Descriptive Representation and Judicial Outcomes in Multi-Ethnic Societies: Online Appendix

OCTOBER 2015

Contents

List of Tables

List of Figures

1 Introduction

In this appendix, we describe several variations of the analysis performed in the main text. Each of these alternate analyses relaxes an assumption made in the main analysis or otherwise alters the modeling strategy in order to test the robustness of the paper's results under different modeling conditions. First, we provide results of the key analyses disaggregated by each of the three District courts. Second, using the pooled data we use model selection to fit parsimonious versions of the models reported in the main text, showing that results

do not change substantively even when potentially spurious predictors are eliminated from the model. Third, we fit a more flexible set of models that allow interactions between predictors in order to demonstrate robustness under weaker assumptions about how predictors may influence outcomes. Fourth, we consider an alternate version of the model for prison sentence length, transforming the response to provide better justification for standard regression assumptions, and show that the results are qualitatively similar under the new specification. Finally, we use propensity score weighting to estimate the causal effects of mixed panels on leniency and harshness, allowing us to relax parametric assumptions required in the regression modeling framework. Overwhelmingly, these robustness checks strengthen the evidence for the effects in the main paper, substantiating their direction and in most cases their general magnitude.

2   Robustness Check I: Court-Specific Analyses

The data used in this paper comes from three different courts. The relative prevalence of Arab judges on panels varies substantially across the those courts. In particular, appeals heard by the court at Nazareth tend to have mixed-ethnicity panels for most cases while appeals in Tel Aviv and Jerusalem tend not to. To evaluate our results' robustness to possible biases introduced by these differences, we examine contingency tables showing leniency and harshness%, and incarceration rates by panel type and defendant ethnicity for the individual courts. Tables 1 and 2 show court-specific tables giving probabilities of leniency and harshness broken down by defendant ethnicity and panel composition as in the main text.

As is evident from the row and column totals in the tables, the sample sizes for specific combinations of defendant ethnicity and panel type in the individual courts tend to be rather small. Despite the variability in the percentages induced by these small sample sizes, the qualitative effects identified in the main paper still hold with few exceptions: in all courts Arabs tend to receive greater leniency from mixed panels, and in most cases they are also less likely to receive a harsher sentence. Sample size constraints make it difficult to undertake court-specific analyses of greater depth. But if the court-specific analyses cannot

4

be relied upon to provide strong confirming evidence of the effects in the main paper, they certainly do not suggest any kind of disconfirming effect.

Table 1: **Defendant appeals with more lenient verdict (%) by court**

Nazareth	Jewish	Arab	Total
All-Jewish	60.0	25.0	9
Mixed-panel	29.2	33.1	213
P-value (Fisher Exact)	0.17	1.00	
Total	94	128	222

Tel Aviv	Jewish	Arab	Total
All-Jewish	38.6	20.0	80
Mixed-panel	46.8	66.7	50
P-value (Fisher Exact)	0.45	0.20	
Total	117	13	130

Jerusalem	Jewish	Arab	Total
All-Jewish	38.5	25.0	88
Mixed-panel	60.0	80.0	10
P-value (Fisher Exact)	0.38	0.03	
Total	57	41	98

Table 2: **Prosecutorial appeals with harsher verdict (%) by court**

Nazareth	Jewish	Arab	Total
All-Jewish	100.0	100.0	3
Mixed-panel	84.6	63.6	24
P-value (Fisher Exact)	1.00	1.00	
Total	14	13	27

Tel Aviv	Jewish	Arab	Total
All-Jewish	81.5	85.7	34
Mixed-panel	85.7	100.0	16
P-value (Fisher Exact)	1.00	1.00	
Total	41	9	50

Jerusalem	Jewish	Arab	Total
All-Jewish	75.0	80.8	54
Mixed-panel	0.0	33.3	5
P-value (Fisher Exact)	0.08	0.14	
Total	30	29	59

3 Robustness Check II: Parsimonious Models

To further examine the robustness of our findings, we fit parsimonious versions of each of the regression models presented in the main text. We used a stepwise regression strategy that started with a small model and added the best predictors from a given pool, one-by-one, until no additional predictor would reduce the model's Akaike Information Criterion (AIC) by at least two points. By requiring at least a two-point reduction in AIC we impose a stringent requirement that each additional predictor have a strong partial effect, and thus cannot be considered spurious. We then computed cluster-robust standard errors and marginal effects from these parsimonious models as for the full models.

The parsimonious models for our key dependent variables leniency and harshness are shown in tables 3 and 4. For each response four different models were fitted, corresponding to the four full models in the main text (and fitted on the same subsets of the data). In each case the model selection procedure began with a model containing only defendant ethnicity, mixed panel, and their interaction and added predictors from the pool of those in the corresponding full regression according to the model selection strategy. As the tables below make clear, the direction and significance of the marginal effects in the main paper are robust to covariate selection. Harshness and leniency do not change significantly for Jewish defendants facing a mixed panel rather than an all-Jewish panel, while Arab defendants experience significantly more leniency and less harshness in this situation. The decrease in probability of harshness for Arabs is estimated at 20% in our preferred model (as determined by AIC), which is slightly smaller than the figure reported in the main text but still suggests a strong panel effect.

The parsimonious models for incarceration are reported in table 5. The resulting marginal effects are similar to those in the full model. Finally, the parsimonious models for prison term are reported in table 6. Once again, the marginal effects take on similar values to those in the full model, suggesting robustness.

Table 3: **Lenient District-court verdict rates for defendant appeals – Parsimonious Logistic Model**. Marginal effects are computed at mean of other predictors given ethnicity.

More Lenient	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	-0.678	(0.429)	0.115	-0.757	(0.423)	0.074	-0.757	(0.423)	0.074	-0.741	(0.437)	0.09
Mixed Panel	0.325	(0.283)	0.252	0.316	(0.322)	0.326	0.316	(0.322)	0.326	0.346	(0.352)	0.327
Mixed Panel × Arab Defendant	0.852	(0.439)	0.053	1.026	(0.438)	0.02	1.026	(0.438)	0.02	1.077	(0.462)	0.02
Female Defendant	0.925	(0.416)	0.027	1.123	(0.456)	0.014	1.123	(0.456)	0.014	1.404	(0.486)	0.004
Nazareth Court	-0.699	(0.343)	0.042	-2.785	(0.558)	0	-2.785	(0.558)	0	-2.141	(0.632)	0.001
Average Judge Experience				-0.193	(0.042)	0	-0.193	(0.042)	0	-0.178	(0.043)	0
Jerusalem Court				-1.304	(0.489)	0.008	-1.304	(0.489)	0.008	-1.218	(0.545)	0.026
Convicted by trial										1.315	(0.494)	0.008
Prosecution requested rejection of appeal										-1.471	(0.451)	0.001
(Intercept)	-0.49	(0.187)	0.009	4.433	(1.152)	0	4.433	(1.152)	0	4.585	(1.246)	0
N. Predictors	5			7			7			9		
AIC Value	583.198			564.97			564.97			510.656		
Observations	450			450			450			450		
Marg. Effect of Mixed Panel (Jewish Def.)	0.073	(0.072)	0.308	0.068	(0.071)	0.344	0.068	(0.071)	0.342	0.064	(0.066)	0.334
Marg. Effect of Mixed Panel (Arab Def.)	0.223	(0.078)	0.005	0.237	(0.074)	0.001	0.237	(0.075)	0.002	0.22	(0.066)	0.001

Table 4: **Harsher District-court verdict rates for prosecutorial appeals – Parsimonious Logistic Model**. Marginal effects are computed at mean of other predictors given defendant ethnicity.

Harsher	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	0.276	(0.718)	0.701	0.276	(0.718)	0.701	0.632	(0.713)	0.377	0.672	(0.717)	0.351
Mixed Panel	0.044	(0.586)	0.94	0.044	(0.586)	0.94	-0.303	(0.631)	0.632	0.135	(0.707)	0.849
Mixed Panel × Arab Defendant	-1.109	(0.902)	0.221	-1.109	(0.902)	0.221	-0.99	(0.947)	0.298	-1.293	(1.185)	0.277
Jewish Victim							2.377	(0.791)	0.003	2.933	(1.175)	0.014
Prosecution requested rejection of appeal										-3.274	(1.353)	0.017
(Intercept)	1.299	(0.437)	0.004	1.299	(0.437)	0.004	0.765	(0.464)	0.101	0.713	(0.467)	0.129
N. Predictors	3			3			4			5		
AIC Value	149.002			149.002			135.635			131.413		
Observations	136			136			136			136		
Marg. Effect of Mixed Panel (Jewish Def.)	0.007	(0.095)	0.938	0.007	(0.093)	0.937	-0.046	(0.097)	0.639	0.019	(0.093)	0.842
Marg. Effect of Mixed Panel (Arab Def.)	-0.204	(0.143)	0.156	-0.204	(0.14)	0.149	-0.236	(0.137)	0.088	-0.201	(0.137)	0.145

Table 5: **Incarceration rates – Parsimonious Logistic Model**. Marginal effects are computed at mean of other predictors given defendant ethnicity.

Incarceration	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	0.861	(0.331)	0.01	1.053	(0.314)	0.001	1.206	(0.304)	0	1.206	(0.304)	0
Mixed Panel	-0.264	(0.289)	0.361	0.1	(0.307)	0.745	0.111	(0.307)	0.718	0.111	(0.307)	0.718
Mixed Panel × Arab Defendant	-1.197	(0.416)	0.004	-1.152	(0.427)	0.007	-1.259	(0.414)	0.002	-1.259	(0.414)	0.002
Prison term, Magistrate Court	0.259	(0.052)	0	0.273	(0.054)	0	0.266	(0.05)	0	0.266	(0.05)	0
Female Defendant				-1.276	(0.5)	0.011						
Average Judge Age				0.082	(0.027)	0.002	0.082	(0.025)	0.001	0.082	(0.025)	0.001
Physical or Sexual Assault							0.636	(0.241)	0.008	0.636	(0.241)	0.008
(Intercept)	-0.652	(0.322)	0.043	-5.834	(1.751)	0.001	-6.253	(1.64)	0	-6.253	(1.64)	0
N. Predictors	4			6			6			6		
AIC Value	421.563			413.027			411.522			411.522		
Observations	544			544			544			544		
Marg. Effect of Mixed Panel (Jewish Def.)	-0.032	(0.038)	0.411	0.012	(0.041)	0.777	0.013	(0.042)	0.758	0.013	(0.042)	0.759
Marg. Effect of Mixed Panel (Arab Def.)	-0.2	(0.047)	0	-0.139	(0.051)	0.007	-0.151	(0.05)	0.003	-0.151	(0.05)	0.003

Table 6: **District-level prison terms – Parsimonious OLS**

Prison term (months)	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	1.809	(1.11)	0.104	1.809	(1.11)	0.104	1.624	(1.014)	0.11	0.816	(0.912)	0.372
Mixed Panel	-1.972	(0.766)	0.01	-1.972	(0.766)	0.01	-1.707	(0.915)	0.063	-1.55	(0.631)	0.014
Mixed Panel × Arab Defendant	-2.268	(1.161)	0.051	-2.268	(1.161)	0.051	-2.38	(1.075)	0.028	-1.183	(1.001)	0.238
Female Defendant	-4.397	(1.518)	0.004	-4.397	(1.518)	0.004	-4.795	(1.462)	0.001	-3.663	(1.583)	0.021
Jerusalem Court	-3.2	(1.05)	0.002	-3.2	(1.05)	0.002	-1.845	(1.315)	0.161	-2.928	(0.877)	0.001
Prison term, Magistrate Court	0.871	(0.025)	0	0.871	(0.025)	0	0.855	(0.026)	0	0.898	(0.028)	0
Average Judge Age							0.291	(0.138)	0.035			
Nazareth Court							3.044	(1.689)	0.072			
Property or Fraud							2.178	(0.559)	0	1.934	(0.522)	0
Convicted by trial										-2.38	(1.191)	0.046
Prosecution asked for harsher sentence										-5.847	(1.128)	0
(Intercept)	4.598	(0.817)	0	4.598	(0.817)	0	-15.306	(9.485)	0.107	7.595	(0.949)	0
N. Predictors	7			7			10			10		
AIC Value	1362.118			1362.118			1348.386			1273.102		
Observations	379			379			379			379		
Marg. Effect of Mixed Panel (Jewish Def.)	-1.972	(0.766)	0.01	-1.972	(0.766)	0.01	-1.707	(0.915)	0.063	-1.55	(0.631)	0.014
Marg. Effect of Mixed Panel (Arab Def.)	-4.24	(1.3)	0.001	-4.24	(1.3)	0.001	-4.087	(1.166)	0.001	-2.733	(1.09)	0.013

