Sentencing Convicted Felons in the United States: A Bayesian Analysis Using Multilevel Covariates

Iain Pardoe
Lundquist College of Business
University of Oregon
Eugene, OR 97403–1208
(ipardoe@lcbmail.uoregon.edu)

Robert R. Weidner and Richard Frase
Institute on Criminal Justice
University of Minnesota Law School
Minneapolis, MN 55455
(weidn002@tc.umn.edu)

Sentencing Convicted Felons in the United States: A Bayesian Analysis Using Multilevel Covariates

Iain Pardoe, Robert R. Weidner, and Richard Frase

Abstract

There is large variation in levels of imprisonment across the United States, with some states' imprisonment rates six times higher than others. Use of prison in sentencing decisions also varies considerably between counties within states; previous research based on counties as the unit of analysis suggests that variables such as crime rate, unemployment level, racial composition, and geographic region account for some of this variation. Other studies, using individual felons as the unit of analysis, demonstrate how demographics, criminal history, case characteristics, and type of offense affect sentence severity. This paper considers the effects of both county-level and individual-level variables on whether or not a convicted felon receives a prison sentence, rather than a jail or non-custodial sentence. We analyze felony court case processing data for 1996 from 30 of the nation's most populous urban counties using a Bayesian hierarchical logistic regression model.

Key Words: Gibbs sampling; Hierarchical model; Logistic regression; Markov chain Monte Carlo; Mixed effects

Felonies Data

May 1996: 9,110 convictions, 4,358 with complete data, 30 counties, 16 states (Bureau of Justice Statistics' State Court Processing Statistics).

Y = 1 if offender received a prison sentence, 0 for jail or non-custodial sentence. 7-45% of offenders in counties went to prison (ave. 30%).

IMAL = 1 for men, 0 for women.

IBLK = 1 if offender is African American, 0 if not.

Type of offense based on offender's most serious conviction (reference category includes weapons, driving-related, and public order offenses):

- ICVS for murder, rape or robbery;
- ICVM for assault or other violent crime;
- ICTR for a drug trafficking offense;
- ICDR for a drug possession offense; and
- ICPR for burglary or theft, i.e. a property offense.

ICJS = 1 for an active criminal justice status, 0 otherwise.

IPFE = 1 for prior felony convictions, 0 otherwise.

IPMI = 1 for prior misdemeanor convictions, 0 otherwise.

IDET = 1 if offender was detained after being charged, 0 if released.

IREV = 1 if offender's pretrial release was revoked, 0 otherwise.

IBAD = 1 if offender's pretrial release was not revoked after a rearrest.

ITRI = 1 if offender was convicted by trial, 0 if convicted by plea.

CARR = county's arrest rate per 10,000 residents in 1996.

CUNR = county's unemployment rate for 1996.

CBLP = % of county's population that was African American in 1996.

CSTH = 1 if county is located in a Southern state, 0 otherwise.

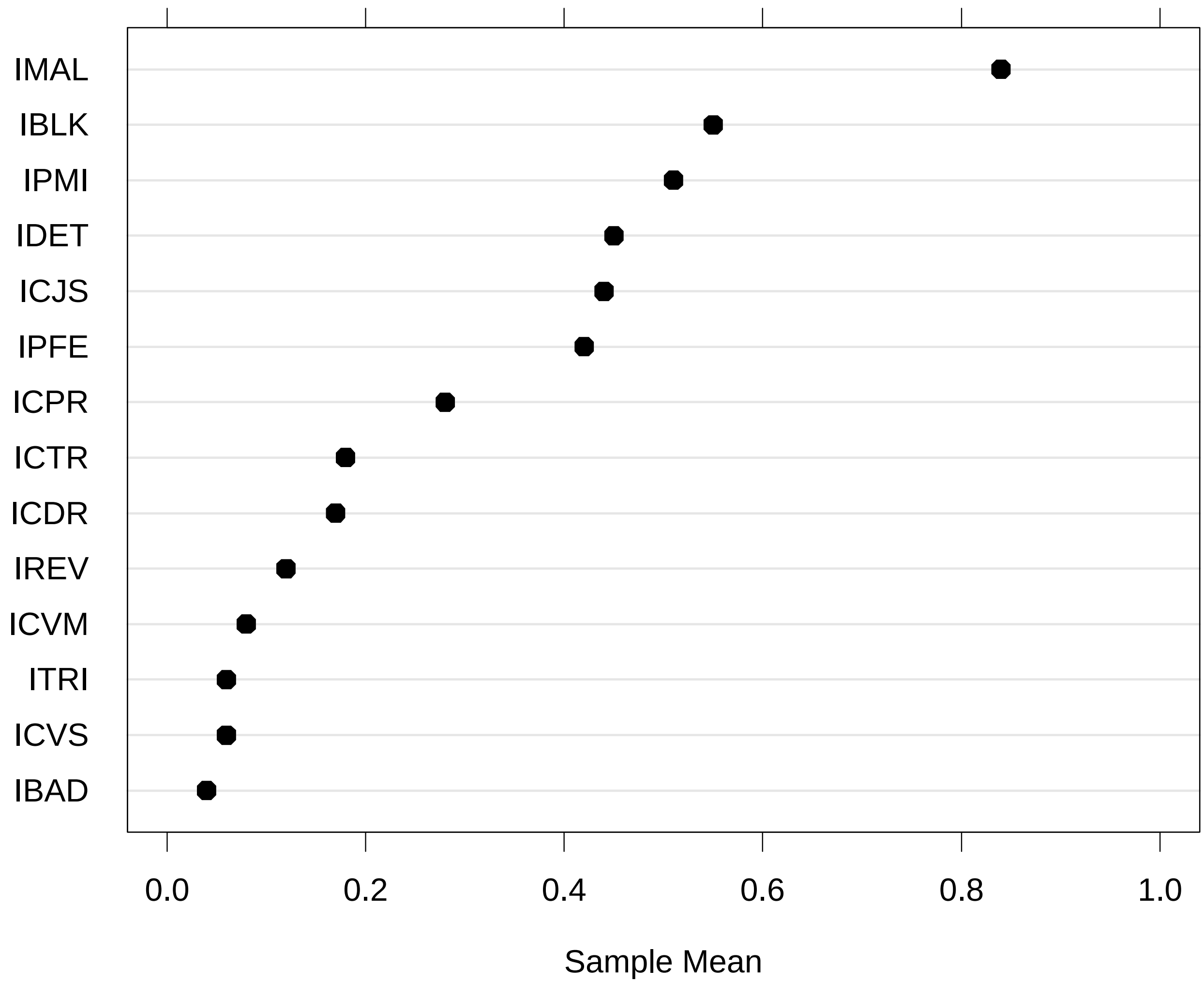

Fig. 1: Sample means for individual-level variables.

