Adjusting for Polling Biases in Venezuela's 2012 Presidential Election

David Rosnick

October 2012

Center for Economic and Policy Research
1611 Connecticut Avenue, NW, Suite 400
Washington, D.C. 20009
202-293-5380
www.cepr.net

Contents

Executive Summary 2
Introduction 2
Historical Poll Results 2
Estimating the Biases Among Polling Companies 4
Estimating the 2012 Election Results from Polling Data 6
Conclusion 8
Addendum: Model and Regression Results 8
References 10

Acknowledgements

The author thanks Mark Weisbrot and Rebecca Ray for helpful comments.

About the Author

David Rosnick is an economist at the Center for Economic and Policy Research, in Washington D.C.

Executive Summary

This paper uses polling data from the various polling firms in Venezuela over the years 2004-2010 to adjust the most recent polling data for polling firm bias in past elections. The regression results show that the average lead of President Hugo Chávez over challenger Henrique Capriles moves from 11.7 percentage points (unadjusted) to a 13.7 percent (adjusted) lead. This would give Capriles a 5.7 percent probability of winning the election.

The adjustment employs a constrained regression model to take into account the bias of participation of the firm in each of the election years that make up the data set (2004, 2006, 2007, 2009, and 2010). The results are robust to excluding any one of the years from the constraint, with the probability of an opposition win ranging from 3.1 to 9 percent, depending on which year is not part of the constraint.

Introduction

According to recent polling of the Venezuelan electorate[1], incumbent president Hugo Rafael Chávez Frías is expected to win by an average of 11.7 points—with pollsters reporting results anywhere from a 19.7 point victory (IVAD) to a 0.8 point loss (C21). But not all polls have the same track record over the previous five elections that took place between 2004 and 2010, and it may help to adjust the various poll results for "house" bias of the polling firm.

Determining the house bias is a tricky matter. Bank of America's Francisco Rodriguez released a short report last month that found Chávez with a 15.9-point lead after adjusting for polling companies' biases, based on their track records.[2] More recently, Iñaki Sagarzazu reported a much closer result—only six points.[3] This paper seeks to help sort out the current polling results.

