Central Analyses and Findings

O'Brien & Grosso Report on Racial Disparities in Juror Selection

- Review of the study design and statistical calculations

  – The MSU jury study design and analyses are thorough and appropriate

- Independent Analysis of the Effect of Black on Strike

- The finding of the effect of "Black" on strike decisions is statistically significant, and robust to variable selection

- Variation of Odds Ratio with Time

  – The odds ratio for "Black", a measure of how much the strike is greater among Black prospective jurors than jurors of other races, has fluctuated over time in the approximate 20 years of the study.

  – Although fluctuating, the odds ratio has consistently been statistically significantly greater than 1 throughout the time period, including at the time of Hasson Bacote's trial in 2009.

Variable Selection – MSU Table 4

TABLE 4
Statewide Logistic Regression Model

		A	B	C	E	F	G		H
		Variable Name	**Variable Description**	**Coefficient**	**S.E.**	**Odds Ratio**	**C.I.**		***p*-value**
1.		Intercept		-1.275	.087	0.280			<.001
2.		**Black**	**Venire member is Black**	0.957	.153	2.605	1.931	3.513	<.001
3.		DP_Reservations	Venire member expressed reservations about the death penalty	2.619	.203	13.716	9.218	20.407	<.001
4.		Hardship	Venire member worried serving would impose a hardship	0.771	.265	2.163	1.288	3.632	.004
5.		JKnewD	Venire member or venire member's immediate family knew the defendant	2.358	.545	10.572	3.634	30.759	<.001
6.		JKnewW	Venire member knew a witness	-0.459	.178	0.632	0.446	0.896	.010
7.		JKnewAtt	Venire member knew one of the attorneys in the case	0.489	.194	1.630	1.114	2.385	.012
8.		LeansState	Venire member expresses view that suggests view favorable to state (e.g., problems with presumption of innocence, right not to testify)	-1.787	.363	0.168	0.082	0.341	<.001
9.		PolicePros	VM or close other has worked with police or prosecutors.	-1.076	.187	0.341	0.237	0.492	<.001
10.		Accused_all	VM or a close other has been accused of criminal wrongdoing	0.494	.139	1.639	1.248	2.153	<.001

$n = 1,719^{47}$, $r^2 = .22$

Variable Selection

- When there are many independent variables, it doesn't make sense to include all of them in the regression

- Deleting the variables that aren't important leads to more precise estimates for the variables that remain, and also helps us avoid problems such as multicollinearity

- "Stepwise regression" is a method of reducing the number of variables by deleting them one at a time

  - Two standard methods of doing this are the "Akaike information criterion" (AIC) and the "Bayesian information criterion" (BIC)

  - Both implemented through the "step" function in R

Expanded Variable Set – AIC Step Method

- All variables from the model + most plausible confounders (18 variables)

  - Black, DP_Reservations, Hardship, JKnewD, JKnewW, JKnewAt, LeansState, PolicePros, Accused_all, JAccused, Helping, JLawEnf, BlueCollar, JVic_All, Defense, CollegeGrad, Gender, Young

- RandomMerge, no data missing in these variables (n=1,169)

- Step function to select variables, yielded 11 variables:

  - Black, DP_Reservations, Hardship, JKnew, JKnewW, JKnewAtt, LeansState, PolicePros, Jaccused, Defense, Young

- Repeated analysis using all cases in MSU RandomMerge data set with full information (n=1,679)

Expanded Variable - AIC Step Model Results

	Estimate	Std. Error	z value	P-Value
(Intercept)	-1.309	0.091	-14.447	2.63E-47
Black1.00	1.001	0.155	6.456	1.03E-10
DP_Reservations1.00	2.658	0.208	12.760	2.73E-37
Hardship1.00	0.842	0.263	3.197	0.00139
JKnewD1.00	2.310	0.556	4.155	3.25E-05
JKnewW1.00	-0.476	0.184	2.587	0.00970
JKnewAtt1.00	0.518	0.199	2.604	0.00921
LeansState1.00	-1.791	0.367	-4.885	1.03E-06
PolicePros1.00	-1.015	0.188	-5.386	7.22E-08
JAccused1.00	0.589	0.190	3.105	0.00190
Defense1.00	1.132	0.631	1.794	0.073
Young1.00	0.606	0.212	2.865	0.00418

Formal Variable Selection

All MSU variables
Omitted identification variables

114 variables
Omitted variables with too many categories (employment), too few responses, too many missing values

1024 Jurors in "RandomMerge" Sample, Eligible to be Struck
74 variables available for possible selection

Backward Selection – Akaike Information Criterion (AIC)

