Income and vote choice in the 2000 Mexican presidential election*

Jeronimo Cortina† Andrew Gelman‡

13 July 2006

Abstract

Using multilevel modeling of state-level economic data and individual-level exit poll data from the 2000 Mexican presidential election, we find that income has a stronger effect in predicting the vote for the conservative party in poorer states than in richer states—a pattern that has also been found in recent U.S. elections. In addition (and unlike in the U.S.), richer states on average tend to support the conservative party at higher rates than poorer states. Our findings are consistent with the 2006 Mexican election, which showed a profound divide between rich and poor states. Income is an important predictor of the vote both at the individual and the state levels.

*We thank the Department of Political Science at the Instituto Tecnologico Autonomo de Mexico (ITAM) and Gert Rosenthal for data, Jeff Lax and Caroline Rosenthal Gelman for helpful comments, and the U.S. National Science Foundation for financial support.

†Department of Political Science and Institute for Social and Economic Research and Policy, Columbia University, New York, jc2062@columbia.edu, www.columbia.edu/∼jc2062

‡Department of Statistics and Department of Political Science, Columbia University, New York, gelman@stat.columbia.edu, www.stat.columbia.edu/∼gelman

1 Introduction

> "The electorate is genuinely divided and the close election underlines it. Many are opting for a change while many are opting for continuity."[1] (Dresser 2006)

> "Yesterday, the electorate confirmed a regional division in which the north and northwest parts of the country favored Felipe Calderón, while the center and south supported Andrés Manuel López Obrador at higher rates."[2] (Reforma 2006)

> "The only thing that the election shows is that social polarization is not a children's story and less an invention. This polarization is a reality.... It is or it seems to be the legitimization of the fight between the rich and the poor."[3] (Alemán 2006)

> "The new map depicts an industrialized north, where business ties to the United States have played an enormous role in the rise of the right-leaning, conservative party, and a more agricultural south that is a hotbed of leftist discontent and anti-globalization sentiment." (McKinley 2006)

The conservative candidate from the National Action Party (PAN) won the most contested presidential election in Mexico's modern times by a margin of 0.6% over the leftist candidate from the Party of the Democratic Revolution (PRD) and almost 14% over the "catchall" candidate from the Institutional Revolutionary Party (PRI). Regardless of what may happen once Mexico's Electoral Court validates the electoral results, one thing is evident: the presidential vote was geographically divided, with the states of the north and center-west supporting the PAN and the states of the center and the south supporting the PRD. In other words, the electoral result was characterized by a divide between rich and poor states. This pattern is strikingly clear, but, as we shall see, it is not a simple aggregation of rich voters supporting the PAN candidate and poor voters supporting the PRD.

What happened in the July 2 presidential election? Did richer voters support the PAN candidate and poorer voters support the PRD? Does living in a rich or poor state change individual vote preferences—that is, does geography matter for voting behavior, after controlling for individual

[1] "El electorado está genuinamente dividido y la elección apretada lo subraya. Muchos optan por el cambio y muchos optan por la continuidad."

[2] "Los electores confirmaron, el día de ayer, una división regional en la que el norte y centro-occidente del País favorecieron a Felipe Calderón, mientras que las regiones centro y sur se manifestaron más por Andrés Manuel López Obrador."

[3] "Lo único que muestra es que la polarización social no es un cuento y menos un invento. Esa polarización es una realidad.... Es o parece ser la legitimación de la lucha de pobres contra ricos."

characteristics? Given that the 2006 exit polls are not yet publicly available,[4] we try to answer these questions by analyzing the relation between income and vote choice at the state and individual level on the outcome of the 2000 Mexican presidential election, which can be considered as the apogee of Mexico's democratic transition[5] that started in the late 1970s with the first comprehensive electoral reform (Becerra, Salazar & Woldenberg 2000, Lujambio 1997, Ochoa-Reza 2004).

Studies of the 2000 presidential election found that political factors such as the content of political campaigns, the notion of regime change, and the pro- and anti-regime divide in the electorate proved to better account for the variation in voting behavior than socio-demographic variables or even the left-right ideological division within the electorate (see the edited volume by Dominguez & Lawson 2004). While income is often included as a control, and the positive link between income and support for the conservative party is almost always noted in multivariate analyses (for example, see Klesner 1995, Dominguez & McCann 1996, Moreno 2003, Dominguez & Lawson 2004), the connection between income and vote choice has not been analyzed taking geography into account.