Figure 1: Changes in predicted probabilities of a more lenient sentence (where the defendant appealed) and in predicted probabilities of a harsher sentence (where the prosecutor appealed) for defendants facing a mixed panel in each of the parsimonious fitted models. The changes in predicted probability are calculated by averaging over changes in probability estimated by the model for each defendant (of the given ethnicity) in the dataset at the observed values of the covariates for this defendant. Thick lines show 90% confidence intervals and thinner lines 95% confidence intervals.
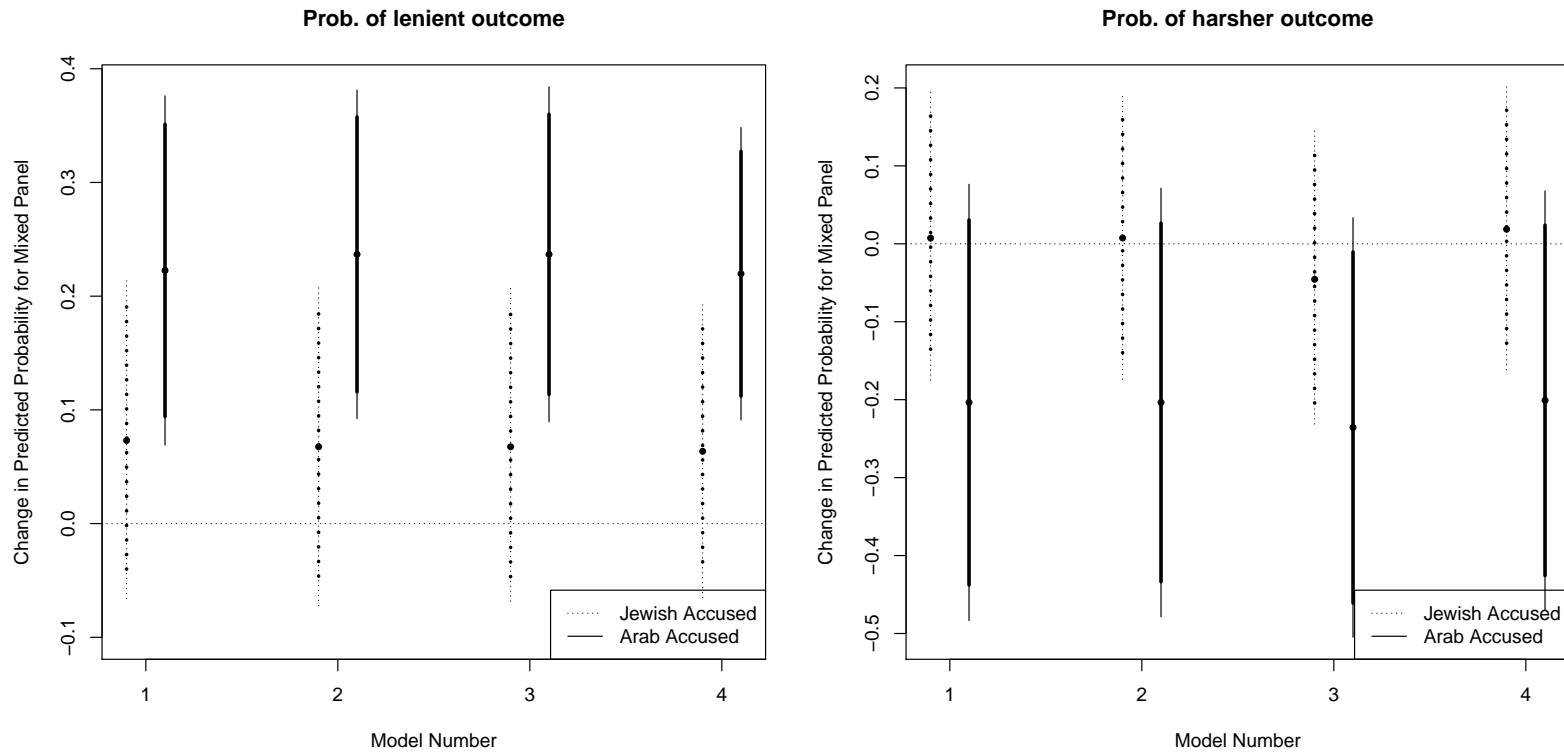
Figure 2: Changes in predicted probabilities of incarceration due to mixed panel and marginal effects of mixed panels on length of prison term for each of the parsimonious fitted models. The prison term model is an ordinary linear regression so the effects for that model are constant. For the incarceration model, changes in predicted probability are calculated by averaging over changes in probability estimated by the model for each defendant (of the given ethnicity) in the dataset at the observed values of the covariates for this defendant. Thick lines show 90% confidence intervals and thinner lines 95% confidence intervals.
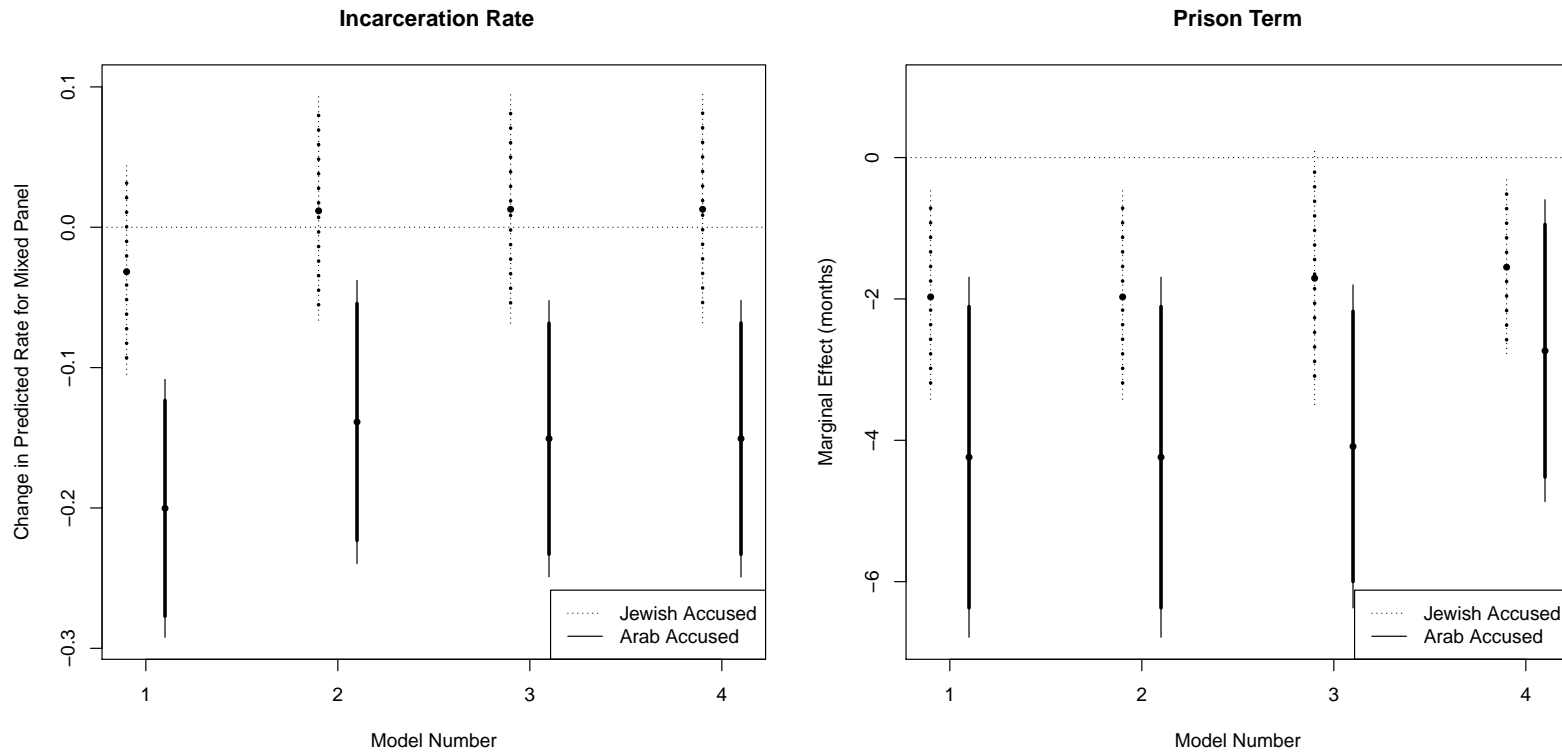
4 Robustness Check III: Interactions

By fitting a model with only one interaction (between mixed panel and defendant ethnicity), we assumed in the main paper that the partial effects of covariates other than defendant ethnicity for the response in question do not change under a mixed panel as opposed to an all-Jewish one. As a robustness check, we relaxed this assumption by refitting our models and including interactions between mixed panel and *all other covariates*. Judging full models containing many interactions to be unwieldy and unreliable, we focused on adapting the parsimonious models of the previous section to allow for interactions. We did so by refitting the models, expanding each pool of potential covariates to include interaction terms of each of these covariates with mixed panel.

Tables 7 and 8 give the interaction models for our key outcomes, leniency and harshness. Most are very similar to the corresponding parsimonious models; in fact, the model selection routine for harshness chose not to accept any interactions into the model. The fourth model for leniency did include an interaction between mixed panel and prosecutorial requests for a harsher sentence. The marginal effects of mixed panel in this model were qualitatively identical to those in the main paper (no change for Jewish defendants and an increase in leniency for Arab defendants). Although the effect sizes for Arab defendants were smaller than in the parsimonious model given above in section 3, they tell a similar story otherwise.

Table 9 gives the interaction models for incarceration. These models differed from the parsimonious no-interaction models in table 5 by consistently including an interaction between mixed panel and length of magistrate-level prison term. The marginal effects show a weaker effect of mixed panels for Arabs than in previous models but agree with them in direction. Therefore these models strengthen the argument made in the main paper about the qualitative effect of mixed panels on incarceration.

Table 10 gives the interaction models for prison term length. Here interactions only entered the third and fourth models. In the third model the chosen interaction was between average judge age and mixed panel. However, this model is suspect because of its abnormally large coefficient values for the intercept and mixed panel, compared to the other models in the table; the likely explanation is that since Arab judges tend to be a little

11

younger than the average Jewish judge, collinearity between the mixed panel and judge age covariates resulted in degenerate values for these covariates and the interaction. More credibly, the fourth model includes an interaction between mixed panel and prosecutorial request for a harsher sentence (much like the fourth model for leniency). Here the marginal effect is slightly weaker than in the other models, but it is still at the cusp of significance and remains negative for defendants of both ethnicitys. Clearly there is still consistent evidence that mixed panels improve outcomes for both ethnicitys and especially for Arabs.

In summary, the interaction models suggest that for most covariates the assumption of no interaction with mixed panel is likely justified, since so few interactions appeared in any regressions. Furthermore, even though certain partial effects may not be invariant under mixed or all-Jewish panels, the marginal effects of mixed panels on outcomes of interest remain similar to the effects given by interaction-free models.

Table 7: **Lenient District-court verdict rates for defendant appeals – Parsimonious Logistic Model with interactions**. Marginal effects are computed at mean of other predictors given defendant ethnicity.