	Estimate	Std. Error	z value	P-Value		Estimate	Std. Error	z value	P-Value
(Intercept)	-1.750	0.172	-10.193	2.31E-24	LawEnforcement1.00	-1.709	1.445	-1.183	0.237
Accused_all1.00	0.605	0.241	2.509	0.0121	LawEnfOther1.00	-15.223	552.512	-0.0276	0.978
Black1.00	0.878	0.205	4.284	1.83E-05	LeansAmbig1.00	-0.711	0.595	-1.195	0.232
DefB1	0.426	0.301	1.419	0.156	LeansState1.00	-2.106	0.545	-3.863	0.000112
DefRM1	-0.549	0.307	-1.791	0.0733	Marital2	1.014	0.217	4.676	2.93E-06
DP_Reservations1.00	2.933	0.258	11.620	3.26E-31	Marital3	0.5976	0.2450	2.4395	0.0147
FamAccused1.00	-1.0187	0.510	-1.998	0.0457	Marital4	-0.0488	0.570	-0.0856	0.932
Hardship1.00	0.746	0.343	2.178	0.0294	Marital5	-0.230	1.973	-0.117	0.907
Homemaker1.00	0.836	0.288	2.905	0.00367	OthAccused_homicide1.00	2.915	0.866	3.364	0.000767
HungJury1.00	-0.288	1.194	-0.242	0.809	OthAccused_homicide9.00	2.229	0.820	2.717	0.00659
JBias_all1.00	1.226	0.609	2.011	0.0443	OthAccused_nonviolent1.00	1.027	0.621	1.654	0.0981
JCivWit1.00	0.702	0.561	1.252	0.211	OthVic1.00	-14.457	1015.348	-0.0142	0.989
JCredPO1.00	-14.505	641.822	-0.0226	0.982	PolicePros1.00	-0.698	0.227	-3.075	0.00211
JExpert1.00	-16.487	1203.505	-0.0137	0.989	SpouseBlueCollar1.00	0.394	0.211	1.867	0.0619
JKnewAtt1.00	0.532	0.255	2.082	0.037					
JKnewD1.00	3.817	1.184	3.223	0.00127					
JKnewParty1.00	-1.337	0.853	-1.568	0.117					
JKnewW1.00	-0.454	0.237	-1.916	0.0553	VeryYoung1.00	0.918	0.396	2.320	0.0204
JNoLife1.00	-1.965	1.167	-1.684	0.0922					

Backward Selection - Bayesian Information Criterion (BIC)

	Estimate	Std. Error	z value	P-Value
(Intercept)	-1.390	0.0926	-15.012	6.15E-51
Accused_all1.00	0.541	0.146	3.696	0.000219
Black1.00	0.882	0.165	5.339	9.33E-08
DP_Reservations1.00	2.581	0.212	12.194	3.34E-34
Homemaker1.00	0.647	0.255	2.533	0.0113
JKnewD1.00	2.434	0.589	4.131	3.61E-05
LawEnfOther1.00	-14.409	331.137	-0.0435	0.965
LeansState1.00	-1.876	0.404	-4.643	3.43E-06
PolicePros1.00	-0.971	0.194	-5.007	5.52E-07

Robustness of the Findings

Variable	MSU Table 4		Expanded Variable AIC Step		Backward Selection AIC		Backward Selection BIC	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Black	2.605	1.931-3.513	2.722	2.009-3.688	2.405	1.610-3.593	2.416	1.748-3.340
DP_Reservations	13.716	9.218-20.407	14.262	9.482-21.451	19.945	12.039-33.043	13.212	8.726-20.006
JKnewD	10.572	3.634-30.759	10.077	3.389-29.963	45.467	4.463-463	11.405	3.594-36.193
LeansState	0.168	0.082-0.341	0.167	0.081-0.342	0.122	0.042-0.354	0.153	0.069-0.338
PolicePros	0.341	0.237-0.492	0.362	0.251-0.524	0.497	0.319-0.776	0.379	0.259-0.554

