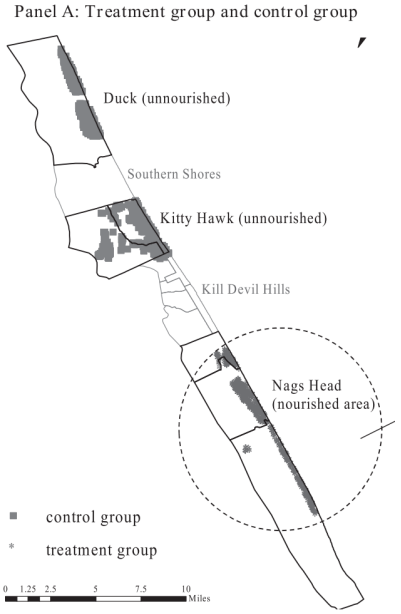
Replication Project

Blain Morin

10/30/2020

Background and Overview

This project replicates the difference-in-differences model estimated in Qiu and Gopalakrishnan's 2018 paper, "Shoreline defense against climate change and capitalized impact of beach nourishment" using a Bayesian framework (link to paper). The paper evaluates the impact of beach nourishment (which is "the process of replacing an eroding section of a beach with sand dredged from inlets or offshore sand reserves") on coastal housing prices. The analysis focuses on on island in the Outer Banks of North Carolina. Here is the map of the area analyzed for reference:



In the analysis, the authors compare the trend in sale prices of homes in Nags Head (the town that received beach nourishment) to the trend of sale prices in Duck and Kitty Hawk (which are towns that didn't receive beach nourishment until a few years after Nags Head). Here, receiving the beach nourishment is considered the treatment. The authors' diff-in-diff model takes the form:

$$ln(P_{ijt}) = \alpha_0 + \alpha_1 X_i + \beta_1 Nourish + \beta_2 PostNourish + \beta_3 Nourish * PostNourish + \eta_j + \zeta_t + \epsilon_{ijt}$$

where P_{ijt} is the price of home i in location j sold in year t. X_i is a vector of control variables for housing characteristics. $Nourish$ is an indicator variable for being in the treatment group. The main coefficient of interest is β_3, which is the average treatment effect of beach nourishment. The authors also include census block group and year fixed effects (η_j and ζ_t respectively).

OLS Estimation Results

In the published paper, the authors use year fixed effects as well as seasonal fixed effects. Including the year fixed effect causes collinearity problems when I try to replicate their analysis (in other words the year fixed effects are collinear with the *PostNorish* covariate, which is an indicator variable that = 1 when the year is greater than 2010). I also tried to filter the data based on the authors description, but am not able to exactly match their sample size. OLS regression results are presented in Table 1.

Table 1: OLS Regression Results

	Dependent variable:
	ln(Sale Price)
# of Bathrooms	0.054^{***} (0.009)
Living Area (100 Sqft)	0.040^{***} (0.003)
Living Area Squared	-0.0004^{***} (0.0001)
Age of Property (Years)	-0.005^{***} (0.0004)
# of Stories	0.091^{***} (0.013)
Distance to Shoreline (10m)	-0.002^{***} (0.0002)
Ocean Front = Yes	0.330^{***} (0.025)
Nourish	-0.135 (0.134)
PostNourish	-0.049^{***} (0.013)
Nourish*PostNourish	-0.004 (0.022)
Constant	12.074^{***} (0.057)
Location Fixed Effects?	Yes
Year Fixed Effects?	No
Observations	2,095
R^2	0.747
Adjusted R^2	0.744
Residual Std. Error	0.230 (df = 2074)
F Statistic	305.602^{***} (df = 20; 2074)
Note:	*p<0.1; **p<0.05; ***p<0.01

Bayesian Estimation Results

In this section, I replicate the analysis using rstan. Here is the specification in the .stan file:

```
data {
  int<lower=0> N;
  int<lower=1> K;
  matrix[N,K] x;
  vector[N] y;
}

parameters {
  vector[K] beta;
  real<lower=0> sigma;
}

model {
  y ~ normal(x * beta, sigma);
}
```

This model was run with 4 chains, each with 2000 iterations. A flat prior was used because of lack of prior information. The x matrix contains all of the housing covariates as well as the year and location fixed effects. Table 2 presents the summary results for the non-fixed effect variables.

