eiCompare: Comparing Ecological Inference Estimates across EI and EI:R×C

by Loren Collingwood, Kassra Oskooii, Sergio Garcia-Rios, and Matt Barreto

Abstract Social scientists and statisticians often use aggregate data to predict individual-level behavior because the latter are not always available. Various statistical techniques have been developed to make inferences from one level (e.g., precinct) to another level (e.g., individual voter) that minimize errors associated with ecological inference. While ecological inference has been shown to be highly problematic in a wide array of scientific fields, many political scientists and analysis employ the techniques when studying voting patterns. Indeed, federal voting rights lawsuits now require such an analysis, yet expert reports are not consistent in which type of ecological inference is used. This is especially the case in the analysis of racially polarized voting when there are multiple candidates and multiple racial groups. **eiCompare** was developed to easily assess two of the more common ecological inference methods: the EI method developed by King (1997), and the EI:R×C method developed by Rosen et al. (2001); Lau et al. (2006). The package facilitates a seamless comparison between these methods so that scholars and legal practitioners can easily assess the two methods and whether they produce similar or disparate findings.

Introduction

Ecological inference is a widely debated methodology for attempting to understand individual, or micro behavior from aggregate data. Ecological inference has come under fire for being unreliable, especially in the fields of biological sciences, ecology, epidemiology, public health and many social sciences. For example, Freedman (1999) explains that when confronted with individual level data, many ecological aggregate estimates in epidemiology have been proven to be wrong. In the field of ecology Martin et al. (2005) expose the problem of zero-inflation in studies of the presence or absence of specific species of different animals and note that ecological techniques can lead to incorrect inference. Greenland (2001) describes the many pitfalls of ecological inference in public health due to the nonrandomization of social context across ecological units of analysis. Elsewhere, Greenland and Robins (1994) have argued that the problem of ecological confounder control leads to biased estimates of risk in epidemiology. Related, Frair et al. (2010) argue that while some ecological analysis can be informative when studying animal habitat preference, existing methods of ecological inference provide imprecise information on variation in the outcome variables and that considerable improvements are necessary. Wakefield (2004) provides a nice comparison of how ecological inference performs across epidemiological versus social scientific research. He concludes that in epidemiological applications individual-level data are required for consistently accurate statistical inference.

However, within the narrow subfield of racial voting patterns in American elections ecological inference is regularly used. This is especially common in scholarly research on the voting rights act where the United States Supreme Courts directly recommended ecological inference analysis as the main statistical method to estimate voting preference by racial group (e.g. *Thornburg v. Gingles 478 U.S. 30, 1986*). Because Courts in the U.S. have so heavily relied on ecological inference, it has gained prominence in political science research. The American Constitution Society for Law and Policy explains that ecological inference is one of the three statistical analyses that must be performed in voting rights research on racial voting patterns.[1] As ecological inference evolved a group of scholars developed the **eiPack** for the software R (R Core Team, 2015) and published an article in *R News* announcing the new package (Lau et al., 2006).

This article does not conclude that ecological inference is appropriate or reliable outside the specific domain of American elections. Indeed, scholars in the fields of epidemiology and public health have correctly pointed out the limitations of individual level inference from aggregate date. However, its application to voting data in the United States represents one area where it may have utility, if model assumptions are met (Tam Cho and Gaines, 2004). Indeed, the main point of our article is not to settle the debate on the accuracy of ecological inference in the sciences writ large, but rather to assess the degree of similarity or difference with respect to two heavily used R packages within the field of political science, **ei** and **eiPack**. Our package, **eiCompare** offers scholars who regularly use ecological inference in analyses of voting patterns the ability to easily compare, contrast and diagnose estimates across two different ecological methods that are recommended statistical techniques in voting rights litigation.

Today, although there is continued debate among social scientists (Greiner, 2007, 2011; Cho, 1998) -

[1] http://www.acslaw.org/sites/default/files/VRI_Guide_to_Section_2_Litigation.pdf

the courts generally rely on two statistical approaches to ecological data. The first, ecological inference (EI), developed by King (1997) is said to be preferred when there are only two racial or ethnic groups, and ideally only two candidates contesting office. However, Wakefield (2004) notes that EI methods can be improved with the use of survey data as Bayesian priors. The second, ecological inference R × C (R×C) developed by Rosen et al. (2001), is said to be preferred when there are multiple racial or ethnic groups, or multiple candidates contesting office. However, it is not clear that when faced with the exact same dataset, they would produce different results. In one case, analysis of the same dataset across multiple ecological approaches found they tend to produce the same conclusion (Grofman and Barreto, 2009). However, others have argued that using King's EI iterative approach with multiple racial groups or multiple candidates will fail and should not be relied on (Ferree, 2004). Still others have gone further and stated that EI cannot be used to analyze multiple racial group or multiple candidate elections, stating that "it biases the analysis for finding racially polarized voting," going on to call this approach "problematic and no valid statistical inferences can be drawn" (Katz, 2014).

