Supporting Information for **Geographic Boundaries and Local Economic Conditions Matter for Views of the Economy**

Contents

1 Data Description

1.1 Individual-Level Data (Gallup)

Daily Gallup data consist of approximately 1,000 respondents surveyed per day, ∼350 days per year. We obtained these data from [REDACTED] library, which licensed the data from Gallup for university researchers, with every respondent (or almost every respondent – see below) geo-coded to their ZIP code of residence. Days without 1,000 respondents include major holidays (President's Day, Memorial Day, July 4th, Labor Day, Thanksgiving, December 23rd through the 25th, December 29th through January 1st), as well as some inconsistent additional days. There doesn't appear to be systematic missingness connected to the economy, as illustrated in Figure 1 by the constant annual samples of roughly 350,000 respondents over the course of the Global Financial Crisis and Great Recession (2009 - 2011). Starting in the final months of 2017, we have increasing missingness due to the acquisition schedule of the [REDACTED] library.

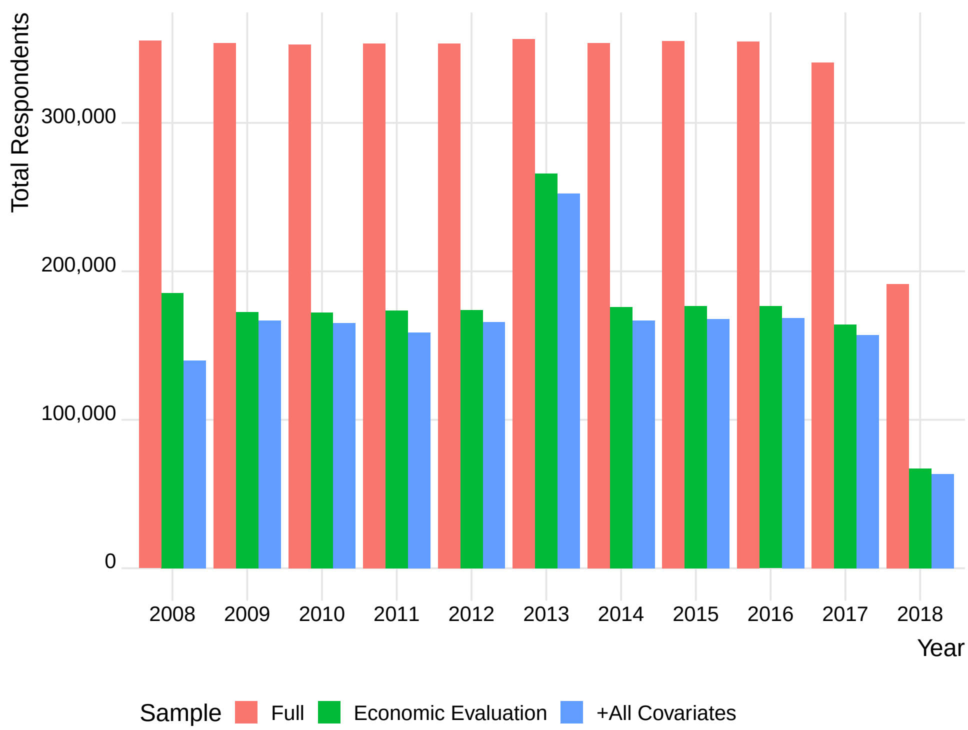

Figure 1: Total respondents (y-axis) by year (x-axis) for the full data (red), those that answered the economic evaluation question (green bar), and those with covariate information for race, age, gender, marital status, educational attainment, household income, and party affiliation (blue bar).

Our manuscript focuses narrowly on a single outcome measure of an individual's evaluation of the economy. Specifically, we use the consistently asked M30 question, which asks respondents to indicate their assessment of current economic conditions. Options include poor, only fair, good, and excellent. We drop respondents who indicated that they don't know, as well as those that weren't asked this question (approximately half of the respondents, with the exception of 2013, as illustrated in Figure 1), reducing annual samples from $\sim$ 350,000 to roughly 175,000 per year. Our main analyses treat this outcome as a continuous measure, but our results are robust to dichotomizing responses into $0 = \{\text{poor, only fair}\}$ and $1 = \{\text{good, excellent}\}$. The shares of respondents indicating each of these choices by year is visualized in Figure 2, revealing generally dismal evaluations until 2011, and increasingly positive evaluations starting in 2017. The last two years of our sample are less reliable due to missingness.

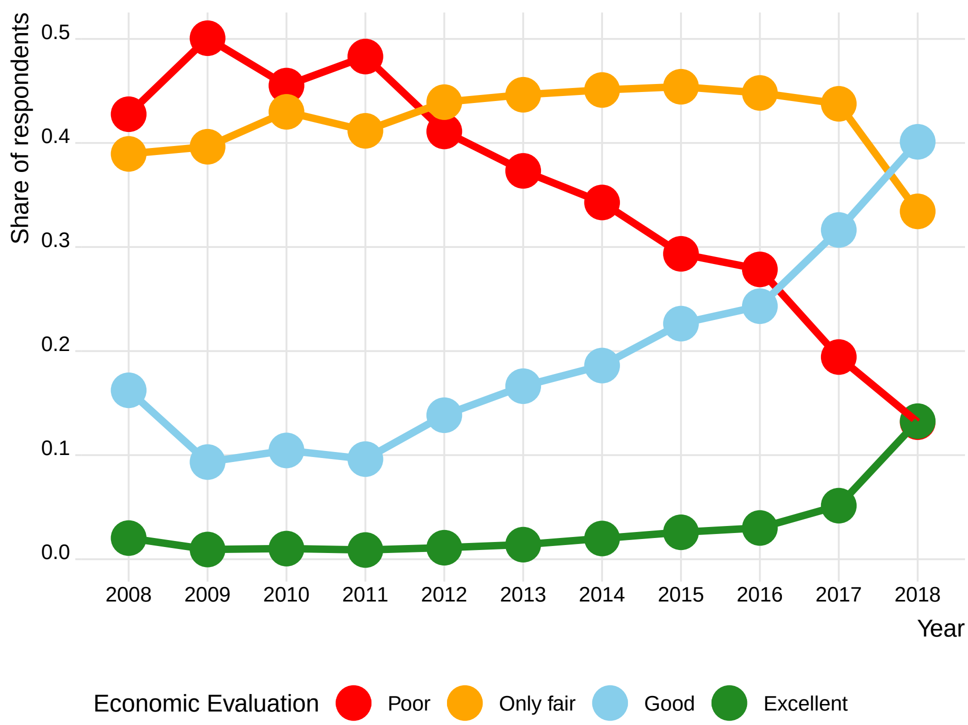

Figure 2: Share of total respondents (y-axis) indicating one of four economic evaluations (poor = red, only fair = orange, good = blue, excellent = green) by year (x-axis).

We are also interested in comparing the predictive power of contextual measures of the economy, aggregated to different geographic units, conditioning on a standard set of continuously recorded demographic covariates. These include the respondent's age, race, gender, marital status, educational attainment, household income, and party affiliation. We bin these covariates as follows:

- Age (6 bins): 18-24, 25-34, 35-44, 45 - 54, 55 - 64, and 65+
- Race/Ethnicity (5 bins): White, Black, Asian, Hispanic, Other (these are the categories provided by Gallup)
- Gender (2 bins): Male, female (these are the categories provided by Gallup)

- Martial status (6 bins): Single / never been married, Married, Separated, Divorced, Widowed, Domestic partnership / Living with partner (not legally married)
- Educational attainment (9 bins): Less than high school diploma, High school degree or diploma, Technical/vocational school, Some college but no degree, Two year associated degree, Four year bachelor's degree, Some post graduate work or schooling but no degree, Postgraduate or professional degree
- Household income (12 categories): Less than \$720, \$720 to \$5,999, \$6,000 to \$11,999, \$12,000 to \$23,999, \$24,000 to \$35,999, \$36,000 to \$47,999, \$48,000 to \$59,999, \$60,000 to \$89,999, \$90,000 to \$119,999, \$120,000 or more, don't know, refused
- Party affiliation (6 values): Republican, Lean Republican, Independent (no lean), Lean Democrat, Democrat, Refused

We retain respondents who responded with don't know or refused for the income and party affiliation categories in order to allow for the strongest test of individual-level covariates that might affect prediction of economic evaluations. Summary statistics of our coverage across race, age, education, and party affiliation is summarized in Figure 3.

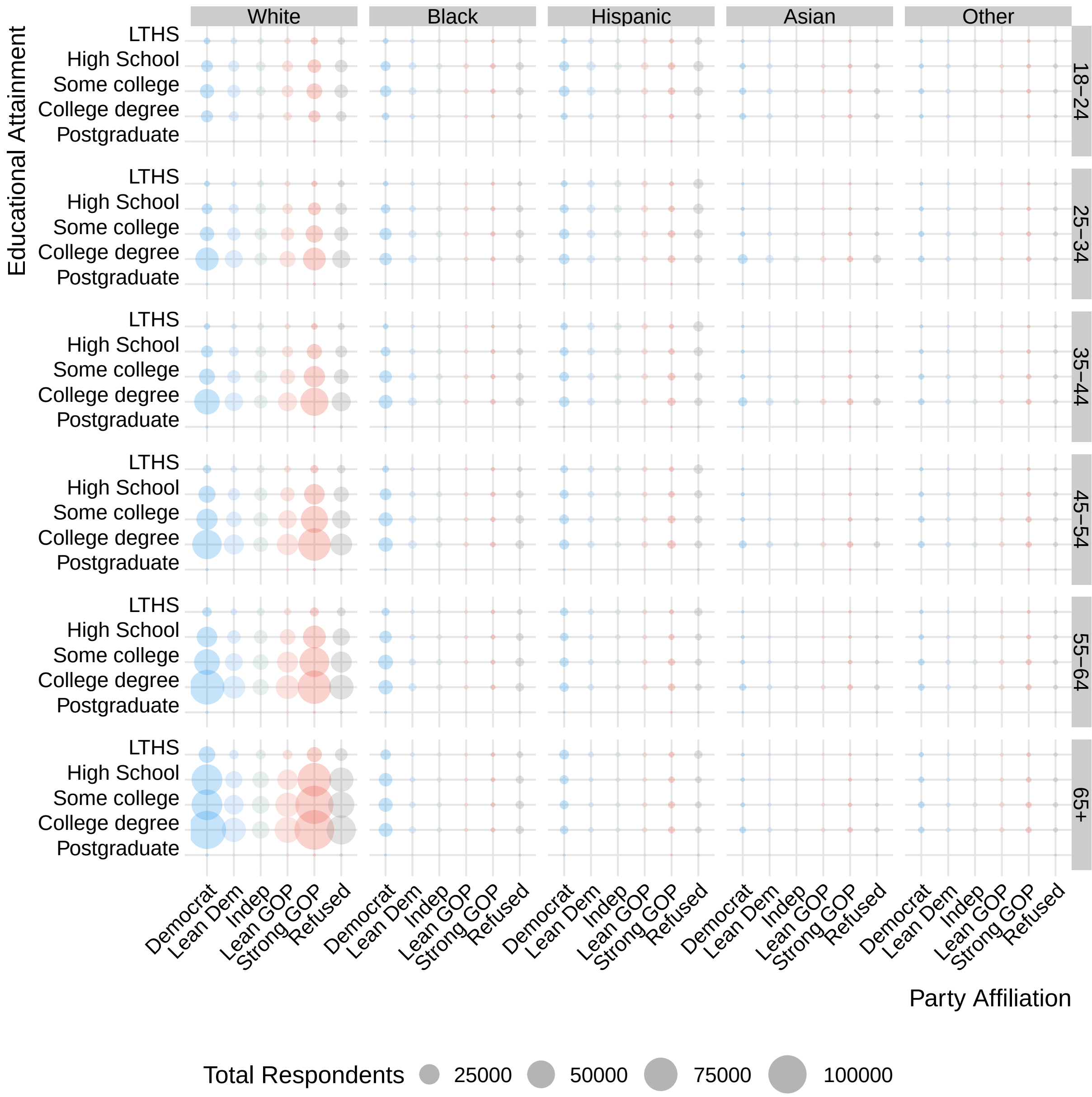

Figure 3: Total respondents by party affiliation (x-axes), educational attainment (y-axes), race (columns), and age group (rows). Educational attainment bins compressed for visual clarity. Points sized by the total number of respondents in the full data associated with each group.

1.2 Partisan composition of the sample

Respondents were asked two questions pertaining to their partisanship, allowing us to separate partisans from partisan leaners. Such a separation is useful because it allows us not to classify leaners as Independents.

