PUBPOL 6090 Problem Set 2

Thelonious Goerz

2023-10-03

Problem 1

1.1

In Table 1, I replicate the main result from Washington (2008). Based on this model specification, the interpretation for $\beta GIRLS$ is that for every one more Female child that a congressperson has there is a 2.37 point increase in their NOW score (which is a score of voting on women's issues) which is statistically significant.

1.2

Washington's identification strategy relies on the assumption 1) that child gender at birth is randomly assigned, 2) that parents are not following a "gender-biased stopping rule for fertility" which essentially states that parents are not changing their fertility behaviors based on child gender to select the gender type or composition of their children, and 3) that conditional on the number of children, the number of female children is a random variable. Additionally, Washington makes another assumption 4) that voters are not selecting constituents based on the gender composition of their children.

To yield a causal estimate, Washington conditions on the number of children and the number of daughters and sons, to isolate the effect. Therefore estimating the relationship between number of female children and voting behavior can be interpreted as causal with the specific interpretation of the coefficient that it is the "effect of daughters relative to sons" because of the linear dependence between children which makes it impossible to discern whether voting patterns are due to more influence from daughters, less from sons, or a combination of both.

1.3

In Table 2, I compare the results of a variety of regression models with different controls to select the most parsimonious model that still yields the true causal effect

In Model 7, I run the base specification from equation 2 in Washington (2008) which includes only child gender and a fixed effect for the total number of children. The point estimate is much larger than the preferred result (Model 1). The effect is likely very high considering that we do not have information gender and democratic vote shares which are both shown to be very predictive of NOW score as evidenced by model 1.

Based on these different specifications, I believe that Model 2 presents the best specification of the causal estimate that Washington identifies. My reasoning is that with regard to women's issues, party affiliation, gender, and race likely determine much of ideological stance toward voting. Indeed, the point estimate is very similar to Model 1 (2.47 compared to 2.30). I also believe that Model 5 is a reasonable specification, because it adds important information on age and religiosity, however, the point estimate is much larger than Model 2.

Model 2 still includes fixed effects for region and total number of children which satisfies Washington's first identifying 2nd identifying assumption. As a result, I conclude that Model 2 is the most parsimonious stand in for Model 1. Assumptions 1 and 3 are evaluated in the next section.

Table 1: Replication of Table 2, Column 1 Result From Washington (2008)

	(1)
Girls	2.30*
	(1.04)
Female	10.85***
	(2.69)
White	1.89
	(3.47)
Republican	−44.89***
	(2.11)
Service Length	0.24
	(0.30)
Service Length Squared	−0.01
	(0.01)
Age	0.66
	(0.80)
Age Squared	−0.01
	(0.01)
No Religion	7.28
	(7.03)
Catholic	−3.97*
	(1.94)
Other Christian Religion	0.78
	(4.61)
Other Religion	10.78**
	(3.82)
Democratic Vote Share	84.26***
	(10.92)
Num.Obs.	430
R2	0.820

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All Models include Child Count and Region Fixed Effects.

Table 2: Replication of Main Result From Washington (2008) With Additional Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Girls	2.30*	2.47*	4.70*	3.09*	2.52*	4.88*	5.34*
	(1.04)	(1.13)	(2.16)	(1.54)	(1.15)	(2.17)	(2.26)
Female	10.85***	10.53***			12.44***		
	(2.69)	(2.86)			(2.88)		
White	1.89	−12.70***					
	(3.47)	(2.86)					
Republican	−44.89***	−54.45***			−57.74***		
	(2.11)	(1.92)			(1.89)		
Service Length	0.24		−0.58			−0.08	
	(0.30)		(0.62)			(0.57)	
Service Length Squared	−0.01		0.03			0.02	
	(0.01)		(0.02)			(0.02)	
Age	0.66		2.09		0.28		
	(0.80)		(1.65)		(0.78)		
Age Squared	−0.01		−0.02		0.00		
	(0.01)		(0.02)		(0.01)		
No Religion	7.28			19.98+			
	(7.03)			(10.39)			
Catholic	−3.97*			1.61	−5.59**		
	(1.94)			(2.85)	(2.06)		
Other Christian Religion	0.78			−3.24			
	(4.61)			(6.83)			
Other Religion	10.78**			21.73***			
	(3.82)			(5.30)			
Democratic Vote Share	84.26***			186.30***			
	(10.92)			(10.27)			
Num.Obs.	430	430	430	430	430	430	430
R2	0.820	0.778	0.200	0.589	0.771	0.190	0.068