As with any methodological advancement, there is a healthy and rigorous debate in the literature. However, very little real election data has been brought to bear in this debate. Ferree (2004) offers a simulation of Black, White, and Latino turnout and voting patterns, and then examines real data from a parliamentary election in South Africa using a proportional representation system. (Grofman and Barreto, 2009) compare an exit poll to precinct election data in Los Angeles, but only compare Goodman's ecological regression against King's EI, using the single-equation versus double-equation approach, and do not examine the R×C approach at all.

Debates Over Ecological Inference

The challenges surrounding ecological inference are well documented. Robinson (2009) pointed out that relying on aggregate data to infer the behavior of individuals can result in the ecological fallacy, and since then scholars have applied different methods to discern more accurately individual correlations from aggregate data. Goodman (1953, 1959) advanced the idea of ecological regression where individual patterns can be drawn from ecological data under certain conditions. However Goodman's logic assumed that group patterns were consistent across each ecological unit, and in reality that may not be the case.

Eventually, systematic analysis revealed that these early methods could be unreliable (King, 1997). Ecological inference is King's (1997) solution to the ecological fallacy problem inherent in aggregate data, and since the late 1990s has been the benchmark method courts use in evaluating racial polarization in voting rights lawsuits, and has been used widely in comparative politics research on group and ethnic voting patterns. Critics claim that King's EI model was designed primarily for situations with just two groups (e.g., blacks and whites; Hispanics and Anglos, etc.). While many geographic areas (e.g., Mississippi, Alabama) still contain essentially two groups and hence pose no threat to traditional EI estimation procedures, the growth of racial groups such as Latinos and Asians have challenged the historical biracial focus on race in the United States (thereby challenging traditional EI model assumptions). Rosen et al. (2001) suggest a rows by columns (R×C) approach which allows for multiple racial groups, and multiple candidates; however, their Bayesian approach suffered computational difficulties and was not employed at a mass level. Since then, computing power has steadily improved, making R×C a realistic solution for many scenarios and accessible packages now exist in R that are widely used. These two methodological approaches are now both regularly used in political science; however, there is no consistent evidence how they perform side-by-side, and are different.

Ferree (2004) critiques King's EI model, arguing that the conditions for iterative estimation (e.g., black vs. non black, white vs. non-white, Hispanic vs. non-Hispanic) can be considerably biased due to aggregation bias and multimodality in the data. In a hypothetical simulation dataset, Ferree shows that combining blacks and whites into a single "non-Hispanic" group in order to estimate Hispanic turnout can vastly overestimate Hispanic turnout, for example. However, the analysis did not provide any clues as to the specific conditions when and how R×C is significantly better or preferred to EI. For example, if there are three racial groups in equal thirds of the electorate, does aggregation bias create more error in EI than a scenario in which two dominant groups comprise 90% and a small group is just 10%? Likewise, is EI's iterative approach to candidates more stable when analyzing three candidates and far less stable when eight candidates contest the election? These questions have not been considered empirically. Instead, the existing scholarship uses simulation data to prove theoretically that EI might create bias and that R×C is preferred. We argue that real election data should be considered in a side-by-side comparison.

Despite some critiques, other political scientists have defended ecological inference and even ecological regression using both simulations and real data. Owen and Grofman (1997) assess whether or not ecological fallacy in ecological regression is a theoretical problem only, a real problem for empirical analysis. In an extensive review, Owen and Grofman conclude that despite the valid theoretical concerns, linear ecological regression still holds up and provides meaningful and accurate estimates of racially polarized voting. A decade later, Grofman and Barreto (2009) again take up the question of how ecological models compare to one another using a combination of simulation, actual election precinct data, and an accompanying individual-level exit poll. Their analysis argues that there is general consistency across all ecological models and that once voter turnout rates are accounted for, ecological regression and King's EI lead scholars to the same results. However, Grofman and Barreto did not consider R×C in their comparison.