The questions on the survey are

- "In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?"
- "As of today, do you lean more to the Democratic Party or the Republican Party?"

As we show in the table below, fewer than 10% of Gallup respondents do not identify with, or lean toward, either of the major parties.

By comparison, in the 2016 wave of the ANES, 13.6% of respondents identified as pure Independents.

Response	Proportion of respondents (%)	Number of respondents
Republican	30.84	429,947
Republican leaners	13.33	185,805
Independents	9.84	137,186
Democratic leaners	13.42	187,122
Democrats	32.56	453,850

Table 1: Party identification of Gallup respondents (among respondents who answered the prompt about the state of the economy

1.3 Geographic Data

Our manuscript uses 7 distinct geographic units at which to aggregate contextual measures. These range from ZIP codes at the smallest, to Census regions at the largest. We obtain daily Gallup data geocoded to the ZIP code level, and we calculate the latitude and longitude of these zip centroids for several of the crosswalks that follow. In aggregating to different units, we rely on different approaches, enumerated below.

ZIP codes

To calculate the latitude and longitude of each ZIP code tabulation area's centroid (ZCTA), we reply on gazetteer files, made publicly available by the Census at www.census.gov/

geographies/reference-files/time-series/geo/gazetteer-files.html. These sources cover the years 2010, 2012-2018. We rely on the 2010 files for 2008-2009, and 2011. We argue that mistakes in the coding at this lowest level introduce measurement error that would make our results weaker at the ZIP code level, making any attenuation bias conservative.

Over this period, there are 24 ZIP codes that do not appear in every year. In addition, there are 3,690 ZIP codes from the Gallup surveys that aren't listed in the gazetteer files, comprising 21,383 total respondents over the decade we analyze.[1] We fill in these missing ZIP codes via three additional datasets.

First, we scrape the zipdatamaps.com/ webpage for each of these 3,690 missing ZIP codes. This reduces the number of missing codes to 748.

Second, we obtain a different crosswalk file from CivicSpace Labs, obtained from public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude. This crosswalk combines the 1990 and 2000 gazetteer files with supplemental sources to provide the most comprehensive coverage of ZIP codes in the United States at the turn of the century. With this crosswalk, the number of unmatched ZIP codes falls to 7. These final 7 ZIP codes are PO boxes and are added manually from zipdatamaps.com/, albeit without population data.

The improvement using out-of-date data likely reflects the self-reported nature of the Gallup respondents who may not be aware of changes to their ZIP code. The downside of relying on this source is that we don't have population data for 741 ZIP codes.

Counties

The second-smallest unit of geography is the county. County FIPS codes are recorded in the Gallup data for all but 10,721 respondents, including several thousand respondents for whom we don't record zip-level data. For those respondents with a county identifier but no ZIP code (68,589 respondents), we assign them to the latitude and longitude of their county's centroid. For those without a county identifier and with a ZIP code (27 respondents), we use a zip-to-county crosswalk file provided by the Department of Housing and Urban Development (HUD) who provide such crosswalks for every quarter between 2010 and 2020 on their website www.huduser.gov/portal/datasets/usps_crosswalk.html. Of the remaining 10k respondents without a county FIPS code, all but one are from the 2015 Gallup data, suggesting that there isn't systematic missingness over space.

Commuting Zones

Commuting zones are geographic units defined by commuting patterns. Specifically, Tolbert Charles and Sizer (1996) used "journey to work" data collected by the Census Bureau

[1]There are also 79,293 respondents who are not given a ZIP code.

to calculate county-to-county travel patterns. They defined commuting zones as groups of counties with a high degree of within-unit travel and a low degree of across-unit travel. We obtain a crosswalk file for these commuting zones from the U.S. Department of Agriculture (data.nal.usda.gov/dataset/commuting-zones-and-labor-market-areas) and then assign survey respondents to their commuting zone based on their county of residence.

Congressional Districts

To assign respondents to their Congressional District of residence, we rely on shapefiles provided by (Lewis et al., 2013) at cdmaps.polisci.ucla.edu/. We use the latitude and longitude of the respondent to project them into the polygon using the `over` function from the `sp` package for `R`. We proceed year by year, assigning individuals to their congressional district based on the borders determined for the appropriate Congress.

Designated Market Areas

Designated market areas (DMAs) describe regions in the United States where residents receive similar broadcasts on television and radio. We rely on two sources of matching ZIP codes to DMAs. The first uses the crosswalk files provided by Sood (2018) (dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IVXEHT&version=7.3) to match ZIP codes with DMAs. There are 108 ZIP codes that don't match the DMA data, as well as those 79,293 respondents without a ZIP code. For these, we use the 2008 shapefile provided by Sood (2018) and project respondents into the appropriate polygon based on their latitude and longitude, again using the `over` function from the `sp` package for `R`.

States and Census Regions

The last two units are straightforward. State identifiers are included in all rows of the Gallup data and we map these to their Census division according to https://en.wikipedia.org/wiki/List_of_regions_of_the_United_States#/media/File:Census_Regions_and_Division_of_the_United_States.svg.

1.4 Contextual Measures & Distributions of Economic Outcomes

The preceding description focused on how we assign Gallup survey respondents to the appropriate geographic units of interest, ranging from ZIP codes up to census regions. For the contextual measures, the process is largely identical, starting from zip-level IRS tax return data, obtained from https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi. However, these data are non-randomly missing at the smallest level, which bears some discussion.

Specifically, small or sparsely populated ZIP codes are likely to have missing data, or be missing completely from some years. Across the full period of analysis, there are 13,557 ZIP codes that are not consistently measured, or roughly a third of all ZIP codes in the United States. However, since these are sparsely populated areas, they correspond to only a 4.3% reduction in the total Gallup sample, or approximately 159,000 respondents. We drop these respondents across all our results, regardless of whether we are aggregating to the ZIP code (where they would be dropped anyway) or to larger geographic units at which we would have contextual measures. Doing so ensures that, while our analysis is unable to incorporate the least populated areas of the United States, it remains an apples-to-apples comparison when we compare the conclusions drawn with zip-level data compared to state-level data.

Our three main contextual measures of interest are calculated as follows. The raw IRS data includes a count of the number of tax returns filed for each zip in the prior year by AGI bin, as well as an estimate of the total aggregate gross income included in these filings, and the same for the unemployment compensation. There are seven AGI bins per ZIP code, corresponding to:

- No AGI Stub
- \$1 to less than \$25,000
- \$25,000 to less than \$50,000
- \$50,000 to less than \$75,000
- \$75,000 to less than \$100,000
- \$100,000 to less than \$200,000
- \$200,000 or more

We calculate the empirical Gini coefficient by taking the cumulative proportion of filings in each AGI bin and the cumulative proportion of amounts in each bin to construct an empirical Lorenz curve. An example of these curves for Washington DC's ZIP codes in 2009 is depicted in Figure 4. We approximate the area under the curve with the following equation:

$$G_z = \frac{B+1}{B} - \sum_{b=1}^{B} (propFilings_b - propFilings_{b-1})(propAGI_b + propAGI_{b-1}) \qquad (1)$$

where B is the total number of bins, indexed by b and are arrange in increasing order.

For each ZIP code, we also calculate the AGI per return and the unemployment compensation per return. The former captures a measure of an area's wealth, while the latter captures an area's reliance on the social safety net. The former two measures are highly skewed, even after Winsorizing outliers beyond 99.9% of the data. As such, we log each, adding 1 to the unemployment per return measure as there are many ZIP codes without any such filings. We also calculate the proportion of the population filing as an additional contextual control. Descriptive statistics of these measures are presented in Figure 5

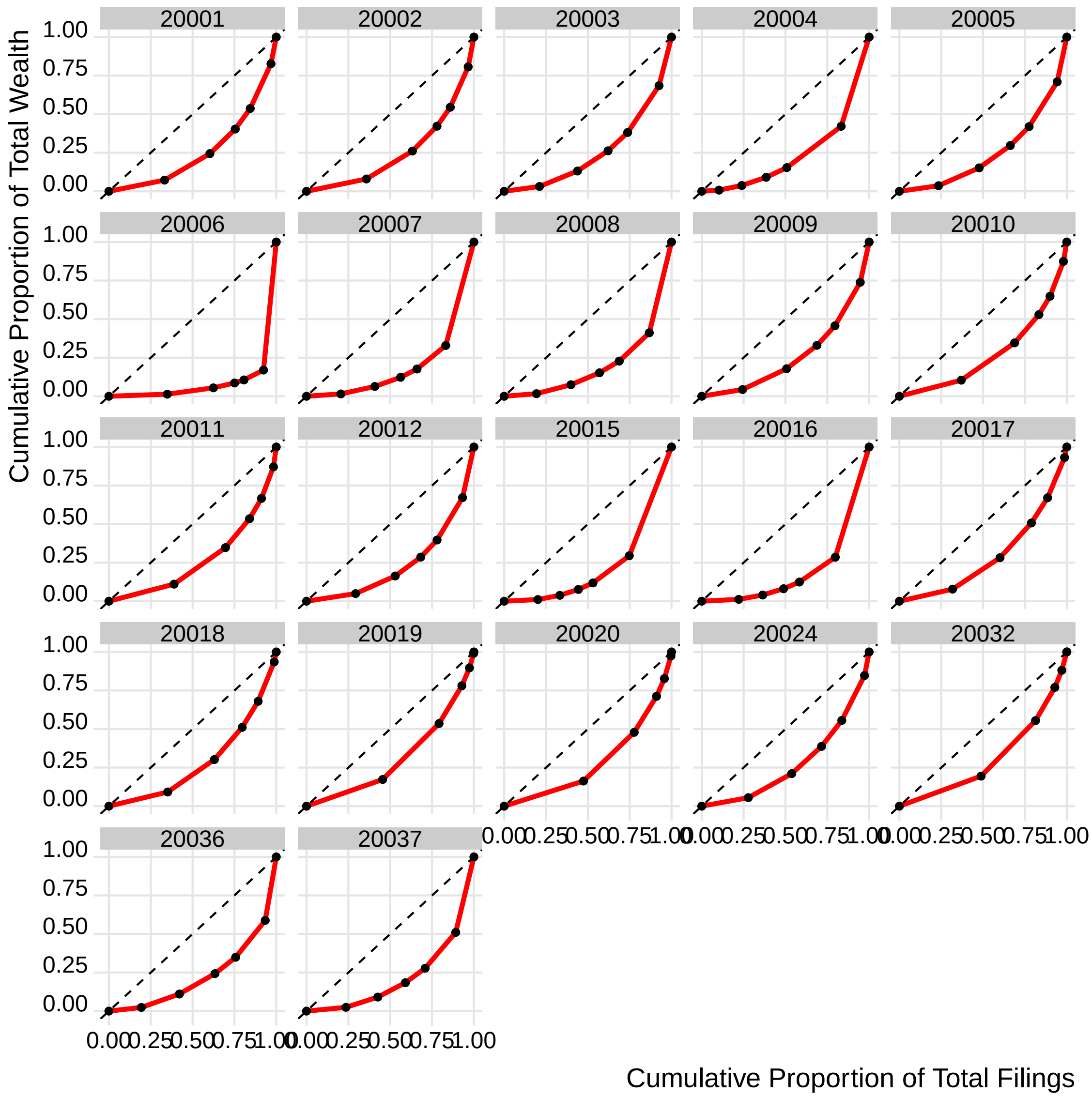

Figure 4: Empirical Lorenz curves for Washington DC's ZIP codes.

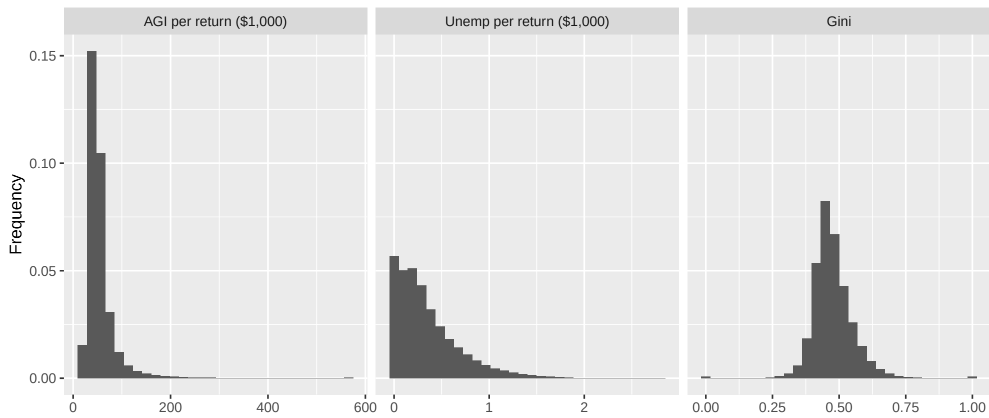

Figure 5: Distributions of contextual measures.