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ All Models include Child Count and Region Fixed Effects. Model 1 is the same model as in Table 1, Models 2-6 present alternative specifications with different groupings of controls based on topic.

Table 3: Association Between Total Children and Proportion Girls

	(1)
Proportion Girls	1.32***
	(0.22)
Num.Obs.	434
R2	0.076

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

1.4

Assumption 1) that child gender is randomly assigned at birth cannot be tested with these data. Assumption 2) that parents are not following a gender-biased fertility stopping rule cannot be evaluated because we do not have data on the gender of the first born child, though Washington does test this possibility in the paper. In Table 3, I estimate the association between proportion of girl children and the ultimate number of children that a congressperson has. The association is positive and significant, which suggests that the an increase in the proportion of girls corresponds to having more children. It may be the case that more female children is associated with having more children in general, which may indicate a gender preference, but it is not clear.

I test assumption 4 by relying on voter characteristics. In Table 4 I present a regression of the relationship between the proportion of female children and various voter population characteristics. Results suggest that there is not a significant association between voter characteristics and fraction of children that are girls. Though there is a significant negative correlation between fraction graduate high school and the number of female children.

1.5

I think that the research design in this paper is credible and clever. I would recommend that it be published, however I would suggest additional analysis subgroup analyses by age-group and service length which may signal more conservative views over time. It may be the case that these effects are consecrated wholly for younger congresspeople. I am also interested to know whether these results are sensitive to congressperson adoption patterns. While I think this is probably a rare phenomenon, if parents can select a child to adopt, this could bias the estimates.

1.6

As I note in 1.4, having one more female child is associated with 1.32 more children. This may indicate that parents may continue having children if they have more girls in hopes of diversifying the gender of their children which may bias the estimates. It also seems that a large reduction in the point estimate from Model 7 to Model 1 in Table 2 is likely due to Republican voting behavior as is evidenced by Table 3 in Washington 2008. I think that it might be the case that there is only an effect on NOW score for democrats but no effect for republicans.

Problem 2

a)

In Table 5, I present OLS regressions on simulated data with classic, robust, and bootstrapped standard errors.

In table 5, I compare regression estimates of Y on X with three different standard errors. The naive variance estimate is much smaller than the robust and bootstrapped SEs. The robust SE is slightly larger which is to be expected and the bootstrapped SE is slightly smaller than the robust SE. Based on these results, we can

Table 4: Association Between Proportion of Girl Children and Voter Characteristics

	(1)
Fraction Female	−1.72
	(1.29)
Fraction White	0.02
	(0.15)
Fraction Graduated HS	−0.84*
	(0.38)
Fraction Graduated College	0.48
	(0.44)
Fraction Support Abortion	0.15
	(0.25)
Fraction Prefer Spending on Services	−0.33
	(0.21)
Fraction Prefer Spending on Defense	0.17
	(0.36)
Fraction Prefer Spending on Crime	−0.40
	(0.26)
Fraction Pro-LGBTQ	−0.04
	(0.36)
Num.Obs.	402
R2	0.033

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5: Comparison Between Classic, Huber-White, and 1000 Replication Bootstrap Standard errors

	Classic SE	Robust SE	Bootstrap SE
(Intercept)	1.940	1.940	1.940
	(0.147)	(0.149)	(0.154)
x	1.217	1.217	1.217
	(0.159)	(0.272)	(0.265)
Num.Obs.	200	200	200
R2	0.228	0.228	0.228

Bootstrap SEs use 100 replications.

