.12	ns	25.75	ns	12.28	**	3.01	ns
10	Included	9.4	-0.19	37.24	1.28	23.43	1.17	12.76	-1.97	4.06	0.29
10	Excluded	9.59		35.96	ns	22.26	ns	14.73	ns	3.77	ns
	Mean difference		0.138		1.72		-0.728		-2.157		0.153

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, $n = 1154$

Table 6: Actual vote of respondents included in and excluded from the sampling frame (ITANES data, post-electoral cross-sectional survey)

		party voted									
		1		2		3		4		5	
		Communist		Olive		Go		National		Northern	
imputation		Refoundation		tree		Italy		Alliance		League	
number	Sample	Share	Diff.	Share	Diff.	Share	Diff.	Share	Diff.	Share	Diff.
1	Included	8.23	-2.5	40.04	10.98	20.02	-7.73	12.52	-1.35	3.66	1.57
1	Excluded	10.73	**	29.06	***	27.75	***	13.87	ns	2.09	*
2	Included	8.38	-1.88	37.57	1.42	21.73	-1.09	12.34	-2.02	3.22	-0.11
2	Excluded	10.26	ns	36.15	ns	22.82	ns	14.36	ns	3.33	ns
3	Included	9.33	1.76	37.79	2.28	21.59	-1.65	13.08	0.81	2.93	-1.25
3	Excluded	7.57	ns	35.51	ns	23.24	ns	12.27	ns	4.18	ns
4	Included	8.96	0.35	37.46	1.07	20.88	-4.68	12.37	-2.07	3.41	0.63
4	Excluded	8.61	ns	36.39	ns	25.56	**	14.44	ns	2.78	ns
5	Included	8.75	-0.52	37.41	0.89	21.7	-1.33	12.95	0.31	3.57	1.32
5	Excluded	9.27	ns	36.52	ns	23.03	ns	12.64	ns	2.25	ns
6	Included	8.18	-2.72	38.09	3.52	21.55	-1.85	11.91	-3.78	3.55	1.16
6	Excluded	10.9	*	34.57	ns	23.4	ns	15.69	**	2.39	ns
7	Included	8.86	-0.06	38.7	6	20.89	-4.52	12.66	-0.85	3.44	0.74
7	Excluded	8.92	ns	32.7	**	25.41	**	13.51	ns	2.7	ns
8	Included	8.6	-1.1	37.83	2.52	21.18	-3.35	13.3	1.71	3.62	1.46
8	Excluded	9.7	ns	35.31	ns	24.53	*	11.59	ns	2.16	*
9	Included	8.85	-0.11	38.52	5.47	21	-4.21	12.51	-1.5	3.13	-0.51
9	Excluded	8.96	ns	33.05	**	25.21	**	14.01	ns	3.64	ns
10	Included	8.68	-0.77	37.63	1.67	21.55	-1.81	12.42	-1.75	3.29	0.14
10	Excluded	9.45		35.96	ns	23.36	ns	14.17	ns	3.15	
	Mean difference		-0.755		3.582		-3.222		-1.049		0.515

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ $n = 1476$

5.3 Weighting

As a check to prove the robustness of our previous analysis to the empirical model adopted, we develop a weighting procedure. The idea behind it is simple. Our statistical matching suggested that coverage error leads to overestimate the votes for the center-left coalition (especially the Olive Tree) and to underestimate the votes for the center-right coalition (especially Go Italy), this implies that correcting for non-coverage error would lead to a new distribution of votes with an decreased estimated share for the center-left and an increased estimated share for the center-right.

For this exercise and without loss of generality, we can treat the ITANES sample as the universe of voters and we can re-weight the estimates of each party's share as follows.