Greiner and Quinn (2010) combine R×C methods with individual level exit poll data, and argue that this hybrid model can be preferable to a straight aggregation model. However, using exit poll data is not always available to all researchers and practitioners. Indeed, in most county or city elections, exit poll data does not exist which is why scholars often attempt to infer voting patterns through aggregate data. Herron and Shotts (2003) also criticize EI estimates when used for second-stage regression - given that error is baked into the second-level regression estimation. However Adolph and King (2003) respond by adjusting the EI procedure to reduce inconsistencies when estimating second-stage regressions. Nevertheless, these issues with EI do not speak specifically to R×C methods.

Greiner and Quinn (2009) extend the 2x2 EI contingency problem to 3x3 and estimate voting preferences simultaneously for three candidates across three racial groups (but using counts instead of percentages). We extend this work by analyzing real-world datasets with sizes greater than 3x3 (multiple candidates and at least three racial groups). In all of this, our main goal is to assess whether using iterative EI or simultaneous R×C approaches change the conclusions social scientists can make from the data.

Finally, some have gone even further in arguing that EI is ill-equipped to handle complex datasets with multiple candidates and multiple racial groups, and that only R×C can produce reliable results (Katz, 2014). In explaining the theoretical reasons why EI cannot accurately process such elections Katz argues "adding additional groups and vote choices to King's (1997) EI is not straightforward," and also adds "given the estimation uncertainty, it may not be possible to infer which candidate is preferred by members of the group." The argument against EI in multiple racial group, or especially multiple candidate elections is that EI takes an iterative approach pitting candidate A versus all others who are not candidate A. If the election features four candidates (A, B, C, D) critics state that you cannot accurately estimate vote choice quantities if you compare the vote for candidate A against the combined vote for B, C, D. The iterative approach would then move on to estimate the vote share for candidate B against the combined vote for A, C, D and so on, so that four separate equations are run. Katz (2014) claims that EI biases the findings in favor of bloc-voting stating "this jerry rigged approach to dealing with more than two vote choices stacks the deck in favor of finding statistical evidence for racially polarized." Given these debates, our package allows scholars to quite easily make side-by-side comparisons and evaluate these competing claims.

While important advancements have been made in ecological inference techniques by King (1997) and Rosen et al. (2001) there is no consistency in which technique is used and how results are presented. What's more, legal experts and social scientists often argue during voting rights lawsuits that one technique is superior to the other, or that their results are more accurate. There is no question that both social scientists and legal experts would greatly benefit from a standardized software package that presents both ecological inference results (EI and R×C) simultaneously and metrics to compare each set of results. Thus, **eiCompare** was designed to compare the most commonly used methods today, EI and R×C, but also incorporates Goodman methods. The package lets analysts seamlessly assess whether EI and R×C estimates are similar (see King (1997) and Rosen et al. (2001) for a methodological description of the techniques). It incorporates functions from **ei** (King and Roberts, 2013) and **eiPack** (Lau et al., 2012) into a new package that relatively quickly compares ecological inference estimates across the two routines.

The package includes several functions that ultimately produce tables of results from the different ecological inference methods. Thus, in the case of racially polarized voting, analysts can quickly assess whether different racial groups preferred different candidates, according to the EI, R×C, and Goodman approaches. **eiCompare** wraps the `ei` procedure (King and Roberts, 2012) into a generalized function, has a variety of table-making functions, and a plotting method that graphically depicts the difference between estimates for the two main EI methods (EI and R×C). Below, we use a working example of a voter precinct dataset in Corona, CA. To use the package, the process is simple: 1) Load the package, the appropriate data, run the EI generalized function, and create an EI table of results, 2) Run the R×C function (from **eiPack**) and create a table of results, 3) Run the Goodman regression generalized function if the user chooses, 4) Combine the results of all the algorithms together into a comparison table, and 5) Plot the comparison results. Before we conclude, we also compare EI and R×C findings against exit poll data from a 2005 Los Angeles mayoral run-off election. The rest of the paper follows this aforementioned outline.