In this paper, we find that, on average, individual income matters more in poorer states than in richer states—a similar pattern as found by Gelman, Shor, Bafumi & Park (2005) in analyzing U.S. electoral data. The difference in voting patterns between rich and poor individuals is greater in rich states than in poor states. At the aggregate level, however, the conservative party (PAN) does better in richer states (in terms of GDP per capita) than in poorer states—unlike in the United States, where the Republicans have in recent years performed better in the poor states.

The rest of this paper is organized as follows. Section 2 summarizes the state-level presidential results for the 2000 and 2006 elections, Sections 3 and 4 describe our methods and results, and we discuss our findings in Section 5.

2 Geography matters: Mexico's political mosaic

Mexico is a country of geographically and ethnically diverse traditions and cultures. Just as the cuisine changes considerably all over the territory, income, state development and individual political preferences change dramatically from one Mexican state to another. For instance, the GDP per

[4] We plan to replicate our analysis with 2006 poll data.

[5] "A democratic transition is complete when sufficient agreement has been reached about political procedures to produce an elected, government, when a government comes to power that is the direct result of a free and popular vote, when this government de facto has the authority to generate new policies, and when the executive, legislative and judicial power generated by the new democracy does not have to share power with other bodies de jure." (Linz & Stepan 1996, p. 3)

2000 election:

2006 election:

Figure 1: States won by each of the political parties in the 2000 and 2006 elections. White represents those states in which the PAN won, light gray represents those states won by the PRI, and dark gray represents the states won by the PRD.

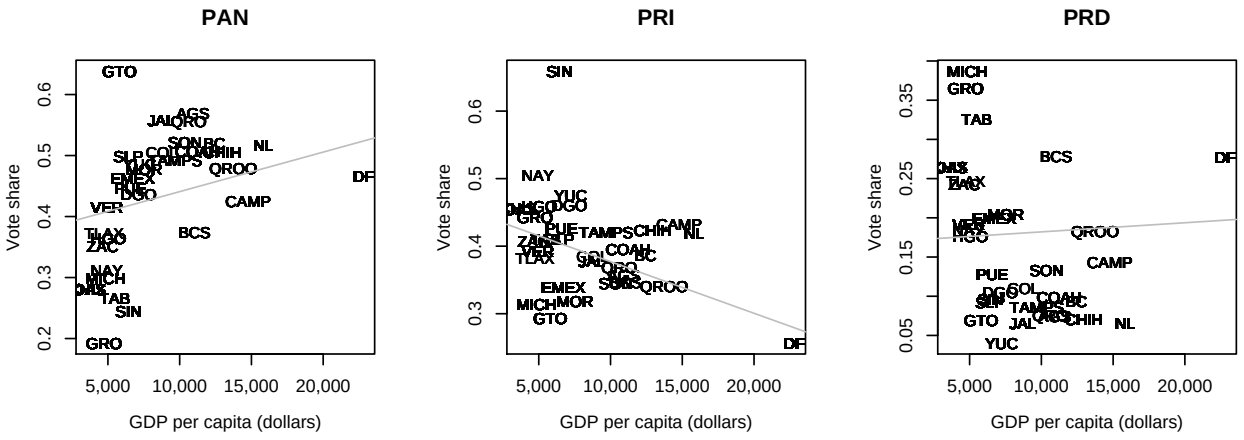

Figure 2: For each of the three major parties, the vote share in each state, plotted vs. per capita GDP. The conservative party (PAN) does better in richer states—in contrast to the United States, where the Republicans do best in the poor states.

capita of the richest state (Mexico, D.F.) is more than six times that of the poorest state (Chiapas). Similar differences are found in other realms such as health and education (PNUD 2003).

Nowadays Mexican politics can be defined as an ideologically polarized tripartite party system. On the left of the political spectrum, we find the PRD, the party of the *"clase popular"* (i.e., the poor); on the right is the PAN, the party of the middle class; and in a blurry center is the PRI, the former ruling party. At the individual level, public opinion and exit poll data show that voters of higher income and socioeconomic status tend to support the PAN while the less affluent tend to support the PRD or the PRI (Moreno 2003, Klesner 2004, Reforma 2006).