More Lenient	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	-0.678	(0.429)	0.115	-0.757	(0.423)	0.074	-0.757	(0.423)	0.074	-0.626	(0.406)	0.124
Mixed Panel	0.325	(0.283)	0.252	0.316	(0.322)	0.326	0.316	(0.322)	0.326	1.82	(0.812)	0.026
Mixed Panel × Arab Defendant	0.852	(0.439)	0.053	1.026	(0.438)	0.02	1.026	(0.438)	0.02	0.872	(0.455)	0.056
Female Defendant	0.925	(0.416)	0.027	1.123	(0.456)	0.014	1.123	(0.456)	0.014	1.225	(0.497)	0.014
Nazareth Court	-0.699	(0.343)	0.042	-2.785	(0.558)	0	-2.785	(0.558)	0			
Average Judge Experience				-0.193	(0.042)	0	-0.193	(0.042)	0			
Jerusalem Court				-1.304	(0.489)	0.008	-1.304	(0.489)	0.008			
Convicted by trial										1.512	(0.472)	0.001
Prosecution requested rejection of appeal										-0.395	(0.38)	0.298
Mixed Panel × Prosec. requested rejection										-2.085	(0.791)	0.009
(Intercept)	-0.49	(0.187)	0.009	4.433	(1.152)	0	4.433	(1.152)	0	-0.773	(0.44)	0.079
N. Predictors	5			7			7			7		
AIC Value	583.198			564.97			564.97			507.828		
Observations	450			450			450			450		
Marg. Effect of Mixed Panel (Jewish Def.)	0.036	(0.035)	0.296	0.034	(0.036)	0.337	0.034	(0.035)	0.333	0.016	(0.029)	0.571
Marg. Effect of Mixed Panel (Arab Def.)	0.151	(0.05)	0.003	0.161	(0.048)	0.001	0.161	(0.048)	0.001	0.12	(0.043)	0.005

Table 8: **Harsher higher-court verdict rates for prosecutorial appeals – Parsimonious Logistic Model with interactions**. Marginal effects are computed at mean of other predictors given defendant ethnicity.

Harsher	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	0.276	(0.718)	0.701	0.276	(0.718)	0.701	0.632	(0.713)	0.377	0.672	(0.717)	0.351
Mixed Panel	0.044	(0.586)	0.94	0.044	(0.586)	0.94	-0.303	(0.631)	0.632	0.135	(0.707)	0.849
Mixed Panel × Arab Defendant	-1.109	(0.902)	0.221	-1.109	(0.902)	0.221	-0.99	(0.947)	0.298	-1.293	(1.185)	0.277
Jewish Victim							2.377	(0.791)	0.003	2.933	(1.175)	0.014
Prosecution requested rejection of appeal										-3.274	(1.353)	0.017
(Intercept)	1.299	(0.437)	0.004	1.299	(0.437)	0.004	0.765	(0.464)	0.101	0.713	(0.467)	0.129
N. Predictors	3			3			4			5		
AIC Value	149.002			149.002			135.635			131.413		
Observations	136			136			136			136		
Marg. Effect of Mixed Panel (Jewish Def.)	0.003	(0.032)	0.938	0.003	(0.032)	0.938	-0.013	(0.029)	0.643	0.006	(0.029)	0.84
Marg. Effect of Mixed Panel (Arab Def.)	-0.064	(0.044)	0.148	-0.064	(0.044)	0.145	-0.069	(0.04)	0.084	-0.056	(0.039)	0.148

14

Table 9: **Incarceration rates – Parsimonious Logistic Model with interactions**. Marginal effects are computed at mean of other predictors given defendant ethnicity.

Incarceration	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	0.746	(0.323)	0.021	0.934	(0.289)	0.001	1.081	(0.286)	0	1.081	(0.286)	0
Mixed Panel	-0.751	(0.535)	0.161	-0.395	(0.439)	0.369	-0.367	(0.432)	0.395	-0.367	(0.432)	0.395
Mixed Panel × Arab Defendant	-1.078	(0.425)	0.011	-1.053	(0.423)	0.013	-1.112	(0.407)	0.007	-1.112	(0.407)	0.007
Prison term, Magistrate Court	0.193	(0.056)	0.001	0.192	(0.057)	0.001	0.197	(0.055)	0	0.197	(0.055)	0
Mixed Panel ×Prison term, Magistrate Court	0.126	(0.086)	0.145	0.151	(0.087)	0.086	0.133	(0.081)	0.101	0.133	(0.081)	0.101
Average Judge Age				0.086	(0.025)	0.001	0.09	(0.024)	0	0.09	(0.024)	0
Physical or Sexual Assault							0.565	(0.229)	0.014	0.565	(0.229)	0.014
(Intercept)	-0.402	(0.365)	0.271	-5.905	(1.635)	0	-6.474	(1.593)	0	-6.474	(1.593)	0
N. Predictors	5			6			7			7		
AIC Value	417.953			410.233			407.514			407.514		
Observations	544			544			544			544		
Marg. Effect of Mixed Panel (Jewish Def.)	-0.014	(0.019)	0.446	0.018	(0.024)	0.469	0.015	(0.023)	0.513	0.015	(0.024)	0.517
Marg. Effect of Mixed Panel (Arab Def.)	-0.129	(0.031)	0	-0.084	(0.036)	0.021	-0.089	(0.034)	0.009	-0.089	(0.035)	0.01

Table 10: **District-level prison terms – Parsimonious OLS with interactions.**

Prison term (months)	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	1.809	(1.11)	0.104	1.809	(1.11)	0.104	1.651	(0.992)	0.097	0.743	(0.927)	0.423
Mixed Panel	-1.972	(0.766)	0.01	-1.972	(0.766)	0.01	19.814	(11.361)	0.082	-3.63	(1.841)	0.049
Mixed Panel × Arab Defendant	-2.268	(1.161)	0.051	-2.268	(1.161)	0.051	-2.255	(1.047)	0.032	-1.211	(1.018)	0.235
Female Defendant	-4.397	(1.518)	0.004	-4.397	(1.518)	0.004	-5.003	(1.424)	0	-3.378	(1.588)	0.034
Jerusalem Court	-3.2	(1.05)	0.002	-3.2	(1.05)	0.002	-1.806	(1.275)	0.157	-3.047	(0.872)	0.001
Prison term, Magistrate Court	0.871	(0.025)	0	0.871	(0.025)	0	0.854	(0.026)	0	0.901	(0.028)	0
Average Judge Age							0.34	(0.153)	0.027			
Property or Fraud							2.205	(0.549)	0	1.987	(0.514)	0
Mixed Panel × Average Judge Age							-0.334	(0.178)	0.062			
Convicted by trial										-2.203	(1.262)	0.082
Prosecution requested rejection of appeal										-7.182	(1.264)	0
Mixed Panel × Prosec. requested rejection										2.931	(1.968)	0.137
(Intercept)	4.598	(0.817)	0	4.598	(0.817)	0	-18.388	(10.314)	0.075	8.363	(1.17)	0
N. Predictors	7			7			10			11		
AIC Value	1362.118			1362.118			1347.425			1269.309		
Observations	379			379			379			379		
Marg. Effect of Mixed Panel (Jewish Def.)	-1.972	(0.87)	0.024	-1.972	(0.887)	0.027	-0.648	(1.116)	0.562	-0.699	(0.695)	0.315
Marg. Effect of Mixed Panel (Arab Def.)	-4.24	(1.319)	0.001	-4.24	(1.321)	0.001	-2.09	(1.332)	0.117	-1.91	(0.985)	0.053

Figure 3: Changes in predicted probabilities of a more lenient sentence (where the defendant appealed) and in predicted probabilities of a harsher sentence (where the prosecutor appealed) for defendants facing a mixed panel in each of the parsimonious fitted models with interactions. The changes in predicted probability are calculated by averaging over changes in probability estimated by the model for each defendant (of the given ethnicity) in the dataset at the observed values of the covariates for this defendant. Thick lines show 90% confidence intervals and thinner lines 95% confidence intervals.
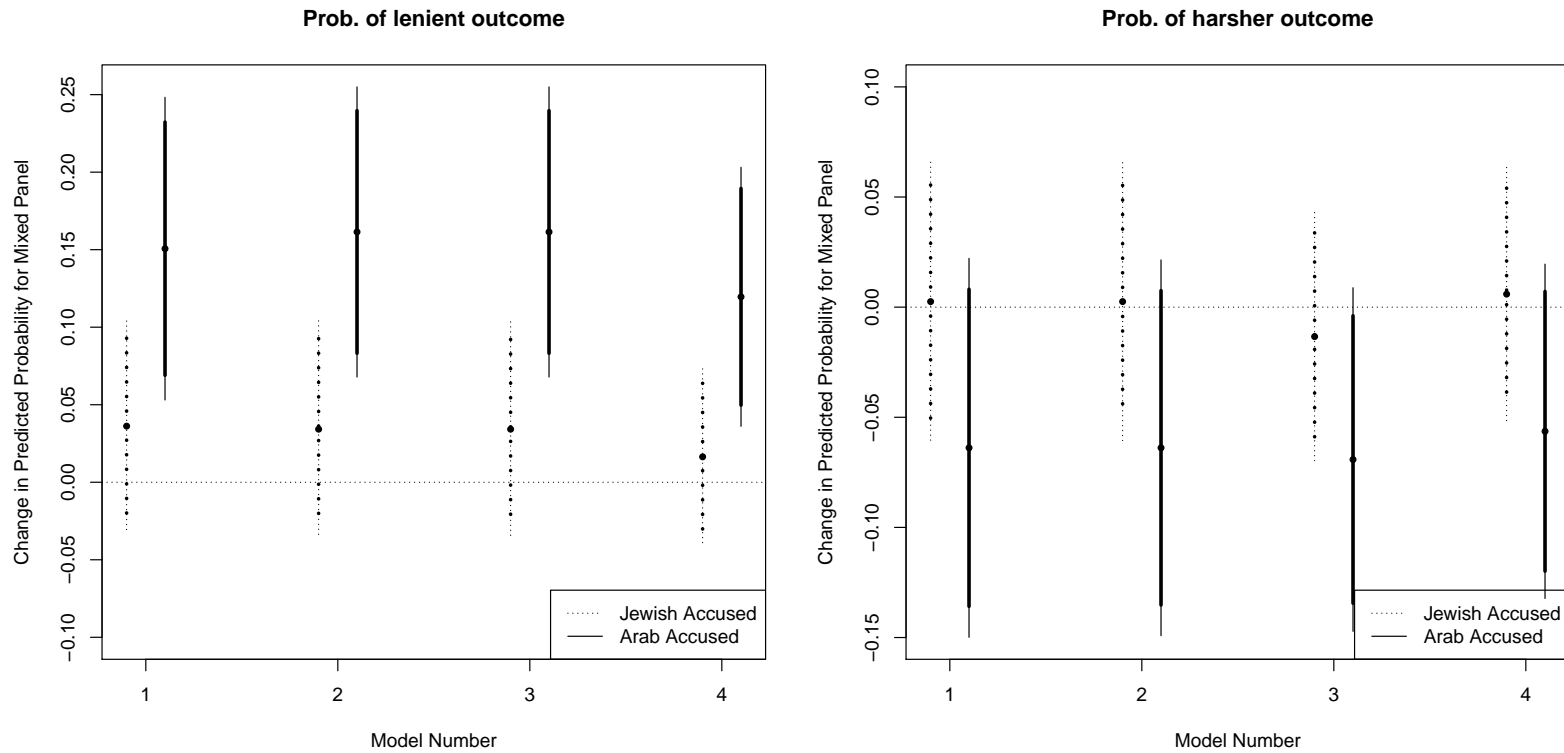
Figure 4: Changes in predicted probabilities of incarceration due to mixed panel and marginal effects of mixed panels on length of prison term for each of the parsimonious fitted models with interactions. The prison term model is an ordinary linear regression so the effects for that model are constant. For the incarceration model, changes in predicted probability are calculated by averaging over changes in probability estimated by the model for each defendant (of the given ethnicity) in the dataset at the observed values of the covariates for this defendant. Thick lines show 90% confidence intervals and thinner lines 95% confidence intervals.
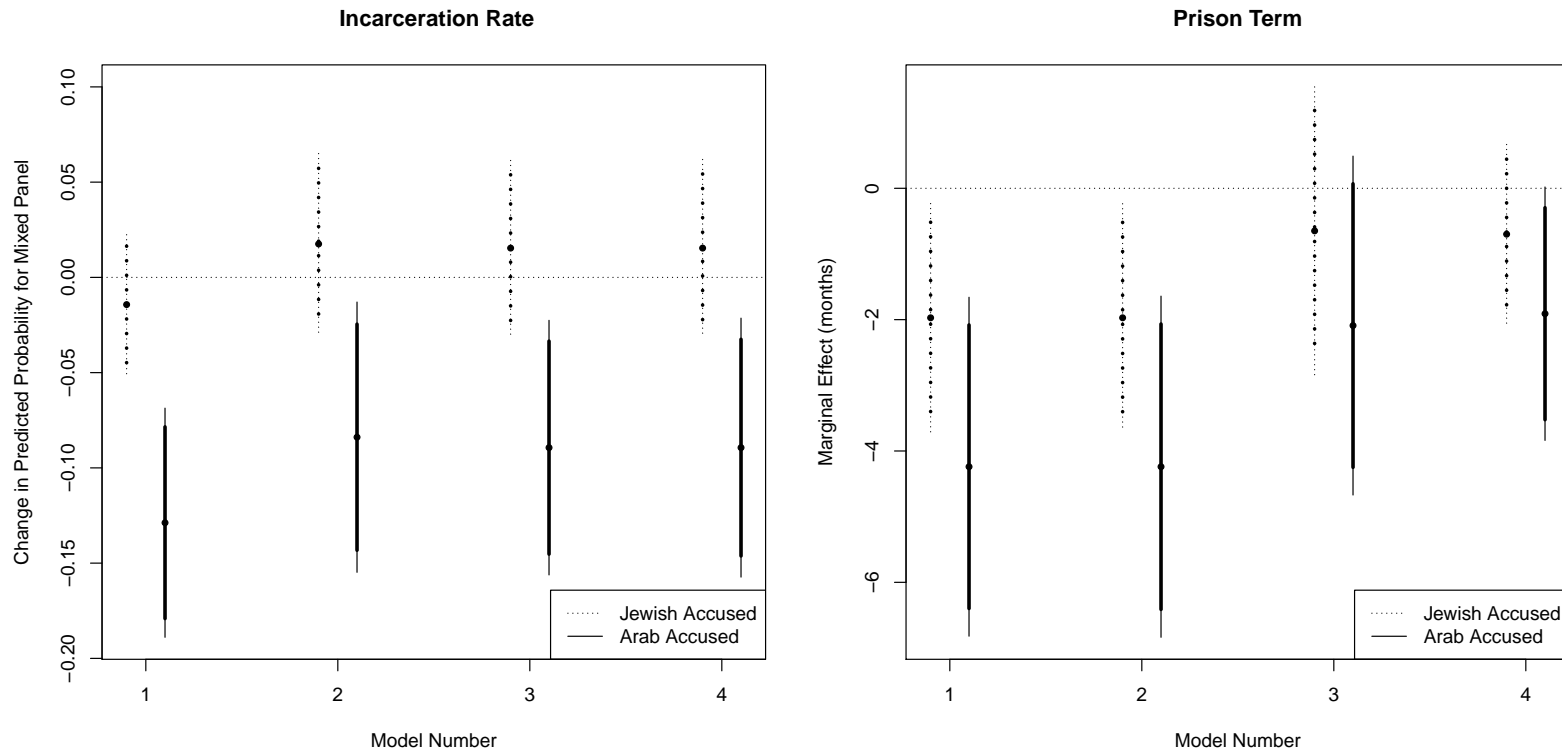
5   Robustness Check IV: Transforming the Response