1. EI Generalize

To begin, we install (`install.packages(``eiCompare'')`) and load the **eiCompare** package (`library(eiCompare)`) from the CRAN repository. First, we load the aggregate-level dataset (`data(cor_06)`) into `R`, in this case a precinct (voting district) dataset from a 2006 election in the city of Corona, CA. Table 1 below displays the first five rows and column headers of the dataset. This dataset includes all the necessary variables to run the code in the `eiCompare` package. The first column is `precinct`, which essentially operates as a unique identifier. The second column, `totvote`, is the total number of votes cast within the precinct. Columns three and four are the two racial groups of whom we seek to determine their mean voting preference. The rest of the columns are the percent of the total vote for each respective candidate.

	precinct	totvote	pct_latino	pct_other	pct_breitenbucher	pct_montanez	pct_spiegel	pct_skipworth
1	22000	942	0.21	0.79	0.20	0.21	0.29	0.30
2	22002	1240	0.16	0.84	0.22	0.22	0.29	0.27
3	22003	1060	0.21	0.79	0.22	0.22	0.30	0.26
4	22004	1280	0.45	0.55	0.18	0.27	0.30	0.24
5	22008	1172	0.31	0.69	0.23	0.25	0.30	0.22
6	22012	1093	0.21	0.79	0.20	0.24	0.32	0.24

Table 1: Precinct dataset of Corona, CA, used for ecological inference. Each row is a precinct, the dataset must have a total column, racial/ethnic percentages of people living in the precinct, and vote percent for each candidate.

We are interested in how the four candidates (Breitenbucher, Montanez, Spiegel, Skipworth) performed with Latino voters and non-Latino voters (mostly non-Hispanic white), so we can asses whether racially polarized voting exists. The process begins with the `ei_est_gen()` function, which is a generalized version of the `ei` function from the **ei** package. Instead of having to estimate EI results for each candidate and each racial group separately, `ei_est_gen()` automates this process.

The `ei_est_gen()` function takes a vector of candidate names (e.g., `c("pct_breitenbucher", "pct_montanez","pct_spiegel","pct_skipworth")`), a character vector of racial group names with a tilde in front of the variable name (e.g., `c("∼pct_latino","∼pct_other")`), a character string of the name of the total column (`"totvote"`) representing the total number of people in the jurisdiction (e.g., registered voters, ballots cast) that is passed to the `ei` function, a `data` call for the `data.frame()` object where the data are stored, and a character string of `table_names` (e.g., `c("EI: Pct Lat","EI: Pct Other")`) that are used to display the results. The function also has four default arguments, `rho`, `sample`, `tomog`, and `density_plot`. The former two can be used to adjust the parameters of the `ei` algorithm. These are especially useful when the initial run does not compile or warnings are produced. The latter two plot out tomography and density plots, respectively into the working directory but are default set to off. These plots can be used to assess the stability – and thus veracity – of the EI procedure (see King and Roberts (2012) and King (1997) for details).

Finally, the `...` argument passes additional arguments onto the `ei()` function from the **ei** package. One final note, given its iterative nature, the `ei_est_gen()` function can take a while to execute. This typically depends on features unique to the dataset, including the number of candidates and groups, the amount of racial/ethnic segregation within the city/area, as well as the number of precincts. This particular example does not take especially long, executing in about a minute on a standard Macbook pro.

```
# LOAD DATA
data(cor_06)
# SET SEED FOR REPRODUCIBILITY
set.seed(294271)
# CREATE CHARACTER VECTORS REQUIRED FOR FUNCTION
cands <- c("pct_breitenbucher","pct_montanez","pct_spiegel", "pct_skipworth")
race_group2 <- c("~ pct_latino", "~ pct_other")
table_names <- c("EI: Pct Lat", "EI: Pct Other")
# RUN EI GENERALIZED FUNCTION
results <- ei_est_gen(cand_vector=cands, race_group = race_group2,
                    total = "totvote", data = cor_06, table_names = table_names)
# LOOK AT TABLE OF RESULTS
results
```

The call to the results object produces a table of results indicating the mean estimated voting preferences for Latinos and non-Latinos within the city of Corona (see Table 2). The results strongly suggest the presence of racially polarized voting, as Latinos prefer Montanez as their number one choice, whereas non-Latinos do not.