Table 6: Comparison of Classic, Robust, and Bootstrap Estimated Standard Errors from 1000 Monte Carlo Simulations

	Mean	SD	Min	Max
Beta	0.99	0.33	0.03	2.16
Bootstrapped SE	0.30	0.07	0.17	0.73
Classic SE	0.17	0.02	0.13	0.27
Robust SE	0.31	0.07	0.18	0.84

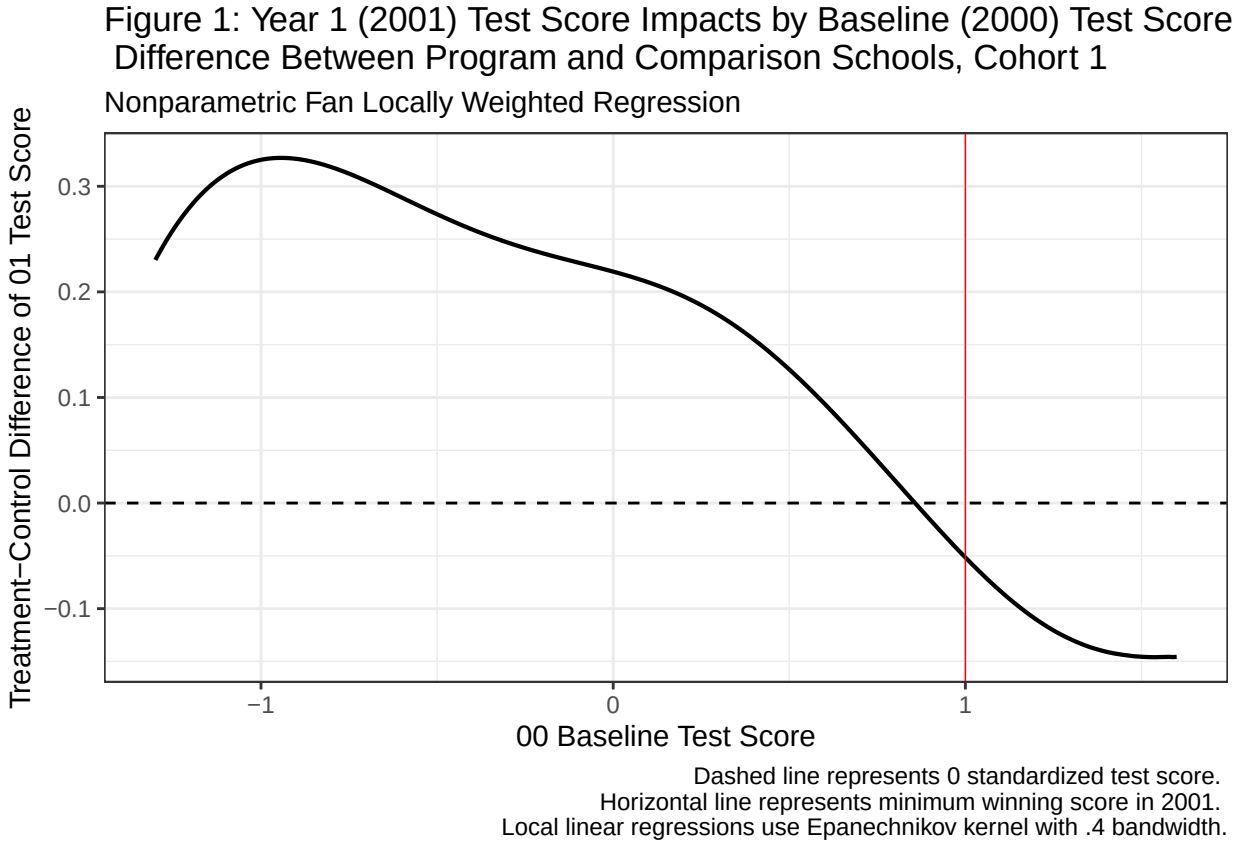
Each monte carlo sample was of size 200 and bootstrap SEs at each draw are based on 1000 resamples.

conclude that the bootstrapped SE recovers the robust Huber-White estimate of the variance rather than the *i.i.d.* estimate.