Using the Multi Purpose survey, we run a probit model to estimate the probability of being included in the sampling frame according to the formula (4). As independent variables, we use the same variables that we previously used for running the multiple imputation models (i.e. age, sex, region of residence, employment status, sector of employment and education). We then compute predicted probabilities of inclusion in the sampling frame for both the Multi Purpose and two of the ITANES surveys (pre-electoral component of the panel and cross-sectional post electoral survey) and we compute a set of weights which are given by the inverse of such probabilities (see Horvitz and Thompson, 1952; Särndal, Swensson, and Wretman, 1992). Finally, we used the weights we created to re-weight the distribution of votes derived from the ITANES surveys and we compared the two distributions by testing the significance of the difference in their first moments.[12]

Table 7 shows the results of this exercise for both the pre-electoral component of the ITANES panel and the cross-sectional post electoral survey. This is meant to capture the effect of weighting on the estimates of both vote intentions (pre-electoral component) and actual vote (cross-sectional post electoral survey). In particular, columns (1) and (4) show the distribution of the votes computed without any correction; columns (2) and (5) report the distribution obtained after applying our weighting procedure correcting for non-coverage bias and columns (3) and (6) show the difference between the first moments of the weighted and the unweighted distribution. As a consequence, columns (3) and (6) show the direction of our correction for non-coverage error. As expected, we find that the weighting procedure

[12]Here we compared any estimated share separately.

Table 7: Weighting political preferences, first moments

	pre electoral survey				post electoral refreshment sample			
	1	2	3		4	5	6	
	unweight.	weighted	difference		unweight.	weighted	difference	
Com. Ref.	9.45	9.60	0.15	ns	8.88	8.61	-0.26	**
Olive tree	36.92	36.37	-0.54	*	37.20	36.64	-0.55	**
Go Italy	23.14	23.53	0.39	ns	22.02	22.75	0.73	**
N. Alliance	13.26	13.47	0.21	ns	12.87	13.30	0.42	**
N. League	3.99	3.83	-0.15	**	3.25	3.11	-0.15	**
N	1,154				1,476			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

leads to an increase in the estimated shares of the votes for the center-right parties (with the exception of the Northern League) and a decrease in those for the center-left. The differences in the first moments are all significant at 5% when the cross-sectional post electoral survey is used. However, when the pre-electoral surveys is used, the only significant differences are those estimated for the Olive Tree (significant at 10%) and for the Northern League (significant at 5%). With the only exception of the Communist Refoundation, this pattern is consistent between the two surveys. We therefore conclude that coverage error may bias the predictions of the polls and, in particular, may account for the overestimation of the left-wing coalition. This is consistent with the previous findings. This analysis also shows that non-coverage error may also lead to an overestimation of the Northern League.

6 Changes of mind and non-response error

So far we have shown that coverage error may play a role in determining the accuracy of the polls (A). From equation (1) we know that A may also depend on respondents changing their minds ($CM|NC, NR$) and non-response ($NR|NC$). Here we address two questions: Do changes of mind and non-response bias the results of the polls and, if they do, in what direction?

To answer these questions we compare the official results of the General Elections (table 8, column 1) with the vote intentions of the respondents who belong to the pre-electoral survey of the ITANES panel component (table 8, column 2). Recall that the sample frame of this survey has a full coverage of the population. The comparison of columns (1) and (2) in Table 8 shows that the ITANES survey overestimated the votes for the center-left coalition by 4 percent points. In absence of coverage error, we interpret this finding as preliminary evidence that changes of mind together with non-response may bias the results of the polls. In line with equation (1), we can write the bias leading to an overestimation of the center-left as:

$$A = NR + CM|NR > 0 \tag{7}$$

Where a positive sign refers to an overestimation of the vote for the center-left coalition.

To disentangle the role played by changes of mind, we look separately at two categories of voters: the late-shifts ($CM_{ls}|NR$) and the late-deciders ($CM_{ld}|NR$). The former are those who declare a different vote before and after the elections, the latter are those who make up their mind at the very last minute.

We first look at the behavior of late-shifts. In columns (4) and (5) of table 8, we compare vote intentions and declared vote for the ITANES respondents who (i) took part in both the pre and post electoral panel component and (ii) have valid cases on the two variables of interest.