Candidate	EI: Pct Lat	EI: Pct Other
pct_breitenbucher	19.68	21.12
se	0.75	0.13
pct_montanez	35.95	20.13
se	0.03	0.08
pct_spiegel	28.43	31.01
se	0.57	0.23
pct_skipworth	18.64	26.84
se	0.71	0.23
Total	102.69	99.10

Table 2: EI mean estimates for Latino and Non-Latino candidate vote preferences in Corona, 2006

2. EI: R×C

The R×C builds off of code from the **eiPack** package, where **eiCompare** simply takes the former's results and puts them into a similar `data.frame()/table()` object similar to the results from the `ei_est_gen()` function. First, the user follows the code from the **eiPack** package (here we use the `ei.reg.bayes()` function), and creates a formula object including all candidates and all groups. The user must ensure that the percentages on both signs of the ∼ symbol add to 1. Thus, the initial `table()` code is a simple data check to ensure that this rule is followed. The R×C model is then run using the `ei.reg.bayes()` model. Users can read the **eiPack** documentation to familiarize themselves with this procedure. Depending on the nature of one's data, the R×C code can take a while to run. Finally, the results are passed onto the `bayes_table_make()` function, along with a vector of candidate names, and a vector of table names, similar to what was passed to `ei_est_gen()`.

```
# CHECK TO MAKE SURE DATA SUMS TO 1 FOR EACH PRECINCT
with(cor_06, pct_latino + pct_other)
with(cor_06, pct_breitenbucher + pct_montanez + pct_spiegel + pct_skipworth)
# SET SEED FOR REPRODUCIBILITY
set.seed(124271)
#RxC GENERATE FORMULA
form <- formula(cbind(pct_breitenbucher,pct_montanez,
        pct_spiegel, pct_skipworth) ~ cbind(pct_latino, pct_other))
# RUN EI:RxC MODEL
ei_bayes <- ei.reg.bayes(form, data=cor_06, sample=10000, truncate=T)
# CREATE TABLE NAMES
table_names <- c("RxC: Pct Lat", "RxC: Pct Other")

# TABLE CREATION
ei_bayes_res <- bayes_table_make(ei_bayes, cand_vector= cands, table_names = table_names)
# LOOK AT TABLE OF RESULTS
ei_bayes_res
```

Candidate	RxC: Pct Lat	RxC: Pct Other
pct_breitenbucher	18.22	21.58
se	1.62	0.53
pct_montanez	34.96	20.44
se	1.72	0.56
pct_spiegel	28.24	31.05
se	1.08	0.35
pct_skipworth	18.61	26.91
se	1.73	0.56
Total	100.03	99.99

Table 3: EI:R×C mean estimates for Latino and Non-Latino candidate vote preferences in Corona, 2006

The results are presented in Table 3, and look remarkably similar to those presented in Table 2. Indeed, the exact same conclusions would be drawn from an analysis of both tables: Latinos prefer Montanez as their first choice and non-latinos prefer Spiegel as their top choice.

3. Goodman Generalize

While many users will skip over the Goodman regression when conducting ecological inference, given the documented issues with the method (Shively, 1969; King, 1997), **eiCompare** nevertheless has a Goodman regression generalized function, similar to the `ei_est_gen()` function. This function takes a character vector of candidate names, a character vector of racial groups, the name of the column, a data object, and a character vector of table names. Because Goodman is simply a linear regression, the execution is very fast.

```
table_names <- c("Good: Pct Lat", "Good: Pct Other")
good <- goodman_generalize(cands, race_group2, "totvote", cor_06, table_names)
good
```

Table 4 shows the Goodman regression results. In this particular case, these results align quite closely with results from the two EI models. All three approaches essentially tell us the same thing.

Candidate	Good: Pct Lat	Good: Pct Other
pct_breitenbucher	17.51	20.34
se	3.18	3.74
pct_montanez	35.00	20.48
se	3.41	4.01
pct_spiegel	28.52	31.61
se	2.16	2.54
pct_skipworth	18.97	27.57
se	3.45	4.05
Total	100.00	100.00

Table 4: Goodman regression estimates for Latino and Non-Latino candidate vote preferences in Corona, 2006

4. Combining Results

The last two sections address the comparison component of the package. The function, `ei_rc_good_table()`, takes the objects from the EI, R×C, and Goodman regression, and puts them into a `data.frame()/table()` object. To simplify comparison, the table adds an EI-R×C column differential for each racial group. This format lets the user quickly assess how the EI and R×C methods stack up against one another. The function takes the following arguments: ei results object (e.g., `results`), an R×C object (e.g., `ei_bayes_res`), and a character vector `groups` (e.g., `c("Latino","Other")`) argument. The `good` argument for the Goodman regression is set to Null, and the `include_good` argument defaults to FALSE. If the user wants to include a Goodman regression in the comparison of results they need to change the latter to TRUE and specify the the `good` argument as the object name from the `goodman_generalize()` call.