Variable	Mean	SD	Min	Max
CARR	582.83	228.04	293	1,349
CUNR	5.62	1.96	3.5	10.6
CBLP	21.85	16.15	1.9	64.8
CSTH	0.20	0.41	0	1

Table 1: Descriptive statistics for county-level variables

Fig. 1 and Table 1 provide univariate summaries of the data.

Hierarchical Logistic Regression Model

We take a Bayesian approach, using a generalization of the model of Wong and Mason (1985). First, the usual logistic regression model is fit to n_j individuals within each of $J = 30$ counties. For the i-th individual in the j-th county, we observe a dichotomous response,

$$Y_{ij} = \begin{cases} 1 & \text{for a prison sentence} \\ 0 & \text{for a non-prison (jail or non-custodial) sentence} \end{cases}$$

Then $Y_{ij}|p_{ij} \sim$ Bernouilli(p_{ij}), where $p_{ij} =$ Pr$(Y_{ij} = 1)$, and

$$\mathrm{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \boldsymbol{X}_i^T \boldsymbol{\beta}_j \tag{1}$$

where $\boldsymbol{X}_i$ respresents K individual-level variables and $\boldsymbol{\beta}_j$ consists of K unknown regression coefficients (specific to the j-th county). Next, to allow the K regression coefficients to be related across counties, assume each coefficient can be explained by up to L county-level variables,

$$\boldsymbol{\beta}_j = \boldsymbol{G}_j \boldsymbol{\eta} + \boldsymbol{\alpha}_j \tag{2}$$

where $\boldsymbol{G}_j$ is a $K \times M$ block-diagonal matrix of L county-level variables, $\boldsymbol{\eta}$ consists of M unknown regression coefficients, and $\boldsymbol{\alpha}_j$ is a $K \times 1$ vector of county-level errors. Combining (1) and (2) leads to

$$\mathrm{logit}(p_{ij}) = \boldsymbol{X}_i^T \boldsymbol{G}_j \boldsymbol{\eta} + \boldsymbol{X}_i^T \boldsymbol{\alpha}_j \tag{3}$$

Conventionally, the η-parameters in (3) are fixed effects while the α-parameters are random effects. Mixed models like this can be fit using specialized computer software such as “MLwiN” and “HLM”. An alternative approach is to put the model into a Bayesian framework.

ESTIMATION

Priors: $\boldsymbol{\eta}$ flat; $\boldsymbol{\alpha}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}^{-1})$, where $\mathbf{0}$ is a K-vector of zeros and $\boldsymbol{\Gamma}^{-1}$ is a $K \times K$ covariance matrix; $\boldsymbol{\Gamma} \sim$ Wishart$(\boldsymbol{R}, K)$, where $\boldsymbol{R}$ can be considered a prior estimate of $\boldsymbol{\Gamma}^{-1}$ based on K observations, and, to represent vague prior knowledge, the Wishart degrees of freedom is set as small as possible to be K.

We used "WinBUGS" to generate posterior $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}_j$ samples. $\boldsymbol{R}$ was specified to have values 0.1 along the diagonal and 0.005 elsewhere. Sensitivity analysis confirmed that the choice of $\boldsymbol{R}$ has little effect on the results. We began with $K = 15$ individual-level and $L = 5$ county-level variables (including intercepts), i.e. 75 terms in total. We ran four chains for 1,500 iterations, discarding 500 burn-in samples from each. With so many η's, many were estimated with considerable imprecision. In particular, nine η's (corresponding to interactions) had posterior SD's at least twice the absolute value of their posterior means; we excluded these interactions from subsequent models. We continued to reduce the number of model terms in this way; our final model included just 30 terms, with all interactions having posterior standard deviations no more than 0.6 times the absolute value of their posterior means.

In generating samples, CARR, CUNR, and CBLP were centered at their sample means and scaled by their sample standard deviations. After running four chains for 13,000 iterations, trace plots showed good mixing and MCMC convergence diagnostics indicated convergence. In particular, the 0.975 quantiles of the "corrected scale reduction factor" (Brooks and Gelman 1998) for the η's were each less than 1.3.

Model Assessment

Before interpreting results, we assessed underlying assumptions of the model. Posterior samples of the $\boldsymbol{\alpha}_j$ can be thought of as residuals, and so lend themselves to the usual kinds of model diagnostics. The fact that they averaged very close to zero across counties is reassuring, but unsurprising. More open to doubt are the normality and exchangeability assumptions. However, normal probability plots revealed no strong abnormalities, and plotting posterior means of the $\boldsymbol{\alpha}_j$ against county-level covariates also revealed no worrisome patterns (plots not shown).

We also carried out a sensitivity analysis for $\boldsymbol{R}$. Specifying $\boldsymbol{R}$ to have values 0.1 along the diagonal and -0.005 elsewhere lead to changes in posterior means for the η's averaging 0.03 in absolute value, with none larger than 0.10. Increasing the elements of $\boldsymbol{R}$ by a factor of 100 lead to changes averaging 0.13 in absolute value, with none larger than 0.37.

Finally, we checked the fit of the model using a generalization of the "Bayes marginal model plot" (BMMP) of Pardoe (2001). Here, the response, Y, is plotted against functions of the predictors, $h(\boldsymbol{X})$. A nonparametric smooth of Y provides a model-free estimate of the mean function in this plot, while a nonparametric smooth of the fitted values provides a comparable model-based estimate. Smooths matching closely for any h provide support for the model; otherwise model inadequacy is indicated. Adding model-based smooths using posterior samples allows this assessment to be made more easily. For example, Fig. 2 is a BMMP with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}}$, where $\hat{\boldsymbol{\eta}}$ is the posterior mean.