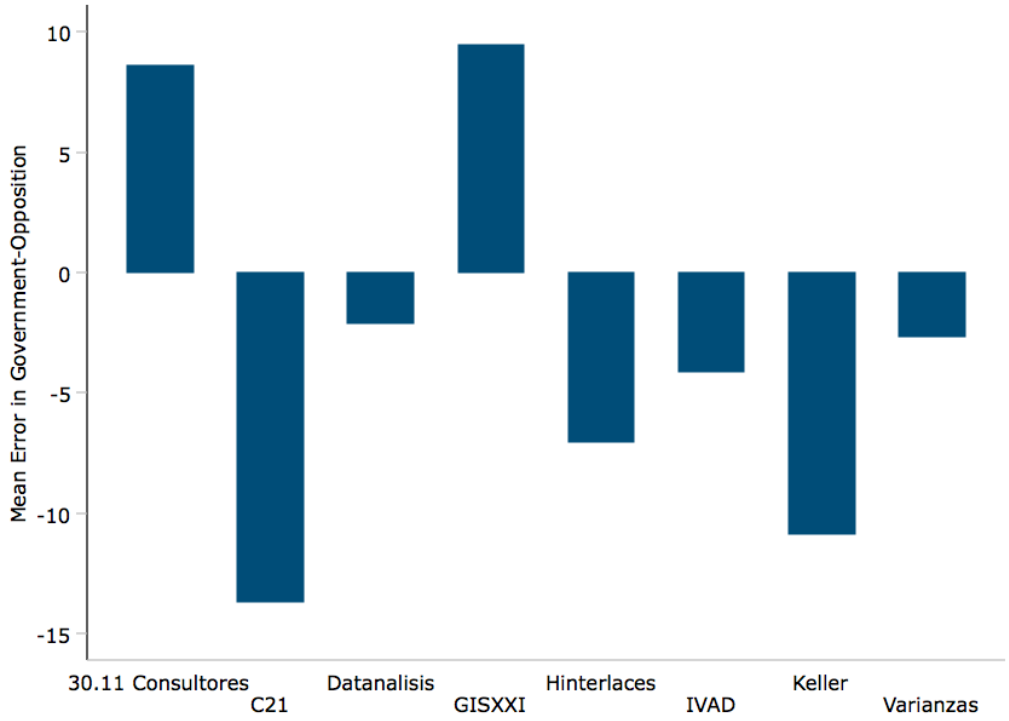
Historical Poll Results

Figure 1 shows the average polling error for eight different polling companies. A positive error implies that a poll reported better performance on the part of the government than the actual election results indicated. Thus, a positive error may indicate pro-government bias, or equivalently that the government under-performed in the election.

It would appear that the opposition outperformed relative to the polling of 30.11 Consultores, for example. However, we have historical data for this polling company in only 2006 and 2007—years in which the government slightly underperformed relative to the average poll. This is seen in **Figure 2**, below.

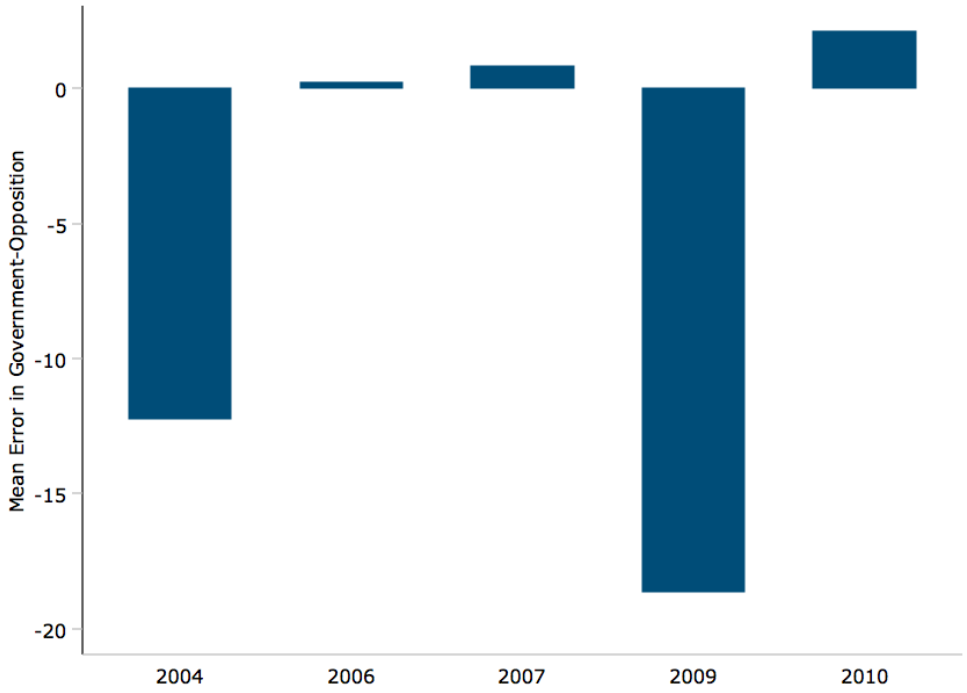
1 Based on the most recent surveys conducted between August 25 and September 25.
2 Rodriguez (2012).
3 Sagarzazu (2012).

Figure 1
Mean Polling Error by Company: 2004-2010



Sources: Sagarzazu (2012) and author's calculations.