Candidate	EI: Pct Lat	RxC: Pct Lat	EI_Diff	EI: Pct Other	RxC: Pct Other	EI_Diff
pct_breitenbucher	19.68	18.22	-1.46	21.12	21.58	0.46
se	0.75	1.62		0.13	0.53	
pct_montanez	35.95	34.96	-0.99	20.13	20.44	0.31
se	0.03	1.72		0.08	0.56	
pct_spiegel	28.43	28.24	-0.19	31.01	31.05	0.04
se	0.57	1.08		0.23	0.35	
pct_skipworth	18.64	18.61	-0.02	26.84	26.91	0.07
se	0.71	1.73		0.23	0.56	
Total	102.69	100.03	-2.66	99.10	99.99	0.88

Table 5: EI and R×C comparisons for Latino and Non-Latino candidate vote preferences in Corona, 2006

The results of `ei_rc_good_table()` is a new class `ei_compare`, which includes a `data.frame()` and groups character vector. This output is ultimately passed to `plot()`.

```
ei_rc_combine <- ei_rc_good_table(results, ei_bayes_res,
                        groups= c("Latino", "Other"))
ei_rc_combine@data
```

```
ei_rc_g_combine <- ei_rc_good_table(results, ei_bayes_res, good,
                          groups= c("Latino", "Other"), include_good=T)
ei_rc_g_combine
```

Table 5 displays the output of a call to the `ei_rc_good_table()` function for the first line of code above. The user must include the code `@data` onto the outputted table name to extract just the table. This table basically summarizes the results of the EI and R×C analyses. Clearly, very little difference emerges between the two methods in this particular instance.

5. Plotting Results

Finally, users can plot the results of the EI, and R×C comparison to more visually determine whether the two methods are similar. Plotting is simple, as plot methods have been developed for the `ei_compare` class. The code below produces the plot depicted in Figure 1.

```
# PLOT COMPARISON -- adjust the axes labels slightly
plot(ei_rc_combine, cex.axis=.5, cex.lab=.7)
```