At the regional level, Figure 1 shows that the PAN regularly has done better in the wealthier parts of the country (center-west and north), and worse in Mexico City and the south (Klesner 2004, p. 105).[6] At this level of aggregation, however, the relationship between income and vote choice is not so clear. For instance, the average GDP per capita[7] in the regions won by the PAN in 2000 was $9,050, while the GDP per capita in those regions where the PAN under-performed was $8,800 (PNUD 2003).

At the state level, a different picture emerges. The average GDP per capita in the states won by the PAN was $10,200 for 2000 and $9,700 in 2006. In contrast, the GDP per capita in those

[6] Adapting Klesner (2004), we define the regions as: *north*: Baja California, Baja California Sur, Coahuila, Chihuahua, Durango, Nuevo Leon, San Luis Ptosi, Sinaloa, Sonora, Tamaulipas, Zacatecas; *center-west*: Aguascalientes, Colima, Guanajuato, Jalisco, Michoacan, Nayarit, Queretaro; *center*: Estado de Mexico, Hidalgo, Morelos, Puebla, Tlaxcala; *south*: Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz, Yucatan. For our analyses, we consider the Federal District (Mexico, D.F.) as a separate state from the rest of the Estado de Mexico.

[7] GDP in 2000 PPP U.S. dollars.

states won by the PRD was $4,800 in 2000 and $7,300 for 2006.[8] Figure 2 shows the details, and the variation within each plot reveals that collapsing multiple states into large regions entails significant loss of information that otherwise may uncover sharper and quite revealing differences between states.

Overall, PAN does better in the richer states and the PRD does better in the poorer states. However, Figure 2 also shows that there are no definite or absolute regional partisan strongholds. In other words, there is more variation between states than is suggested by current literature. One way to account for this variation between states, going beyond the inclusion of indicator variables for each state, is to use multilevel modeling. This statistical technique allows us to understand the relation between income and vote among individuals and states simultaneously.

3 Methods

Multilevel modeling allows us to estimate patterns of variation within and between groups (in this case, states), taking into account the hierarchical nature of the data (individuals within states) and also the specific characteristics of each state by allowing their intercepts and slopes to vary. (See, e.g., Snijders & Bosker (1999) for a general overview of multilevel models, and Gelman et al. (2005) for the particular example of income and voting.)

Our central model is a varying-intercept, varying-slope model predicting vote choice from individual income and GDP per capita, which we fit to data from the Grupo Reforma 2000 national exit poll (Reforma, El Norte & Mural 2000) excluding those respondents who did not report for whom they voted or who supported parties other than the PAN, PRI, or PRD. This left us with 2,540 responses, with sample sizes within states ranging from 9 in Tlaxcala to 339 in Estado de Mexico. The multilevel model allows us to estimate the income-voting relation in each state, with the estimates for the larger states coming largely from their own data and the estimates for smaller states relying more of the state-level regression model.[9]

The model for individual voters i is

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i, \text{ for } i = 1, \ldots, n, \tag{1}$$

[8] The 2006 figures are based on the preliminary results reported by IFE.

[9] More specifically, each estimated state-level coefficient in a multilevel model is a weighted average of the unpooled estimate for the state and the completely pooled estimate using individual and state-level predictors.

where y_i represents vote choice (1 = PRD, 2 = PRI, 3 = PAN), and x_i represents household income on a standardized scale.[10] The continuous model is clearly an approximation on this discrete scale, and so, as discussed below, we also fit logistic regressions to predict the vote for the parties of the right and the left.[11]

Since we are interested in comparing states with different wealth, we include GDP per capita within each state as a state-level predictor. The group-level intercepts and slopes are modeled as,

$$\alpha_j = \gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha, \text{ for } j = 1, \ldots, 32$$

$$\beta_j = \gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta, \text{ for } j = 1, \ldots, 32, \tag{2}$$

where u_j is the GDP per capita in state j, and the errors ϵ_j^α, ϵ_j^α have mean 0, variances σ_α^2, σ_β^2, and correlation ρ, all of which are estimated from the data when combined with the individual model. We also let the general levels for the intercepts and slopes (the parameters γ_0^α and γ_0^β) vary by region (north, center-west, center, south, and Mexico City), so that the model allows systematic variation by region and among states within regions.