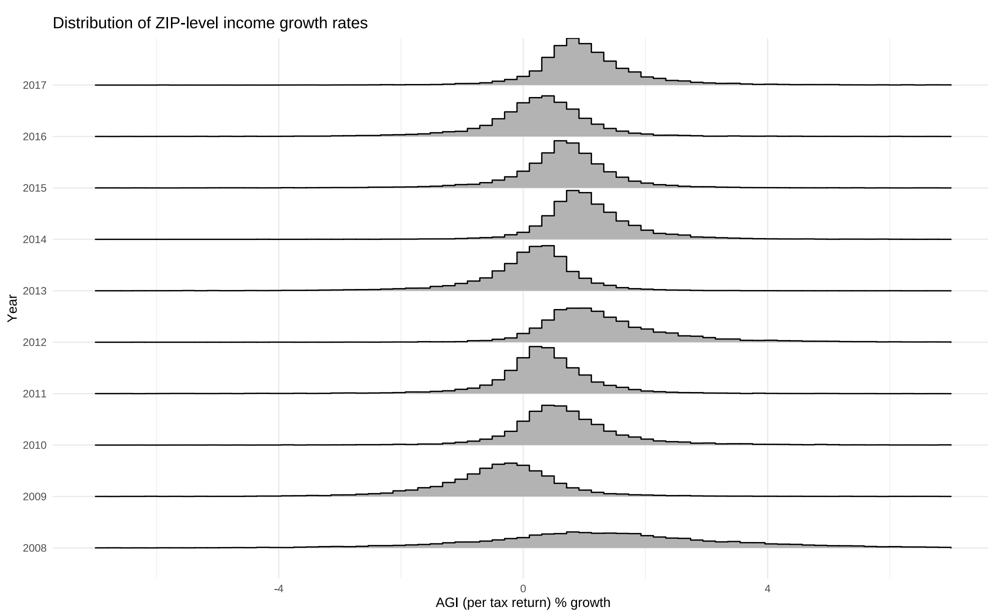

Figure 6: Annual income growth rates at the ZIP code level in each year of analysis.

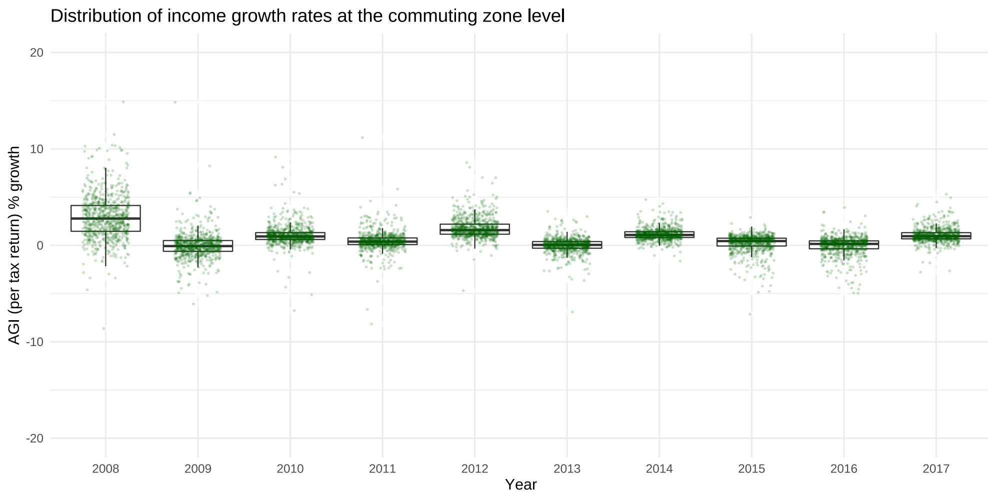

Figure 7: Annual income growth rates at the commuting zone level in each year of analysis.

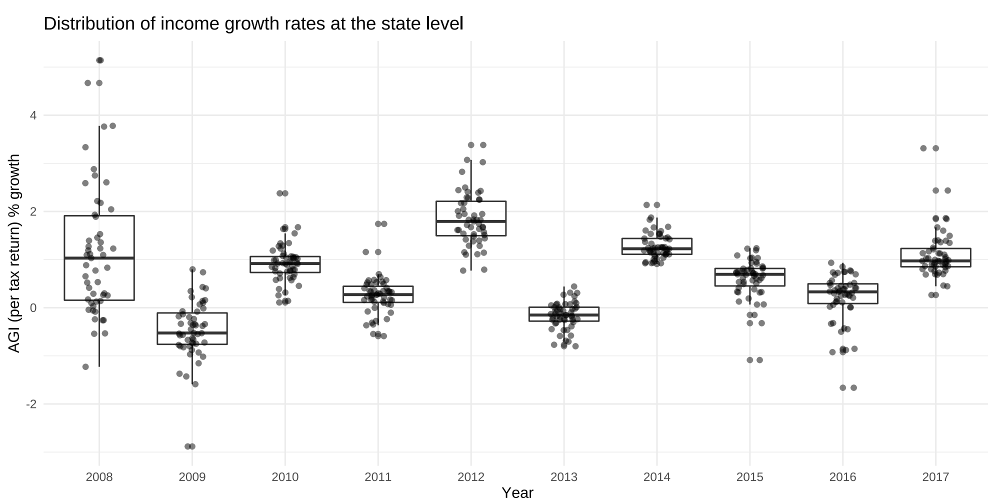

Figure 8: Annual income growth rates at the state level in each year of analysis.

2 Robustness

The following sections present supplementary analysis for the main findings. These range from different units of geographic aggregation for the variable importance densities, to different temporal transformations of the contextual variables, to hyper-parameter tuning for the random forests.

2.1 Variable Importance

Our main analysis examined which predictors were most important in predicting an individual's evaluation of the economy where contextual measures were aggregated to the commuting zone. In Figure 9 below, we replicate these results at the state level of aggregation, and compare robustness across different types of dummies. As illustrated, the choice of unit of aggregation only matters for contextual predictors and, to a much less significant degree, the most important partisan predictor (Democrat).

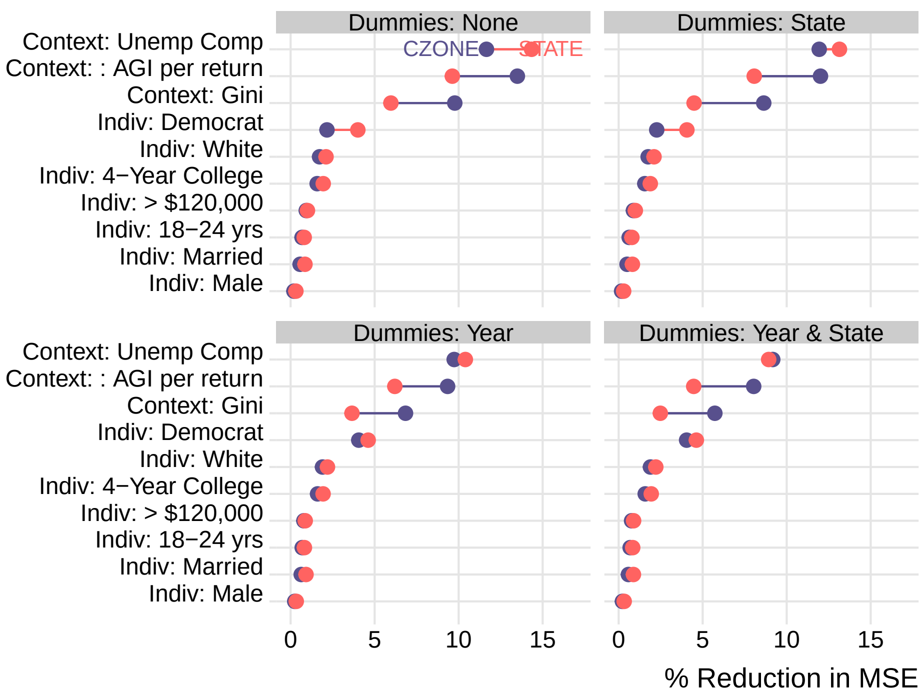

Figure 9: Variable importance of contextual and individual predictors aggregated to the commuting zone (purple) and the state (red). Only the most important categories among the individual-level predictors are shown. Dummies for state and year are omitted for clarity.

Figure 10 focuses on the contextual measures, aggregated across all available geographic units. Both figures highlight the substitution effect of larger units of aggregation when we don't include state fixed effects, particularly for unemployment compensation.

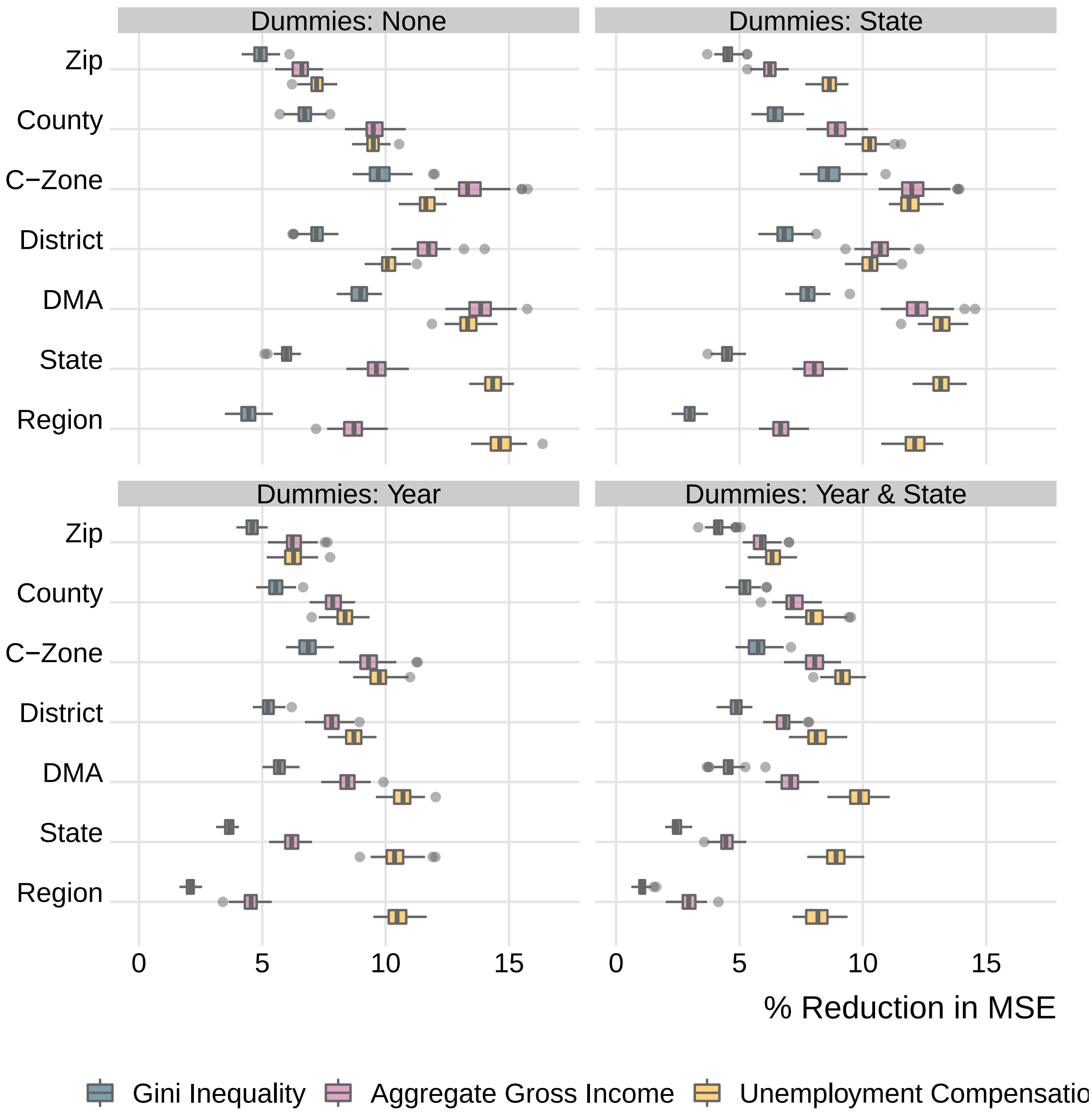

Figure 10: Variable importance of contextual predictors aggregated to different geographical units (y-axes).