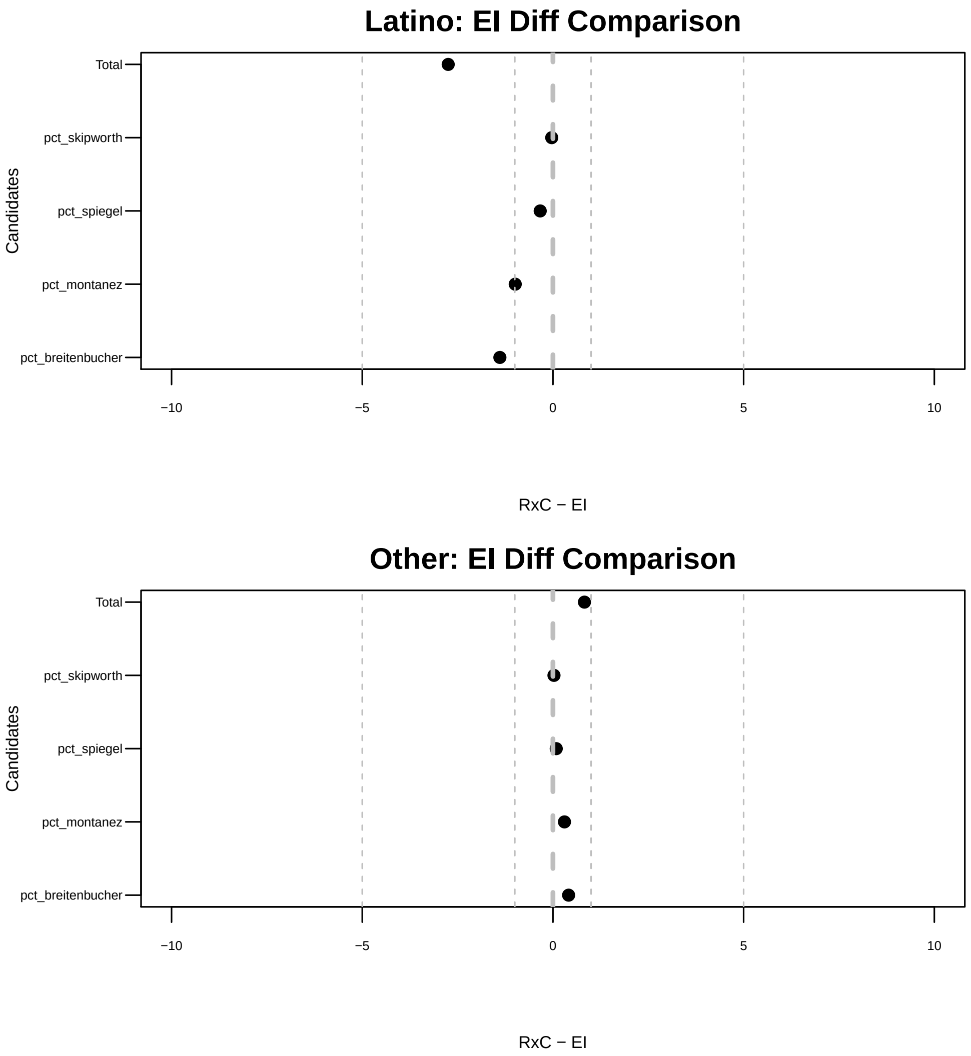

Figure 1: Comparison of EI and R×C methods for Corona 06 precinct data

Comparing Ecological and Individual-Level Data

One possible question remains, whether or not ecological estimates line up with individual level estimates. Many studies have pointed out that ecological fallacy and aggregation bias can produce ecological inference results that are highly questionable. In this section we implement the **eiCompare** package for a mayoral election in a multiethnic setting in which an individual-level exit poll survey was also administered. The **eiCompare** package provides EI and R×C results for the 2005 Los Angeles mayoral runoff election between Antonio Villaraigosa and James Hahn, and we also add results for the Los Angeles Times exit poll. Results are displayed in Table 6.

	EI: AV	EI: JH	RxC: AV	RxC: JH	Exit: AV	Exit: JH	MOE
White	45	54	48	52	50	50	+/- 2.5
Black	58	40	50	50	48	52	+/-4.2
Latino	82	17	81	19	84	16	+/-3.6
Asian	48	51	47	53	44	56	+/-6.1

Table 6: Percent voting for Antonio Villaraigosa (AV) and James Hahn (JH) by ethnic group. Comparison between EI, R×C, and exit poll methods, Los Angeles mayoral election runoff, May 2005. Exit poll taken from Los Angeles Times.

The results presented in Table 6 demonstrate that not only do EI and R×C produce remarkably consistent results, but they very closely match the individual level estimates for the Los Angeles Times. The EI R×C estimates are all with the confidence range of the individual level data reported by the exit poll.

Summary

eiCompare is a new package that builds on the work of King and others that attempts to address the ecological inference problem of making individual-level assessments based on aggregate-level data. As we have reviewed above, there is considerable debate in the sciences about the utility and accuracy of ecological techniques. Despite these well documented questions, ecological inference is widely used in political science and will continue to grow in importance when the constitutionally mandated redistricting in 2021 occurs. The redistricting cycle will bring with it extensive academic, legislative, and legal research using ecological inference to assess racial voting patterns across all 50 states.

While this new package does not develop a new method, *per se*, it improves analysts' ability to quickly compare different commonly used EI algorithms to assess the veracity of the methods and also produce tables of their findings. While R×C has been touted as the method necessary in situations with multiple groups and multiple candidates, the results do not always demonstrate face validity. In these scenarios – and others – analysts may want to incorporate original EI methods so they can compare how the two approaches stack up. Ultimately, this approach provides a needed assessment between two commonly used methods in voting behavior research.

Bibliography

C. Adolph and G. King. Analyzing second-stage ecological regressions: Comment on Herron and Shotts. *Political Analysis*, 11(1):65–76, 2003. [p3]

W. K. T. Cho. Iff the assumption fits?: A comment on the King ecological inference solution. *Political Analysis*, 7(1):143–163, 1998. [p1]

K. E. Ferree. Iterative approaches to R×C ecological inference problems: where they can go wrong and one quick fix. *Political Analysis*, 12(2):143–159, 2004. [p2]

J. L. Frair, J. Fieberg, M. Hebblewhite, F. Cagnacci, N. J. DeCesare, and L. Pedrotti. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1550):2187–2200, 2010. [p1]

D. A. Freedman. Ecological inference and the ecological fallacy. *International Encyclopedia of the social & Behavioral sciences*, 6:4027–4030, 1999. [p1]