In addition, we examine the estimated intercepts α_j and slopes β_j when including other predictors: state inequality, sex, age categories, the type of locality (urban, mixed or rural), the main reason for voting they voted (for a change, for the candidate, civic duty, custom, the least of evils, party loyalty, campaign promises, or other reasons),[12] and religion (Roman Catholic, evangelical Christian, other).

Our next step is to analyze each party by itself, modeling the probability of supporting the PAN and the PRD candidate, controlling for individual income and state GDP per capita. We fit a multilevel varying-intercept, varying-slope logistic regression of the form

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta_{j[i]} x_i), \text{ for } i = 1, \ldots, n, \tag{3}$$

adapting model (1) to the logistic scale. The state intercepts and slopes are modeled as in (2). We fit two different logistic regressions: PAN versus all others, and PRD versus all others.

For the linear model (1), positive slopes β_j correspond to richer voters within states supporting the PAN. For the logistic models (3), we would expect positive slopes β_j for the PAN model

[10] The survey had nine income categories (in pesos per month): \$0–\$1,000, \$1,001–\$2,000, \$2,001–\$4,000, \$4,001–\$6,000, \$6,001–\$8,000, \$8,001–\$12,000, \$12,001–\$16,000, \$16,001–\$20,000, and more than \$20,000. We coded these as 1–9 and then standardized by subtracting the mean and dividing by two standard deviations (Gelman 2006).

[11] We are also exploring multinomial logit and probit models, which involve some technical challenges since we are fitting these models in a multilevel context. We do not expect the multinomial discrete models to change the substantive findings.

[12] We include the reason for voting because of its relevance in determining vote choice, as mentioned in Section 1.

(corresponding to richer voters being more likely to support the PAN candidate) and negative slopes for the PRD model. We summarize the models by plotting the curves $\alpha_j + \beta_j x$ (for the linear model) and $\text{logit}^{-1}(\alpha_j + \beta_j x)$ (for the logistic models) for each of the 32 states, and by plotting the estimated intercepts α_j and estimated slopes β_j vs. u_j, the state-level GDP per capita.

We fit the models using Bugs (Spiegelhalter, Thomas, Best & Lunn 2003, Sturtz, Ligges & Gelman 2005) and the lmer function in R (R Development Core Team 2006, Bates 2005), following the approach of Gelman et al. (2005).

4 Results

4.1 Individuals within states

We first present the results of fitting the linear model (1) predicting vote choice (on a 1–3 scale) given individual income. Figure 3 shows the fitted lines from the multilevel model, with the states ordered from poorest to richest. The lines tend to be steeper in the poorer states: for example, the average slope in the five poorest (Chiapas, Oaxaca, Zacatecas, Guerrero, and Tlaxcala) is 0.40, and the average slope for the 5 richest (Chihuahua, Quintana Roo, Campeche, Nuevo Leon, and Mexico D.F.) is only 0.26.

Our next step is to add state inequality, gender, age, type of locality, the main reason for which voters say they voted, and religion. The coefficients for individual and state-level income show similar patterns as before, so for the remaining analyses we only use income and GDP per capita as predictors, since we are interested in studying the differences between the affluent and less affluent voters. Even if the effects of income and GDP per capita had been explained by other predictors, the correlations would still be real, in the sense of representing real differences between rich and poor voters, and rich and poor states.

Now, to ascertain if income matters more in poor states than in rich states just as in the U.S., we plot the estimated state intercepts and slopes as a function of the average state GDP per capita. To explore these results further, we display in Figure 4 the intercept α_j and the slope β_j for the 32 states including Mexico City,[13] plotted vs. state income. On average, richer states have higher intercepts and lower slopes than poor states (with the pattern especially clear outside of Mexico City, which is a clear outlier as the richest state, with voting patterns more typical of poorer areas).

[13] As noted earlier, we consider Mexico City as its own region in the multilevel model, thus allowing its intercept and slope to differ from what otherwise might be expected given its per-capita GDP.