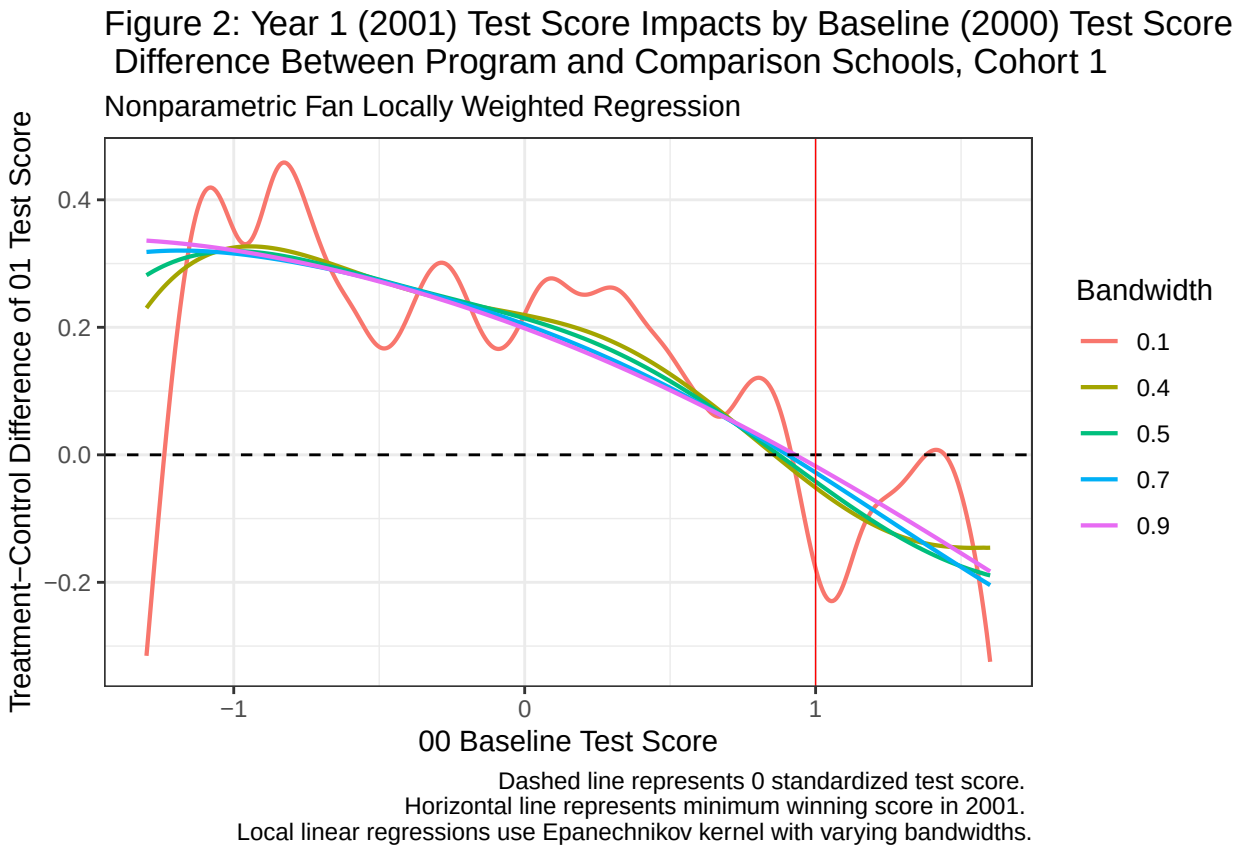
b)

In Table 6, I display monte carlo simulations of β and estimated standard errors using the naive, robust, and bootstrapped variance estimates. As we can see, after 1000 simulations, the estimate of $\hat{\beta}$ is unbiased, with a "true" standard error of 0.34. The naive variance estimate underestimates the true standard error by just under half the real standard error. The robust standard error, 0.32 corrects for heteroskedasticity and recovers the desirable inference properties were there no heteroskedasticity. The bootstrapped SE approximates the robust SE and is very close to that estimate at 0.31.