The distribution of nonzero prison terms in the dataset has a strong positive skew, as is evident from the first histogram in figure 5. As a result, the assumption of normally-distributed errors required for drawing inferences about regression coefficients is not likely to be satisfied in an OLS with raw nonzero prison term as the response. To check that our results are robust to biases associated with this incorrect assumption, we transformed the prison term response by taking square roots and refit the regression. To verify the quality of the model, we computed ten-fold cross-validation error for the square root model and compared it to the analogous value for the original model. The cross-validation exercise showed that the square root model performed only slightly worse than the OLS model in prediction. In addition to the square root transform, we considered using a log transform, a cube root or a fourth root. Since these models did not perform well relative the OLS in the cross-validation exercise, we limit our analysis to the square root models.

Full and parsimonious models for the square root of the positive prison term are described in tables 11 and 12. The marginal effects listed in these tables are given on the scale of the raw prison term (i.e. in months); these effects were computed at the mean (or median for dummy variables) of the other covariates. The calculation of these effects was non-standard and a mathematical derivation for the formula we used is given in the last section of this appendix. We computed standard errors (which in turn were used to find t-statistics and p-values) using the bootstrap method.

The marginal effects of a mixed panel for Arab defendants on prison term in these models, which range from a decrease in 2 to a decrease in 3 months, are very similar to those found in the other models for prison term. They are almost always significant, though in three models they fall slightly below the 90% significance level. This is likely due in part to small sample size, since zero-value prison terms are not included there are fewer than 400 observations and to the more conservative nature of bootstrap standard errors. We note that marginal effects for Jewish defendants were somewhat larger and stronger than in other models for prison term. In sum, the transform square root models provide further evidence that Arab defendants tend to receive shorter sentences when facing mixed panels.

18

Table 11: **Positive district-level prison terms – Full OLS with square-root response**. Marginal effects are given in terms of prison term (at mean of other predictors given ethnicity).

Square root of prison term	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	-0.142	(0.121)	0.24	-0.154	(0.117)	0.191	-0.144	(0.114)	0.208	-0.202	(0.106)	0.058
Mixed Panel	-0.416	(0.146)	0.005	-0.407	(0.15)	0.007	-0.386	(0.15)	0.01	-0.363	(0.142)	0.011
Mixed Panel \times Arab Defendant	0.069	(0.135)	0.608	0.07	(0.134)	0.604	0.054	(0.134)	0.685	0.16	(0.124)	0.2
Prison term, Magistrate Court	0.092	(0.004)	0	0.092	(0.004)	0	0.089	(0.004)	0	0.093	(0.004)	0
Nazareth Court	0.269	(0.158)	0.089	0.619	(0.258)	0.017	0.545	(0.255)	0.033	0.49	(0.216)	0.024
Jerusalem Court	-0.314	(0.111)	0.005	-0.162	(0.182)	0.374	-0.193	(0.176)	0.273	-0.212	(0.148)	0.153
Female Defendant	-0.435	(0.189)	0.022	-0.412	(0.182)	0.025	-0.506	(0.193)	0.009	-0.382	(0.211)	0.071
Previous Criminal Record	-0.048	(0.066)	0.466	-0.039	(0.064)	0.546	-0.059	(0.073)	0.417	-0.054	(0.071)	0.447
Average Judge Age				0.012	(0.015)	0.408	0.011	(0.016)	0.469	0.012	(0.015)	0.437
Average Judge Experience				0.022	(0.02)	0.278	0.024	(0.021)	0.27	0.015	(0.021)	0.491
Female Judges				-0.051	(0.081)	0.524	-0.067	(0.073)	0.354	-0.052	(0.067)	0.436
Physical or Sexual Assault							0.094	(0.103)	0.36	0.1	(0.098)	0.311
Property or Fraud							0.403	(0.112)	0	0.344	(0.109)	0.002
Fiscal, Economic, or Business							0.142	(0.143)	0.32	0.141	(0.135)	0.298
Jewish Victim							-0.003	(0.096)	0.977	0.084	(0.081)	0.299
Prosecution requested rejection of appeal										-0.536	(0.124)	0
Convicted by guilty plea										-0.011	(0.086)	0.898
Convicted by trial										-0.135	(0.128)	0.295
(Intercept)	2.6	(0.103)	0	1.314	(0.898)	0.144	1.195	(0.907)	0.189	1.655	(0.781)	0.035
N. Predictors	9			12			16			19		
AIC Value	-258.306			-257.543			-270.825			-311.037		
Observations	379			379			379			379		
Marg. Effect of Mixed Panel (Jewish Def.)	-3.331	(1.063)	0.002	-3.265	(1.102)	0.003	-3.095	(1.064)	0.004	-2.906	(1.036)	0.005
Marg. Effect of Mixed Panel (Arab Def.)	-2.816	(1.254)	0.025	-2.755	(1.254)	0.029	-2.693	(1.238)	0.03	-1.721	(1.233)	0.164

Table 12: **Positive district-level prison terms – Parsimonious OLS with square-root response**. Marginal effects are given in terms of prison term (at mean of other predictors given ethnicity).

Square root of prison term	Est_1	SE_1	Pval_1	Est_2	SE_2	Pval_2	Est_3	SE_3	Pval_3	Est_4	SE_4	Pval_4
Arab Defendant	-0.144	(0.12)	0.232	-0.139	(0.118)	0.24	-0.068	(0.124)	0.585	-0.207	(0.117)	0.078
Mixed Panel	-0.417	(0.145)	0.004	-0.381	(0.144)	0.008	-0.307	(0.101)	0.002	-0.387	(0.135)	0.004
Mixed Panel × Arab Defendant	0.079	(0.131)	0.546	0.052	(0.134)	0.698	0.086	(0.133)	0.519	0.165	(0.129)	0.2
Female Defendant	-0.425	(0.188)	0.025				-0.583	(0.173)	0.001	-0.414	(0.2)	0.039
Jerusalem Court	-0.318	(0.113)	0.005	-0.188	(0.135)	0.164	-0.361	(0.102)	0	-0.315	(0.101)	0.002
Nazareth Court	0.259	(0.158)	0.102	0.558	(0.214)	0.01				0.257	(0.16)	0.108
Prison term, Magistrate Court	0.092	(0.004)	0	0.092	(0.004)	0	0.088	(0.004)	0	0.092	(0.004)	0
Average Judge Age				0.024	(0.014)	0.078						
Property or Fraud							0.355	(0.068)	0	0.315	(0.063)	0
Regulatory							-0.257	(0.138)	0.063			
Prosecution requested rejection of appeal										-0.523	(0.13)	0
(Intercept)	2.576	(0.101)	0	0.917	(0.962)	0.341	2.538	(0.101)	0	2.78	(0.125)	0
N. Predictors	8			8			9			10		
AIC Value	-259.898			-259.855			-278.225			-318.471		
Observations	379			379			379			379		
Marg. Effect of Mixed Panel (Jewish Def.)	-3.342	(1.061)	0.002	-3.051	(1.085)	0.005	-2.504	(0.828)	0.003	-3.098	(0.97)	0.002
Marg. Effect of Mixed Panel (Arab Def.)	-2.748	(1.227)	0.026	-2.663	(1.202)	0.027	-1.573	(0.987)	0.112	-1.88	(1.169)	0.109

