THE WALL STREET JOURNAL.

For Math Whizzes, The Election Means A Quadrillion Options

* * *

Close Race Has Programmers Predicting the Outcome; 'I'm Just Some Geek'

By Charles Forelle

To prepare for next week's election, Lawrence N. Allen taught himself the Matlab statistical programming language and built a database of 1,700 state polls pulled off the Internet. His program runs a "likelihood analysis" on 15 closely contested battleground states. It takes 50 minutes to run on an old computer he got in return for a bunch of parts from a broken laptop.

The unemployed computer programmer in Oakland, Calif., identifies his politics as "to the left of standard Democratic candidates" and says he flirted with voting for Ralph Nader in 2000 before opting for libertarian Harry Browne. His calculations, made on Oct. 20, give Mr. Bush a 78.1% chance of victory.

Mr. Allen says he drew inspiration from Sam Wang, an assistant professor of molecular biology and neuroscience at Princeton University, who devised a computer program to analyze state polls and step through all the possible outcomes of 22 supposed battleground states.

There are 4,194,304 of them. (That's 2 to the 22nd power: two possible choices—Bush or Kerry—in 22 states.)

As of yesterday evening, Mr. Wang's "median outcome" was a razor-thin majority for Mr. Bush—279 votes in the decisive Electoral College, versus Mr. Kerry's 259, not counting undecided voters. But if the results followed historical patterns in which undecided voters generally break for the challenger, the Massachusetts senator would wind up with 307 electoral votes and the Oval Office, Prof. Wang says, based on his computations.

Messrs. Allen and Wang are among an elite cadre of political amateurs unleashing the tools of statistics and mathematics on an extraordinarily close presidential race.

Andrea Moro, an economics professor at the University of Minnesota, uses a type of simulation known as the Monte Carlo method to calculate the probable outcome. John Denker, a physicist and former AT&T Corp. and Bell Labs researcher, parses the Electoral College with a gigantic Microsoft Excel spreadsheet populated with reams of
Please Turn to Page A14, Column 4

Election Options for Math Whizzes

Continued From First Page
polling data that the Founding Fathers couldn't have imagined.

"Sometimes, brute force has an elegance all its own," says Mr. Denker, who describes himself as "polymathic mad scientist."

Mr. Wang says, "Electoral prognostication is just exploding on the Internet." He fastidiously updates data and posts comments to his Web site, election.princeton.edu. At 3:30 a.m., Oct. 22, he had this to say about the big state known for its hurricanes and hanging chads: "If Bush wins Florida, his win probability is 88%; if he loses, it's only 20%."

The Internet explosion was primed by the rise of blogging and the ready availability of state polls, manna for numerically inclined political junkies. Now, there is an abundance of Web sites dedicated to drawing electoral maps in red, for Republican, and blue, for Democrat.
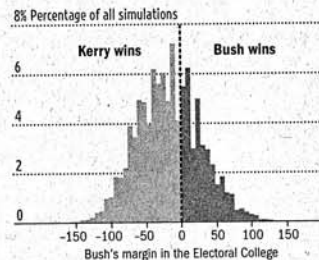
But the statistical modelers contend that isn't enough to go one-by-one through the states and call them for Mr. Bush or Mr. Kerry. That, they say, misses substantial nuances that are greatly magnified by the large number of states in play. For example, a candidate polling slim margins of victory in a number of small states is less likely to win them all than an opponent who has larger leads in fewer, but larger, states. The distinction is hard to represent on a color-coded map; it is more easily captured by statistical software.

So the new wave of amateur prognosticators employ a technique widely used in the physical sciences, known as likelihood analysis or "probabilistic modeling." The idea is to understand complex events by breaking them into simpler, discrete events and assessing the probabilities of those events' actually occurring. Physicists, for instance, express the positions of subatomic particles as probabilities. Astronomers and cosmologists use likelihood analysis to generate estimates for quantities such as the age and expansion rate of the universe.