Figure 2
Mean Polling Error by Election Year



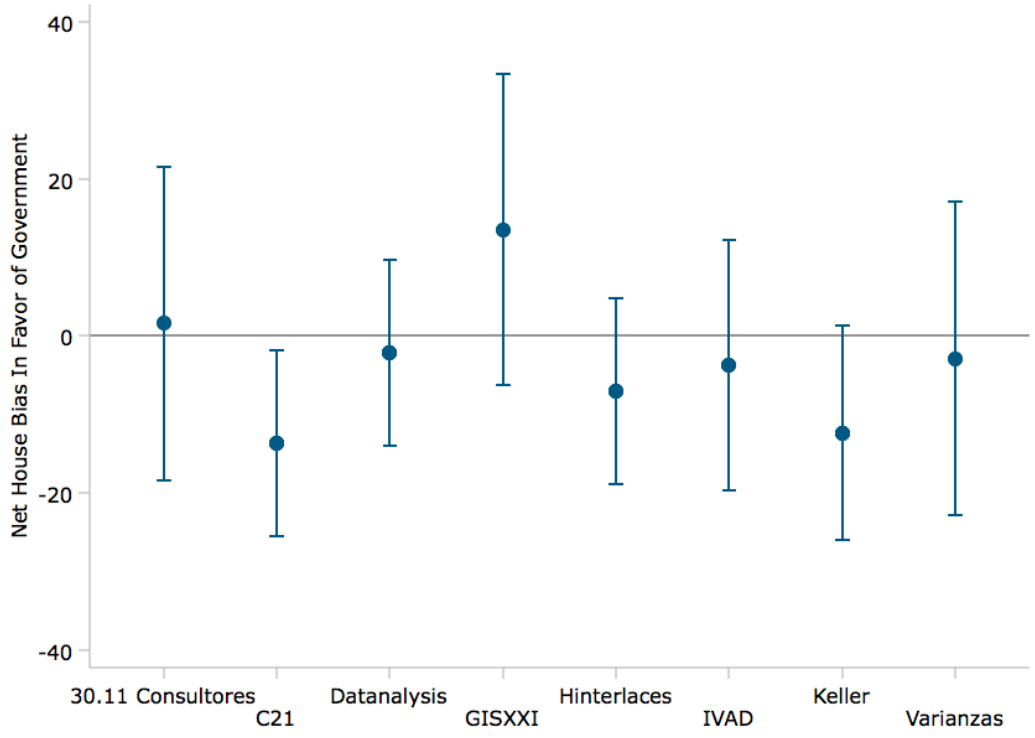
Sources: Sagarzazu (2012) and author's calculations.

Estimating the Biases Among Polling Companies

In order to estimate any "house" bias, then, we must control for any errors that might be systematic to any given election. This greatly complicates matters. Naïvely, we might want to simply regress the net error in favor of the government against indicator variables both for polling company and for year. However, this would report to us only the bias relative to some assumed baseline. That is, we might discover that Datanalisis polls 11.6 points higher for the government than does C21, but that does not tell us how to adjust either poll. Thus, we must find some outside constraint to find our level. For example, we might believe that Varianzas is unbiased and adjust all other polls correspondingly. Unfortunately, determining the bias of Varianzas is precisely our interest. If we knew that Varianzas was unbiased, we could simply use their polls to predict the outcome of the election.

Instead, we assume an "ideal" pollster would not be biased (on average) over the course of all five elections. If the house effect is zero for a single "ideal" pollster, then the average error over time should also be zero. In our simple model, this would imply that the average year effect must be zero, regardless of the house effect. We thus constrain the model so that all five election-year effects do sum to zero. Figure 3 shows the resulting estimates of house bias including 95% confidence intervals.[4]