Figure 5: Histograms of raw positive prison terms, and prison terms under a square root transform. The transformation largely eliminates the skewness of the distribution, providing a better basis for the assumption of normal errors required in interpreting regression coefficients.
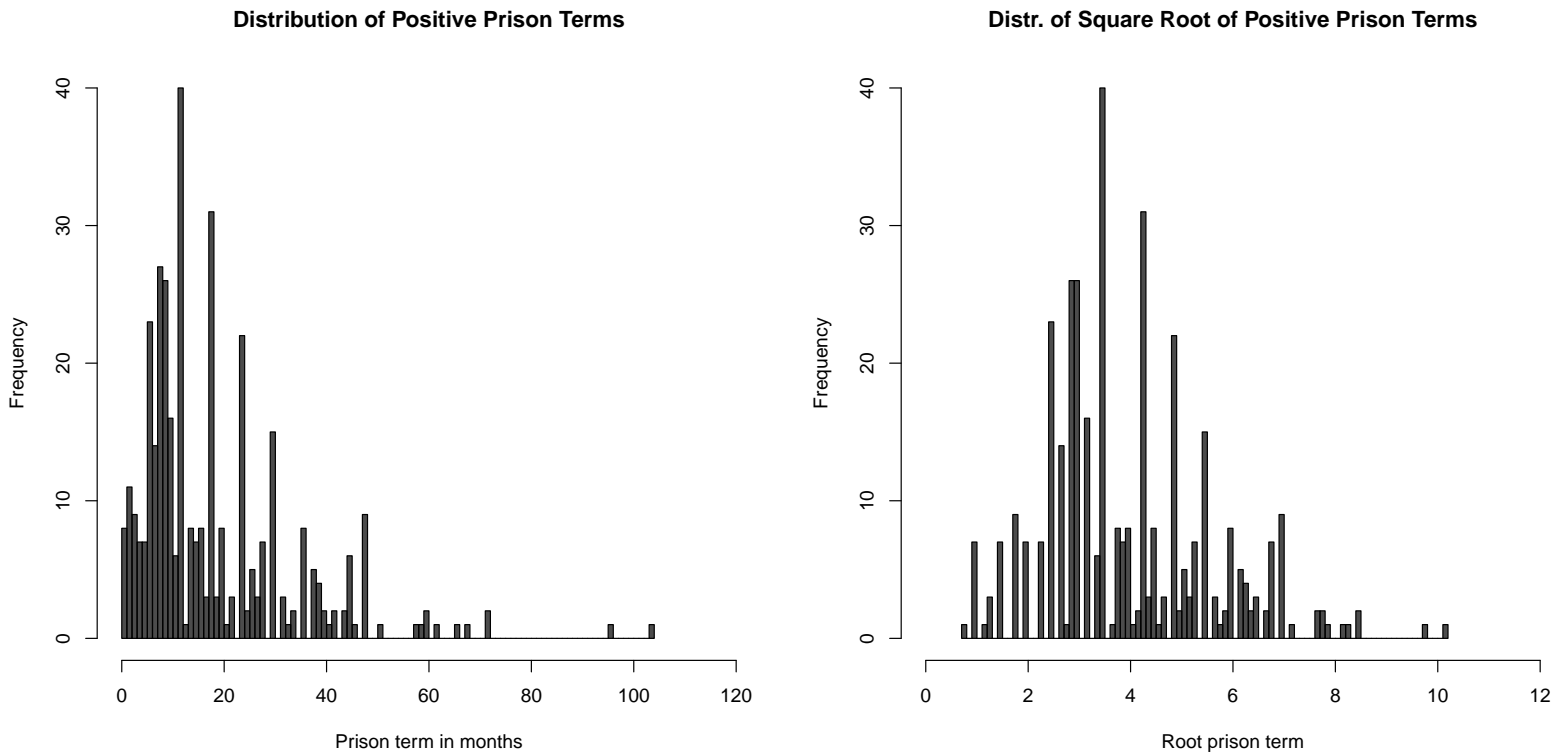
Figure 6: Marginal effects of facing a mixed panel on length of prison term (for cases with nonzero prison terms) for both the parsimonious (left-hand side) and full (right-hand side) OLS models fitted on the square root of the response. The marginal effects are non-constant and have been calculated by averaging over marginal effects estimated by the model for each defendant (of the given ethnicity) in the dataset at the observed values of the covariates for this defendant.. The solid lines represent 90% confidence intervals and the dashed lines 95% confidence intervals.
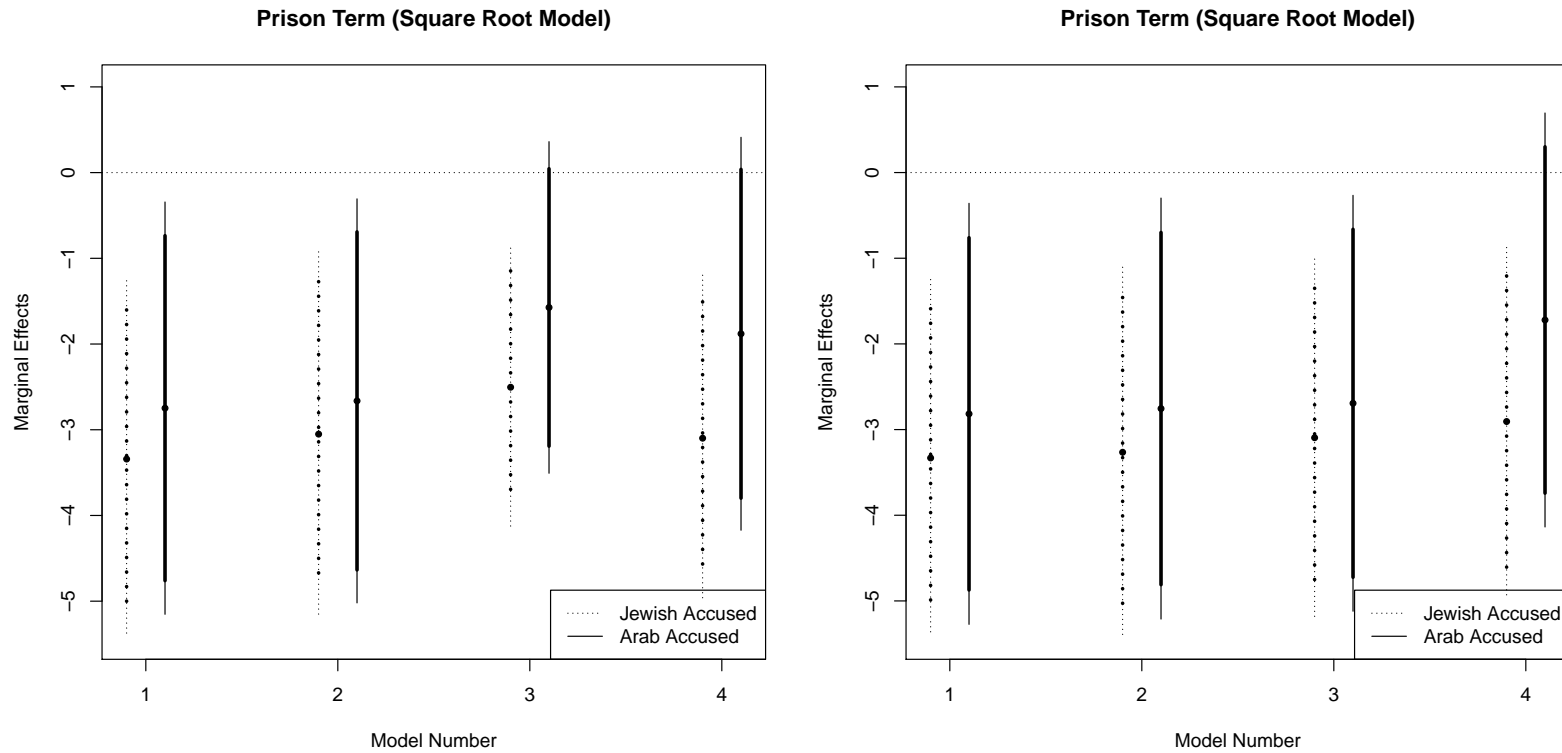
6 Robustness Check V: Propensity Score Weighting

As an alternative to the parametric models considered up to this point, we performed a non-parametric analysis using propensity score weighting. While propensity score weighting or matching is not satisfactory as a main line of analysis — unequal distribution of mixed panels across courts makes it impossible to balance court effects adequately — it provides an additional robustness check. We used the twang package in R, which fits propensity scores using generalized boosted regression, for this analysis. In particular, we fit propensity scores (for the propensity to face a mixed panel rather than an all-Jewish one) within the following slices of the data:

- cases with Arab defendants in which the defense appealed

- cases with Jewish defendants in which the defense appealed

- cases with Arab defendants in which the prosecution appealed

- cases with Jewish defendants in which the prosecution appealed

These subsets of the data allow analysis of the leniency and harshness rates for Arabs and Jews parallel to the one given by the regression model described in the main text. As inputs to the generalized boosted regression algorithm for fitting the propensity score, we gave the full set of covariates used in the largest regression models in the main paper (including length of prison term assigned by the magistrate court). Tables 14 through 21 give balance measures on these covariates for each of the four subsets of the data, both before and after propensity score weighting, and figures 7 through 10 show diagnostics for the fitted scores. While the distributions of fitted propensity scores for those facing mixed panels and those facing all-Jewish panels do not overlap as much as would be ideal and post-weighting balance is not perfect, the weighting approach does improve balance on many important covariates in every subset of the data.

After re-weighting, we performed a logistic regression of the relevant outcome (either leniency or harshness) against the mixed panel indicator in each subset of the data. Table 13 summarizes the fitted effects of the mixed panel obtained by this method. The results

agree exactly with the findings in the main paper: mixed panels cause an estimated increase in leniency of 16% for Arabs and an estimated decrease in harshness of 24%. The results for the harshness outcome are significant at the 90% level, whereas for leniency the p-value falls just below significance levels. In contrast, the effects for Jewish defendants are much smaller with higher p-values, essentially null results. Thus under an entirely different estimation approach, the results of the main paper still hold.

Table 13: **Outcome for each propensity score weighting.**

Outcome	Defendant ethnicity	Estimate	Std. Error	t value	$\Pr(> |t|)$
Leniency	Arab	0.16	0.12	1.32	0.19
Leniency	Jew	-0.05	0.08	-0.62	0.54
Harshness	Arab	-0.24	0.14	-1.71	0.09
Harshness	Jew	-0.09	0.12	-0.71	0.48

Table 14: **Balance before weighting for defendant appeals with Arab defendants.**

Defendant Appeals: Arab Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	11.83	(13.18)	18.02	(19.84)	-0.40	-2.06	0.04
Nazareth Court	0.94	(0.24)	0.08	(0.27)	1.88	19.64	0.00
Jerusalem Court	0.04	(0.19)	0.72	(0.45)	-1.63	-10.36	0.00
Female Defendant	0.05	(0.22)	0.06	(0.24)	-0.03	-0.18	0.86
Previous Criminal Record	0.57	(0.50)	0.74	(0.44)	-0.34	-2.02	0.03
Average Judge Age	55.62	(3.67)	61.8	(4.45)	-1.29	-8.81	0.00
Average Judge Experience	15.7	(2.63)	19.91	(4.79)	-1.09	-5.92	0.00
Female Judges	1.36	(0.55)	1.08	(0.70)	0.45	2.54	0.01
Physical or Sexual Assault	0.42	(0.50)	0.48	(0.50)	-0.13	-0.76	0.45
Property or Fraud	0.50	(0.50)	0.50	(0.50)	0.00	0.00	1.00
Fiscal, Economic, or Business	0.17	(0.37)	0.20	(0.40)	-0.09	-0.51	0.61
Jewish Victim	0.40	(0.49)	0.40	(0.48)	0.00	0.00	0.86
Prosecution requested rejection of appeal	0.84	(0.36)	0.70	(0.46)	0.35	1.92	0.06
Convicted by guilty plea	0.40	(0.49)	0.62	(0.49)	-0.44	-2.70	0.01
Convicted by trial	0.09	(0.29)	0.22	(0.42)	-0.39	-2.02	0.04

Table 15: **Balance after weighting for defendant appeals with Arab defendants.**

Defendant Appeals: Arab Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	11.77	(12.87)	10.27	(17.34)	0.10	0.48	0.64
Nazareth Court	0.87	(0.34)	0.47	(0.50)	0.85	2.84	0.00
Jerusalem Court	0.09	(0.29)	0.41	(0.50)	-0.76	-2.81	0.00
Female Defendant	0.06	(0.23)	0.14	(0.35)	-0.29	-0.76	0.45
Previous Criminal Record	0.59	(0.49)	0.42	(0.05)	0.35	1.48	0.14
Average Judge Age	56.24	(4.25)	59.88	(4.14)	-0.80	-4.08	0.00
Average Judge Experience	16.01	(3.02)	17.11	(4.93)	-0.28	-1.05	0.30
Female Judges	1.30	(0.62)	1.04	(0.54)	0.43	2.87	0.00
Physical or Sexual Assault	0.43	(0.50)	0.39	(0.49)	0.09	0.35	0.73
Property or Fraud	0.48	(0.50)	0.61	(0.49)	-0.25	-1.01	0.31
Fiscal, Economic, or Business	0.16	(0.37)	0.11	(0.32)	0.14	0.87	0.38
Jewish Victim	0.39	(0.49)	0.53	(0.48)	-0.29	-1.13	0.28
Prosecution requested rejection of appeal	0.85	(0.36)	0.73	(0.45)	0.30	1.06	0.29
Convicted by guilty plea	0.41	(0.49)	0.46	(0.50)	-0.12	-0.44	0.66
Convicted by trial	0.12	(0.32)	0.12	(0.33)	-0.02	-0.11	0.92

Table 16: **Balance before weighting for defendant appeals with Jewish defendants.**

Defendant Appeals: Jewish Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	15.54	(16.70)	15.22	(15.10)	0.02	0.17	0.87
Nazareth Court	0.63	(0.48)	0.04	(0.20)	1.24	13.38	0.00
Jerusalem Court	0.04	(0.19)	0.41	(0.49)	-0.91	-8.06	0.00
Female Defendant	0.02	(0.15)	0.09	(0.27)	-0.28	-2.28	0.02
Previous Criminal Record	0.69	(0.47)	0.59	(0.49)	0.20	1.66	0.11
Average Judge Age	58.18	(5.36)	63.55	(4.89)	-0.93	-8.59	0.00
Average Judge Experience	18.52	(5.36)	22.50	(5.17)	-0.71	-6.20	0.00
Female Judges	1.39	(0.58)	1.45	(0.74)	-0.09	-0.72	0.47
Physical or Sexual Assault	0.53	(0.50)	0.46	(0.50)	0.13	1.10	0.27
Property or Fraud	0.45	(0.50)	0.37	(0.48)	0.16	1.28	0.20
Fiscal, Economic, or Business	0.16	(0.37)	0.28	(0.45)	-0.29	-2.37	0.02
Jewish Victim	0.64	(0.47)	0.42	(0.48)	0.44	3.69	0.00
Prosecution requested rejection of appeal	0.81	(0.40)	0.72	(0.45)	0.21	1.69	0.09
Convicted by guilty plea	0.39	(0.49)	0.49	(0.50)	-0.21	-1.70	0.10
Convicted by trial	0.21	(0.41)	0.32	(0.46)	-0.26	-2.10	0.04