The blue smooth of the data passes through the center of the red

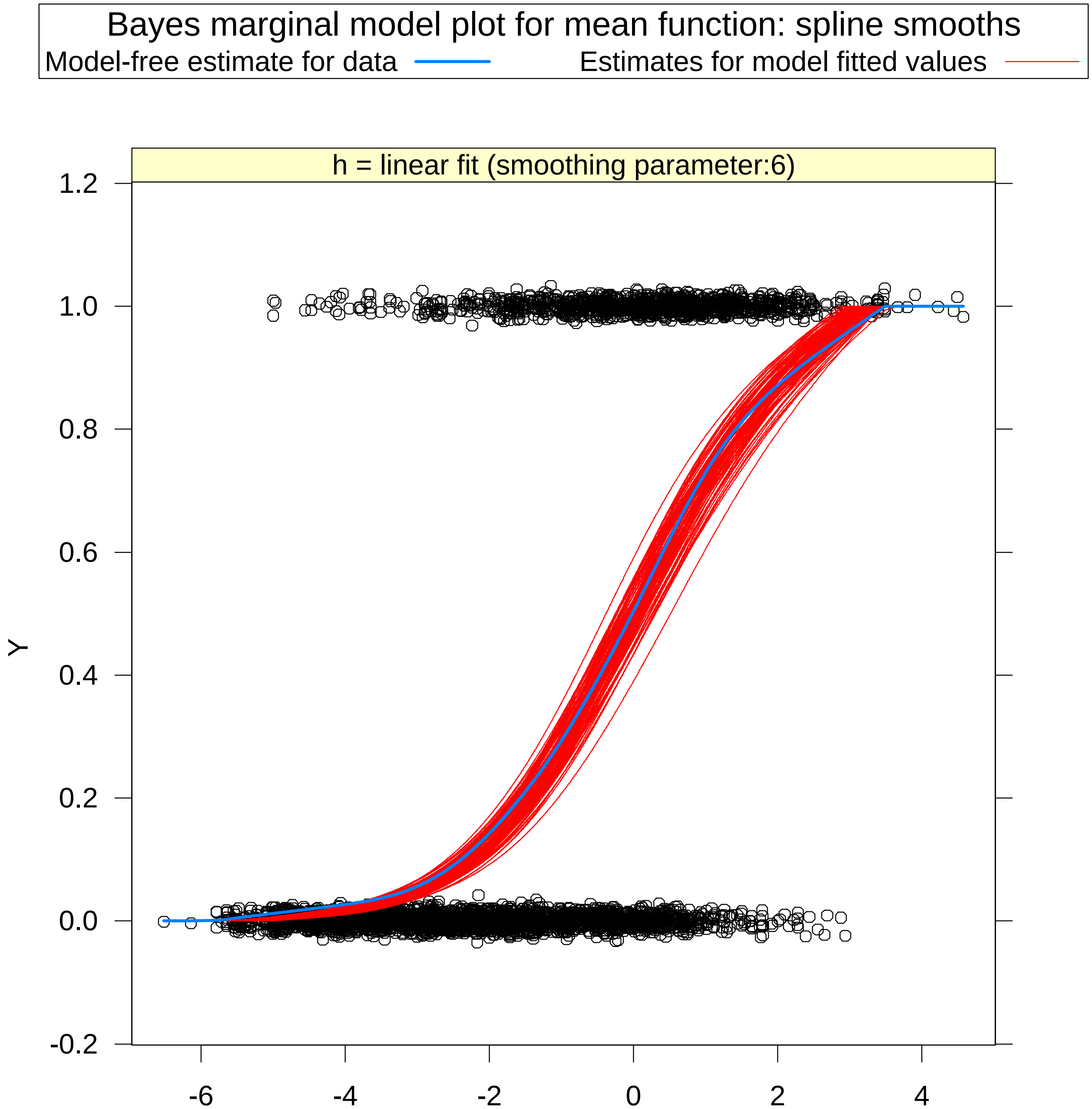

Fig. 2: Bayes marginal model plot with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}}$, where $\hat{\boldsymbol{\eta}}$ is the posterior mean. The data have been jittered to aid visualization of relative density and the smooths are smoothing splines with six effective degrees of freedom.

band of model-based smooths of $1/(1 + \exp(-\boldsymbol{X}_i^T \boldsymbol{G}_j \boldsymbol{\eta}^*))$, where $\boldsymbol{\eta}^*$ are 100 posterior samples. So, there is no indication of lack-of-fit from this plot, or indeed from similar plots with other h-functions. Further discussion of model-checking plots for hierarchical logistic regression is provided in [work to be presented at JSM 2002 in New York!].

RESULTS

Table 2: Posterior summaries for 10,000 samples (500 sample burn-in and only every fifth sample retained), with stars indicating HPD intervals that exclude zero.

Term	Mean	SD	95% HPD interval		exp(Mean)
CARR	0.218	0.130	−0.047	0.467	1.243
CUNR	−0.132	0.163	−0.442	0.196	0.876
CBLP	0.008	0.198	−0.389	0.388	1.008
CSTH	0.847	0.450	−0.028	1.730	2.333
IMAL	0.538	0.163	0.223	0.855*	1.712
IBLK	−0.022	0.121	−0.247	0.224	0.978
ICVS	2.864	0.255	2.362	3.374*	17.532
CUNR.ICVS	0.370	0.207	−0.041	0.781	1.447
ICVM	2.176	0.249	1.695	2.671*	8.811
ICTR	1.941	0.266	1.397	2.462*	6.966
CARR.ICTR	0.383	0.119	0.149	0.618*	1.466
ICDR	0.617	0.330	−0.103	1.228	1.854
ICPR	1.176	0.243	0.704	1.646*	3.241
ICJS	0.514	0.127	0.278	0.779*	1.672
IPFE	0.634	0.143	1.354	1.910*	5.124
CUNR.IPFE	0.501	0.139	0.234	0.782*	1.651
CBLP.IPFE	−0.311	0.150	−0.612	−0.024*	0.733
CSTH.IPFE	−0.663	0.345	−1.322	0.006	0.515
IPMI	−0.077	0.139	−0.343	0.206	0.926
CBLP.IPMI	−0.324	0.137	−0.581	−0.054*	0.724
CSTH.IPMI	0.578	0.334	−0.085	1.219	1.783
IDET	2.261	0.170	1.941	2.610*	9.593
CSTH.IDET	−0.776	0.374	−1.491	−0.029*	0.460
IREV	1.699	0.199	1.284	2.066*	5.468
CSTH.IREV	−1.222	0.587	−2.441	−0.132*	0.295
IBAD	0.565	0.325	−0.090	1.185	1.759
CARR.IBAD	0.559	0.272	0.026	1.101*	1.749
CBLP.IBAD	−1.049	0.450	−1.964	−0.215*	0.350
ITRI	1.191	0.265	0.674	1.715*	3.290