Figure 3
Estimated House Biases

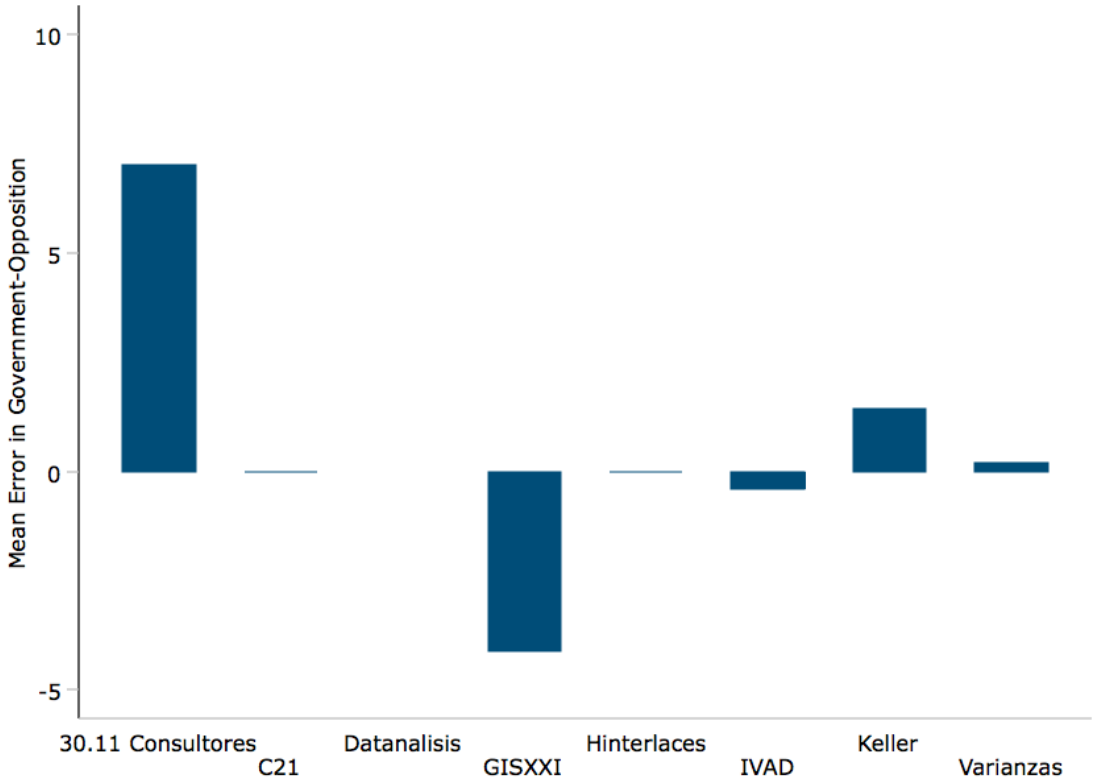


4 The regression results are shown in Table 3 in the Addendum.

Sources: Sagarzazu (2012) and author's calculations.
These results suggest that C21 and possibly Keller show some significant polling bias against the government, while GISXXI may show some bias in favor.[5]

As a reality check, we compute the average poll error after adjusting for house bias. If we have properly adjusted for house bias, then the errors are all election-year effects, which sum to zero. Hence, any remaining error in the mean implies a bias resulting from participation in only select elections rather than all five. For example, we see in **Figure 4** that 30.11 Consultores did in fact poll in elections where the government fared poorly.

Figure 4
Participation Bias Among Polling Companies



Sources: Sagarzazu (2012) and author's calculations.

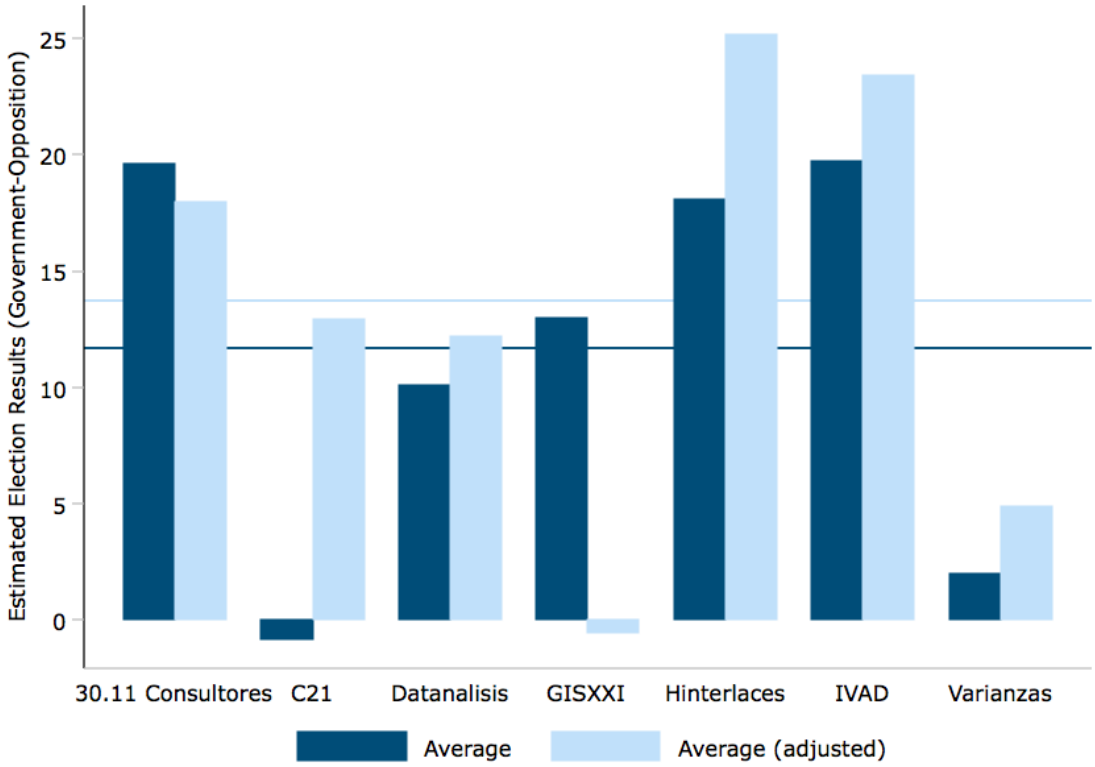
C21, Datanalisis, and Hinterlaces each polled in all five elections—so *after adjusting for their house biases*, they appear to be "ideal." Likewise, GISXXI participated in only two elections, of which 2009 was one—a year in which available polling greatly underestimated support for the government.