Problem 3



In Figure 1, I present a visualization of the treatment-control difference in test scores for Girls in the first experimental cohort. Broadly, the punchline from the plot suggests that the treatment effect was most effective for girls with lower 2000 baseline test scores. As test-scores approach 1, the treatment effect declines in intensity.



In Figure 2, I re-estimate results in plot one with differing bandwidths of the locally weighted regression. A bandwidth of 0.4 is preferred, however bandwidths of 0.5 and 0.9 provide similar overall conclusions. A bandwidth of 0.2 provides more specific information on the underlying variation in treatment effect but is a little less flexible. For a bandwidth of 0.2 it appears that scores around -1 and scores around 1.5 have noticeable increased 2001 test scores, which are not revealed by the smoothed 0.7 bandwidth. The results are slightly different, though much less general. Nevertheless, I still think that the 0.4 bandwidth characterizes the overall trend best.

3.b

In Figure 3, I present Figure 1 with bootstrapped 95% confidence intervals of the difference between the local regressions of the treatment and control groups.

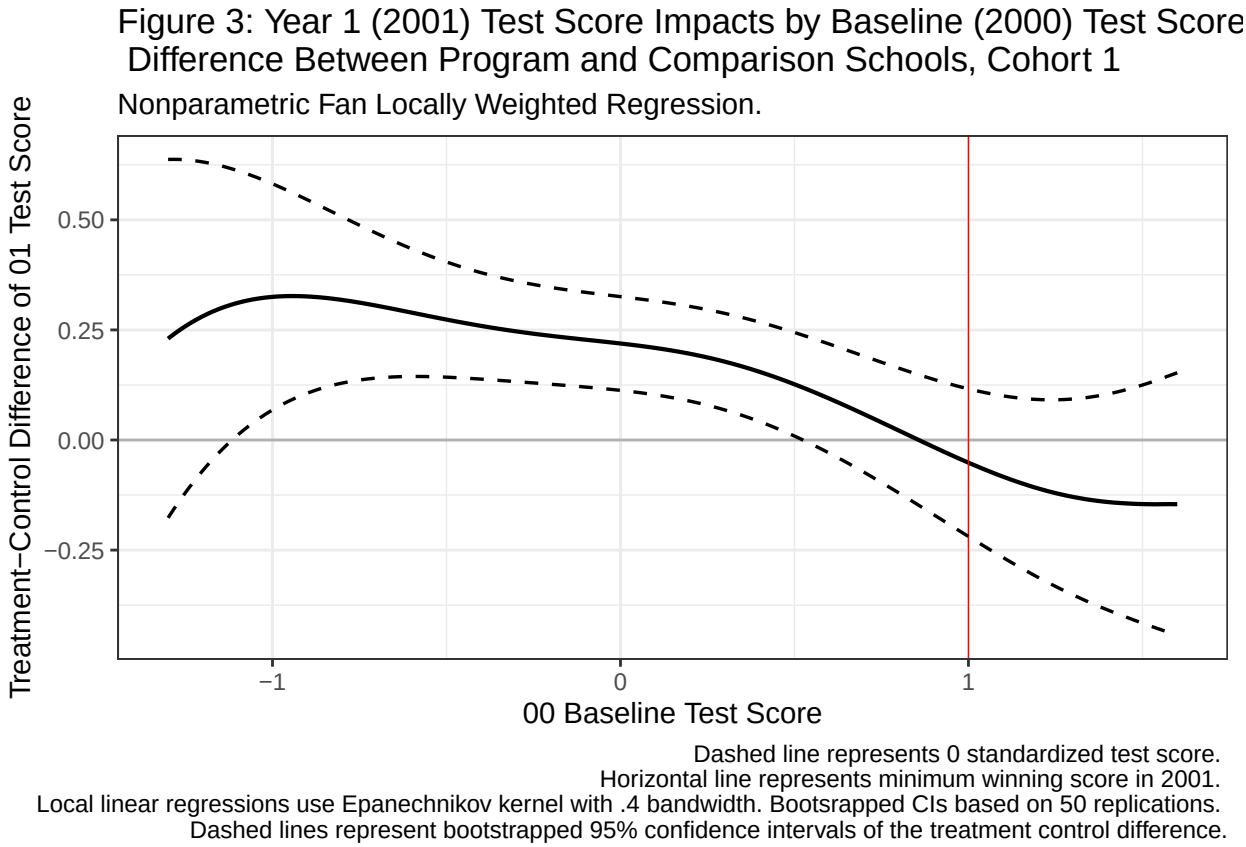


Figure 3 indicates that for students with a baseline test score of -1 to about 0.5 their test score increase in 2001 was significantly different than 0 because the dashed confidence bands do not cross 0. However, for the extreme ends of the distribution, th bands do cross 0 indicating that the effect is not statistically different.

3.c

Based on Figure 3, results indicate that the majority of the test score gains as a result of the treatment were for students in the middle of the distribution while high and low scorers' increases were not significantly different. Because of the insignificant difference for the high and low baseline scorers, these results indicate that there were not particularly negative externalities for low-scoring students. In other words, because high and low test-scorers both have similar treatment effects, we cannot reject the hypothesis that they are different. While the paper argues that the externalities for low test-scorers are positive based on the longitudinal data, using single cohort data leads to a different conclusion.

Problem 4

4.a

In Table 7, I present RMSE and MAPE goodness of fit statistics for in and out of sample data for a variety of models.

4.b