Table 17: **Balance after weighting for defendant appeals with Jewish defendants.**

Defendant Appeals: Jewish Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	16.68	(17.03)	15.74	(15.66)	0.06	0.34	0.73
Nazareth Court	0.52	(0.50)	0.23	(0.42)	0.59	3.02	0.00
Jerusalem Court	0.11	(0.32)	0.31	(0.47)	-0.48	-2.97	0.00
Female Defendant	0.03	(0.15)	0.07	(0.25)	-0.21	-1.73	0.09
Previous Criminal Record	0.70	(0.47)	0.68	(0.47)	0.05	0.34	0.66
Average Judge Age	59.29	(5.57)	61.78	(5.50)	-0.44	-2.56	0.01
Average Judge Experience	19.03	(5.41)	20.60	(6.00)	-0.27	-1.52	0.13
Female Judges	1.25	(0.67)	1.38	(0.72)	-0.17	-1.14	0.25
Physical or Sexual Assault	0.53	(0.50)	0.48	(0.50)	0.10	0.61	0.54
Property or Fraud	0.44	(0.50)	0.41	(0.49)	0.06	0.38	0.70
Fiscal, Economic, or Business	0.18	(0.38)	0.24	(0.43)	-0.16	-1.14	0.25
Jewish Victim	0.61	(0.48)	0.46	(0.49)	0.31	1.95	0.06
Prosecution requested rejection of appeal	0.81	(0.39)	0.66	(0.47)	0.33	1.95	0.05
Convicted by guilty plea	0.41	(0.49)	0.51	(0.50)	-0.21	-1.34	0.19
Convicted by trial	0.21	(0.41)	0.25	(0.43)	-0.10	-0.69	0.49

Table 18: **Balance before weighting for prosecutorial appeals with Arab defendants.**

Prosecutor Appeals: Arab Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	11.83	(13.18)	18.02	(19.84)	-0.40	-2.06	0.04
Nazareth Court	0.94	(0.24)	0.08	(0.27)	1.88	19.64	0.00
Jerusalem Court	0.04	(0.19)	0.72	(0.45)	-1.63	-10.36	0.00
Female Defendant	0.05	(0.22)	0.06	(0.24)	-0.03	-0.18	0.86
Previous Criminal Record	0.57	(0.50)	0.74	(0.44)	-0.34	-2.20	0.03
Average Judge Age	55.62	(3.67)	61.80	(4.45)	-1.29	-8.81	0.00
Average Judge Experience	15.70	(2.63)	19.91	(4.79)	-1.09	-5.92	0.00
Female Judges	1.36	(0.55)	1.08	(0.70)	0.45	2.54	0.01
Physical or Sexual Assault	0.42	(0.50)	0.48	(0.50)	-0.13	-0.76	0.45
Property or Fraud	0.50	(0.50)	0.50	(0.50)	0.00	0.00	1.00
Fiscal, Economic, or Business	0.17	(0.37)	0.20	(0.40)	-0.09	-0.51	0.61
Jewish Victim	0.40	(0.49)	0.40	(0.48)	0.00	0.00	0.86
Prosecution requested rejection of appeal	0.84	(0.36)	0.70	(0.46)	0.35	1.92	0.06
Convicted by guilty plea	0.40	(0.49)	0.62	(0.49)	-0.44	-2.70	0.01
Convicted by trial	0.09	(0.29)	0.22	(0.42)	-0.39	-2.02	0.04

Table 19: **Balance after weighting for prosecutorial appeals with Arab defendants.**

Prosecutor Appeals: Arab Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	11.77	(12.87)	10.27	(17.34)	0.10	0.48	0.64
Nazareth Court	0.87	(0.34)	0.47	(0.50)	0.85	2.84	0.00
Jerusalem Court	0.09	(0.29)	0.41	(0.50)	-0.76	-2.81	0.00
Female Defendant	0.06	(0.23)	0.14	(0.35)	-0.29	-0.76	0.45
Previous Criminal Record	0.59	(0.49)	0.42	(0.50)	0.35	1.48	0.14
Average Judge Age	56.24	(4.25)	59.88	(4.14)	-0.80	-4.08	0.00
Average Judge Experience	16.01	(3.02)	17.11	(4.93)	-0.28	-1.05	0.30
Female Judges	1.30	(0.62)	1.04	(0.54)	0.43	2.87	0.00
Physical or Sexual Assault	0.43	(0.50)	0.39	(0.49)	0.09	0.35	0.73
Property or Fraud	0.48	(0.50)	0.61	(0.49)	-0.25	-1.01	0.31
Fiscal, Economic, or Business	0.16	(0.37)	0.11	(0.32)	0.14	0.87	0.38
Jewish Victim	0.39	(0.49)	0.53	(0.48)	-0.29	-1.13	0.28
Prosecution requested rejection of appeal	0.85	(0.36)	0.73	(0.45)	0.30	1.06	0.29
Convicted by guilty plea	0.41	(0.49)	0.46	(0.50)	-0.12	-0.44	0.66
Convicted by trial	0.12	(0.32)	0.12	(0.33)	-0.02	-0.11	0.92

Table 20: **Balance before weighting for prosecutorial appeals with Jewish defendants.**

Prosecutor Appeals: Jewish Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	11.90	(16.34)	6.11	(9.87)	0.46	1.77	0.08
Nazareth Court	0.45	(0.51)	0.02	(0.13)	1.15	4.55	0.00
Jerusalem Court	0.07	(0.26)	0.50	(0.50)	-0.90	-5.24	0.00
Female Defendant	0.00	(0.03)	0.03	(0.15)	-0.20	-1.20	0.25
Previous Criminal Record	0.59	(0.51)	0.52	(0.50)	0.15	0.65	0.53
Average Judge Age	60.68	(5.36)	64.61	(3.75)	-0.83	-3.56	0.00
Average Judge Experience	20.68	(5.44)	22.61	(3.82)	-0.43	-1.73	0.09
Female Judges	1.55	(0.63)	1.25	(0.79)	0.40	1.92	0.06
Physical or Sexual Assault	0.45	(0.51)	0.46	(0.50)	-0.03	-0.14	0.89
Property or Fraud	0.41	(0.50)	0.29	(0.46)	0.27	1.16	0.25
Fiscal, Economic, or Business	0.41	(0.50)	0.32	(0.47)	0.19	0.83	0.41
Jewish Victim	0.54	(0.50)	0.36	(0.46)	0.38	1.63	0.11
Prosecution requested rejection of appeal	0.14	(0.35)	0.00	(0.00)	0.65	2.14	0.04
Convicted by guilty plea	0.34	(0.48)	0.57	(0.50)	-0.45	-2.05	0.04
Convicted by trial	0.34	(0.48)	0.28	(0.44)	0.15	0.64	0.52

Table 21: **Balance after weighting for prosecutorial appeals with Jewish defendants.**

Prosecutor Appeals: Jewish Defendants	Mean Mixed	SE Mixed	Mean All-Jewish	SE All-Jewish	Std. Effect	T-stat	P-value
Prison term, Magistrate Court	10.68	(16.92)	5.93	(9.63)	0.36	1.29	0.20
Nazareth Court	0.35	(0.49)	0.04	(0.21)	0.81	3.05	0.00
Jerusalem Court	0.18	(0.39)	0.48	(0.50)	-0.62	-2.29	0.02
Female Defendant	0.00	(0.03)	0.03	(0.14)	-0.22	-1.26	0.22
Previous Criminal Record	0.62	(0.50)	0.53	(0.50)	0.17	0.69	0.50
Average Judge Age	61.81	(5.37)	64.36	(3.89)	-0.54	-2.07	0.04
Average Judge Experience	20.88	(5.06)	22.41	(4.06)	-0.34	-1.39	0.17
Female Judges	1.37	(0.79)	1.23	(0.77)	0.18	0.61	0.54
Physical or Sexual Assault	0.48	(0.51)	0.48	(0.50)	0.01	0.02	0.98
Property or Fraud	0.38	(0.50)	0.27	(0.45)	0.24	0.98	0.33
Fiscal, Economic, or Business	0.35	(0.49)	0.32	(0.47)	0.07	0.29	0.77
Jewish Victim	0.46	(0.50)	0.38	(0.47)	0.18	0.72	0.48
Prosecution requested rejection of appeal	0.11	(0.32)	0.00	(0.00)	0.52	2.03	0.05
Convicted by guilty plea	0.27	(0.45)	0.57	(0.50)	-0.60	-2.78	0.01
Convicted by trial	0.33	(0.48)	0.29	(0.45)	0.10	0.39	0.70

Figure 7: Diagnostic Plots for weighting of Arab defendants for leniency analysis. The leftmost plot shows overlap in the propensity scores for the treated (mixed panel) and control (all-Jewish panel) groups, the middle panel shows the change in balance due to weighting for each of the balance variables with effect size less than 3 (solid points represent statistically significant differences), and the rightmost plot compares the pre-weighting and post-weighting p-values to the uniform distribution.
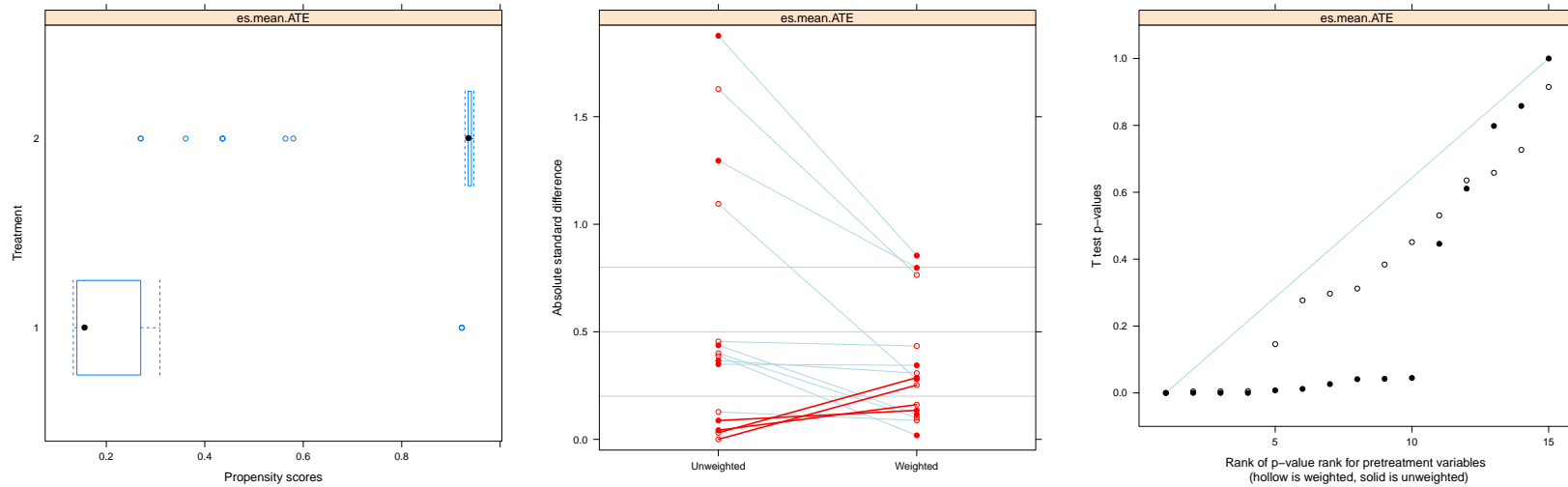
Figure 8: Diagnostic Plots for weighting of Jewish defendants for leniency analysis. The leftmost plot shows overlap in the propensity scores for the treated (mixed panel) and control (all-Jewish panel) groups, the middle panel shows the change in balance due to weighting for each of the balance variables with effect size less than 3 (solid points represent statistically significant differences), and the rightmost plot compares the pre-weighting and post-weighting p-values to the uniform distribution.
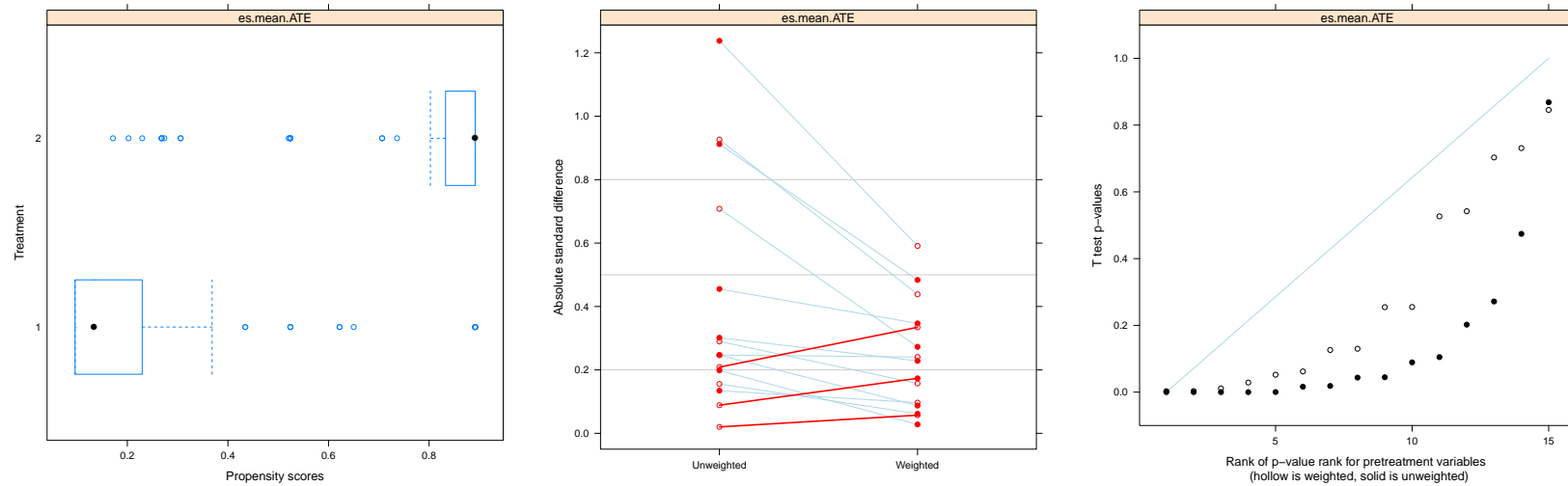
Figure 9: Diagnostic Plots for weighting of Arab defendants for harshness analysis. The leftmost plot shows overlap in the propensity scores for the treated (mixed panel) and control (all-Jewish panel) groups, the middle panel shows the change in balance due to weighting for each of the balance variables with effect size less than 3 (solid points represent statistically significant differences), and the rightmost plot compares the pre-weighting and post-weighting p-values to the uniform distribution.

