súmate

Av. Francisco de Miranda, Centro Plaza, Torre A, Piso 5, Los Palos Grande, Caracas, Telf: 58 (212) 5156215
Correo-e: prensa@sumate.org / Pág. Web: www.sumate.org

Caracas, September 17, 2004.

NPS_ 80

The Carter report on the Hausmann-Rigobon Analysis is bereft of any statistical basis.

Yesterday, the Carter Center issued a Report on the representative character of the Data for the Second Audit and the correlation between the number of original recall petition signatures and the actual number of 'YES' votes polled in the Presidential Recall Referendum of August 15; in this report, the Center states its position regarding the Hausmann/Rigobon Report.

The Carter Center Report rejects the conclusions reached by Professors Ricardo Hausmann [of Harvard] and Roberto Rigobon [of M.I.T.] which were: 1) that the post-referendum audit of the results of the Presidential Recall Referendum in Venezuela, observed by the OAS and the Carter Center, was not carried out on a sample [properly] representing the universe of the all the automated polling stations; and the sample [used for it] was not randomly selected; and 2) the statistical proof that they, Hausmann and Rigobon, developed to [ascertain whether there had been any] manipulation [of results] gave a positive result. In both cases, Hausmann and Rigobon developed statistical analyses underpinning their conclusion that the probability of over 99% of there having been manipulation involved in the random selection mentioned above as also the results of the Presidential Recall Referendum itself.

In stating their position, the Carter Center offer a Statistical Report that, notwithstanding its lack of professional statistical input, states [baldly] that: 1) the correlation of the petition signatures and the votes in both the overall sample and the [sample for the] audit are identical; and 2) the mean [????] of the votes in the audited sample and that of the [actual] voting are similar.

In this regard, Professor Rigobon observes in the first case, being the assertion that the 'correlation between petition signatures and votes is identical,' as follows:
"The first argument is statistically wrong; I would explain as follows: 1) The correlation of a variable with itself is exactly equal to one; 2) the correlation of a variable with 10% (ten percent) of itself is also one; and 3) the correlation of two variables remains unaltered if you multiply them by positive numbers. In other words, the correlation between the petition signatures and the 'YES' votes in the real data is exactly the same as the correlation between the signatures and the fraudulent votes, bearing in mind that the fraudulent votes are equal to the 'YES' votes, less 10 or 20 or 90 percent. That said, whilst the correlations are identical, the total number of votes is not. By way of example, we can consider two people working in a company and on the same salary, say Bs.1,000,000 [per month], and receiving a yearly rise of 20%: these two people will always have identical incomes because their salaries are correlated in 20% yearly increases. But, in that same office, if there is a person earning Bs. 500,000 [per month], being half the amount mentioned before, whose yearly raises are also 20% , then that person's salary is also correlated with the other two because all have the same 20% yearly increase, regardless of their earning twice the amount. In other words, the correlations are identical but the salaries are not.

As for the second argument, being 'The mean of the audited sample is similar to the mean of the votes overall.' This is simply statistical sophistry. And statistically wrong. I would explain how it was possible to have perpetrated a fraud showing the same the same [or a similar] mean between the audited and unaudited samples. Now, insofar as it is an example, we won't use the Venezuelan Recall Referendum but rather, a fictitious presidential election in Florida.

Let us suppose that, in Florida, there exists a voter distribution such that half the automated polling centers have a democratic majority and the other half have a republican majority. So, half and half. How do we know that there are some polling stations that are democrat and others that are republican? [In the event,] no problem:

we know the results of previous voting episodes and also how many democrats and republicans are registered in each automated polling station. That said; let us now assume there is a fraud planned for half the democrat polling stations, being a quarter of the overall number of polling stations. [Let us further assume that] In these polling stations, the voting machines convert ten percent (10%) of democrat votes into republican ones. And, to keep things simple, let us suppose that this shift of democrat to republican is enough to change the polling station from democrat to republican. We would [reasonably] qualify such an undertaking as 'fraud.'

The result of this election would give ¾ of the centers as republican and ¼ as democrat. Now the problem arises in figuring just how to generate a random sample such as to ensure that the fraud will remain undetected by the international observers' audit. To see what is going on here, it is vital to know that computer programs for generating random samples need an initial number that is inputted to the computer. The program generates the random sequence from this number; it is known as the 'seed.'

The Carter Center says that the National Electoral Council's program [always] generates exactly the same sample as long as it starts from the same seed. This would mean that it would not be difficult for someone to generate a set of random numbers before [???] the elections with a greater 'representiveness' of centers where, naturally, republicans win; in other words, showing ¾ republican and ¼ democrat. All I have to do is to remember the 'seed' and I enter it when the time comes to select the polling stations for carrying out the audit. The conclusion is: in democrat-majority centers not included in the audit, I can tweak the results to reduce votes for the democrats. This means that the end result of the elections will reflect republicans winning at 75% of polling stations -- as will the audit. This example is interesting because the Carter Center would have no problem in asserting that the audit sample in question was random.

Av. Francisco de Miranda, Centro Plaza, Torre A, Piso 5, Los Palos Grande, Caracas, Telf: 58 (212) 5156215
Correo-e: prensa@sumate.org / Pág. Web: www.sumate.org

Look too at the example of the Kino [type of lottery] (do US readers have enough familiarity with 'Kino' raffles? Is there a better model for the purpose?) In this case, every Sunday, 25 [numbered] balls are put into machine for it to choose 15. Now, if [a particular group of] 15 are 'loaded' and so heavier, they will be the ones that always come out. But, if the unscrupulous gentleman who did the loading knows where the boxes [containing the tickets] with those ['random'] numbers were sent, he will show up, buy them and win the jackpot!

You can't demonstrate fraud by comparisons of means or correlations. Truth be told, this is really elementary stuff. You must compare means and CONDITIONAL correlations. What shows fraud in the above example isn't that the means are different but rather that they are conditioned by the sample information available, being how many polling stations are democrat sand how many are republican. So, the audited sample will exhibit a closer correspondence with that information than would the unaudited sample. Which is precisely what Ricardo Hausmann and I showed."

Professor Rigobon concludes his response thus, "It is possible that a flawed manipulation would fail in the dimensions taken by the Carter center but let's not underestimate the people who carried it out."