Based on these sets of models, I would choose the Lasso and Post-Lasso models because of their out of sample RMSE and MAPE. The LASSO model uses only selected coefficients and the post-LASSO model uses selected coefficients in addition to age-county-year variables. Both perform well out of sample and particularly the post-LASSO model performs well with the addition of theoretically relevant variables.

Table 7: Comparison of Different Prediction Models for Estimating Child Population Counts

Model	IS RMSE	OOS RMSE	IS MAPE	OOS MAPE
Naive Model (OLS)	506.00	824.86	189.63	212.29
Global Sample Average (Intercept)	3494.80	6205.28	1303.08	1543.44
Kitchen Sink Regression	151.41	224.83	45.44	50.97
LASSO Regression (w/ Min Lambda)	152.70	226.63	44.52	49.96
Post-Lasso (w/ Focal Predictors)	152.64	226.67	44.43	49.84
OLS with State FEs and Nonlinear Age	387.08	7203.03	141.21	1897.33
OLS with Nonlinear Age	505.69	7192.71	194.45	1841.55

Naive Model includes births, deaths, and migration as features. Global Sample Average Model includes only an intercept. LASSO regression model estimated with optimal training data lambda determined through cross validation. IS = In Sample, OOS = Out of Sample.

4.c

In and out of sample RMSE and MAPE do vary by model. The Naive model, which incorporates births,deaths, and migration, is parsimonious and performs reasonably well on MAPE out of sample but do not perform well in RMSE. Across models, OOS fit is usually worse than in-sample fit, with the exception of the kitchen sink and LASSO models. The models with the lowest OOS RMSE and MAPE were more parsimonious which can reduce overfitting or adding in-sample specific noise. Additionally, the LASSO models fit with cross-validation to select lambdas may perform better because of a more principled feature selection. In models that are more complex and add in all information possible, performance may be worse.

4.d

I think that the strength of using LASSO and other data-driven variable selection methods makes sense for modeling in settings where data are high dimensional. Additionally, when forecasting or prediction is the desired outcome, LASSO performs particularly well at maximizing OOS goodness of fit. If the goal of analysis is prediction, I think that OLS and other traditional methods provide more substantively and theoretically relevant estimates. Additionally, when conducting statistical inference, it is possible that machine learning methods may select data that predict well but are not substantively informative. Thus, the variables chosen may be nonsensical relative to the actual problem of study. That said, I believe that ML methods do add greatly to modeling and provide powerful tools to perform robustness checks and analyze model performance.