5 The bias found for GISXXI was relatively large, but not statistically significant.

Estimating the 2012 Election Results from Polling Data

Not knowing what kind of election bias there might be in 2012, we may now adjust the latest polling results for the implied house biases, with the results in **Figure 5**, below.[6] Figure 5 shows that after adjusting for bias, Chávez's lead grows from 11.7 points to 13.7. Surprisingly, C21 and GISXXI almost exactly reverse roles after adjusting.

Figure 5
Polling Averages—Unadjusted and Adjusted



Sources: Sagarzazu (2012) and author's calculations.

With the estimated house effects in hand, we may ask what effect on our results came from assuming that the average year effect is actually zero. To investigate the sensitivity, we repeat the analysis, constraining the average of only four elections rather than all five. The fifth effect in each case is unconstrained—implying that our "ideal" pollster did not participate in that year.[7] **Table 1** shows the unadjusted poll average, and the various adjusted averages.

6 No result is reported for Keller, as no recent survey from that firm was included in the source data.

7 Or at least, that the unconstrained year is unusual and its effect should not be directly included in evaluation of unbiased polling. In a sense, we grade or "ideal" pollster "on a curve" to take into account a possible one-off error in the election model.

Table 1
Adjustment Results

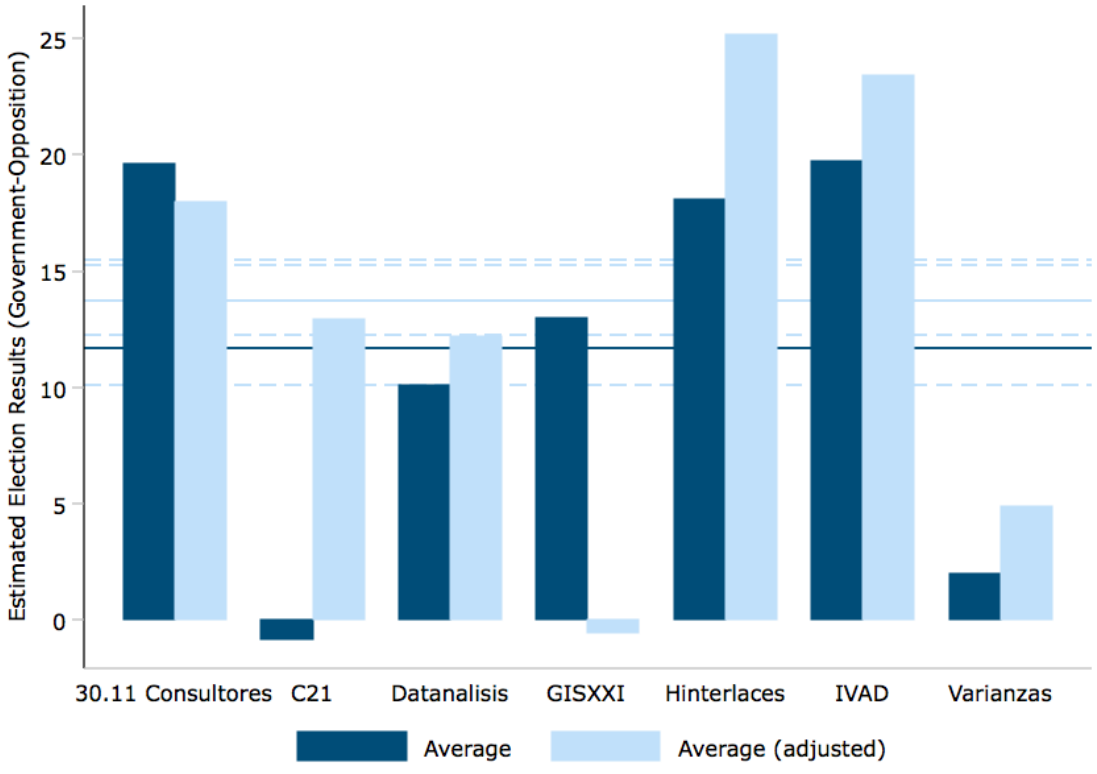
Method		Average 2012 Poll Result (Net Government-Opposition)
Unadjusted		+11.7
Fully-constrained adjustment		+13.7
Partially-constrained adjustment		
	2004 unconstrained	+12.3
	2006 unconstrained	+15.5
	2007 unconstrained	+15.5
	2009 unconstrained	+10.1
	2010 unconstrained	+15.2

Sources: Sagarzazu (2012) and author's calculations.

Table 1 shows that the basic adjustment is particularly sensitive in particular to the 2009 polling results. Nevertheless, the adjustments show the government with a lead of 10-15 points.