Figure 10: Diagnostic Plots for weighting of Jewish defendants for harshness analysis. The leftmost plot shows overlap in the propensity scores for the treated (mixed panel) and control (all-Jewish panel) groups, the middle panel shows the change in balance due to weighting for each of the balance variables with effect size less than 3 (solid points represent statistically significant differences), and the rightmost plot compares the pre-weighting and post-weighting p-values to the uniform distribution.
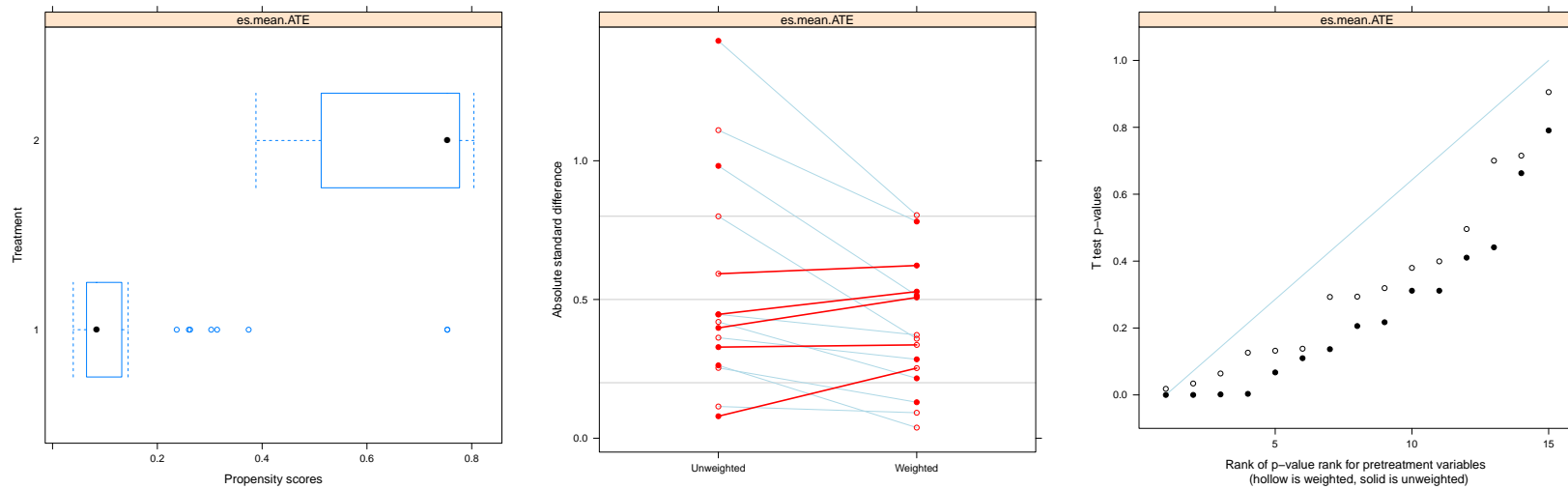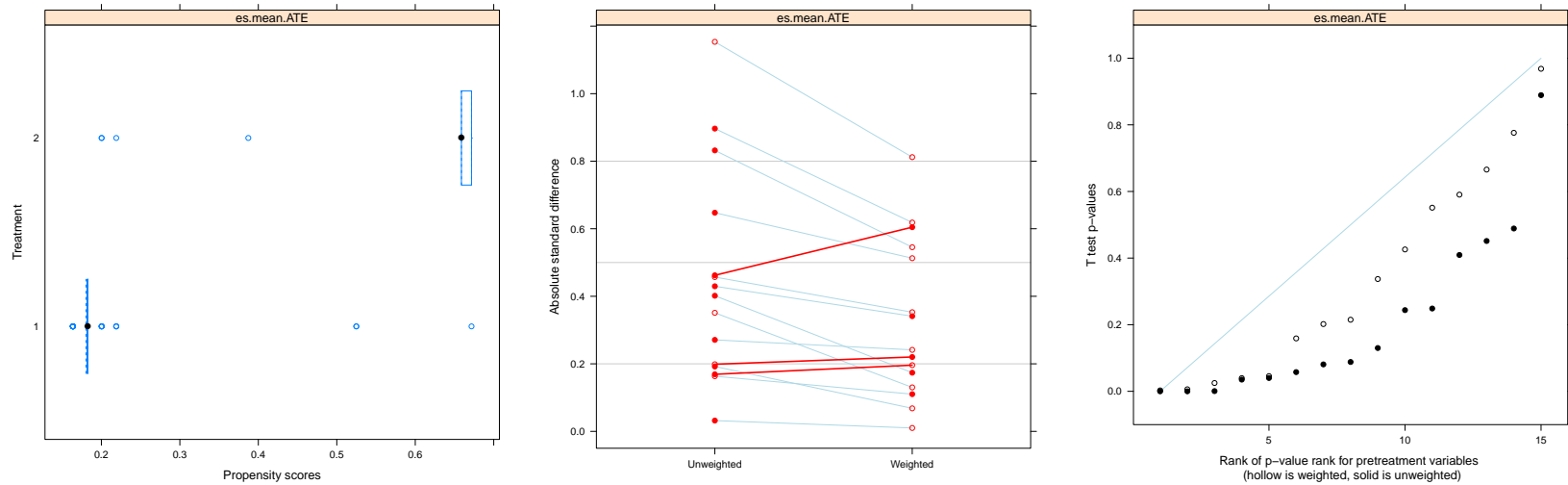
7 Text Analysis

We utilize innovations in automated text analysis to assess whether the publicly available judicial opinions can shed light on potential causal mechanisms for the ethnicity-based panel effects identified in this study. As we argue in the main text, though we cannot observe the deliberation or arguments among district court judges, the deliberative process may be reflected in written opinions. The idea is that if minority judges provide a different perspective based on their life experiences and their better understanding of the consequences leading minorities to commit crimes—as assumed by deliberation arguments—then this perspective may be reflected in the arguments provided by judges to justify the outcome of the judicial process.

To test this proposition we conduct a form of text analysis known as multinomial inverse regression (MINR) (Taddy, 2012). MINR utilizes unique properties of the multinomial distribution to obtain a low-dimensional projection of the correlation between count data in a document (e.g., word or phrase counts) and an associated outcome variable. As shown by Taddy (2012), the inverse distribution of text given a "sentiment" can be used to obtain sufficient reduction scores, which, when conditioned-upon, render the outcome variable conditionally independent of the original count data.

As an initial step, we preprocess the text as is customary in content-centric text analysis using the "bag of words" assumption to ensure that words and phrases are identified by substantive lexical meaning rather than idiosyncratic suffixes, punctuation or numerical characters (Grimmer and Stewart, 2013). We extract the raw text from the HTML opinion files; remove non-alef-bet characters; apply the MILA XML parser (Mitts and Mitts, 2013) to extract dictionary forms of Hebrew words, eliminate common "stop" words in Hebrew using a list prepared by Mitts (2013), tokenize the raw text into unigrams (1-word), bigrams (2-word phrases), and trigrams (3-word phrase); and generate a document-term-matrix consisting of the frequencies of each unigram, bigram or trigram in each document. We were able to obtain the full-text of all but one of the opinions in the dataset, yielding a document-term-matrix of 543 documents and approximately 30,500 terms in total.

To identify potential causal mechanisms for the interaction term mixed panels with

Arab defendant, we estimate MINR using a new "outcome" variable that is set to 1 for cases with mixed panels equals 1, and Arab defendant equals 1 (see Data Subset A). We then run a separate analysis for (a) defendant appeals (182 cases in which the panel decides on leniency), (b) prosecution appeals (51 cases the panel decides on harshness), and (c) the union of defendant and prosecution (all 220 cases with an Arab defendant in which the panel decides on incarceration). Finally, we perform the above estimations with no covariates to identify simple correlations between text and the outcome as well as all of the covariates to identify the link between text and the outcome after partialing out the effect of court- and case-specific attributes. This yields a total of 18 estimations: 3 token types (unigrams, bigrams, and trigrams) by 3 outcome variables by 2 sets of covariates (with and without controls).

Data Subset A

Outcome = 1: Arab defendant = 1 & Mixed Panel = 1

Outcome = 0: Arab defendant = 1 & Mixed Panel = 0

For brevity we display in Tables 22–24 the top 30 words (unigrams) associated with mixed-panels and All-Jewish panels by "score," i.e., correlation with the MINR outcome variable, for regression models with controls for each of the three types of cases, described above.[1] We color code words that arguably point to the defendant background or to circumstances that should be associated with more lenient ruling, we use red for words that arguably associated with harsher sentencing, and green for words that likely are related to appeal justification, but their direction is more ambiguous (e.g., "apparently"). As evidenced in Tables 22–24, we do not find much evidence in support for a deliberation effect.

