|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Online Appendices**  **Table A1: Descriptive Statistics\*** | | | | | | |
|  | Full Dataset  N = 2,056 | | Obs. | Restricted Dataset  N = 1,429 + | | Obs. |
| Democrat on Ballot | | 75% | 1,492 | 80% | | 1,079 |
| Republican on Ballot | | 79% | 1,574 | 82% | | 1,116 |
| Black Candidate | | 16% | 308 | 14% | | 183 |
| Black Democrat | | 15% | 294 | 14% | | 172 |
| Black Republican | | 1% | 25 | 1% | | 20 |
| Black Incumbent | | 10% | 197 | 8% | | 111 |
| Latino Candidate | | 12% | 243 | 16% | | 219 |
| Latino Democrat | | 10% | 192 | 13% | | 175 |
| Latino Republican | | 4% | 86 | 6% | | 78 |
| Latino Incumbent | | 7% | 135 | 8% | | 120 |
| Upper Chamber | | 23% |  | 24% | |  |
| Incumbent | | 76% |  | 72% | |  |
| **District Population in Restricted Dataset** | | **Median** | **Mean** | **Std. Dev.** | **Min.** | **Max.** |
| White CVAP | | 73% | 66% | 24% | 1% | 99% |
| Latino CVAP | | 10% | 16% | 18% | 0% | 94% |
| Black CVAP | | 6% | 12% | 16% | 0% | 96% |
| Asian CVAP | | 2% | 3% | 6% | 0% | 52% |
| All Minority CVAP | | 27% | 34% | 25% | 2% | 99% |
| \* A number of districts have both a Latino Democrat and Latino Republican, or a black Democrat and black Republican on the ballot, accounting for the discrepancy between the partisan and racial categories. + We lose more than 600 districts when we include our measure of district partisanship in the regression results, which explains the differences in the N between the two columns. | | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table A2: Where Are Primary Candidates Running (TX, CA)?** | | | |
|  | White VAP | Latino VAP | Afr.-Am. VAP |
| White Candidates | 57% | 25% | 9% |
| Latino Candidates | 27% | 60% | 5% |
| African-American  Candidates | 27% | 36% | 27% |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table A3: Coarsened Exact Matching for Black Dems** | | | | |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| *Initial Imbalance* | L1 | Mean | Min. | 25% | 50% | 75% | Max. |
| *L1* = .77 |  |  |  |  |  |  |  |
| **Black CVAP** | 0.71 | 0.34 | 0.02 | 0.22 | 0.45 | 0.46 | 0.18 |
| **% Dem. partisanship** | 0.52 | 0.19 | 0.08 | 0.21 | 0.26 | 0.19 | 0.03 |
| **South** | 0.24 | 0.24 | 0 | 0 | 1 | 0 | 0 |
|  |  |  |  |  |  |  |  |
| *After Matching* | L1 | Mean | Min. | 25% | 50% | 75% | Max. |
| *L1* = .32 |  |  |  |  |  |  |  |
| **Black CVAP** | 0.12 | 0.00 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 |
| **% Dem. partisanship** | 0.05 | 0.00 | 0.01 | 0.01 | -0.01 | 0.00 | 0.01 |
| **South** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

|  |  |  |
| --- | --- | --- |
| Number of strata: 96 | | |
| Number of matched strata: 41 | | |
|  |  |  |
|  | No Black Dem | Black Dem |
| All | 907 | 172 |
| Matched | 544 | 93 |
| Unmatched | 363 | 79 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table A4: Coarsened Exact Matching for Latino Dems** | | | | |  |  |  |
|  |  |  |  |  |  |  |  |
| *Initial Imbalance* | L1 | Mean | Min. | 25% | 50% | 75% | Max. |
| *L1* = .77 |  |  |  |  |  |  |  |
| **Latino CVAP** | 0.72 | 0.36 | 0.01 | 0.32 | 0.42 | 0.44 | 0.11 |
| **% Dem. partisanship** | 0.37 | 0.10 | 0.07 | 0.14 | 0.15 | 0.10 | -0.03 |
|  |  |  |  |  |  |  |  |
| *After Matching* | L1 | Mean | Min. | 25% | 50% | 75% | Max. |
| *L1* = .40 |  |  |  |  |  |  |  |
| **Latino CVAP** | 0.10 | 0.00 | 0.01 | 0.00 | 0.00 | -0.01 | 0.02 |
| **% Dem. partisanship** | 0.15 | 0.00 | -0.01 | 0.01 | 0.01 | 0.01 | -0.03 |

|  |  |  |
| --- | --- | --- |
| Number of strata: 79 | | |
| Number of matched strata: 43 | | |
|  | No Latino Dem | Latino Dem |
| All | 904 | 175 |
| Matched | 450 | 124 |
| Unmatched | 454 | 51 |

|  |  |  |
| --- | --- | --- |
| **Table A5: Are Minority Democrats Higher “Quality” Candidates?** | | |
| In all districts | Avg. Total Receipts  (Std. Error) | Democrat Winners Only  (Std. Error) |
| Black Democrats (N= 222) | 126k (12k)\* | 138k (13k) |
| Latino Democrats (N = 183) | 254k (28k) | 259k (30k) |
| White Democrats (N = 813) | 196k (13k) | 280k (20k) |
|  |  |  |
| In majority White districts with no incumbent | | |
| Black Democrats (N = 17) | 126k (58k) | 204k (104k) |
| Latino Democrats (N = 14) | 140k (55k) | 146k (70k) |
| White Democrats (N = 193) | 201k (25k) | 329k (44k) |

\* Denotes a statistically significant difference from White Democrats at the 95% confidence

level. The data are from the DIME dataset (Bonica 2013) and cover roughly 70% of our

sample due to missing values from the DIME dataset.