Figure 6 is identical to Figure 5, but the dashed lines show the different averages of Table 1 depending on which of the five elections is unconstrained.

Figure 6
Polling Averages- Including Partially-Constrained Models



Sources: Sagarzazu (2012) and author's calculations.

Finally, we may estimate-- based on historical year effects—the likelihood of such a lead holding up. **Table 2** estimates the chance of an opposition victory under the assumption that an *ordinary* random election-year effect occurs.

Table 2

Method	Average 2012 Poll Result (Net Government-Opposition)	Standard Deviation in Constrained Year Effects	Ratio (z)	Chance of Opposition Victory[8]
Unadjusted	+11.7			
Fully-constrained adjustment	+13.7	8.68	1.58	5.7%
Partially-constrained adjustment				
2004 unconstrained	+12.3	9.15	1.34	9.0
2006 unconstrained	+15.5	8.88	1.75	4.0
2007 unconstrained	+15.5	8.86	1.75	4.0
2009 unconstrained	+10.1	5.42	1.86	3.1
2010 unconstrained	+15.2	9.07	1.68	4.6

Sources: Sagarzazu (2012) and author's calculations.

Conclusion

Adjusting polls for possible house bias appears to increase expectations for the government in the 2012 election—with a 13.7 point lead. However, this result is sensitive to the establishment of a neutral baseline. In particular, Chávez's lead may be somewhat smaller (10.1 points) if the 2009 balloting is considered somewhat anomalous. However, if the 2009 ballot is considered anomalous, there is considerably less variation in the ordinary year effect, making a 10-point lead more secure.

Addendum: Model and Regression Results

In this paper, we regress 28 poll errors (relative to official results) against indicator variables for each polling company and each vote. That is, the error in the poll performed by company i for the ballot in year t is given simply by

$$E_{i,t} = \pi_i + \gamma_t + \varepsilon_{i,t}$$

where π_i is the house bias of polling company i, γ_t is the effect of year t and $\varepsilon_{i,t}$ is a random error associated with the specific observation. Not every combination of company and year is represented in the data.