Table 2: Bayesian Estimation Results

term	mean	sd	n_eff	Rhat
Constant	12.0744	0.0568	2429.586	1.0004
# of Bathrooms	0.0537	0.0089	7334.114	1.0003
Living Area (100 Sqft)	0.0396	0.0031	6928.108	0.9999
Living Area Squared	-0.0004	0.0001	6438.678	1.0000
Age of Property (Years)	-0.0046	0.0004	7807.745	1.0000
# of Stories	0.0908	0.0126	6784.899	0.9997
Distance to Shoreline (10m)	-0.0019	0.0002	9085.058	1.0000
Ocean Front = Yes	0.3300	0.0255	6598.649	0.9998
Nourish	-0.1348	0.1353	2255.250	1.0015
PostNourish	-0.0494	0.0134	6139.730	1.0003
Nourish*PostNourish	-0.0042	0.0223	5977.502	1.0004
Sigma	0.2298	0.0035	6754.194	1.0005
lp_	2033.2651	3.3415	3119.229	1.0005

Looking at the summary results in Table 2, the effective sample and chain mixing (seen in the n_eff and Rhat columns, respectively) are sufficient for all variables. The estimated coefficients are also similar to the OLS specification.

Bayesian Diagnostics

Figure 1: MCMC Tracplots

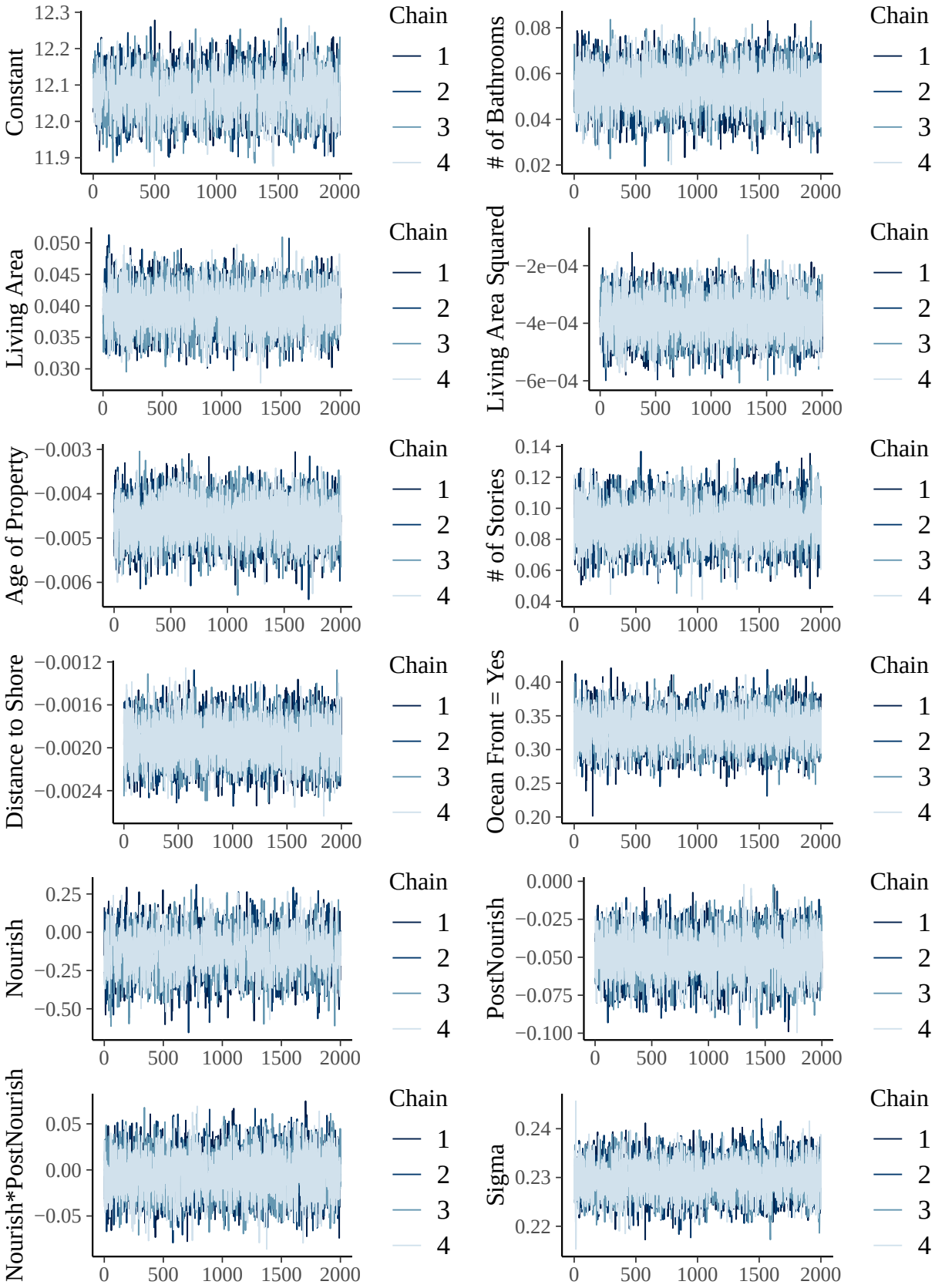


Figure 1 plots the MCMC chains for the variables of interest. As expected, the chains are mixing well and are bouncing around a narrow area.

Estimate Comparison