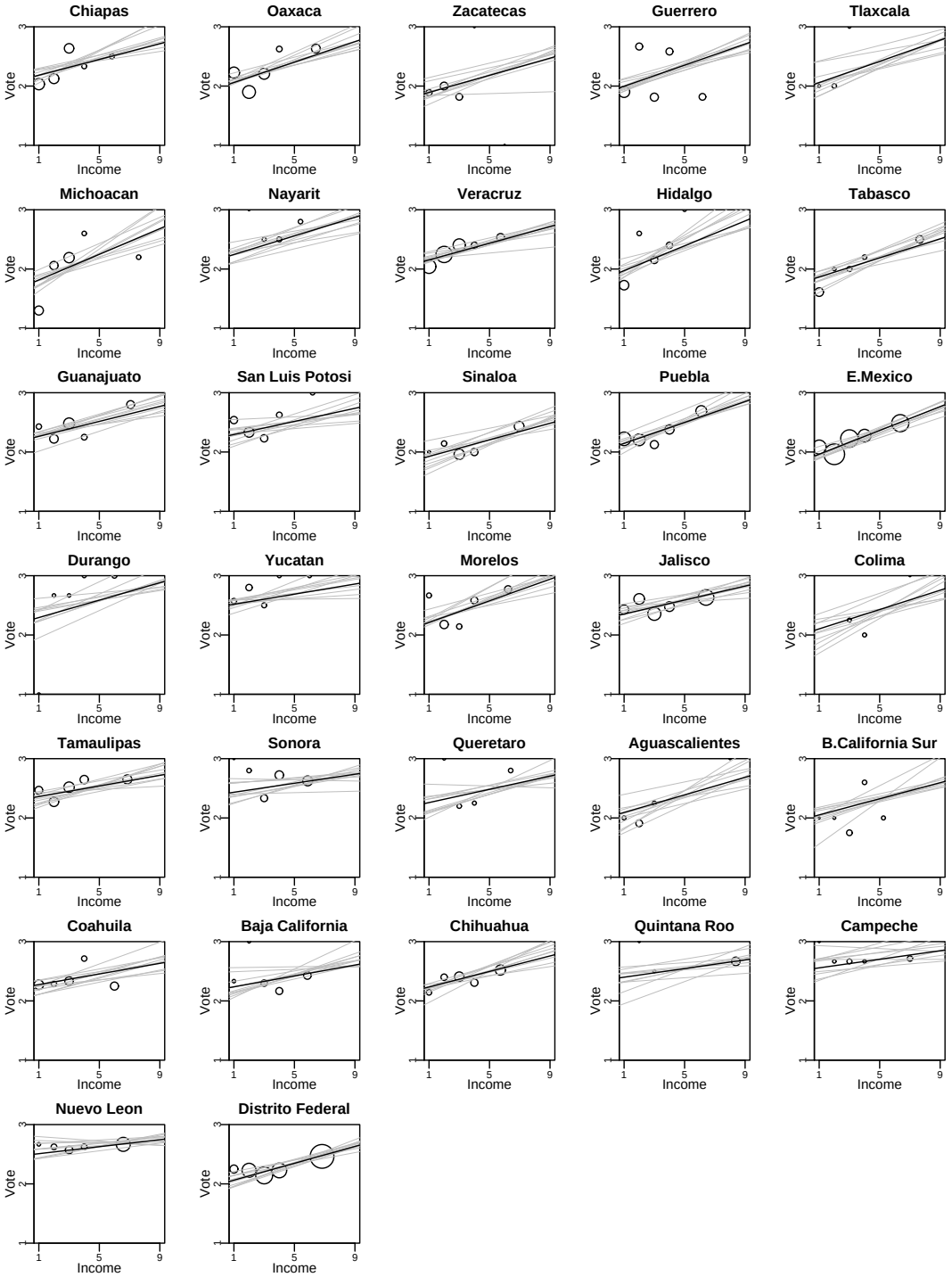

Figure 3: Estimated regression lines $y = \alpha_j + \beta_j x$ of expected vote choice (1=PRD, 2=PRI, 3=PAN) on income, for each of the 32 states j, ordered from poorest to richest state. In all states, income is positively correlated with voting for more conservative parties, with the relation between income and voting being strongest in the poorest states.

For each state, the dark and light lines show the estimated regression line (based on the posterior median of the coefficients) and uncertainty (posterior simulation draws). The circles show the relative proportion of individuals in each income category in the survey (with categories 5–9 combined into a single circle because of small sample sizes). The area of each circle is proportional to the number of respondents it represents.