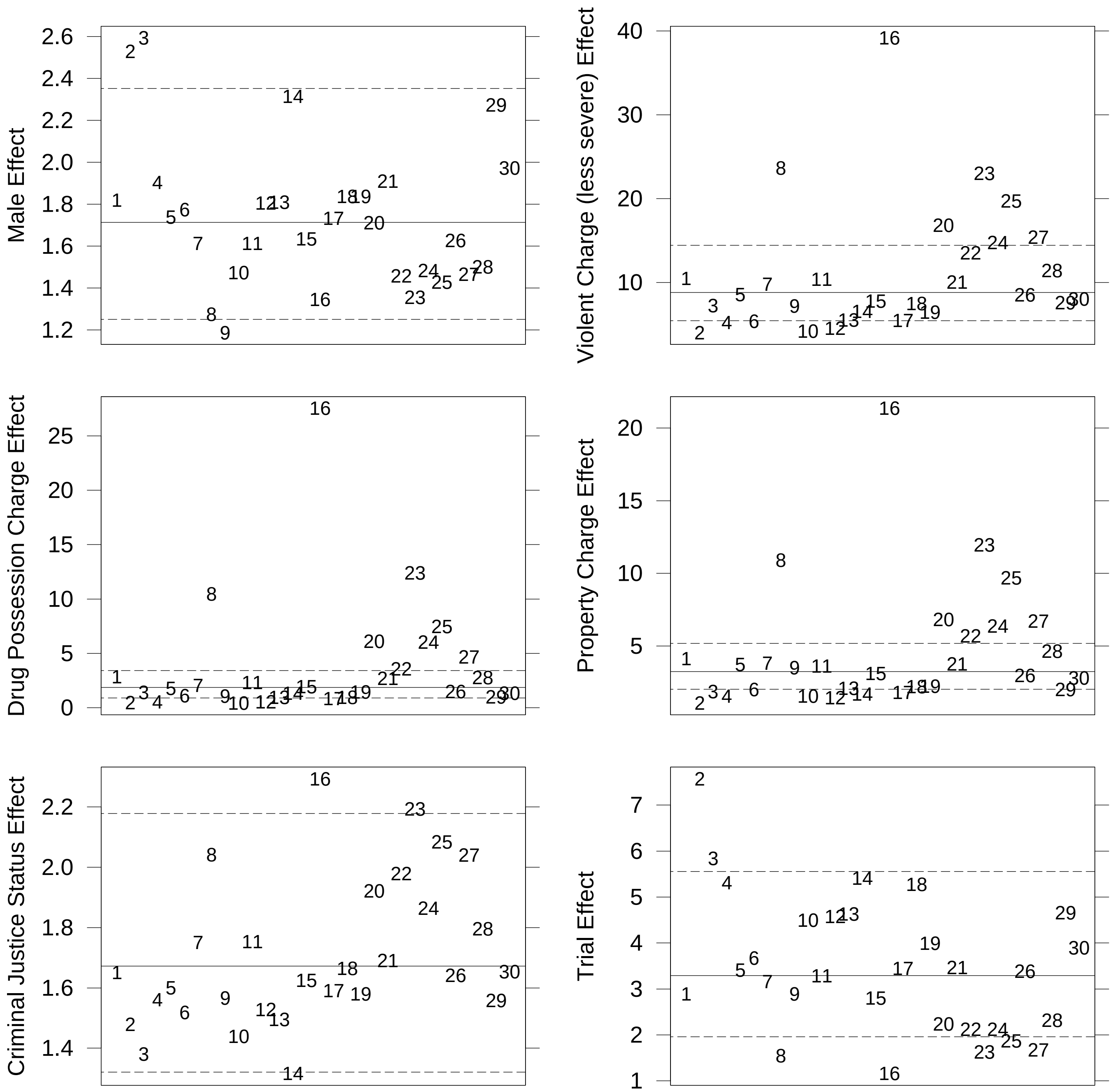

Fig. 3: Estimated odds ratios (solid lines), 95% HPD end-points (dashed lines), and county estimates (numbers), for IMAL, ICVM, ICDR, ICPR, ICJS, *and* ITRI.

Consider IMAL and Fig. 3. Solid line is the estimated odds ratio, i.e. exponentiated posterior mean of the appropriate η. Dashed lines are the exponentiated end-points of the 95% HPD interval. The numbered points are the estimated odds ratios within each county, i.e. exponentiated posterior mean of the sum of the appopriate η and α_j.