Odds Ratio Over Time
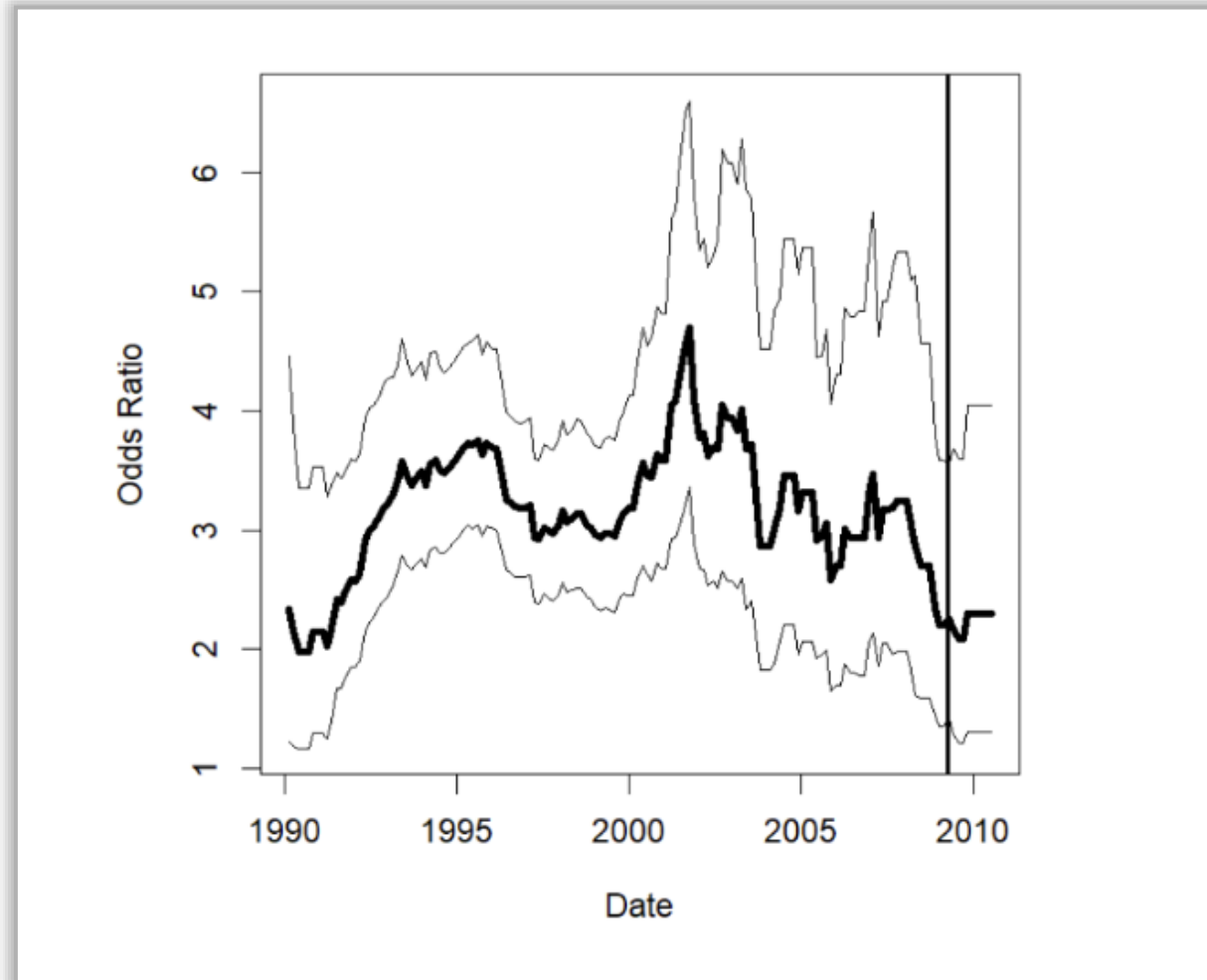
where n is the number of binary observations and $\hat{L}_0$ is the maximized likelihood under the null. The numerator can be seen as a ratio of the relative likelihood with the $1/n$ power having the effect of a geometric mean on the observations. The denominator simply normalizes so that $0 \le R^2 \le 1$. For example, for the current model, the R^2 is:

```
lmodr <- glm(chd ~ age + height  +bmi + sdp + chol + dibep + cigs +
    ↪ arcus, family=binomial, wcgs)
(1-exp((lmodr$dev-lmodr$null)/3140))/(1-exp(-lmodr$null/3140))
[1] 0.14315
```

This gives the impression of a fairly poor fit when judged from the experience of linear models. However, this is misleading. In a standard linear model, it is possible for the observed and fitted values to be very close showing a strong fit and an R^2 close to one. This simply isn't possible for binary response models given the natural variation. It is quite common to see low values of Naglekerke's and other R^2 substitutes even when the model is good. For this reason, it may be best to avoid this statistic except perhaps for the purpose of comparing compatible models.

Fiarway (2016),, Extending the Linear Model with R

Conclusions of Juror Selection Testimony

- MSU (Michigan State) study used logistic regression – "Black" very highly significant predictor with or without other variables

- Other analyses showed same effect at district, county and prosecutor level

- My analyses: looked at several other ways of doing the analysis but conclusions remained the same

- R^2 is relatively small but we argued this doesn't matter

- One more complicated analysis: "random effects analysis" allow for the "Black" effect to vary randomly between cases

- This does indeed demonstrate a case to case variability, but the overall effect is still strongly significant (see next slide)

Statewide Analysis Combining Fixed and Random Effects