Table 8: political preferences: actual and estimated data

		ITANES survey			
	Official Data	Pre-electoral Panel Component	Post-electoral Cross-sectional Survey	Balanced Panel	
	(1) Actual shares	(2) Voting intentions	(3) Declared vote	(4) Voting intentions	(5) Declared vote
Communist Refoundation Party	5.80	9.43	7.58	9.37	8
Italian Communists	2.3	1.78	1.95	1.44	1.42
Federation of the Greens	2.1	1.89	2.09	2.08	1.94
Olive tree	31.3	36.67	40.06	35.59	39.45
Rose in the first	2.6	1.38	2.54	1.99	2.21
Italy of Values	2.3	0.89	1.49	1.05	1.39
Udeur	1.4	0	1	0	0.47
Italian Democratic Socialist Party		0.42		*0.5*	
Pensioners' party.	0.9	0.35	0.22	0	0.22
Total center left	48.7	52.81	56.93	52.02	55.1
Union of Christian Democrats	6.8	5.08	6.84	5.26	6.97
Forza Italia	23.7	24.01	20.71	24.56	20.88
National Alliance	12.3	12.91	11.21	13.17	12.02
New christian democrats and de michelis' socialists	0.7	0.3	0.25	0.31	0
Northern League	4.6	4.16	3.53	4.36	4.39
Social Alternative with Alessandra Mussolini			0.43		0.53
Total center right	48.1	46.46	42.54	47.66	44.26
Other party	0.9	0.73	0.09	0.33	0.09
Total		1,157	1,079	763	763

The comparison of columns (4) and (5) shows a difference of 3 percentage points between respondents who intended to vote for the left-wing coalition and respondents who declared to have voted for the same coalition. In absence of non-coverage error, we can interpret this finding as preliminary evidence that (i) voters may change their mind between the time the polls are carried out and the election day and (ii) for late-shifts the direction of this change may be from the center-right to the center-left, i.e.,

$$CM_{ls}|NR < 0 \tag{8}$$

Where a negative sign refers to an underestimation of the vote for the center-left coalition.[13]

Let us look at the behavior of late-deciders who made up their mind two weeks before the elections are held (this coincides with the embargo period). Our aim is twofold, to assess whether late-deciders vote differently from the rest of the voters and whether their voting behavior may introduce bias in the polls. The ITANES survey contains a question asking the respondents when they decided who to vote for, therefore it is possible to observe the voting behavior of late-deciders.

We use a multinomial logit to estimate the correlation between being a late-decider and voting for the center-left or the center-right coalition. We run the model on two samples, the cross-sectional post electoral survey (column 1 in Table 9) and the pooled sample of all the post electoral components i.e. the post-electoral cross sectional survey and the post electoral component of the panel (column 2). We used this second sample as a robustness check in order to fully exploit the information on actual voting.

[13]One may argue that this result is the consequence of a bias towards the winning coalition, however, the 2006 Italian General Elections are a nice case in this respect because no clear winner emerged. Indeed, the number of votes obtained by the center-left coalition, was much lower that what expected on the basis of both exit polls and pre-electoral polls and the results of the elections came as a bad surprise to the left-wing parties.

Table 9: Late-deciders and voting behavior

	1		2	
	Post electoral survey		Pooled post-electoral sample	
		$P > \|z\|$		$P > \|z\|$
Center-right	1.062	0.32	1.074	0.5
	(0.208)		(0.154)	
Other party	2.956	1.550	2.449	1.31
	(2.066)		(1.678)	
N	1,481			2,560

Source: Multi Purpose Survey, weighted data. Reference category: center-left. Standard error in parenthesis.

Table 9 reports the results of such a model (expressed in relative risk ratios). The coefficients are not statistically significant. This suggests that there are no differences between the political preferences of late deciders and the preferences of those voters whose decision regarding what party to vote for was formed before the embargo period.

In formulas.

$$CM_{ld}|NR = 0 \tag{9}$$

and

$$CM|NR = CM_{ls}|NR + CM_{ld}|NR <= 0 \tag{10}$$

To conclude, we have found preliminary evidence that suggests that respondents who change their minds and, in particular, those voters who change their mind between the time the polls are carried out and the election day may bias the predictions of the polls. However, they do not seem to be responsible for the overestimation of the center-left coalition, indeed, the bias that they may introduce seems to go in the opposite direction, leading to an underestimation of the left-wing coalition.