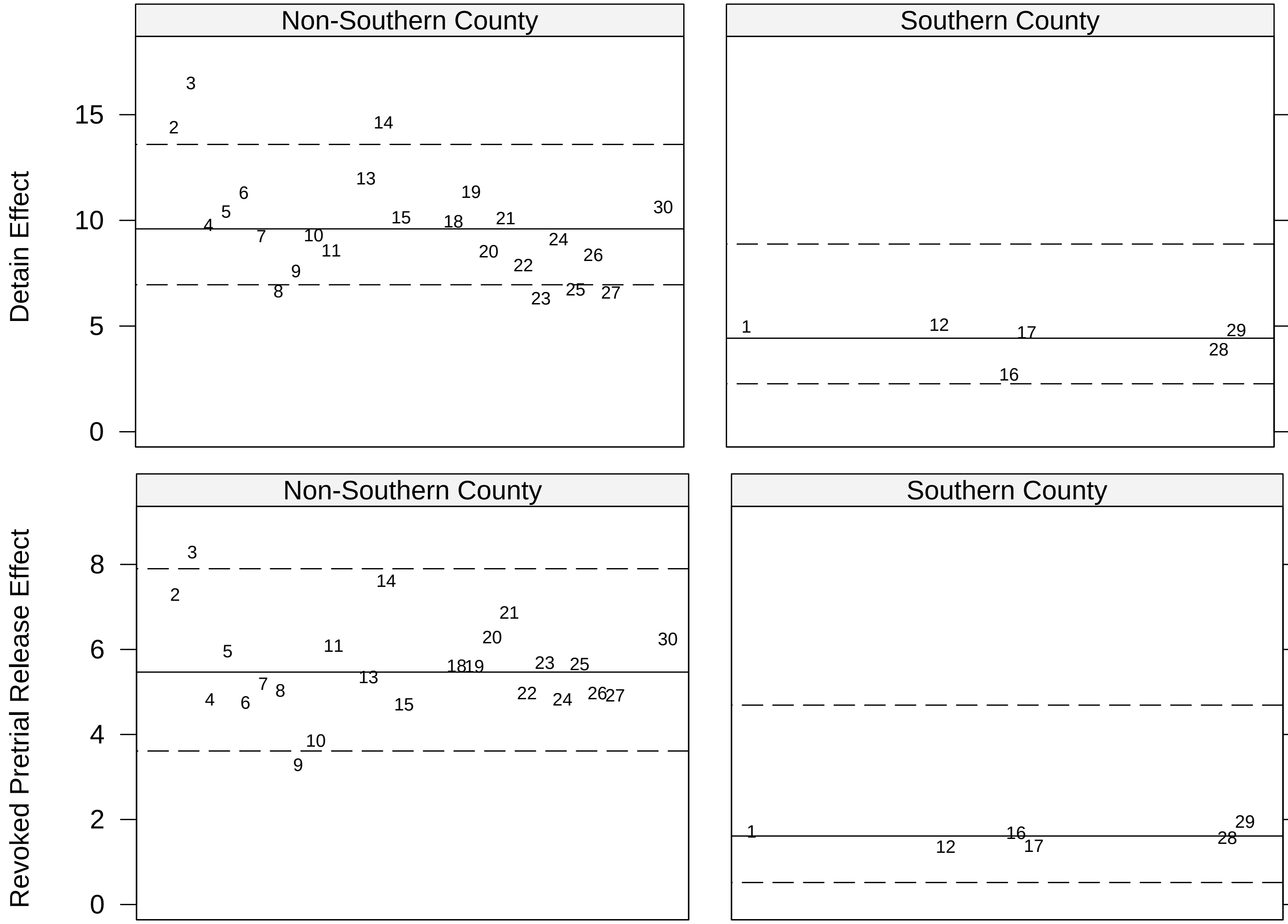

Fig. 4: Estimated odds ratios (solid lines), 95% HPD end-points (dashed lines), and county estimates (numbers), for IDET *and* IREV*, in Non-Southern and Southern counties.*

Interaction terms involve all four county-level variables; those involving only CSTH are the easiest to interpret. For example, posterior mean of the η for the CSTH.IDET interaction is negative, so being detained pretrial (IDET) has a lesser effect on odds of receiving a prison sentence in the South, $\exp(2.261 - 0.776) = 4.415$, than elsewhere, $\exp(2.261) = 9.593$. Fig. 4, constructed similarly to Fig. 3 above but taking into account the CSTH interactions, illustrates the effects for IDET and IREV.