2.2 Temporal Transformations

Our main analyses focused on the relationship between economic evaluations and the annual percent change in contextual predictors. Below, we examine the robustness of these results to alternative transformations, including:

- Lagged measures from the prior year and prior five years (lag1 & lag5)

- Growth rates over one and five years (pctChg1 & pctChg5)
- Annual change and change across five years (chg1 & chg5)
- A moving average (mavWgt) of annual growth rates (denoted ΔX_t to represent the change between $X_t - X_{t-1}$), where more recent changes are weighted more heavily according to a solution proposed in Wlezien (2015). The formula we employ to transform a given contextual feature is:

$$mavWgt_t = \Delta X_t * 0.5 + \Delta X_{t-1} * 0.2 + \Delta X_{t-2} * 0.15 + \Delta X_{t-3} * 0.1 + \Delta X_{t-4} * 0.05 \quad (2)$$

We plot the impact of these alternative temporal transformations on the prediction errors of our main model in Figure 11, illustrating the robustness of our substantive conclusions.

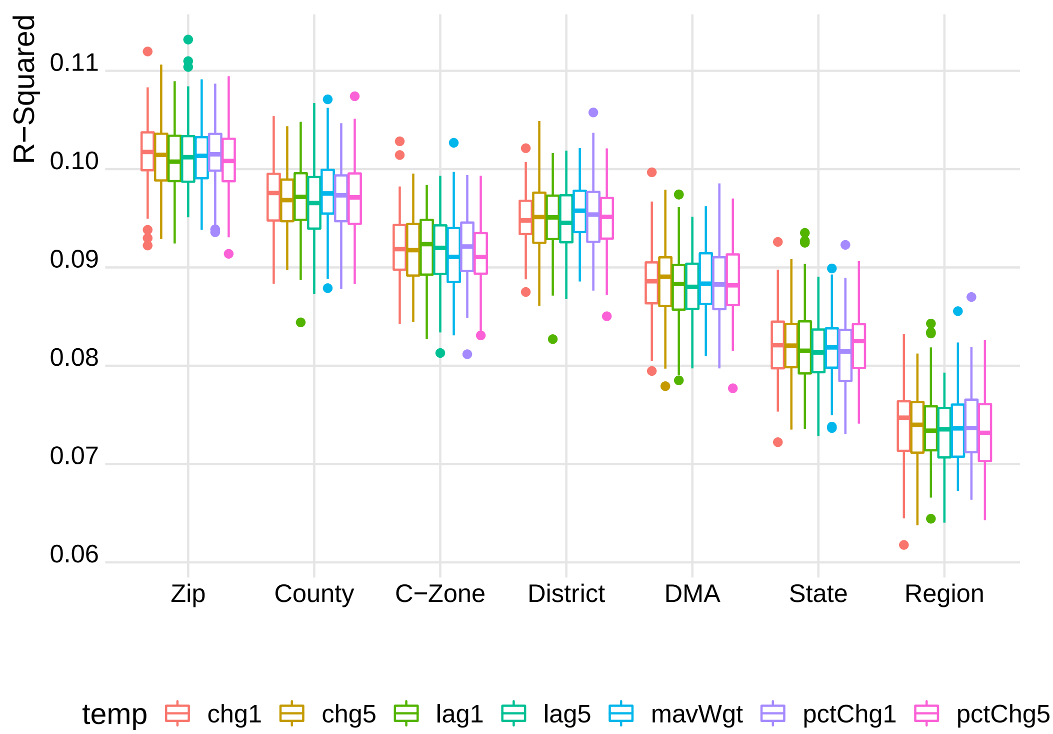

Figure 11: Model performance (measured as RMSE on a 1-4 scale of economic evaluations, y-axis) across different units of geographic aggregation (x-axis) and different choices of temporal transformation (colors).

Looking at the variable importance results (Figure 12) reveals similar lack of evidence of a meaningful difference across these choices.

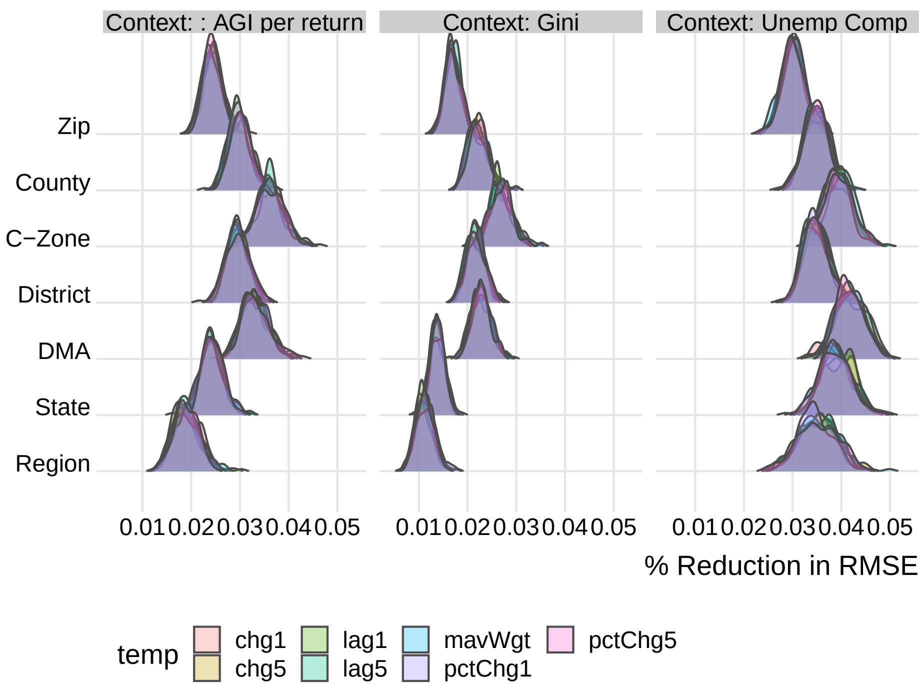

Figure 12: Variable importance for contextual predictors (% reduction in RMSE, x-axes) plotted across different units of aggregation (y-axes) using different temporal transformations (colors).

2.3 Alternative specifications of linear models

2.3.1 Inclusion of YEAR vs. GEO fixed effects

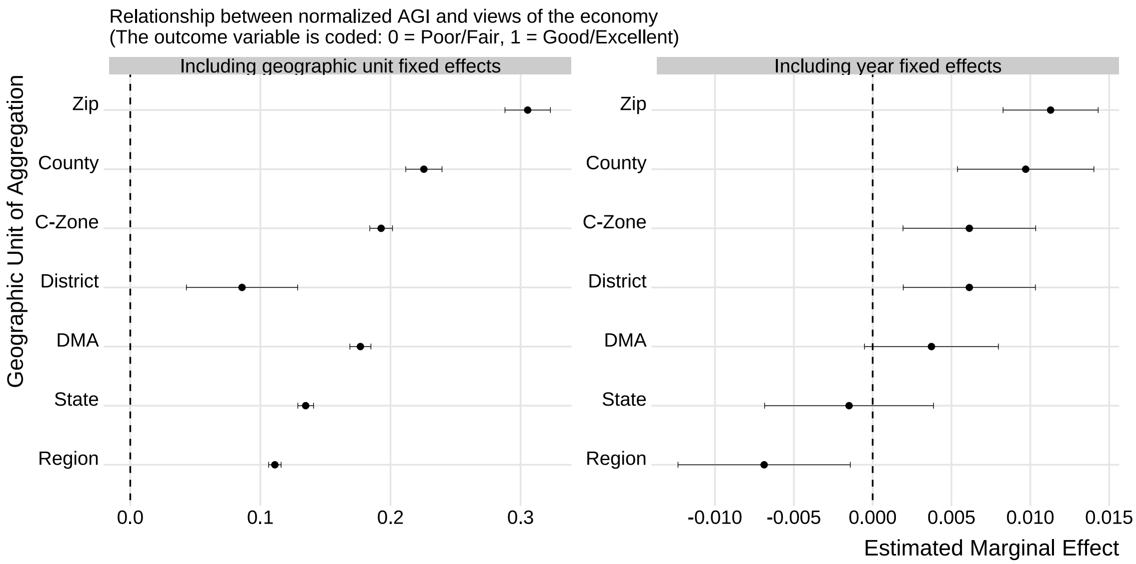

Figure 13: Each row corresponds to one model. The points are are correlations between aggregate gross income (AGI) and positive views of the economy measured at the individual level, conditioning on respondents' demographic characteristics (including partisanship). The units at which we aggregate contextual income are indicated on the y-axis.

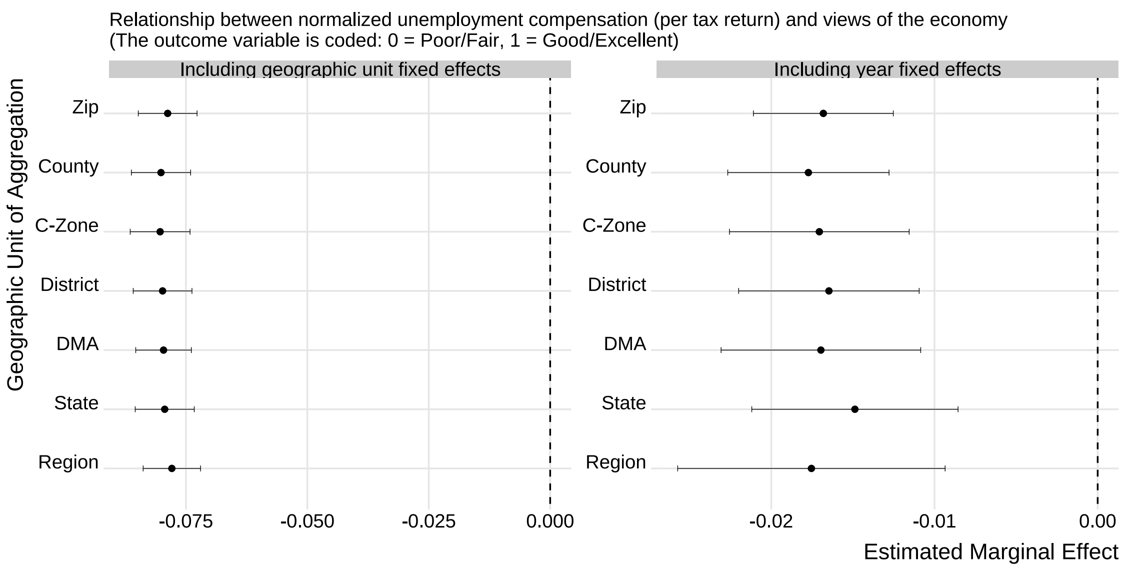

Figure 14: Each row corresponds to one model. The points are are correlations between unemployment compensation and positive views of the economy measured at the individual level, conditioning on respondents' demographic characteristics (including partisanship). The units at which we aggregate the contextual variable are indicated on the y-axis.

2.3.2 Modifying the set of control variables

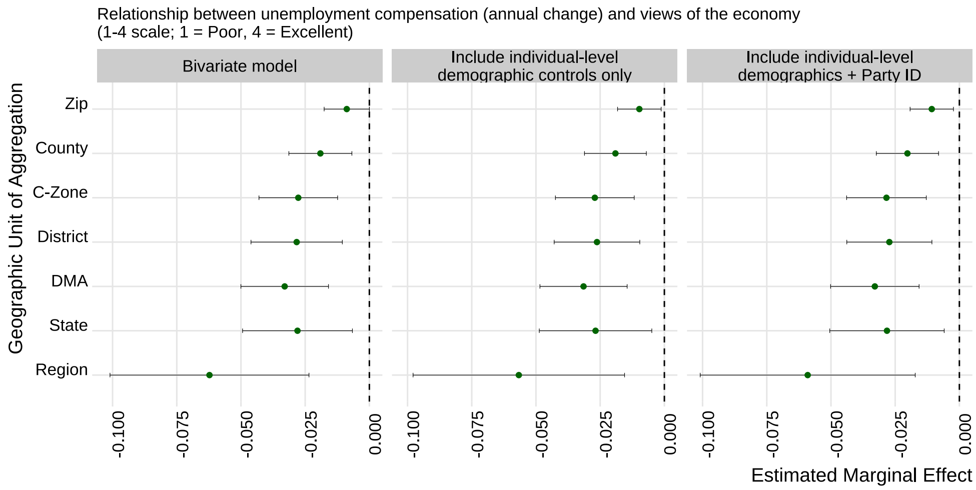

Figure 15: Correlations between changes in the unemployment compensation change (per tax return) and positive views of the economy measured at the individual level. In the first panel, the models from which coefficients are pulled do not include any control variables. In the second panel, we condition on respondents' demographic characteristics. In the right-hand-side panel we also include individual-level partisanship. The contextual predictor is aggregated at distinct levels, which are indicated on the y-axis.