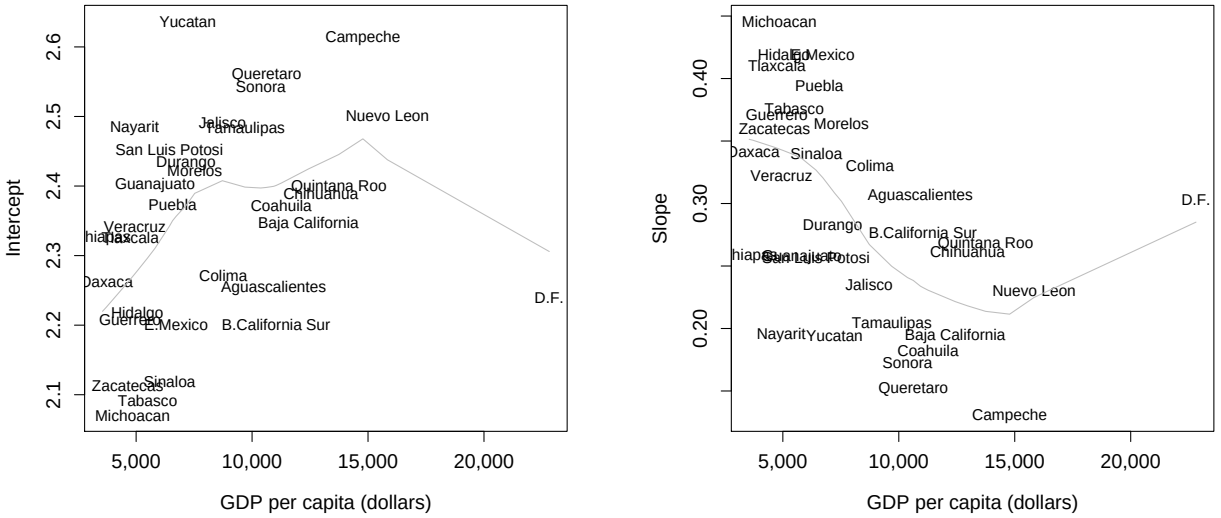

Figure 4: Estimated (a) intercepts α_j and (b) slopes β_j for the linear model, plotted vs. GDP per capita (scaled to 2000 U.S. dollars) for the 32 states including the capital (D.F.). Results for the individual states appear in Figure 3. The curves show lowess fits (Cleveland 1979). Intercepts tend to be higher and slopes tend to be lower in the richer states. When the model is fitted excluding D.F., the positive correlation between intercepts and per-capita GDP in the 31 states, and the negative correlation between slopes and per-capita GDP, are even stronger.

The higher intercepts (as shown in Figure 4a) tell us that a voter of average income is more likely to support the conservative candidate if he or she lives in a richer state. Thus, the differences between rich and poor states are not simply aggregates of differences in individual incomes.

Figure 4b shows that, similarly to the U.S. (see Figure 13 in Gelman et al. (2005)), income matters more in poorer states than in richer states. Poor voters in poorer states are expected to vote at higher rates for the PRD and PRI (the parties on the left) than poor voters in richer states. This can be seen in Figure 3, where, proportionally, more of the poor voters support the PRD and PRI in Chiapas, Oaxaca, Zacatecas, Guerrero, and Tlaxcala, than in Chihuahua, Quintana Roo, Campeche, Nuevo Leon, and Mexico City.

4.2 PAN or PRD?

Considering all three major parties, we have found a positive correlation between income and conservative voting. Moreover, this relationship is, on average, stronger in poorer states than in richer states. We now consider the logistic regressions, first considering the rightmost major party (PAN) compared to all others, then the leftmost (PRD) compared to all others. Figure 5 shows the estimated slopes for the 32 states, plotted vs. GDP per capita, for each of the two logistic