Consider ICVS and Fig. 5. Green line shows estimated odds ratio for

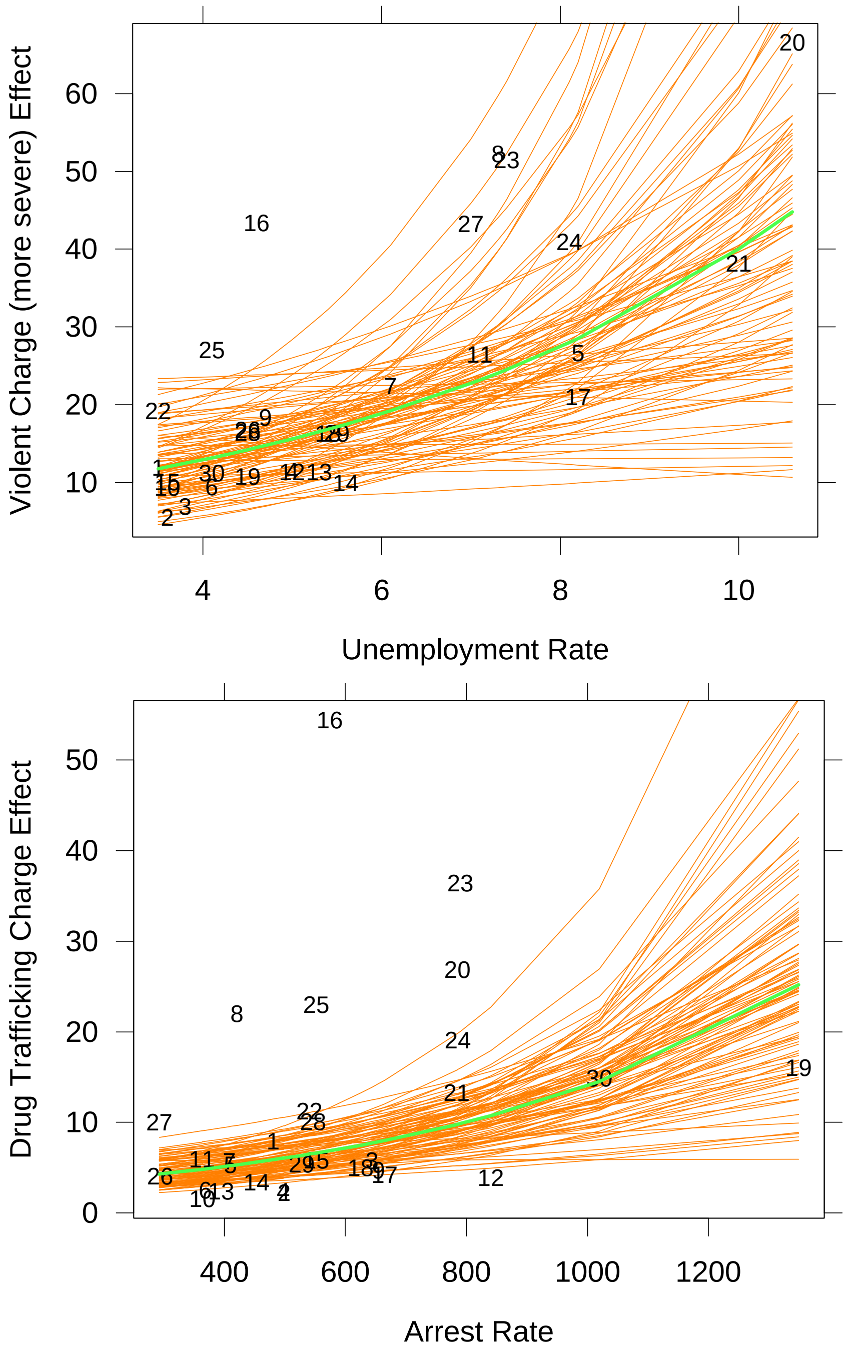

Fig. 5: Estimated odds ratios (green lines), posterior samples (orange lines), and county estimates (numbers), for ICVS *and* ICTR *as functions of* CUNR *and* CARR*.*

severe violent charges increasing from $\exp(2.864 - (1.078 \times 0.370)) = 11.772$ when unemployment is at its minimum to $\exp(2.864 + (2.538 \times 0.370)) = 44.778$ when unemployment is at its maximum. Orange lines represent 100 posterior samples for the ICVS and CUNR.ICVS coefficients. Numbers represent estimated odds ratios for each county.

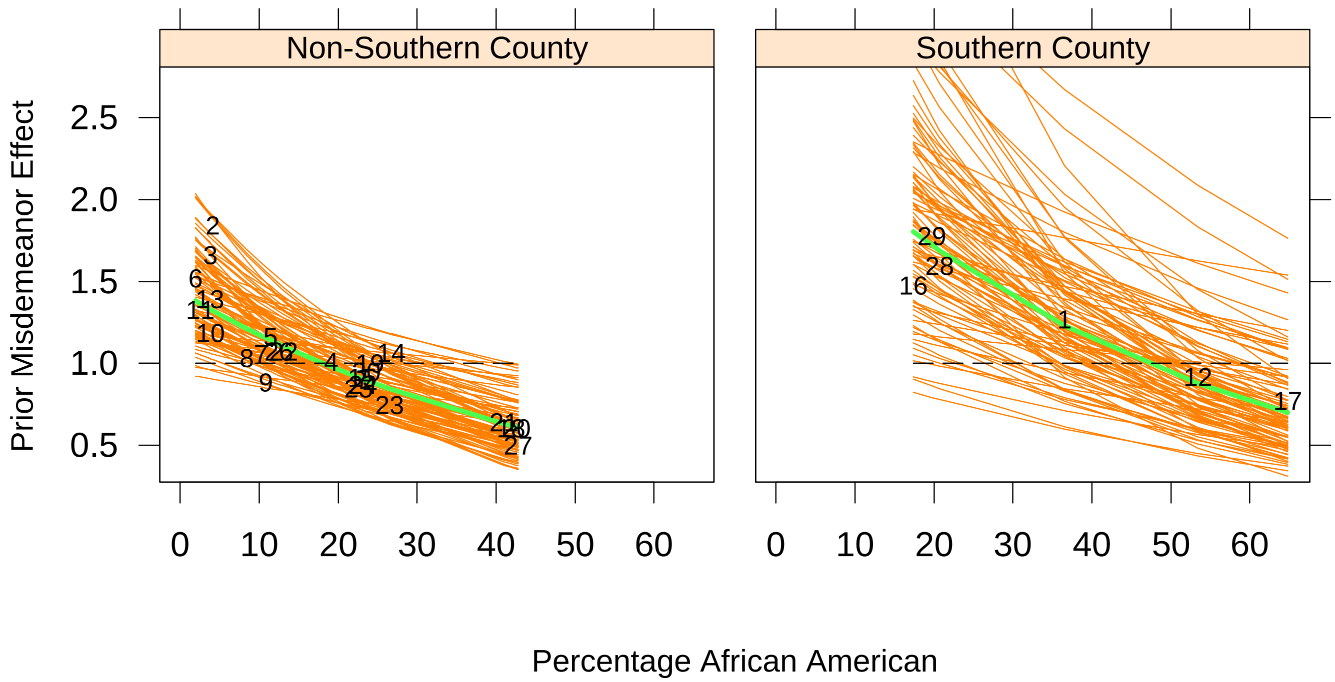

Fig. 6: Estimated odds ratios for IPMI *(green lines), posterior samples (orange lines), and county estimates (numbers), as a function of* CBLP *in Non-Southern and Southern counties. Horizontal dashed line represents "no effect".*