In the fully constrained model, $\Gamma \equiv \gamma_{2004} + \gamma_{2006} + \gamma_{2007} + \gamma_{2009} + \gamma_{2010} = 0$, while in

8 Defined here simply as the probability of a standard normal draw greater than the ratio z.

partially constrained model, the term corresponding to the chosen year (T) is not included in the constraint. (Equivalently, we may say $\Gamma = \gamma_T$.)

Table 3 shows the results of the regression model. Also noted is the standard deviation of the year effects that are included in the constraint. This represents the ordinary variation in the year-to-year results for which an unbiased pollster cannot account. It is implicit in the partial constraint models that the chosen year is-- while not ordinary in that same respect—not necessarily predictable either.

Table 3
Regression Results

	Fully Constrained	Partially Constrained				
		2004	2006	2007	2009	2010
30.11 Consultores	1.62 (0.16)	3.06 (0.31)	-0.12 (-0.01)	-0.14 (-0.01)	5.21 (0.52)	0.08 (0.01)
C21	-13.71 (-2.27)*	-12.26 (-1.96)*	-15.45 (-2.49)*	-15.47 (-2.50)*	-10.11 (-1.64)	-15.24 (-2.47)*
Datanalisis	-2.12 (-0.35)	-0.68 (-0.11)	-3.86 (-0.62)	-3.88 (-0.63)	1.47 (0.24)	-3.66 (-0.59)
GISXXI	13.57 (1.34)	15.01 (1.50)	11.82 (1.18)	11.80 (1.17)	17.16 (1.63)	12.03 (1.15)
Hinterlaces	-7.04 (-1.17)	-5.60 (-0.90)	-8.78 (-1.42)	-8.80 (-1.42)	-3.45 (-0.56)	-8.58 (-1.39)
IVAD	-3.73 (-0.46)	-2.29 (-0.29)	-5.47 (-0.68)	-5.50 (-0.65)	-0.14 (-0.02)	-5.27 (-0.63)
Keller	-12.32 (-1.78)#	-10.88 (-1.61)	-14.07 (-1.96)#	-14.09 (-1.97)#	-8.73 (-1.22)	-13.86 (-1.95)#
Varianzas	-2.87 (-0.28)	-1.43 (-0.13)	-4.61 (-0.46)	-4.63 (-0.46)	-0.73 (0.07)	-4.40 (-0.42)
2004	-5.78 (-0.91)	-7.22 (-0.91)	-4.03 (-0.66)	-4.01 (-0.66)	-9.37 (-1.54)	-4.24 (-0.68)
2006	6.97 (1.22)	5.52 (1.02)	8.71 (1.22)	8.73 (1.54)	3.37 (0.61)	8.50 (1.55)
2007	7.05 (1.34)	5.61 (1.13)	8.79 (1.68)	8.81 (1.34)	3.45 (0.67)	8.59 (1.68)
2009	-14.39 (2.72)*	-15.83 (-3.17)**	-12.64 (-2.49)*	-12.62 (-2.44)*	-17.98 (-2.72)*	-12.85 (-2.44)*
2010	6.14 (1.24)	4.70 (0.97)	7.88 (1.66)	7.91 (1.64)	2.55 (0.51)	7.68 (1.24)
Standard deviation in constrained year effects	8.68 (2.65)**	9.15 (2.90)**	8.88 (2.81)**	8.86 (2.89)**	5.42+	9.07+

Sources: Sagarzazu (2012) and author's calculations.
Notes: t-statistics in parenthesis,
Significant at 10% level
* Significant at 5% level
** Significant at 1% level
+ Standard error could not be estimated

References

Rodriguez, Francisco. 2012. "An Investor's Guide to Venezuela's Election." GEM Economics. Bank of America/Merrill Lynch. September 13.

Sagarzazu, Iñaki. 2012. "Venezuelan Pollsters, their Records and the 2012 Race—September Update." Venezuelan Politics and Human Rights Blog. Washington Office on Latin America. October 2. http://venezuelablog.tumblr.com/post/32767451785/venezuelan-pollsters-their-records-and-the-2012