Evaluating the role of non-response is complicated as information on voting behavior of non-disclosers is usually not available. Our study is not an exception. For a preliminary investigation of the impact of non-response on bias we compare the original results of the General Elections (Table 8, column 1) with the declared vote of respondents belonging to the cross-sectional post-electoral component of the ITANES survey (column 3). Recall that,

under this analytical strategy, non-coverage error (NC) and changes of mind ($CM|NR$) are equal to zero.[14] The comparison of columns (1) and (3) in table 8 shows that the ITANES post-electoral survey overestimates the vote for the center-left coalition by 8 percent points. We can exclude that this difference is due to non-coverage error (given the full coverage of the ITANES survey) or to respondents' changes of mind (since both the original and the survey data are collected after the elections). We therefore interpret this finding as preliminary evidence for non response bias and we conclude that: (i) the effect of non-response may be large (ii) non-disclosers are likely to vote for the center-right coalition (iii) non-response leads to an overestimation of the center-left coalition.

In brief:

$$NR >> 0 \tag{11}$$

Where two positive signs refer to an overestimation of the vote for the center-left coalition, which, together with (10), is in line with (7).

In conclusion, our analysis suggests that the mechanism leading to bias in the polls are complex. We have shown that changes of mind may lead to an underestimate of the votes for the center-left coalition whereas non response may lead to an overestimate of the votes to the center-left coalition.

[14] $CM|NR$ is equal to zero because we are using post-electoral data.

7 Final remarks

The aim of this paper is to adopt a total survey error perspective to assess the impact of different sources of survey error on the accuracy of the predictions of the polls. By looking at the case of the pre-election polls carried out for the 2006 Italian General Elections, we find that both coverage and non-response error may lead to biased estimates of respondents' voting intentions and, in particular, they may lead to an overestimate of the votes for the center-left coalition.

Our main focus is the analysis of non-coverage error and its effects on bias. To evaluate the role of non-coverage error, we compare voting intentions of respondents included in and excluded from the sampling frame (the Directory of Landline Phones) using different analytical and methodological approaches. We find robust evidence that respondents included in the sampling frame are more likely to vote for parties belonging to the center-left coalition. In the presence of a high level of non-coverage error (in 2006 nearly 30 percent of Italians are excluded from the sampling frame), we argue that non-coverage error may lead to non-coverage bias. While evaluating the role of non-coverage error on the accuracy of the predictions of the polls, this paper also contributes to the advancement of the current knowledge in the field from a methodological prospective. We propose to use statistical matching and weighting procedures for inclusion in the sampling frame to overcome the problems of the lack of data on respondents excluded from the sampling frame.

When assessing the role of non-response, our analysis of the role of non-disclosers seems to suggest that their impact on the accuracy of the polls may be relevant and that non disclosers may lead to an overestimate of the center-left coalition. Overall, we speculate that non-response may lead to an overestimates of the votes for the center-left coalition. These findings are consistent with the direction of the bias of the pre-election polls as documented in the work of Callegaro and Gasperoni (2008).

Finally, we investigate the role of changes of minds, by looking separately at late-shifts and late-deciders. The mechanisms that lead to bias are complex. We find indications that respondents may have changed their mind (late-shifts) between the time the polls were carried out and the elections were held. As voting intentions seem to have 'moved' from the center-right to the center-left of the political spectrum, the predictions of the polls may have underestimated the votes for the center-left coalition. Late deciders,

on the contrary, do not seem to have had an impact on accuracy. However, while the analysis of non-coverage and non-response error is likely to describe mechanisms that work irrespective of the election considered, the analysis of the changes of mind is likely to be specific elections under study. In fact whether people change their minds may depend on the last minute political campaign.

To reduce the bias introduced by non-coverage error in the Italian polls, there are at least two viable strategies. First, the introduction of random digit dialling (RDD) may provide a useful contribution in this direction. In this paper, we did not distinguish between the different types of households excluded from the sampling frame, mainly households that, owning a landline phone, opted out of the Directory of (Registered) Landline Phones and cell-only households. RDD of landline phones would contribute to the reduction of that part of non-coverage error due to the exclusion of the former group of respondents whereas a more sophisticated sampling frame based on RDD of landline and cell phone numbers (dual frame) may contribute to the reduction of the overall non-coverage bias. With this respect, we encourage Italian polling agencies to be 'methodologically innovative and courageous' and explore the feasibility of alternative sampling methods. It is also worth mentioning that, due to the increasing diffusion of Internet, web polls may be a possibility to explore in the future.