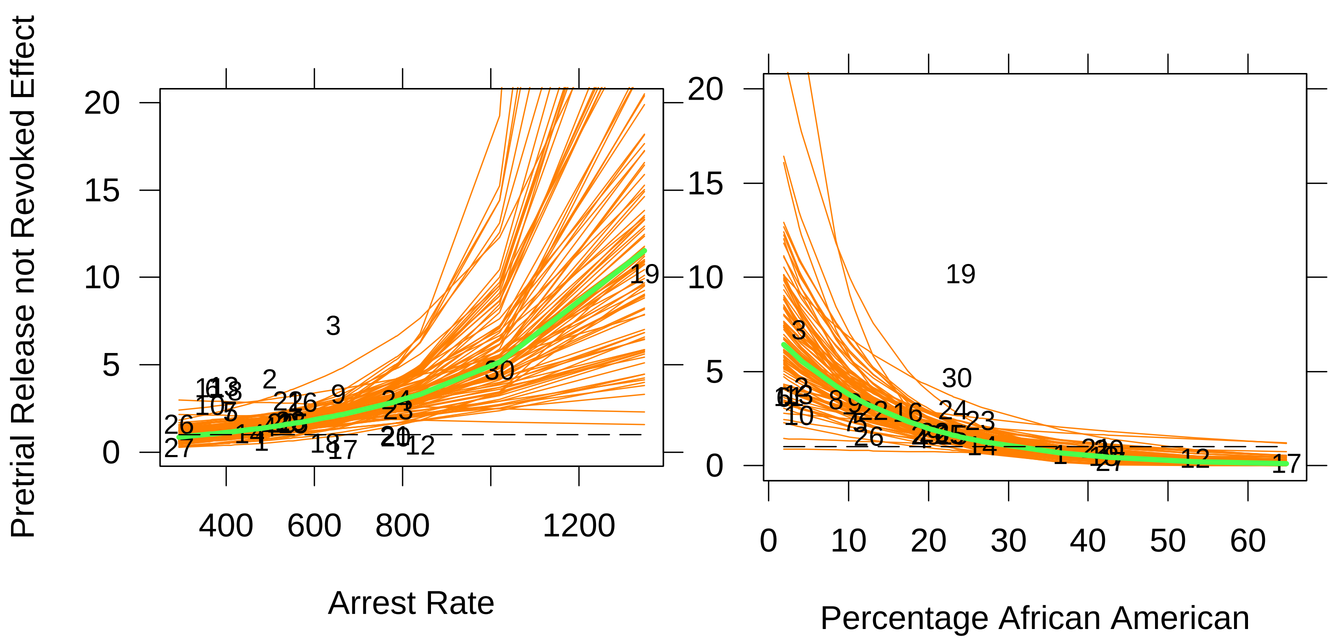

Fig. 7: Estimated odds ratios for IBAD *(green lines), posterior samples (orange lines), and county estimates (numbers), as functions of* CARR *and* CBLP*.*

Fig. 6 for IPMI combines the features of Fig. 4 and Fig. 5, while Fig. 7 for IBAD generalizes Fig. 5.

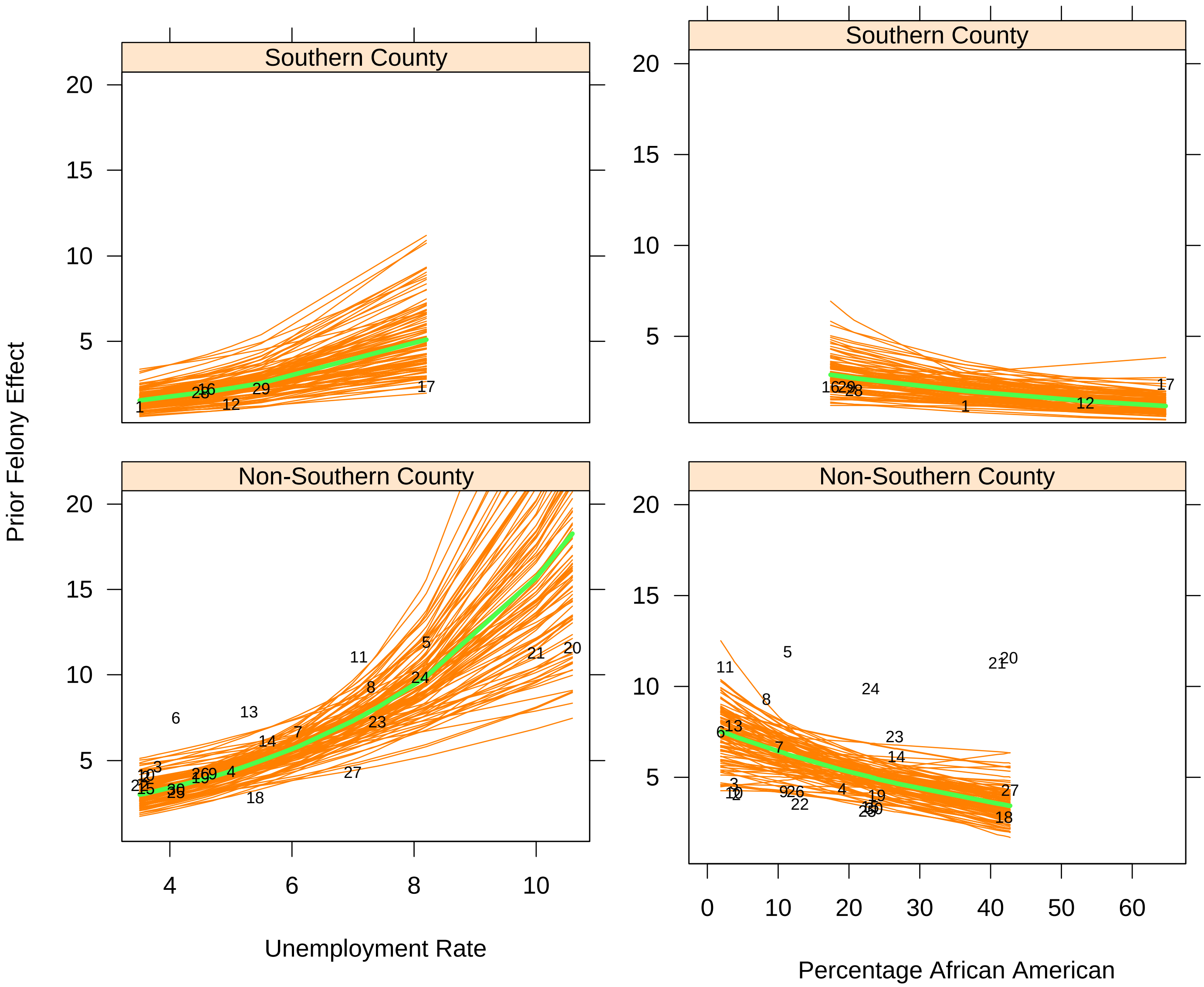

Fig. 8: Estimated odds ratio for IPFE *(green lines), posterior samples (orange lines), and county estimates (numbers), as functions of* CUNR *and* CBLP *in Non-Southern and Southern counties.*

Finally, there are three interactions involving IPFE: with CUNR, CBLP and CSTH. Interpretation of the effects of IPFE can be illustrated by combining the features of Fig. 6 and Fig. 7 in Fig. 8. Prior felonies increase the odds of receiving a prison sentence, but that effect varies widely with unemployment rate (it increases as unemployment increases), percentage African American (it decreases as percentage increases), and whether county is in the South (it is smaller in the South).