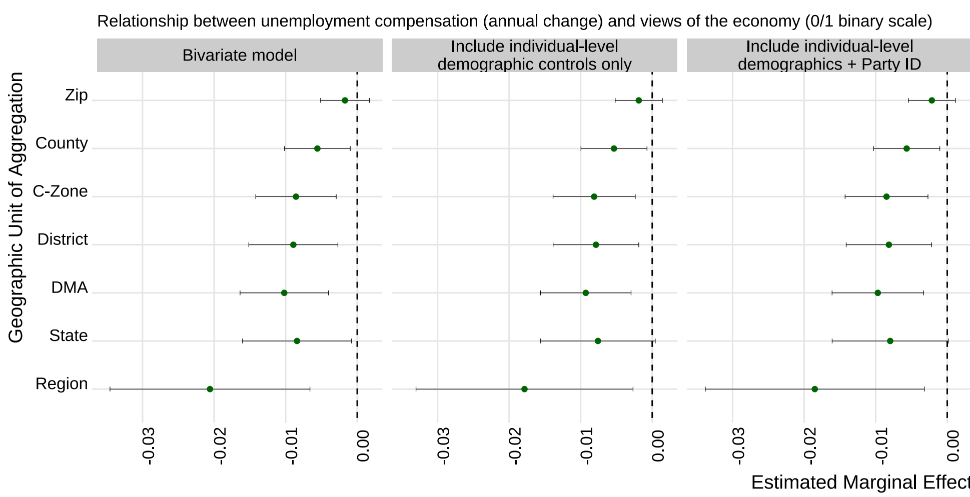

Figure 16: Correlations between changes in the unemployment compensation change (per tax return) and positive views of the economy measured at the individual level. The outcome variable is set to one if economic evaluations are "good" or "excellent". In the first panel, the models from which coefficients are pulled do not include any control variables. In the second panel, we condition on respondents' demographic characteristics. In the right-hand-side panel we also include individual-level partisanship. The contextual predictor is aggregated at distinct levels, which are indicated on the y-axis.

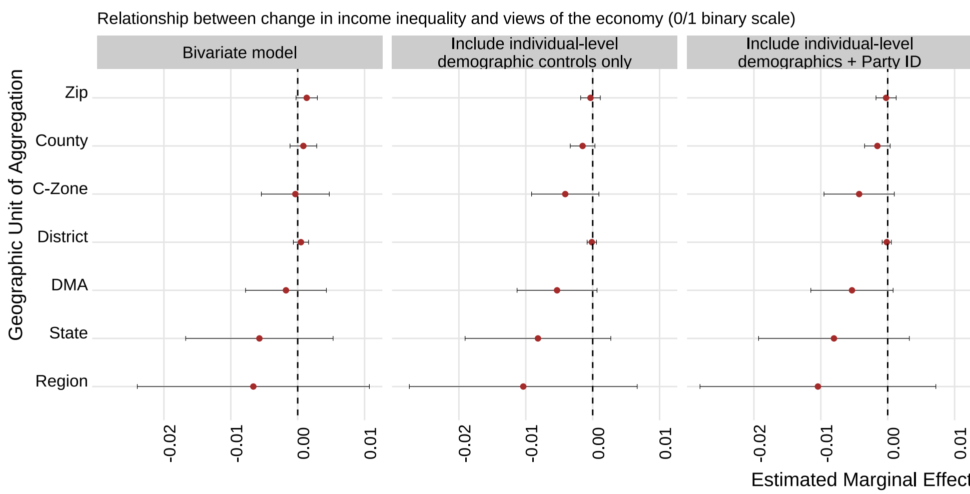

Figure 17: Correlations between changes in the unemployment compensation change (per tax return) and positive views of the economy measured at the individual level. The outcome variable is set to one if economic evaluations are "good" or "excellent". In the first panel, the models from which coefficients are pulled do not include any control variables. In the second panel, we condition on respondents' demographic characteristics. In the right-hand-side panel we also include individual-level partisanship. The contextual predictor is aggregated at distinct levels, which are indicated on the y-axis.

2.3.3 A comparison of a single-level vs. multi-level model

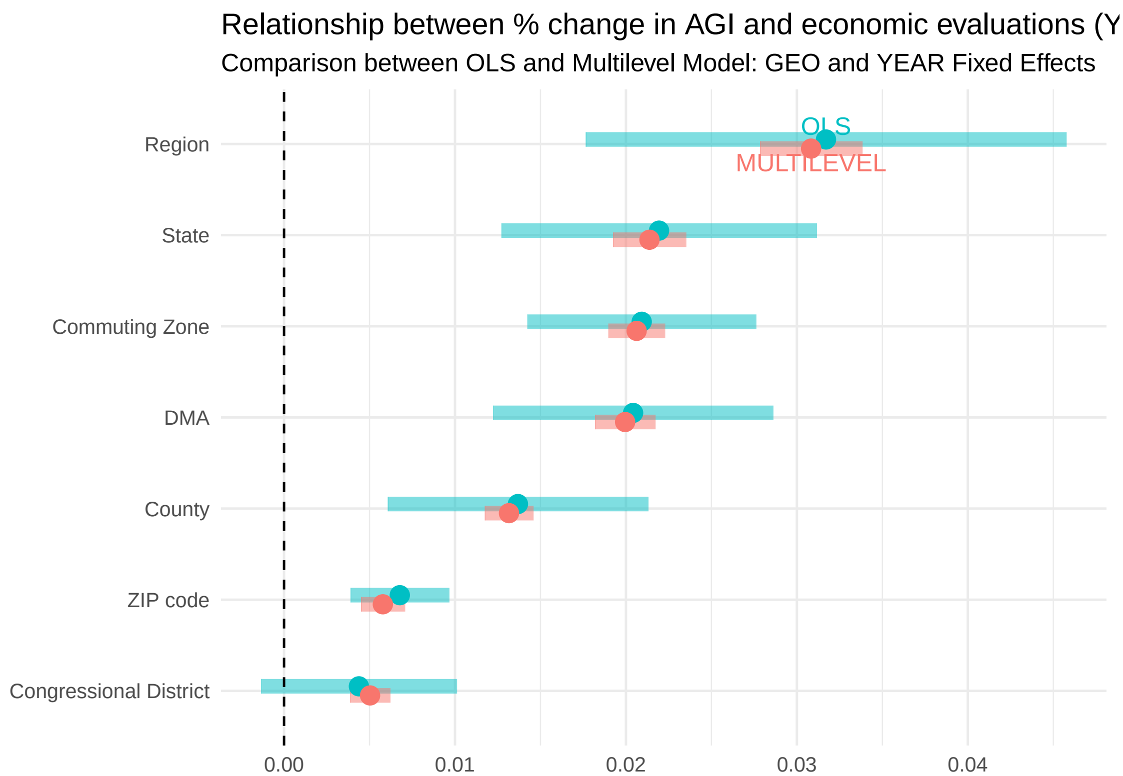

Figure 18: Each row corresponds to one model. The points are are correlations between annual % change in AGI per return and positive views of the economy measured at the individual level, conditioning on respondents' demographic characteristics (including partisanship). The units at which we aggregate the contextual variable are indicated on the y-axis. Coefficients from a multilevel version of the model are displayed in red.

2.4 Random Forest Hyper-parameter Tuning

Our main results for variable importance use random forests implemented via the `ranger` package for `R`, with the following hyper-parameters:

- `num.trees` = 200
- `mtry` = $\sqrt{p}$ where p is the number of predictors. In practice this yields `mtry` = 19
- `min.node.size` = 10

These parameters were chosen based on 5-fold cross-validation using the grid-search implementation provided by the `caret` package. Figure 19 plots the root mean squared error for the smallest (ZIP code) and largest (census region) geographies for different combinations of the hyper-parameters. These models predict the 1-4 economic evaluation outcome as a function of individual-level covariates including age, gender, marital status, race, income, education, and party affiliation, and context-level covariates including unemployment compensation per return, AGI per return, and the Gini, aggregated to either the zip (red) or region (blue).

We opt for a larger `mtry` value to ensure we are able to accurately measure variable importance for each of the contextual predictors, which correspond to only three out of the 57 total predictors (6 categories for age + 2 categories for gender + 6 categories for marital status + 5 categories for race/ethnicity + 13 categories for income + 5 categories for educational attainment + 6 categories for party affiliation + 11 categories for year).[2] Higher values of the `mtry` parameter are shown to reduce the variable importance measures of weak predictors, as discussed in Grömping (2009). In addition, we reduce the number of trees to 20, based on Chipman et al. (2010) and Bleich et al. (2014), who demonstrate that fewer trees requires predictors to compete with each other for inclusion, yielding better variable importance results.

[2] A point of clarification: although 5 party ID categories are measured on the survey, we also kept 3% of respondents with missing party ID, and the status of 'missing PID' was simply treated a sixth possible PID indicator variable.

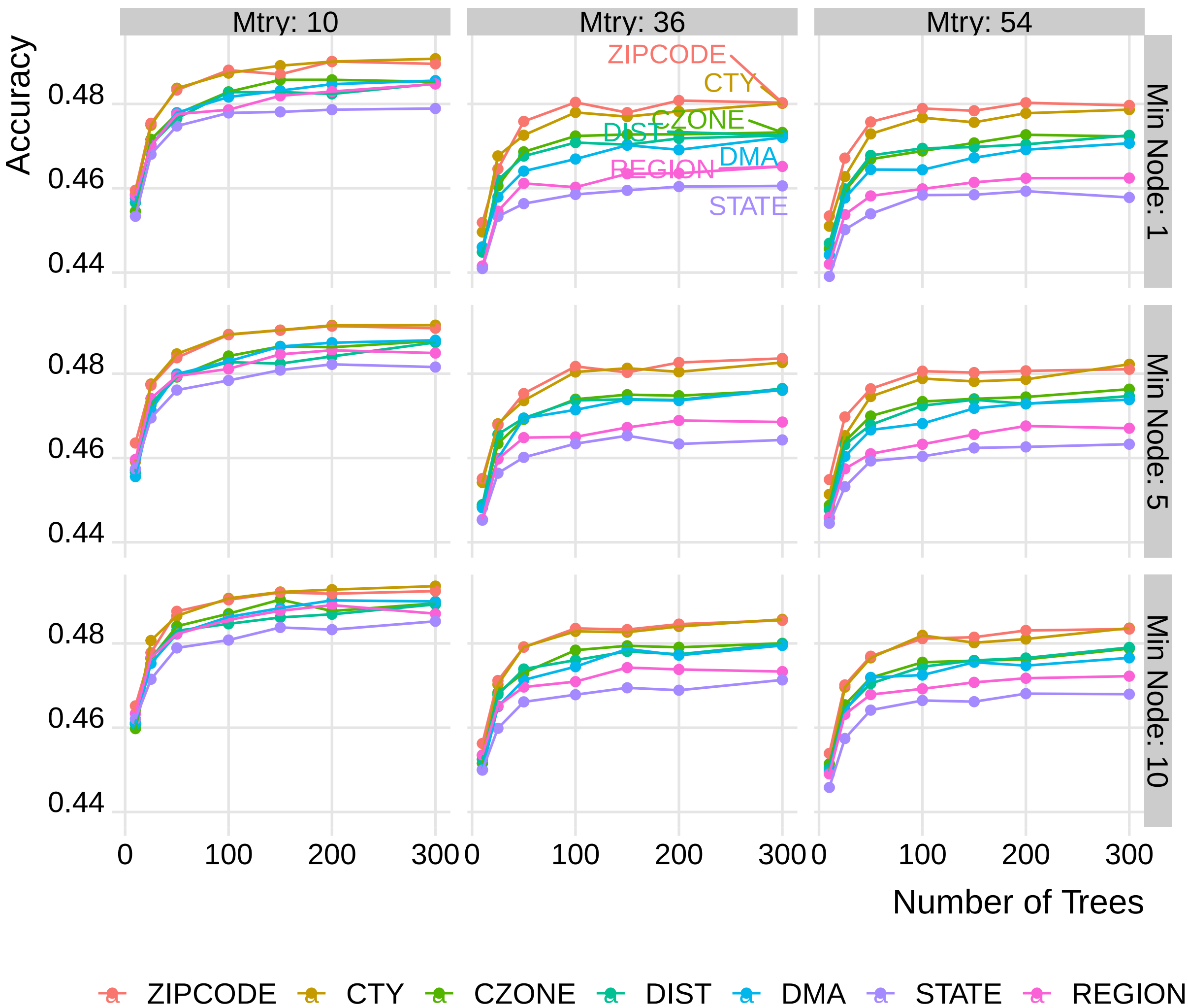

Figure 19: Random forest hyper-parameter tuning results from 5-fold cross-validation via the `caret` package for `R`. Columns correspond to different values of the `mtry` hyper-parameter, rows correspond to different values of the `min.node.size` hyper-parameter, and the x-axes capture different values of the `num.trees` hyper-parameter. `mtry` values based on the square root, 1/3 and 0.5 of the total number of predictors.