An alternative to RDD is post-stratification by using weights correcting for differential inclusion probabilities in the sampling frame. As we have shown in the second part of the paper, this procedure may be successful in reducing the magnitude of non-coverage bias. Due to the already mentioned poor quality of the documentation on the design of the polls that is deposited with the Italian *Presidenza del Consiglio dei Ministri*(Cabinet of Ministers), we are not in the position to provide details on the weighting strategy that pollsters usually adopt. However, informal correspondence held with major Italian pollsters leads us to conclude that some forms of weighting procedures are sometimes used, although these procedures are based on a limited set of variables (mainly region of residence, age and sex) which are unlikely to fully explain the variation in the probability of being included in the sampling frame. Crucially, this depends on the availability of external sources of data having full coverage of voters and containing information predicting inclusion (not having opted out of the Directory of Landline Phones). Obviously, if implemented correctly, this strategy could also help reduce non-response bias, thus limiting the effect of two different sources of error.

The most difficult problem to be solved remains finding methods to predict the behavior of late-shift and late-deciders. This is especially true in the case of Italy where swing voters seem to play an important role in influencing the political agenda (see Profeta, 2007). However, there is a major methodological challenge, i.e. that it is not easy to distinguish between late-shift/late-deciders and people miss-reporting or non-disclosing their true political orientation. There is currently very little knowledge on this and good longitudinal data and paradata are needed to separate the effect of changes of minds from the effect of poor reporting. This is a potentially very interesting strand of research for political scientists, researchers interesting in survey methods as well as poll agencies.

Appendix. Comparability of the Multipurpose and ITANES surveys

Table 10: Geographical distribution

Region	Multi Purpose	ITANES pre-electoral	ITANES post-electoral
Piemonte, Valle d'Aosta	7.9	7.6	7.7
Lombardia	16.4	15.2	15.9
Liguria	2.9	2.8	2.9
Veneto	8.1	9.3	8.7
Trentino Alto Adige	1.6	1.7	1.7
Friuli Venezia Giulia	2.2	1.2	1.3
Emilia Romagna	7.4	7.4	7.3
Toscana	6.3	5.8	5.8
Umbria	1.5	2.3	2.1
Marche	2.7	1.8	2.1
Lazio	9.0	9.2	9.2
Abruzzo, Molise	2.8	2.4	2.3
Campania	9.2	9.9	9.5
Basilicata	1	1.3	1.3
Puglia	6.8	7.4	7.5
Calabria	3.3	3.4	3.3
Sicilia	8.1	8.5	8.4
Sardegna	2.8	3.0	2.9
n	39,274	1,996	2,005

Weighted data

Table 11: Continuous variables

	Multi Purpose		ITANES pre-electoral		ITANES post-electoral	
Variable	mean	sd	mean	sd	mean	sd
Age	48.8	18.2	49	17.29	48.8	18.21
Hours worked	18.0	21.2	19.8	21.8	19.4	21.4
n	39,274		1,996		2,005	

Weighted data

Table 12: Categorical variables

Variable	Multi Purpose	ITANES pre-electoral	ITANES post-electoral
Male	0.5	0.5	0.5
Employment status			
Employed	46.1	49.1	48.2
Looking for a new job	3.1	2.9	3.0
Looking for the first job	3.5	1.9	1.6
At home	16.2	15.5	14
Student	5.4	5.3	6.7
Unable to work	1.1	0.3	0.1
Retired	23	24.8	26
Other	1.7	0.3	0.4
Sector of employment			
No sector	49.1	47.8	48.4
Primary sector	2.6	1.8	2.2
Secondary sector	14.6	15.9	15.8
Tertiary sector	33.7	34.6	33.7
Education			
No education	4.7	3.3	3.2
Primary school	20.9	18.4	19.6
Lower secondary school	30.4	28.3	26.2
2 or 3 years high school	5.7	6.1	7.2
4 or 5 years high school	27.9	35	35.1
University degree	10.5	9	8.7
n	39,274	1,996	2,005

Weighted data

References

AAPOR (2009): "An evaluation of the methodology of the 2008 Pre-election Primary Poll. Revised version," Discussion paper, AAPOR.