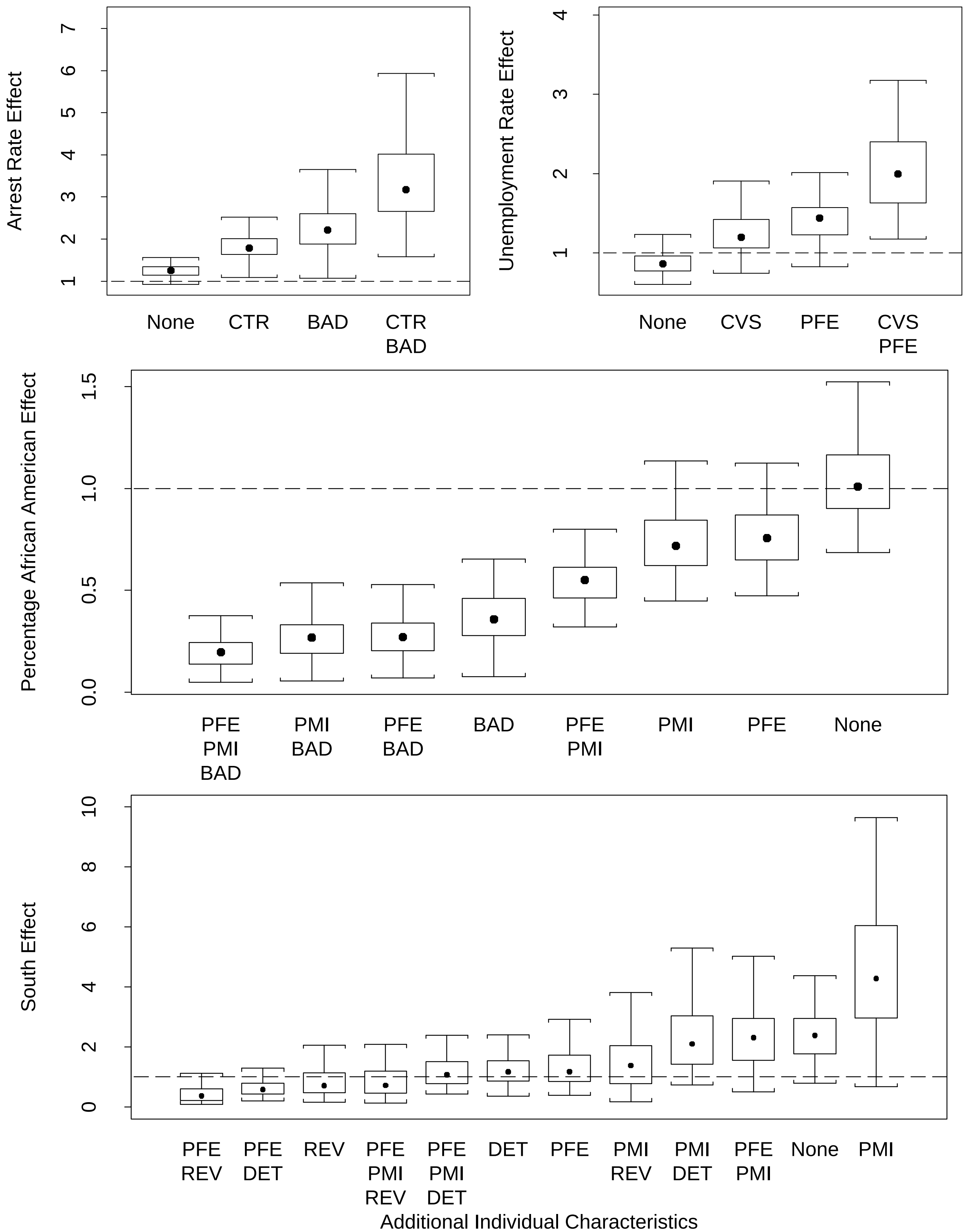

Fig. 9: Posterior samples expressed as odds ratios for one SD increases in CARR, CUNR, *and* CBLP, *and Southern region effects for different categories of individual.*

Fig. 9 shows effects from the perspective of county-level variables.

Discussion

This study advances understanding of how individual-level and county-level variables combine to affect punishment severity, using Bayesian multilevel modeling applied to sentencing data from 30 of the 75 most populous counties in 16 states. These counties are clearly a distinct group as there are more than 3,100 counties in the United States. Nevertheless, these counties have a disproportionate impact on the use of criminal justice system resources (e.g., prison and jail bed space) and the number of offenders affected. In 1996, the 75 most populous counties accounted for 37% of the U.S. population, 50% of all reported serious violent crime in the U.S., and 43% of all felony convictions.

Use of only large urban counties may account for some of the unexpected results obtained, e.g., prior studies suggested that CBLP and CSTH would have a positive effect on sentence severity, whereas our analysis found the opposite. Possible explanations include increased political power for African Americans in large urban counties making racial bias less likely, and large Southern counties differing from medium and small ones in terms of sentencing practices.

Other contextual factors whose effects on sentencing decisions merit further study include size of jurisdiction, urbanization, applicable laws (e.g., mandatory prison terms), political conservatism, level of bureaucratization, and case processing styles (e.g., percent of the caseload disposed by trial). Also, analysis of alternative outcome measures such as sentence length (or actual time served) and the use of prison sentences relative to jail sentences could have important policy relevance.