L. A. Goodman. Ecological regressions and behavior of individuals. *American sociological review*, 1953. [p2]

L. A. Goodman. Some alternatives to ecological correlation. *American Journal of Sociology*, pages 610–625, 1959. [p2]

S. Greenland. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International journal of epidemiology*, 30(6):1343–1350, 2001. [p1]

S. Greenland and J. Robins. Invited commentary: ecologic studies — biases, misconceptions, and counterexamples. *American Journal of Epidemiology*, 139(8):747–760, 1994. [p1]

D. J. Greiner. Ecological inference in voting rights act disputes: Where are we now, and where do we want to be? *Jurimetrics*, pages 115–167, 2007. [p1]

D. J. Greiner. The quantitative empirics of redistricting litigation: Knowledge, threats to knowledge, and the need for less districting. *Yale Law & Policy Review*, 29(2):527–542, 2011. [p1]

D. J. Greiner and K. M. Quinn. Exit polling and racial bloc voting: Combining individual-level and R×C ecological data. *The Annals of Applied Statistics*, pages 1774–1796, 2010. [p3]

J. D. Greiner and K. M. Quinn. R× c ecological inference: bounds, correlations, flexibility and transparency of assumptions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172 (1):67–81, 2009. [p3]

B. Grofman and M. A. Barreto. A reply to Zax's (2002) critique of Grofman and Migalski (1988) double-equation approaches to ecological inference when the independent variable is misspecified. *Sociological Methods & Research*, 37(4):599–617, 2009. [p2, 3]

M. C. Herron and K. W. Shotts. Using ecological inference point estimates as dependent variables in second-stage linear regressions. *Political Analysis*, 11(1):44–64, 2003. [p3]

J. N. Katz. Expert report on voting in the city of Whittier. March 5, 2014. [p2, 3]

G. King. A solution to the ecological inference problem, 1997. [p1, 2, 3, 4, 6]

G. King and M. Roberts. Ei: a (n r) program for ecological inference. *Harvard University. Retrieved from http://gking. harvard. edu/files/ei. pdf*, 2012. [p3, 4]

G. King and M. Roberts. *ei: ei*, 2013. URL `http://gking.harvard.edu/zelig`. R package version 1.3. [p3]

O. Lau, R. T. Moore, and M. Kellermann. eipack: R×C ecological inference and higher-dimension data management. *New Functions for Multivariate Analysis*, 18(1):43, 2006. [p1]

O. Lau, R. T. Moore, and M. Kellermann. *eiPack: eiPack: Ecological Inference and Higher-Dimension Data Management*, 2012. URL `http://CRAN.R-project.org/package=eiPack`. R package version 0.1-7. [p3]

T. G. Martin, B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology letters*, 8(11):1235–1246, 2005. [p1]

G. Owen and B. Grofman. Estimating the likelihood of fallacious ecological inference: linear ecological regression in the presence of context effects. *Political Geography*, 16(8):675–690, 1997. [p2]

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`. [p1]

W. S. Robinson. Ecological correlations and the behavior of individuals. *International journal of epidemiology*, 38(2):337–341, 2009. [p2]

O. Rosen, W. Jiang, G. King, and M. A. Tanner. Bayesian and frequentist inference for ecological inference: The R×C case. *Statistica Neerlandica*, 55(2):134–156, 2001. [p1, 2, 3]

W. P. Shively. "Ecological" inference: the use of aggregate data to study individuals. *American Political Science Review*, 63(04):1183–1196, 1969. [p6]

W. K. Tam Cho and B. J. Gaines. The limits of ecological inference: The case of split-ticket voting. *American Journal of Political Science*, 48(1):152–171, 2004. [p1]

J. Wakefield. Ecological inference for 2× 2 tables (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(3):385–445, 2004. [p1, 2]

Loren Collingwood
Department of Political Science
University of California, Riverside
900 University Avenue
Riverside, CA 92521
USA
loren.collingwood@ucr.edu

Kassra Oskooii
Department of Political Science
University of Washington
101 Gowen Hall
Seattle, WA 98195
USA
kassrao@uw.edu

Sergio Garcia-Rios
Department of Government
Cornell University
212 White Hall
Ithaca, NY 14853
USA
garcia.rios@cornell.edu

Matt A. Barreto
Department of Political Science
University of California, Los Angeles
Bunche Hall 3284
Los Angeles, CA 90095
USA
barretom@ucla.edu