——— (2010): "New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers," Discussion paper, AAPOR.

Barnes, S., M. Kaase, K. Allerback, B. Farah, F. Heunks, R. Inglehart, M. Jennings, H. Klingemann, A. Marsh, and L. Rosenmay (1979): *Political Action: Mass Participation in Five Western Democracies*. Beverly Hills. Sage.

Biemer, P. (2010): "Total Survey Error: Design, Implementation, and Evaluation," *Public Opinion Quarterly*, 74 (5), 817–848.

Blumberg, S., and J. Luke (2007): "Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults," *Public Opinion Quarterly*, 71, 5, 734–749.

Boyle, J., M. Bucuvalas, L. Piekarski, and A. Weiss (2009): "Zero Banks: Coverage Error and Bias in Rdd Samples Based on Hundred Banks with Listed Numbers," *Public Opinion Quarterly*, 73, 729 – 750.

Busse, B., and M. Fuchs (2011): "The components of landline telephone survey coverage bias. The relative importance of no-phone and mobile-only populations," *Quality & Quantity*, forthcoming, http://www.springerlink.com/content/9mn541176hh82ux4/.

Callegaro, M., and G. Gasperoni (2008): "Accuracy of Pre-election Polls for the 2006 Italian Parliamentary Election: Too close to call," *International Journal of Public Opinion Research*, 20, 2, 148–170.

Callegaro, M., and T. Poggio (2004): "Espansione della Telefonia Mobile ed errore di copertura nelle inchieste telefoniche," *Polis*, XVIII, 3, 477–506.

Diamanti, I. (2005): *Bianco, rosso, verde... e azzurro. Mappe e colori dell'Italia politica*. il Mulino.

D'Orazio, M., M. Di Zio, and M. Scanu (2006): *Statistical Matching. Theory and Practice*. Wiley, New York.

Durand, C. (2008): "The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2002 Catastrophe?," *International Journal of Public Opinion Research*, 20, 3, 275–298.

Durand, C., A. Blais, and M. Larochelle (2004): "The polls in the 2002 French Presidential Election: An Autopsy.," *Public Opinion Quarterly*, 68, 4, 602–622.

Durand, C., A. Blais, and S. Vachon (2001): "A late Campaign Swing or a Failure of the Polls? The Case of the 1998 Quebec Elections," *Public Opinion Quarterly*, 65, 108–123.

——— (2002): "Accounting for Biases in Election Surveys: The Case of the 1998 Quebec Elections," *Journal of Official Statistic*, 18, 1, 25–44.

Eurobarometer (2010): "E-Communications Household Survey-Summary," Discussion paper, European Commmission.

Fuchs, M. (2009): "The relative coverage bias in landline telephone samples across Europe. The impact of the mobile-only population," Discussion paper, mimeo.

Gasperoni, G., and M. Callegaro (2007): "Non cantare vittoria". La capacità predittiva dei sondaggi preelettorali pubblicati in occasione delle elezioni politiche italiane del 2001 e del 2006," *Polis*, XXI, 3, 463–488.

Gasperoni, G., and M. Callegaro (2008): "Un miglioramento immeritato? La capacità predittiva dei sondaggi preelettorali e le elezioni politiche del 2008," *Polis*, XXII, 3, 483–506.

Groves, R. M., F. J. Fowler, J. M. Couper, M. P.and Lepkowski, E. Singer, and R. Tourangeau (2009): *Survey Methodology, 2nd.* Wiley, New York.

Groves, R. M., and L. Lyberg (2010): "Total Survey Error. Past, present and future," *Public Opinion Quartely*, 75(5), 849–879.

Häder, M., S. Häder, and M. e. Kuhne (2011): *Telephone Surveys in Europe: Research and Practice*. Springer.