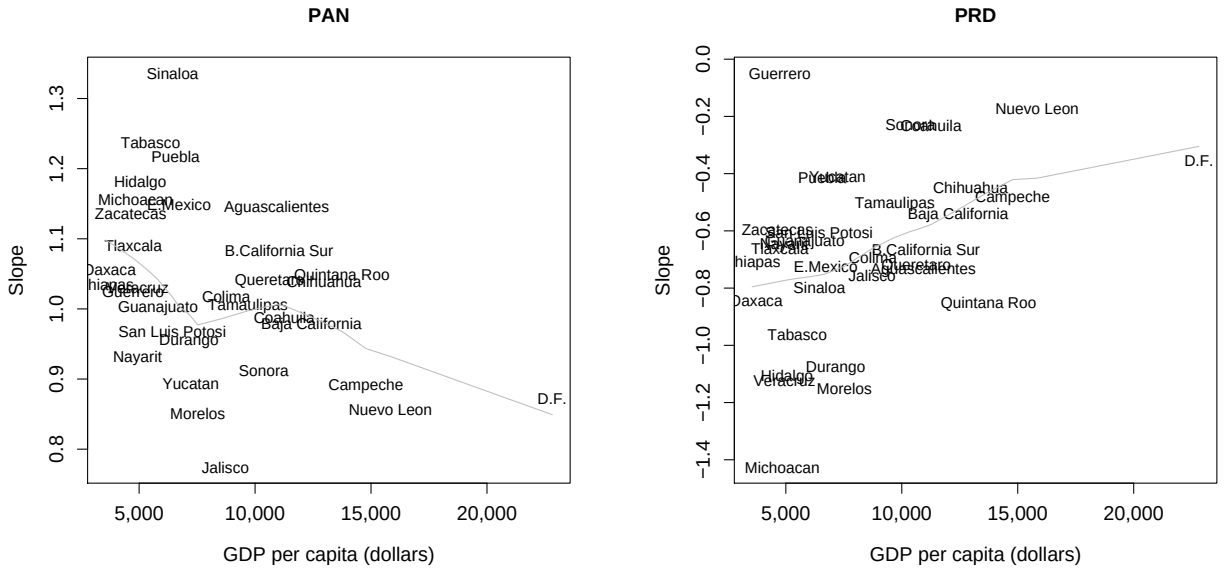

Figure 5: Estimated slopes β_j from the logistic regression models of income predicting vote choice: (a) for the PAN compared to all other parties, (b) for the PRD compared to all other parties. (a) Slopes are very positive—within any state, richer voters are *more* likely to support the PAN—and the slopes for the richer states are closer to zero, so that income is most strongly predictive of vote choice in the poorer states. (b) Slopes are negative—within any state, richer voters are *less* likely to support the PRD—and the slopes for the richer states are closer to zero; again, income is most strongly predictive of vote choice in the poorer states.

regressions. Once again, we see that income matters more in poorer states than in richer states, with quite a bit of variation between states in the role of income in predicting the vote.

The slopes are positive in Figure 5a and negative in Figure 5b, which makes sense given the opposite orientations of the PAN and PRD. In addition, the absolute levels of the slope are much higher for the PAN model (compare the vertical axes of the two graphs in Figure 5), indicating that income is a stronger predictor of PAN vote than PRD vote. This is a subtlety of the multiparty system, in which the three major parties are not aligned on a single dimension.

5 Discussion

We have found the following patterns:

1. Rich states tend to support the conservative party (the PAN) at higher rates than poor states, an opposite pattern from that found in the United States. There are no definite or absolute regional partisan strongholds; that is, there is more variation between states and within regions than what current literature may suggest.

2. In all states, the PAN does better among higher-income voters, but poor voters in richer states tend to support the PAN at higher rates than poor voters in poorer states. That is, income is less important as a predictor in rich states than in poor states.

How can we understand these patterns at the individual and state levels? One plausible explanation has to do with each state's social structure. Those states with more conservative structures are going to be those that, on average, tend to support the PAN at higher rates than those states with less conservative social structures. For instance, historically the PAN has had a close identification with the Catholic church and with its social Christian message. The PAN has done better in those *municipios*[14] with a higher percentage of Catholics than in those *municipios* with lower concentrations of Catholics (Moreno 2003).

In a similar vein, in terms of GDP per capita, poorer states on average tend to be more rural and slightly more conservative than richer states. Individuals in poorer states, especially the wealthy, may be more conservative on average than those wealthy individuals in rich states who may be less conservative and more cosmopolitan; hence, the PAN electoral platform may be more appealing for rich voters living in poorer states than for rich voters living in richer states.