2.5 Modeling aggregate changes

We aggregate economic evaluations at two levels (counties and states in Tables 2 and 3 respectively) to calculate *changes* in economic evaluations and regress them either on changes in economic conditions or on their static measures. We see again that the coefficients are sensitive to the geographic unit of aggregation.

Table 2: Modeling aggregate changes in economic evaluations (county-level)

	Δ Average views of the economy (1-4 scale)			
	(1)	(2)	(3)	(4)
% Change AGI per return (standardized)	−0.001			
	(0.002)			
% Change unemp. comp. per return (standardized)		−0.028		
		(0.004)		
AGI per return (standardized)			0.004	
			(0.002)	
Unemp. comp. per return (standardized)				0.003
				(0.003)
Year FEs	Yes	Yes	Yes	Yes
Observations	3,493	3,493	3,493	3,493
R^2	0.381	0.388	0.382	0.381

Note: Counties with at least 200 Gallup respondents are included.

Table 3: Modeling aggregate changes in economic evaluations (state-level)

	Δ Average views of the economy (1-4 scale)			
	(1)	(2)	(3)	(4)
% Change AGI per return (standardized)	0.011			
	(0.004)			
% Change unemp. comp. per return (standardized)		−0.037		
		(0.006)		
AGI per return (standardized)			0.004	
			(0.003)	
Unemp. comp. per return (standardized)				0.008
				(0.004)
Year FEs	Yes	Yes	Yes	Yes
Observations	459	459	459	459
R^2	0.747	0.762	0.744	0.746

3 Permutation Tests

Our analysis relies on permutation tests of variable importance. The core intuition of any variable importance test hinges on the penalty paid by a predictive model when a covariate is removed or – in the context of a permutation test – has its relationship with the outcome broken through random reshuffling of its elements.

This is not the only tool available to researchers interested in describing variable importance in a more robust manner than simply adding predictors to a base model and comparing changes in RMSE or AIC/BIC. Faster implementations that rely on random forests use "impurity" techniques. These methods measure variable importance as the reduction in impurity (i.e. mis-classified units in a child node) associated with using a given variable as a splitting rule. The variable's overall importance is the weighted sum of these impurity reductions across all instances in which it is used as a splitting rule in the random forest, scaled by the number of trees.

Impurity-based measures of variable importance are known to favor variables with many possible split points, leading to erroneous conclusions that unrelated continuous measures are more important than prognostic discrete measures. Permutation tests, while more computationally expensive, are insulated from this issue.

To demonstrate, we adopt a simulated scenario proposed by a reviewer in which a continuous covariate $\Delta unemp$ is drawn from a standard normal distribution, a dichotomous covariate *party* is drawn from a binomial distribution with $p = 0.5$, and the outcome y is defined by:

$$y_i = \Delta unemp_i - 2 * \Delta unemp_i * party_i + \varepsilon_i \tag{3}$$

As simulated, y and Δ *unemp* are positively correlated for members of $party = 1$ and negatively correlated for members of $party = 0$. Furthermore, both $\Delta unemp$ and *party* are crucial to understanding variation in y_i, meaning that both should be equally ranked by a useful measure of variable importance.

We simulate this data-generating process 1,000 times and each time model y with a random forest. In one instance of the data, we estimate variable importance using both the impurity-based measure and the permutation test approach, saving the results.[3] We plot the resulting variable importance metrics in Figure 20, shows that impurity-based methods understate the importance of unemp and also illustrates the insulation of our preferred permutation method (displayed in the right-most panel) from biases stemming from variable interactions.

[3] We also investigate the performance of an impurity-corrected approach proposed by Nembrini, König and Wright (2018) who developed it to be unbiased in the number of categories.

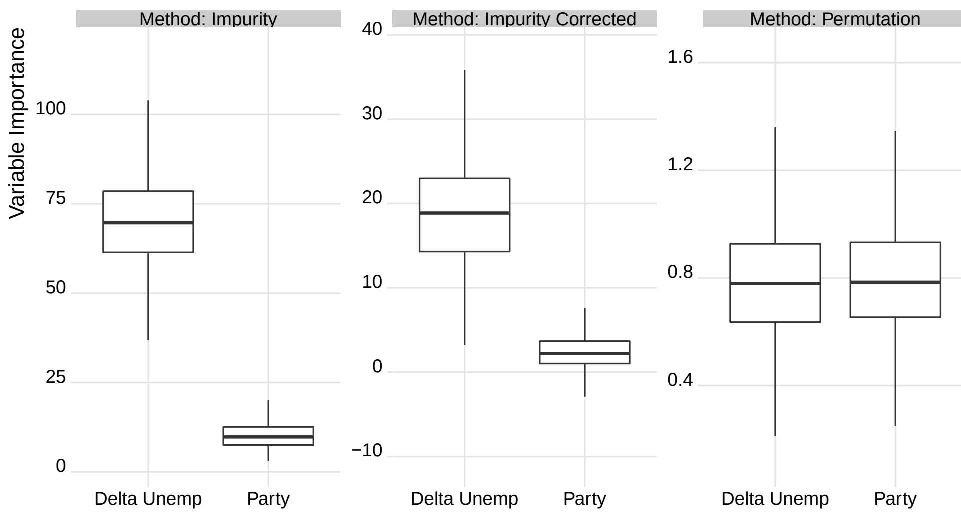

Figure 20: Variable importance from 1,000 simulations for impurity (left panel), impurity corrected (center panel), and permutation methods (right panel).

4 Existing literature

In this section we report a collection of relevant findings and arguments without restricting our attention to political science.

Sensitivity of correlation magnitudes to unit sizes were already investigated in Gehlke and Biehl (1934) who artificially manipulated the spatial resolution of their data on social outcomes and living costs. Specifically, they correlated juvenile delinquency with median rent prices in groups of Census tracts in Cleveland. Successively dividing the same 252 tracts into a smaller number of groups (thus increasing the average area in each step), they found that the correlation between the variables of interest was larger when the area of hypothetical territories increased. This led them to a warning that "relatively high correlation[s] might conceivably occur by census tracts when the traits so studied [are] completely dissociated in the individuals or families of those traits."

In sociology, (Robinson, 1950) highlighted that the correlation between the proportion of foreign-born population and illiteracy rates was negative at the level of Census regions; however, he then showed that the correlation between foreign birth and illiteracy is *positive* at the individual level. It was this paper that heightened awareness of issues surrounding ecological inferences, and it was followed by a rich literature on phenomena closely related to the ecological fallacy, such as the Simpson's Paradox, aggregation bias, and the modifiable areal unit problem (MAUP), among others.

A typology of fallacies that can be committed by researchers who use both contextual and individual-level observations is outlined in Diez-Roux (1998). Blakely and Woodward (2000) elaborate on the list and provide graphical representations of three mechanisms for ecological effects: cross-level effect modification, direct cross-level effect, and the indirect cross-level effect.

When an outcome at the individual level is not observable (e.g. due to secrecy of the ballot) then ecological regressions are among the few available research designs, especially if survey data cannot be collected as it relates to historical events. However, one study of voting offers a poignant example of the associated risks. Schoenberger and Segal (1971) show that, in 77 Southern congressional districts, the Wallace vote share was higher in districts with greater proportion of Black population. They observe that "[i]t would be a fallacy—ecological, logical, sociological and political—to infer from these data that blacks in the South provided a major source of Wallace support."

Researchers have also observed that it can be "natural to personify the states" (Gelman, Shor, Bafumi and Park, 2007) and this makes it tempting to make the assumption, for example, that Republicans have the support of poor voters, because Republican presidential candidates tend to win in poorer states. Although poorer states do tend to support Republicans, the authors also show that Republicans win richer counties only in some states; crucially, at the individual level, the relationship between income and voting Republican is *positive*, contrary to what could be (wrongly) inferred from ecological regressions.

We note that theory-driven ecological regressions have been fruitfully applied in political science: for instance, Healy and Lenz (2017) regress the democratic vote share on measures of economic hardship measured either at the ZIP code level (changes in mortgage delinquencies between 2006 and 2008) or, separately, at the county level (growth in employment and wages). Other examples include studies by Trounstine (2020) and Hill, Hopkins and Huber (2019) both of which use data at the precinct level. The former study explicitly states that the objective it to understand the behavior of precincts while the latter suggests that aggregated voting behavior can yield insights into voter's motivations.

Related to our work, there are several recent papers exploring the relationship economic assessments of respondents and the real economy. In a Danish context, Bisgaard, Dinesen and Sonderskov (2016) used the exact geographic location of survey respondents to examine economic perceptions as a function of proximate unemployment. They show, first, that the correlation between municipality unemployment and unemployment in a narrowly defined radius (ranging from 80 to 2,500 meters) is surprisingly low. Then they illustrate the downsides of using predictors based on pre-defined areas, such as municipality rates (which sometimes do not even correlate with economic perceptions.).

Ang et al. (2021) report correlations between national or county-level economic variables and economic evaluations among a panel of U.S. respondents. In the sample of The American Panel Survey (TAPS) participants, the correlation between the *national* employment rate and economic evaluations was positive and significant. However, county employment rate was not correlated with evaluations of either national or household economic conditions. The authors' primary objective was to quantify the influence of partisanship on economic evaluations; here we simply note an interesting tension about how correlations between the situation in the labor market and reported evaluations depend on the spatial aggregation units.

These results suggest that what counts as "local" in citizens' minds may only loosely match what statistical agencies choose to measure. However, researchers who examine the impact of contextual information typically use data aggregated at a single administrative unit (e.g. counties, districts, or states). This has led some scholars to argue that "people's perceptions of their environment do not resemble governmental units" (Wong et al., 2012, p.1153). In related work, Wong, Bowers, Rubenson, Fredrickson and Rundlett (2018) show that when subjects draw maps of their local communities, then some perceived community attributes (for example the ethnic composition of a given geographic unit) do not reflect official government statistics.

An important recent contribution on issues arising from administrative misclassification is Nemerever and Rogers (2021): they report that approximately 20% of rural Americans reside in counties that are classified as metropolitan. They show that if researchers classify respondents as rural when they don't reside in an urban cluster, then regression results are sensitive to this classification choice. Specifically, if respondents are classified on the basis of ZIP code Tabulation area-level characteristics, then inclusion of a re-conceptualized control variable (for rural status) can affect other regression coefficients of interest.

Various approaches for addressing ecological inference issues have been proposed. King (1997) proposed an ecological inference technique (for extensions and modifications, see King, Rosen and Tanner, 2004) by replacing the constancy assumption of the ecological regression (where individuals from a given group are assumed to behave in the same way in each geographic area) with the idea that propensities to behave in a particular way are drawn from a truncated bivariate normal distribution in the case of two groups.

Because correlations can be expected to change as a function of the unit size, some researchers choose to split their sample. For example, Eberle et al. (2020) split provinces by either population or area and verify whether coefficients are different for the set of smaller vs. larger geographic units. Other authors (Wong, 2001) recommend estimating spatial cumulative distribution functions in order to compare the prevalence (or co-occurrence) of particular variables across regions instead of correlating measurements at a fixed spatial resolution.

Moreover, there are cases when the appropriate geographic unit is known (e.g. when a policy is controlled by state legislators) and when individual-level data of interest (e.g. views on state policies) is available. Consider Shirley and Gelman (2015) who find that individual-level effects on the support for the death penalty vary across states and regions and, accordingly, they make the case for interacting demographic, geographic and time variables. In this context, the contextual variables measured at the state level receive immediate theoretical support for inclusion in statistical models.

Another recent paper highlighting the issues arising from unclear boundaries is Spater (Forthcoming). His proposed solution is to use social network contact data to accurately measure exposure to outgroup members. It is likely that future research will increasingly take advantage of (anonymized and potentially aggregated) mobility data to uncover the shapes of organic communities. The spatial boundaries revealed by people's everyday behavior are then likely to be promising candidates for improved measurement of contextual variables (Athey et al., 2020; Moore and Reeves, 2020).