****

****

Predicted probabilities for Latino and white Democrats in “competitive districts” after matching on Latino candidacy. “Competitive districts” are those that do not have an incumbent running and have at least one Democrat and one Republican on the ballot.

**Figure A3: Figure 3 with uncertainty estimates.**

****

This is a reproduction of the predictions of a Democratic victory from figure three in the

paper but uncertainty estimates have been added. These 95% confidence intervals are

informative, but the first differences are a better indicator of significant differences across

white and minority partisans and are provided in the paper in figures four and five.

That is, even though these confidence intervals overlap in places, we cannot be sure that the predictions are *not* significantly different from one another (a "true null") unless we compute

the first differences with error. The error estimates are provided here for readers who might find them interesting.

**Appendix B: A Methodological Issue: Coding Race and Ethnicity**

To be sure, one of the reasons why scholars to date have done little to examine the supply side of minority representation has been due to the absence of candidate-level data. Two recent databases – SLER and LEAP – have begun to solve this problem, but the issue of coding race of candidates often remains. Outside of a few southern states required by the VRA to capture racial and ethnic background of candidates (Alabama, Florida, Georgia, Louisiana, North Carolina, South Carolina, and Pennsylvania) few options exist for scholars interested in obtaining racial identifiers for candidates.

Coding race of candidates is complicated for a number of reasons, the first of which is race is a subjective, conceptual concept, and therefore open to multiple interpretations (Omi & Winant 1994). As others have noted, this often means that in practice, a person’s racial identification depends on context, who is identifying the race (respondent or “other”), and the way the question is asked (see Masuoka 2011). Given these complications, social scientists interested in studying race are left with a number of options, each with a beset with likely error.

One possibility is using Census surname lists, which are generated by the Census after each iteration of their long survey (see e.g. Barreto, Segura & Woods 2004; Michelson 2003). The Census is able to generate probabilities of last names mapping to specific races using their self-identified racial data, but a number of studies have shown that for each group, the probabilities are less accurate than previously assumed due to multi-racial identities and the practice of changing surnames after marriage.

Another possibility is using geocoding, segregation data, and name frequencies to yield a probabilistic estimate of an individual’s area (see e.g. Enos 2010; Fraga 2014). This method is particularly useful when estimated individual voters in a particular place, but more problematic for candidate race coding, since we often do not have candidate’s home address, and they often represent a large area.

In this project, we use expert coding along with self-identification. Using candidate websites, Facebook pages, newspaper articles or videos, we code candidate race/ethnicity based on surnames, pictures and biographical information. In addition, the coding for Latinos was aided by the National Association of Latino Elected Officials (NALEO), which provides a pre-election list of Latino candidates (NALEO 2012). If there was uncertainty about the candidate’s race or ethnicity, the authors used news accounts, background information or any other piece of information available. We did not code someone as Latino or African-American unless there was clear and near certain evidence that the person belonged to that group. For example, Rick G. Perales (R) in Ohio’s 73rd House district was a candidate whose background was scrutinized because of his surname. While we are fairly confident that Mr. Perales has a Latino ethnic background, he is never described as such, nor does he self-identify as Latino in his campaign material. He does not belong to any Latino political organizations and he is not included on NALEO’s list of Latino candidates. Thus he is coded as “white non-Latino” in order to avoid a false positive. This coding rule works against our hypothesis of finding successful minority candidates in white districts, as some candidates like Mr. Perales, who won in a district that is 87% white VAP, does not get counted as a Latino winning in a white district. We also worked directly with NALEO to contact a couple of extremely tough cases. NALEO was able to verify, via of self-identification, these very tough cases, for example Ray Rodrigues in FL HD 76 (who does not identify as Latino).

**Works Cited**

Barreto, Matt A., Gary M. Segura, and Nathan D. Woods. 2004. "The mobilizing effect of majority-minority districts on Latino turnout." *American Political Science Review* 98.1: 65-75.

Enos, Ryan D. 2010. "What tearing down public housing projects teaches us about the effect of racial threat on political participation." *Unpublished. Retrieved from http://ryandenos.com/papers/chicago\_threat. pdf* .

Masuoka, Natalie. 2011. "The “Multiracial” Option: Social Group Identity and Changing Patterns of Racial Categorization." *American Politics Research* 39.1: 176-204.

Michelson, Melissa R. 2003. "Getting out the Latino vote: How door-to-door canvassing influences voter turnout in rural central California." *Political Behavior* 25.3: 247-263.