Overall, our analysis indicates that richer states tend to support the PAN candidate at higher rates than poorer states. However, by applying multilevel modeling techniques we were able to show that there is much more variation between states than when they are collapsed in large regions. Moreover, while income is positively related with the PAN vote (as previous analyses have shown), its impact seems to be stronger within poorer states than within medium and richer states.

References

Alemán, Ricardo. 2006. "Itinerario politico: no gana pero arrebata." *El Universal* 3 July: first section.

Bates, Douglas. 2005. "Fitting linear models in R using the lme4 package." *R News* 5(1):27–30.

Becerra, Ricardo, Pedro Salazar & Jose Woldenberg. 2000. *La Mecanica del Cambio Politico en Mexico*. Mexico, D.F.: Cal y Arena.

Cleveland, William S. 1979. "Robust locally weighted regression and smoothing scatterplots." *Journal of the American Statistical Association* 74:829–836.

[14] A *municipio* is the equivalent of a county in the United States.

Dominguez, Jorge I. & Chappell Lawson, eds. 2004. *Mexico's Pivotal Democratic Election*. Stanford and La Jolla: Stanford University Press and Center for U.S.-Mexican Studies, UCSD.

Dominguez, Jorge I. & James A. McCann. 1996. *Democratizing Mexico: Public Opinion and Electoral Choices*. Baltimore: Johns Hopkins University Press.

Dresser, Denise. 2006. "Ocho lecciones." *Reforma* 3 July: 21.

Gelman, Andrew. 2006. Scaling regression inputs by dividing by two standard deviations. Technical report, Department of Political Science, Columbia University.

Gelman, Andrew, Boris Shor, Joseph Bafumi & David Park. 2005. Rich state, poor state, red state, blue state: what's the matter with Connecticut? Technical report, Department of Political Science, Columbia University.

Klesner, Joseph L. 1995. "The 1994 Mexican elections: manifestation of a divided society?" *Mexican Studies/Estudios Mexicanos* 11(1):137–149.

Klesner, Joseph L. 2004. The structure of the Mexican electorate. In *Mexico's Pivotal Democratic Election: Candidates, Voters, and the Presidential Campaign of 2000*, ed. Jorge I. Dominguez & Chappell Lawson. Stanford and La Jolla: Stanford University Press and Center for U.S.-Mexican Studies, UCSD pp. 91–122.

Linz, Juan J. & Alfred C. Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore: Johns Hopkins University Press.

Lujambio, Alonso. 1997. Estudio introductorio. In *Gobiernos Divididos en la Federacion Mexicana*, ed. Alonso Lujambio. Mexico, D.F.: Colegio Nacional de Ciencias Politicas y Administracion Publica, Universidad Autonoma Metropolitana, Instituto Federal Electoral.

McKinley, James C., Jr. 2006. "Mexico faces its own red-blue standoff." *New York Times* 9 July: Week in Review, 3.

Moreno, Alejandro. 2003. *El Votante Mexicano: Democracia, Actitudes Politicas y Conducta Electroal*. Mexico, D.F.: Fondo de Cultura Economica.

Ochoa-Reza, Enrique. 2004. Multiple arenas of struggle: federalism and Mexico's transition to democracy. In *Federalism and Democracy in Latin America*, ed. Edward Gibson. Baltimore: Johns Hopkins University Press.

PNUD, Programa de las Naciones Unidas para el Desarrollo. 2003. Informe sobre desarrollo humano: Mexico 2002. Report of Organizacion de las Naciones Unidas.

R Development Core Team. 2006. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, `www.r-project.org`.

Reforma. 2006. "Pintan en dos a la Republica." *Reforma* 3 July.

Reforma, El Norte & Mural. 2000. *Exit Poll Nacional: Elecciones Federales del 2 de Julio de 2000*. Banco de Datos del Instituto Tecnologico Autonomo de Mexico (ITAM).

Snijders, Tom A. B. & Roel J. Bosker. 1999. *Multilevel Analysis*. London: Sage.

Spiegelhalter, David, Andrew Thomas, Nicky Best & Dave Lunn. 2003. *WinBUGS (Bayesian Inference Using Gibbs Sampling)*. MRC Biostatistics Unit, `www.mrc-bsu.cam.ac.uk/bugs`.

Sturtz, Sibylle, Uwe Ligges & Andrew Gelman. 2005. "R2WinBUGS: a package for running WinBUGS from R." *Journal of Statistical Software* 12(3).