5 Re-analysis of existing work on contextual effects

In the following, we re-analyze two recent publications that use contextual measures aggregated to a certain unit of aggregation. The first is work by Green and McElwee (2019) who use ZIP-level data to evaluate the relative influence of economic conditions and racial attitudes in explaining support for Donald Trump in the 2016 election. The authors show that ZIP-level economic conditions are significantly related to Trump vote among white voter file-matched respondents in the 2016 Cooperative Congressional Election Survey (CCES). Furthermore, they find that economic distress – measured as the log of the 2015 ZIP-level unemployment insurance (UI) – also predicts non-voting, but does not predict voting for third-party ("minor") candidates in the 2016 presidential election.

We replicate their findings from Table 1 in their *Perspectives on Politics* article (`https://doi.org/10.1017/S1537592718003365`) in Figure 21. Their coefficients are given in red, while the same estimates aggregating 2015 UI to other units including the county, congressional district (using the 113th Congress shape files), designated market area (DMA), and state yield potentially conflicting conclusions. While the substantive conclusions about the support for Trump persist across alternative units of aggregation, the results for nonvoters (a significant positive relationship) and supporters of non-mainstream candidates (a null result) are sensitive to this choice. These decisions clearly carry important implications for how we understand Trump's 2016 victory, and we commend these authors for their attention to this important choice.[4]

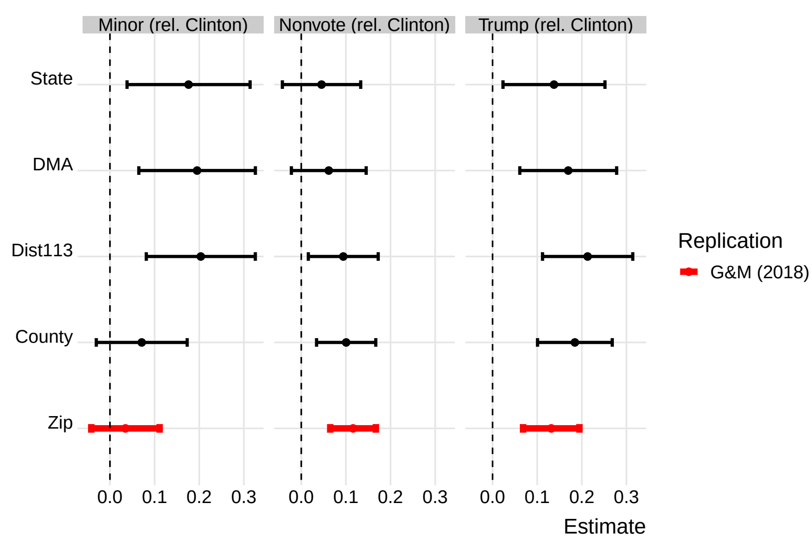

Figure 21: Replication of Table 1 from Green and McElwee (2019). Coefficients (x-axis) show the magnitude of the correlation between "local" UI and voting 3 types of voting behavior. Geographic aggregation of the measure of the economy varies used in a given model is shown on the y-axis.

[4] Green and McElwee (2019) explicitly state their considerations for data selection, noting that: "ZIP codes may more closely approximate the "community" level than counties, which are often large enough to contain multiple communities with highly variable economic conditions."

Our second re-analysis attempts to replicate the results published by Ansolabehere, Meredith and Snowberg (2014), who examine how individuals assess the national unemployment rate as a function of the "local" unemployment rate, where "local" refers to their state of residence. Their choice of the state as the unit of aggregation is twice-motivated. The first motivation is theoretical, emphasizing that monthly state unemployment rates are covered by the media, making them a particularly salient of "mecro"-economic information. The second motivation is practical and reflects the constraints faced by many political science researchers ("from a practical perspective, state is the only geographic variable consistently reported in all of the data sources we use").

We focus on replicating Table 3 from their paper, available at `https://onlinelibrary.wiley.com/doi/full/10.1111/ecpo.12040`, which relies on a publicly available module from the 2008 Cooperative Congressional Election Survey (CCES) to obtain these results. However, their precise replication materials are not publicly shared. As such, we are not able to perfectly recover the coefficients listed in Table 3 of their paper, likely due to undocumented weighting choices, although we are able to obtain roughly similar point estimates for their predictors of interest.

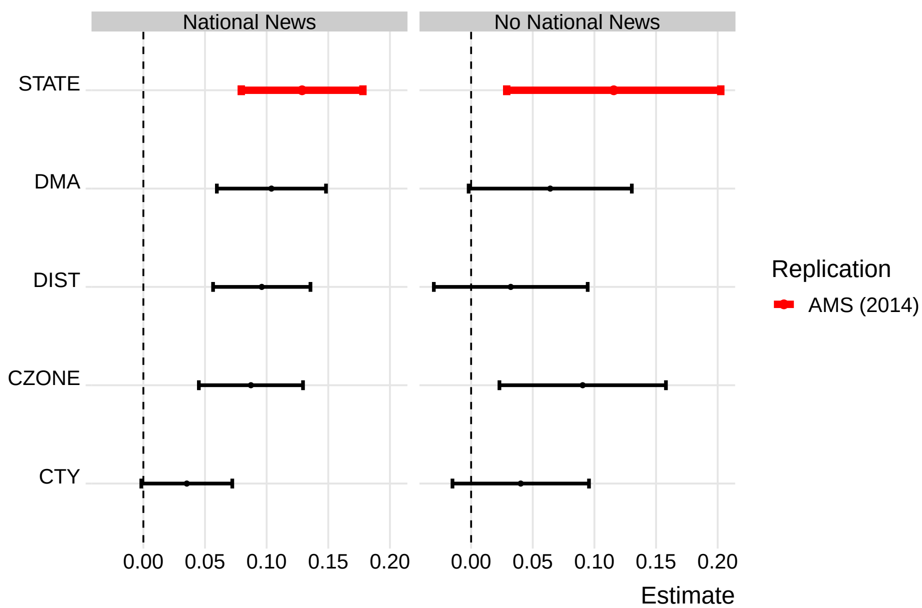

Figure 22: Replication of Table 3 in Ansolabehere, Meredith and Snowberg (2014). Coefficients (x-axis) capture the correlation between a respondent's estimate of the national unemployment rate as a function of their local unemployment, with localness defined by geographic units of aggregation on the y-axis.

We are interested in the sensitivity of their findings to different choices of the geographic unit at which the unemployment rate is calculated. Their main results suggest that the state unemployment rate is an important predictor of assessments of the national unemployment rate among those who pay attention to national news, but not for those who don't. In our replication, summarized in Figure 22, we find even stronger results when aggregating the unemployment rate to the state level among those who pay attention to the news (0.13 in our replication against 0.09 in their Table 3). However, this conclusion is sensitive to the choice of geographic unit, declining at smaller units to the point of being only marginally significant at the 90% level of confidence when measured at the county (left column of Figure 22). Similarly, we find stronger results among those who do not watch national news programs (0.11 in our replication against 0.10 in their Table 3).

The authors do provide some theoretical motivation for their choice of the state as the unit at which respondents are most likely to be informed about the national unemployment rate. Nevertheless, our replication suggests that the empirical results – particularly the confidence in their findings – are sensitive to these choices. Taken together, we argue these re-analyses highlight the importance of careful attention to the question of which geographic unit is most sensible for any empirical study.

6 Partial Dependencies

Examining relationships of substantive interest need not rely on restricting linear regression models. With random forests, one can simulate the outcome by setting certain covariates of interest to various values and predicting the model. In the following plots, we conduct this analysis, setting AGI per return to its quintiles and simulating economic evaluations on a 1-4 scale, disaggregated by year and party ID. All other predictors are set to their means. Figure 23 examines the relationship between economic evaluations and local wealth, aggregated to the commuting zone and region for each year, reinforcing the conclusion from the linear regression results that these choices have non-trivial implications.

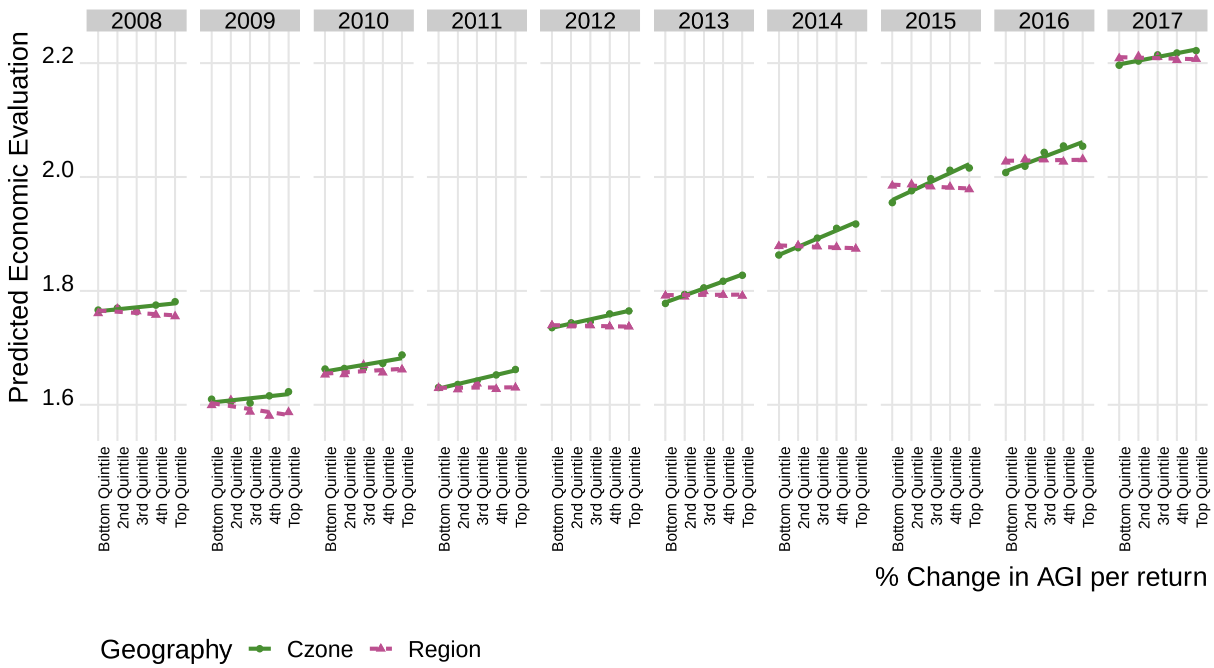

Figure 23: Partial dependence results by year and unit of aggregation (commuting zones in green, regions in purple). AGI per return is binned to its quintiles and outcomes are simulated for each year

We can also look at this relationships by party over time (Figure 24). We find little evidence of heterogeneous relationships by party affiliation, but striking evidence of partisan cheerleading as those of the same party as the president increase their evaluations seemingly overnight.

Finally, we also examine the substantive results from the random forest models via two more robust methods: accumulated local effects (ALEs) and individual conditional expectations (ICEs). These tools for describing relationships are more robust than partial dependencies which are unreliable where features in a machine learning model are correlated. Figure 25 presents the ALE measures for unemployment claims per return aggregated to the county (red) and state (teal) by year. As illustrated, the strength of this relationship is strongest