Horvitz, D., and D. Thompson (1952): "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47(260), 663–685.

Jowell, R., B. Hedges, P. Lynn, G. Farrant, and A. Heath (1993): "The 1992 British Elections: The Failure of the Polls," *Public Opinion Quarterly*, 57, 238–263.

Kamakura, W., and M. Wedel (1997): "Statistical data fusion," *Journal of Marketing Research*, 34, 485–498.

Keeter, S., C. Kennedy, A. Clark, T. Tompson, and M. Mokrzycki (2007): "What's missing from national landline RDD surveys? The impact of the growing cell-only population," *Public Opinion Quarterly*, 71, 5, 772–7952.

Little, R. (1988): "Missing data adjustments in large surveys," *Journal of Business and Economic Statistics*, 6, 287–296.

Luskin, R. (1990): "Explaining political sophistication," *Political Behavior*, 12, 331–361.

Magalhães, P. C. (2005): "Pre-election polls in Portugal: Accuracy, Bias and Sources of Error, 1991-2004," *International Journal of Public Opinion Research*, 17, 4, 399–421.

Merkle, D., G. Langer, and D. Lambert (2011): *Public Opinion in a Globalized World*chap. Methodological Issues in Pre-election Polling. Lessons from ABC News' 32-Night Tracking Poll. Springer.

Mokrzycki, M., S. Keeter, and C. Kennedy (2009): "Cell-Phone-Only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias," *Public Opinion Quarterly*, 73(5), 845–865.

O'Muircheartaigh, C., and P. Lynn (1997): "Editorial: The 1997 UK Pre-election Polls," *Journal of the Royal Statistical Society Series A*, 160, 381–388.

Pisati, M. (2008): "Sondaggi insondabili. Alcune note metodologiche sui sondaggi politici diffusi dai mezzi di comunicazione di massa," *Sociologica*, April.

Profeta, P. (2007): "Political support and tax reforms with an application to Italy," *Public Choice*, 131 1-2, 141–155.

Rassler, S. (2004): "Data Fusion: Identification Problems, Validity, and Multiple Imputation," *Austrian Journal of Statistics*, 33, 153–171.

Renssen, R. H. (1998): "Use of statistical matching techniques in calibration estimation," *Survey Methodology*, 24, 171–183.

Rodgers, W. L. (1984): "An evaluation of statistical matching," *Journal of Business and Economic Statistics, 1984, 2, pp. 91-102.*, 2, 91–102.

Rubin, D. (1986): "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations," *Journal of Business & Economic Statistics*, No. 1, 87–94.

——— (1987): *Multiple imputation for non response in surveys*. Wiley, New York.

Rubin, D. (1996): "Multiple imputation after 18 years," *Journal of the American Statistical Association*, 91, 473–490.

Sala, E., and L. Fumagalli (2011): "Don't talk to strangers. Civic engagement and non-coverage bias for pre-electoral polls in Italy," .

Schenker, N., and J. M. G. Taylor (1996): "Partially parametric techniques for multiple imputation," *Computational Statistics & Data Analysis*, 22, 425–446.

Seligson, M. (1980): "Trust, Efficacy and Modes of Political Participation: A Study of Costa Rican Peasants," *British Journal of Political Science*, 10, 75–98.

Singh, A., H. Mantel, M. Kinack, and G. Rowe (1993): "Statistical matching: use of auxiliary information as an alternative to the conditional independence assumption," *Survey Methodology*, 19, 59–79.

Smith, T. M. F. (1996): "Public Opinion Polls: The UK General Election, 1992," *Journal of the Royal statistical Society. Series A.*, Vol. 159, No. 3, 535–545.

Särndal, C. E., B. Swensson, and J. Wretman (1992): *Model-Assisted Survey Sampling*. New York: Springer-Verlag.

Van der Laan, P. (2000): "Integrating administrative registers and household surveys," *Netherland Official Statistics*, 15, 7–15.

Vicente, P., and E. Reis (2009): "The Mobile-only Population in Portugal and Its Impact in a Dual Frame Telephone Survey," *Survey Research Methods*, 3, 2, 105–111.