Mr. Denker started building his model in August, after watching a TV commentator botch an explanation of probable outcomes. "I'm watching the news and slapping my forehead and saying, how can these guys be so silly?"

Matthew Hubbard, a math lecturer who teaches at California State University in Hayward, notes that many pre-election polls four years ago incorrectly predicted that Mr. Bush would win the popular vote. That led Mr. Hubbard to believe the data weren't being processed properly by the media. "We weren't really getting what I thought was particularly interesting or particu-

larly good information," he says. "So many people, when they talk on TV and write in newspapers, their mathematics are so bad."

So this year, Mr. Hubbard, a computer programmer who once wrote games for Atari, created his own model, with its own Web site, that chomps through 16.8 million possibilities in the Electoral College in 72 seconds. His Oct. 23 prediction gave Mr. Kerry a 73.9% chance of reaching the winning threshold of 270 electoral votes, with Mr. Bush at 24.6%. He rated the probability of a 269-269 tie at 1.6%.

Mr. Hubbard, who is a Democrat, isn't placing his bets quite yet, since the numbers have been shifting over the past several weeks. "I've played enough backgammon and poker to know that you don't celebrate too early," he says.

So it is with this agonizingly close election, where the slightest tips and swings have broad influence, and the combination of possible results is vast. Each of the models uses slightly different polling databases and different methods for assessing the errors in those polls, but all follow a similar scheme.

A statistical formula can transform one or more poll results and margins of error into a probability of victory. To take a simple example that avoids the complications of undecided voters and independent candidates, assume that Mr. Bush garners 55% and Mr. Kerry 45% in a poll with a three percentage-point margin of error. The wide gap and low margin yield a probability of more than 99% that Mr. Bush wins. A poll that's 51%-49% for Mr. Bush, with a four-point margin, on the other hand, would yield only a 68.5% probability of victory.

Armed with the probabilities of victory in each state, the computers go to work crunching the probability of each separate Electoral College outcome. In a two-party system, there are quadrillions

of them. Cutting the number of states down to a dozen or two battlegrounds makes the number more manageable— between about 30,000 and 16 million. (That ignores the fact that two or three states may split their electoral votes.)

The wide variation in the expected outcomes is due largely to the differences in selecting what polls to use, how to treat undecided voters, and what method is used to derive the statistical error in poll results. The modelers say, too, that they don't have a way to take into account "external factors" not relating to the sampling error in poll-taking—including turnout and polling bias.

As a domain, politics is far messier than physics, says Alan Abramowitz, a professor of political science at Emory University and an expert on polling. The data are "subject to all kinds of error," he says. "Trying to impose some of these very high-powered methods may be overwhelming the data." What's more, state polls generally draw from a smaller sample than national polls and are as a result less reliable.

The race is so close and so difficult to assess that Mr. Wang is split with himself: His model predicts a very narrow Bush win when he makes no assumptions about the behavior of undecided voters, and a larger Kerry victory if he allocates more undecideds to the Massachusetts senator.

When, last week, he switched his calculation to include the Kerry-boosting undecided formula, he got scores of protests from readers of all political persuasions who found the commingling of current survey data with historical trends to be an affront to the model's purity.

He was slightly taken aback by a flood of e-mailed criticism. "I'm just some geek posting numbers," he says.

Numbers Game

8% Percentage of all simulations

Kerry wins | Bush wins

Bush's margin in the Electoral College

Source: Andrea Moro, University of Minnesota

- In a speculative exercise, economist Andrea Moro used polls yesterday to assess the probability of a Bush victory in each state. Then he simulated the election 100,000 times to see who would win, based on those state probabilities.
- This chart shows what percentage of the outcomes gave Bush a given margin in the Electoral College. Positive margins—Bush wins—are to the right of the "zero" line.
- The results show more outcomes on the Kerry side of the line—hence, a somewhat more likely Kerry win. But other complex models predict a Bush win, and slight shifts in any important state could change the projected winner.

WSJ.com Journal Link: *Get a state-by-state look at the election and track the latest national polls, at* **WSJ.com.** *Available to nonsubscribers.*