The above specification of the new 'outcome' variable (Data Subset A) has the advantage of a clear causal interpretation as well as a relative large sample size. On the other hand, it also suffers from a clear drawback; that cases in which the appeal has been accepted are thought to be part of the same "sentiment" as cases in which the appeal has been rejected. We therefore construct two alternative subsets of the data: Subset B con-

---

[1] Additional results can be obtained from the authors upon request.

sists solely of those cases where the original outcome variable equals 1 (i.e., only cases in which the Arab defendant's appeal has been accepted), and Subset B consists of all cases, including Jewish defendants

Data Subset B

Outcome = 1: Arab defendant = 1 & lenient = 1 & Mixed Panel = 1
Outcome = 0: Arab defendant = 1 & lenient = 1 & Mixed Panel = 0

Data Subset C

Outcome = 1: Arab defendant = 1 & lenient = 1 & Mixed Panel = 1
Outcome = 0: all cases

Unfortunately, Subset A yields a very small subset of the data. Subset B provides a much larger n for the comparison group where the outcome variable is coded 0, but textual correlations may not be attributed solely to the interaction term. It is thus important to note that in these alternative specifications, the top words and phrases are not descriptively suggestive of causal mechanisms. Again, we find that the words and phrases in these alternative specifications are not informative, suggesting that the estimation process is unable to identify meaningful words that may be associated with specific outcomes.

Finally, we compliment the qualitative text analysis with a more qualitative approach, examining the full text of the 543 available opinions. We find that Israeli appellate courts generally supply an extremely cursory description of the facts of the case, if at all. Several of the opinions are no more than a few sentences where the court simply reduces the judgment without even discussing the case. Others briefly describe the facts but do not provide descriptive rationales or justifications beyond emphasizing the defendant's medical condition or harshness of the sentence. The absence of substantial descriptive text may help explain why our automated text analysis was unable to advance much our understanding of the potential causal mechanisms for the panel effects identified in the paper.

Table 22: Lenient - defendant appeals

	All-Jewish	Score	Mixed	Score
1	Tevet (month)	-0.0079	Prisoner	0.0083
2	Kept	-0.0075	Gap	0.0073
3	Municipality	-0.0069	Hide	0.0066
4	Hague	-0.0068	Story	0.0059
5	Included	-0.0066	Confessed	0.0058
6	Vacated	-0.0065	Deviation	0.0057
7	Hospitalized	-0.0063	Development	0.0056
8	Wire	-0.0062	Sabath	0.0053
9	Doctor	-0.0061	Turns out	0.0052
10	Factory	-0.0061	Plane	0.0052
11	Blow	-0.0060	Point	0.0050
12	Implementation	-0.0060	Ownerless	0.0047
13	Negation	-0.0055	Cash	0.0045
14	Read out	-0.0054	Detainee	0.0043
15	Placed	-0.0054	Direction	0.0042
16	Union	-0.0053	Stab	0.0042
17	Cowshed	-0.0052	Comment	0.0041
18	Put	-0.0052	Translator	0.0039
19	Calf	-0.0050	Carved	0.0037
20	Daud	-0.0050	Turn	0.0035
21	Was taken	-0.0050	Nothing	0.0031
22	Beach	-0.0050	Took over	0.0031
23	Affair	-0.0050	Fixed	0.0030
24	Lung	-0.0050	Apparently	0.0030
25	Creep	-0.0050	Filling	0.0030
26	Secured	-0.0050	Carpenter	0.0030
27	Kinnerert	-0.0050	Stand up	0.0029
28	Effect	-0.0049	Disorder	0.0028
29	Forced	-0.0048	Cohen	0.0027
30	Attracted	-0.0047	Threat	0.0027

Table 23: Harsher - prosecution appeals

	All-Jewish	Score	Mixed	Score
1	Contributed	-0.1740	Mizrachi	0.3205
2	Cancelled	-0.1456	Held	0.1760
3	Traveler	-0.1438	Was asked	0.1334
4	Road	-0.1353	Point	0.1035
5	Drove	-0.1346	Disappearance	0.1002
6	Agreed	-0.1310	Weapon	0.0881
7	Death	-0.1292	Served time	0.0687
8	Accident	-0.1259	Prize	0.0658
9	Blamed	-0.1257	Situation	0.0646
10	Paid	-0.1245	Second	0.0621
11	Truck	-0.1242	Usage	0.0600
12	Driver	-0.1234	Possession	0.0596
13	Shvat (month)	-0.1231	Plentiful	0.0545
14	Cause	-0.1206	Devoted	0.0541
15	Serfdom	-0.1191	Moses	0.0537
16	Room	-0.1081	Short	0.0512
17	Nazareth	-0.1074	Payment	0.0502
18	Company	-0.1065	Added value	0.0502
19	Movement	-0.1012	Risk	0.0501
20	Reassured	-0.0959	More	0.0445
21	Omri	-0.0926	Period	0.0442
22	Disguise	-0.0925	Void	0.0431
23	Charisma	-0.0925	Detainee	0.0430
24	Slipped away	-0.0925	Spread	0.0413
25	Raped	-0.0925	Chronology	0.0413
26	Couple	-0.0912	Committee	0.0413
27	Replaced	-0.0906	Variable	0.0413
28	Attempted	-0.0890	Displaced	0.0413
29	Meeting	-0.0888	Overlap	0.0413
30	Juvenile	-0.0888	On	0.0413

Table 24: Incarceration - all cases

	All-Jewish	Score	Mixed	Score
1	Resided at	-0.0133	Plane	0.0169
2	Value	-0.0126	Prisoner	0.0144
3	Wire	-0.0125	Lawyer	0.0136
4	Union	-0.0110	Story	0.0121
5	Municipality	-0.0106	List	0.0117
6	Responded	-0.0104	Gap	0.0116
7	Failure	-0.0101	Said	0.0109
8	Planned	-0.0098	Equipped	0.0107
9	Vacated	-0.0095	Point	0.0105
10	Serfdom	-0.0094	Development	0.0098
11	Trained	-0.0092	Influence	0.0098
12	Slowed	-0.0092	Turns out	0.0095
13	Divorced	-0.0091	Body part	0.0094
14	Was read	-0.0091	Ownerless	0.0090
15	Gang	-0.0086	Detainee	0.0086
16	Advertisement	-0.0086	Circumvented	0.0086
17	Hospitalized	-0.0086	Sabath	0.0080
18	Guarded	-0.0086	Translator	0.0072
19	Functionality	-0.0085	Combined	0.0070
20	Cowshed	-0.0083	Took over	0.0069
21	Omri	-0.0082	Cash	0.0069
22	Hermon Mt.	-0.0082	Comment	0.0069
23	Involvement	-0.0081	Hidden	0.0066
24	Calf	-0.0080	Stone	0.0066
25	Universe	-0.0080	Police car	0.0065
26	Kinneret (lake)	-0.0080	Forgery	0.0064
27	Secured	-0.0080	Confessed	0.0064
28	Was led	-0.0080	Acceptance	0.0064
29	Beach	-0.0080	Disorder	0.0063
30	Affair	-0.0080	Tuned around	0.0062

8 Descriptive Statistics

Variable	$\bar{x}$	s	Min	$\tilde{x}$	Max	n
Female Judges	1.3	0.7	0	1	3	544
Arab Defendant	0.4	0.5	0	0	1	544
Female Defendant	0	0.2	-0.2	0	1	544
Previous Criminal Record	0.6	0.5	-0.64	1	1.5	544
Prison term, Magistrate Court	13	15.4	0	8	104	544
Sent to Prison, District Court	0.7	0.5	0	1	1	544
Prison term,: District Court	12.7	15.1	0	9	104	544
Convicted by guilty plea	0.5	0.5	-0.06	0	1.34	544
Convicted by trial	0.2	0.4	-0.2	0	1	544
Physical or Sexual Assault	0.5	0.5	0	0	1	544
Property or Fraud	0.4	0.5	0	0	1	544
Fiscal, Economic, or Business	0.2	0.4	0	0	1	544
Jewish Victim	0.44	0.5	-0.32	0	1.46	544
Prosecution requested rejection of appeal	0.6	0.5	-0.04	1	1.28	544
Defendant appealed	0.8	0.4	0	1	1	544
Prosecution appealed	0.2	0.4	0	0	1	544
More lenient higher-court verdict	0.3	0.5	0	0	1	544
Harsher higher-court verdict	0.2	0.4	0	0	1	544
Nazareth Court	0.4	0.5	0	0	1	544
Tel Aviv Court	0.3	0.5	0	0	1	544
Jerusalem Court	0.3	0.4	0	0	1	544
Average Judge Age	60	5.7	49.7	58.7	71	544
Average Judge Experience	19.4	5.3	10	17.3	28.3	544

Table 25: Descriptive Statistics after Imputation: District Appeals Sample, 2007-2011. Note that our imputation approach may assign continuous values to missing values in binary covariates to improve estimation. This explains the negative values listed as minima for several variables in this table, for instance.

9 Technical supplement: marginal effect calculation for square root model

The square-root response model for prison term has the following form:

$$\sqrt{Y} = \beta_A \mathbf{1}_A + \beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M + \boldsymbol{\beta} \mathbf{X} + \epsilon$$

where Y is the prison term in months, $\mathbf{1}_A$ is an indicator of an Arab defendant with regression coefficient β_A, $\mathbf{1}_M$ is an indicator of a mixed panel with coefficient β_M, $\beta_{A:M}$ is the interaction coefficient, $\mathbf{X}$ is a column of additional predictors (including an intercept column) with coefficients $\boldsymbol{\beta}$, and ϵ is a zero-mean Gaussian error term with variance σ^2. We can write the model in terms of Y as follows:

$$
\begin{aligned}
Y &= \left(\beta_A \mathbf{1}_A + \beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M + \boldsymbol{\beta} \mathbf{X} + \epsilon \right)^2 \\
&= \epsilon^2 + 2\epsilon \left(\beta_A \mathbf{1}_A + \beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M + \boldsymbol{\beta} \mathbf{X} \right) \\
&\quad + (\boldsymbol{\beta} \mathbf{X})^2 + \boldsymbol{\beta} \mathbf{X} \left(\beta_A \mathbf{1}_A + \beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M \right) \\
&\quad + \beta_A^2 \mathbf{1}_A + 2\beta_A \left(\beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M \right) \\
&\quad + \beta_M^2 \mathbf{1}_M + 2\beta_M \beta_{A:M} \mathbf{1}_A \mathbf{1}_M + \beta_{A:M}^2 \mathbf{1}_A \mathbf{1}_M \\
&= \mathbf{1}_A \left(2\beta_A \boldsymbol{\beta} \mathbf{X} + \beta_A^2 \right) + \mathbf{1}_M \left(2\beta_M \boldsymbol{\beta} \mathbf{X} + \beta_M^2 \right) \\
&\quad + \mathbf{1}_A \mathbf{1}_M \left(2\beta_{A:M} \boldsymbol{\beta} \mathbf{X} + 2\beta_{A:M} \beta_A + 2\beta_{A:M} \beta_M + \beta_{A:M}^2 \right) \\
&\quad + (\boldsymbol{\beta} \mathbf{X})^2 + 2\epsilon \left(\beta_A \mathbf{1}_A + \beta_M \mathbf{1}_M + \beta_{A:M} \mathbf{1}_A \mathbf{1}_M + \boldsymbol{\beta} \mathbf{X} \right) + \epsilon^2 .
\end{aligned}
$$

To derive the marginal effect ME of a mixed panel on the value of Y conditional on ethnicity, we compute

$$ME(\mathbf{1}_A) = E\left(Y | \mathbf{1}_M = 1, \mathbf{1}_A, \overline{\mathbf{X}}(\mathbf{1}_A) \right) - E\left(Y | \mathbf{1}_M = 0, \mathbf{1}_A, \overline{\mathbf{X}}(\mathbf{1}_A) \right)$$

where $\overline{\mathbf{X}}(\mathbf{1}_A)$ is the mean of the other covariates conditional on ethnicity. For Jews this evaluates to

$$ME(0) = 2\beta_M \boldsymbol{\beta} \mathbf{X} + \beta_M^2 ,$$

and for Arabs it evaluates to

$$ME(1) = 2\beta_M\boldsymbol{\beta}\mathbf{X} + \beta_M^2 + 2\beta_{A:M}\boldsymbol{\beta}\mathbf{X} + 2\beta_{A:M}\beta_A + 2\beta_{A:M}\beta_M + \beta_{A:M}^2.$$

References

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):269–297.

Mitts, Joshua and Tamar Mitts. 2013. "An XML Assembler for MILA Hebrew Dictionary Entries." *(R code)* .

Mitts, Tamar. 2013. "The Cultural Legacies of Conflict: An Automated Content Analysis of Google N-gram in Hebrew." *Mimeo* .

Taddy, Matt. 2012. "Multinomial Inverse Regression for Text Analysis." *Journal of the American Statistical Association* 108(503):755–770.