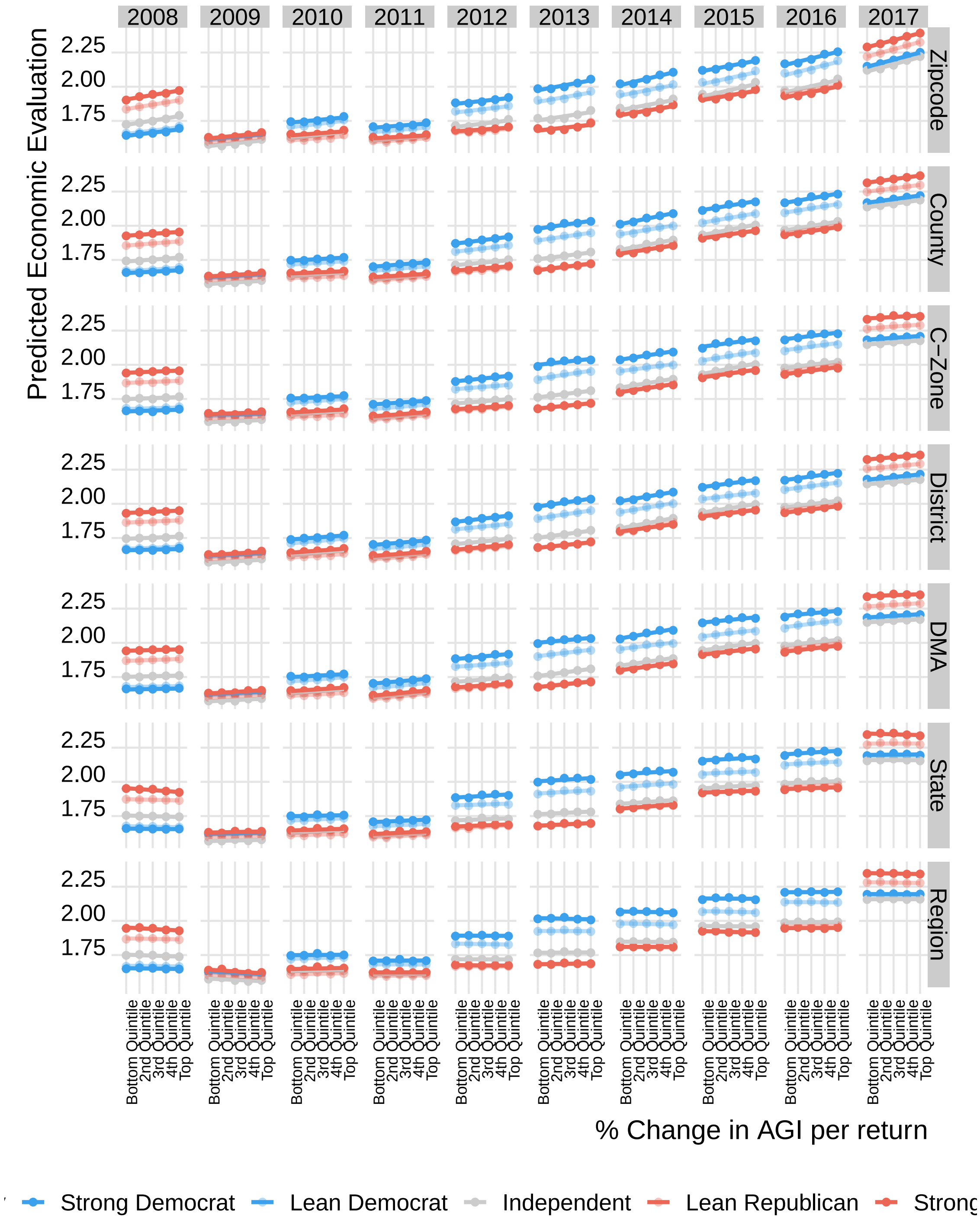

Figure 24: Partial dependence results by year and unit of aggregation (commuting zones in green, regions in purple). AGI per return is binned to its quintiles and outcomes are simulated for each year

between 2008 and 2012, corresponding to hardest years during the Recession. As in the partial dependence plots above, these relationships are attenuated when aggregating local unemployment claims per return to the state, compared to the county.

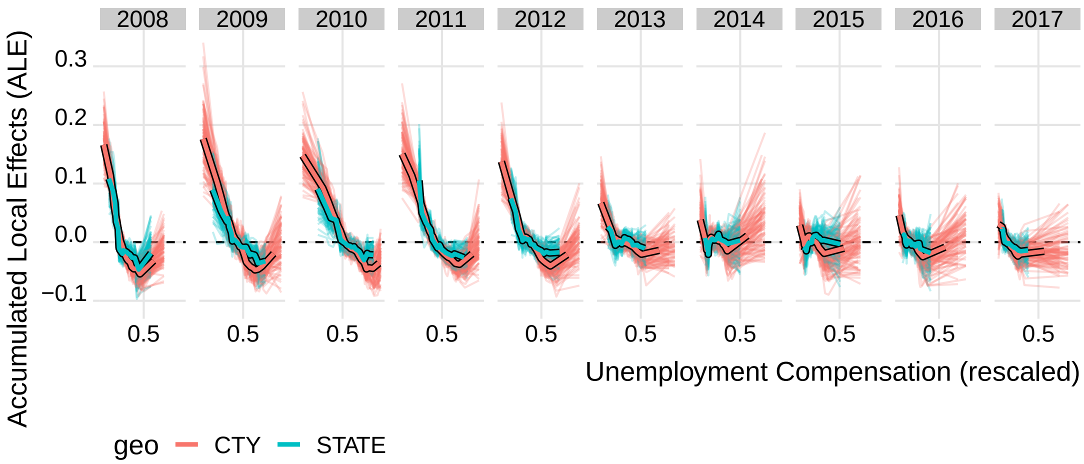

Figure 25: ALE estimates (y-axes) of rescaled unemployment claims per return (x-axes) aggregated to the county (red) and state (teal) by year (facets).

We find similar patterns in the ICE plots in Figure 26, where we instead predict variation in

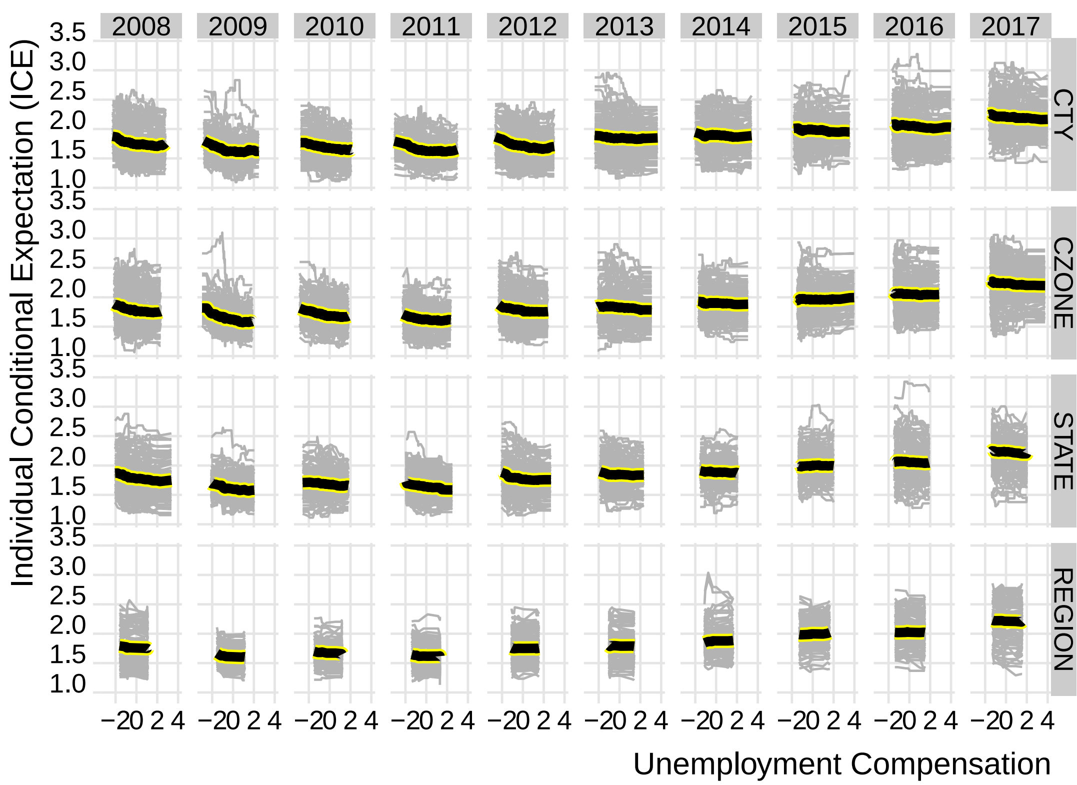

Figure 26: ICE estimates (y-axes) of unemployment claims per return (x-axes) by year (columns) and geographic unit of aggregation (rows).

References

Ang, Zoe, Andrew Reeves, Jon C Rogowski and Arjun Vishwanath. 2021. "Partisanship, Economic Assessments, and Presidential Accountability." *Working Paper* .

Ansolabehere, Stephen, Marc Meredith and Erik Snowberg. 2014. "Mecro-economic voting: Local information and micro-perceptions of the macro-economy." *Economics & Politics* 26(3):380–410.

Athey, Susan, Billy Ferguson, Matthew Gentzkow and Tobias Schmidt. 2020. "Experienced Segregation." *Working Paper* .

Bisgaard, Martin, P T Dinesen and K M Sonderskov. 2016. "Reconsidering the Neighborhood Effect: Does Exposure to Residential Unemployment Influence Voters' Perceptions of the National Economy?" *The Journal of Politics* 78(3):719–732.

Blakely, T A and A J Woodward. 2000. "Ecological effects in multi-level studies." *Journal of epidemiology and community health* 54(5):367–74.

Bleich, Justin, Adam Kapelner, Edward I George and Shane T Jensen. 2014. "Variable selection for BART: an application to gene regulation." *The Annals of Applied Statistics* pp. 1750–1781.

Chipman, Hugh A, Edward I George, Robert E McCulloch et al. 2010. "BART: Bayesian additive regression trees." *The Annals of Applied Statistics* 4(1):266–298.

Diez-Roux, Ana V. 1998. "Bringing context back into epidemiology: Variables and fallacies in multilevel analysis." *American Joumal of Public Health* 88(2):216–222.

Eberle, U. J., J. V. Henderson, D. Rohner and K. Schmidheiny. 2020. "Ethnolinguistic diversity and urban agglomeration." *Proceedings of the National Academy of Sciences* 117(28):16250–16257.

Gehlke, Charles E and Katherine Biehl. 1934. "Certain effects of grouping upon the size of the correlation coefficient in census tract material." *Journal of the American Statistical Association* 29(185A):169–170.

Gelman, Andrew, Boris Shor, Joseph Bafumi and David Park. 2007. "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?" *Quarterly Journal of Political Science* 2:345–367.

Green, Jon and Sean McElwee. 2019. "The differential effects of economic conditions and racial attitudes in the election of Donald Trump." *Perspectives on Politics* 17(2):358–379.

Grömping, Ulrike. 2009. "Variable importance assessment in regression: linear regression versus random forest." *The American Statistician* 63(4):308–319.

Healy, Andrew and Gabriel S Lenz. 2017. "Presidential Voting and the Local Economy: Evidence from Two Population-Based Data Sets." *The Journal of Politics* 79(4):1419–1432.

Hill, Seth J, Daniel J Hopkins and Gregory A Huber. 2019. "Local demographic changes and US presidential voting, 2012 to 2016." *Proceedings of the National Academy of Sciences* 116(50):25023–25028.

King, Gary. 1997. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton University Press.

King, Gary, Ori Rosen and Martin Abba Tanner. 2004. *Ecological Inference: New Methodological Strategies*. Cambridge University Press.

Lewis, Jeffrey B, Brandon DeVine, Lincoln Pitcher and Kenneth C Martis. 2013. "Digital boundary definitions of united states congressional districts, 1789-2012." *Data file and code book* .

Moore, Ryan T. and Andrew Reeves. 2020. "Defining racial and ethnic context with geolocation data." *Political Science Research and Methods* 8(4):780–794.

Nembrini, Stefano, Inke R König and Marvin N Wright. 2018. "The revival of the Gini importance?" *Bioinformatics* 34(21):3711–3718.

Nemerever, Zoe and Melissa Rogers. 2021. "Measuring the Rural Continuum in Political Science." *Political Analysis* p. Published Online (First View).

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15:351–357.

Schoenberger, Robert A. and David R. Segal. 1971. "The Ecology of Dissent: The Southern Wallace Vote in 1968." *Midwest Journal of Political Science* 15:583–586.

Shirley, Kenneth E. and Andrew Gelman. 2015. "Hierarchical models for estimating state and demographic trends in US death penalty public opinion." *Journal of the Royal Statistical Society. Series A* 178(1):1–28.

Sood, Gaurav. 2018. "Geographic information on designated media markets." *URL: https://doi. org/10.7910/DVN/IVXEHT* .

Spater, Jeremy. Forthcoming. "Exposure and Preferences: Evidence from Indian Slums." *American Journal of Political Science* .
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12570*

Tolbert Charles, M and Molly Sizer. 1996. "US Commuting Zones and Labor Market Areas: a 1990 Update." *US Department of Agriculture Economic Research Service Staff Paper* (9614).

Trounstine, Jessica. 2020. "The Geography of Inequality: How Land Use Regulation Produces Segregation." *doi.org* 114(2):443–455.

Wlezien, Christopher. 2015. "The myopic voter? The economy and US presidential elections." *Electoral Studies* 39:195–204.

Wong, Cara, Jake Bowers, Daniel Rubenson, Mark Fredrickson and Ashlea Rundlett. 2018. "Maps in People's Heads: Assessing a New Measure of Context." *Political Science Research and Methods* 8(1):160 – 168.

Wong, Cara, Jake Bowers, Tarah Williams and Katherine Drake Simmons. 2012. "Bringing the person back in: Boundaries, perceptions, and the measurement of racial context." *The Journal of Politics* 74(4):1153–1170.

Wong, David W. S. 2001. "Location-Specific Cumulative Distribution Function (LSCDF): An Alternative to Spatial Correlation Analysis." *Geographical Analysis* 33.