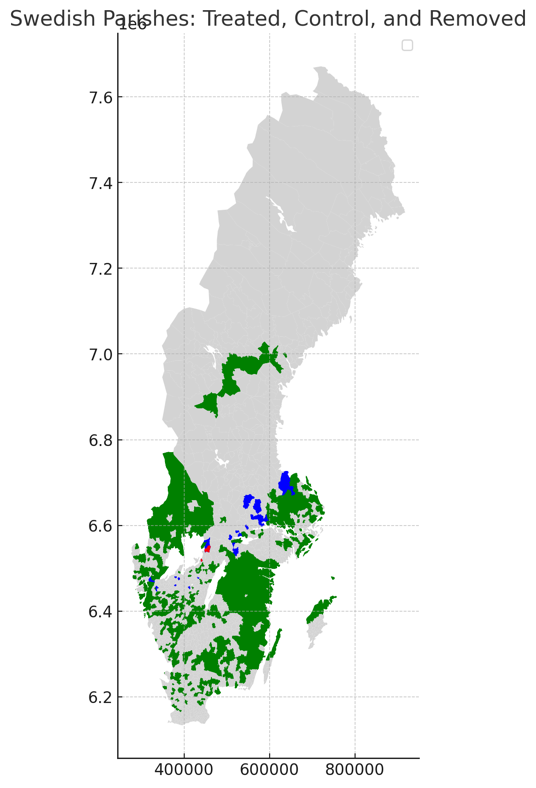
# Solving the problem of balancing

We start with theoretical treatment and control group based on electrification along the western line:



Then, in reality, we have only the parishes that we observe in the 1930 census (ignore colours):



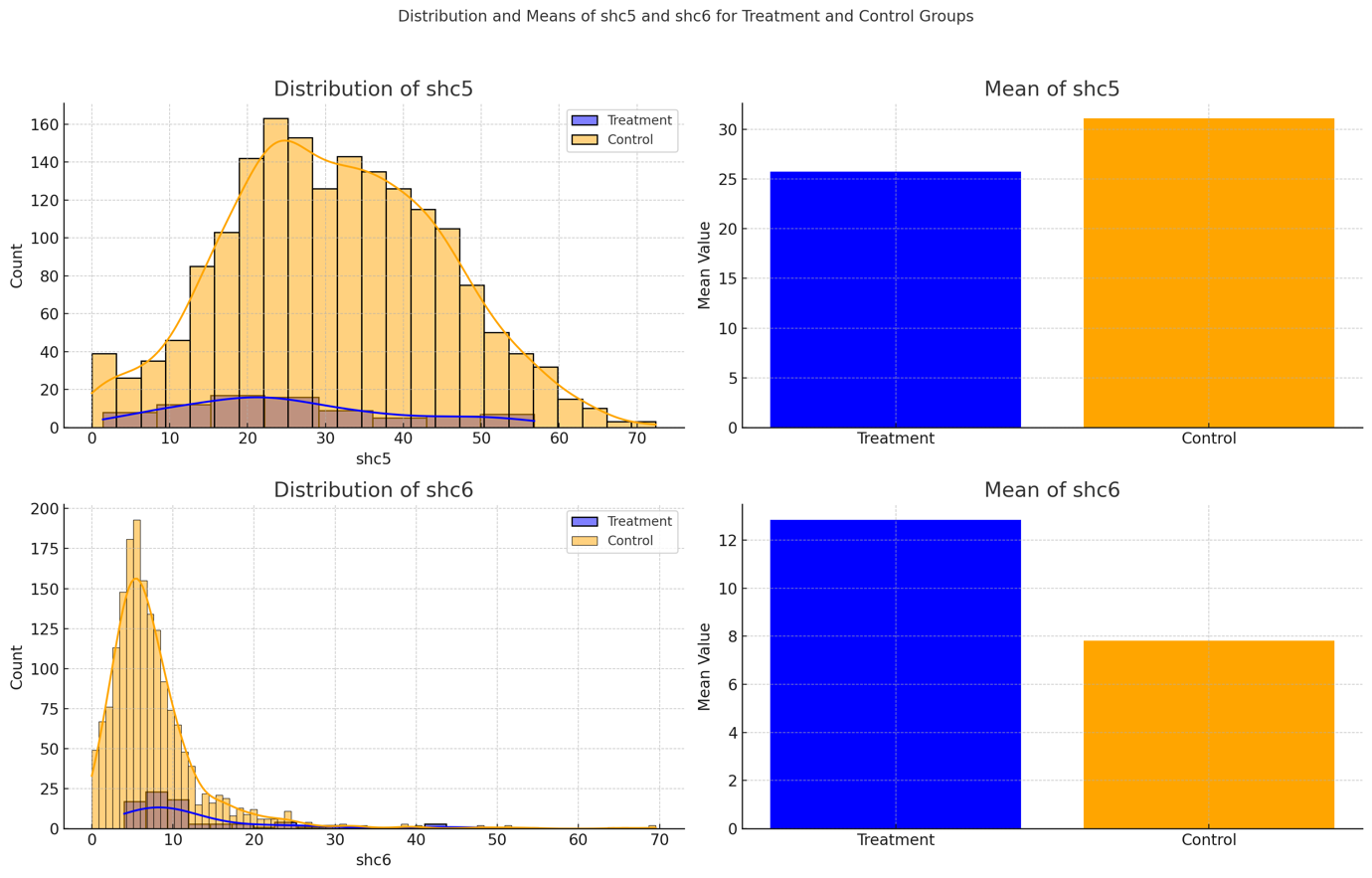
## Balancing in 1890 and 1900.

We run a t-test on the share of workers in different HISCLASS groups in 1890 and 1900, as well as the labour force in each year among the treatment and control groups that we do observe in 1930.

The problem here is that there are fewer farmers and fishermen and more low skilled workers in the treatment parishes than the control parishes:

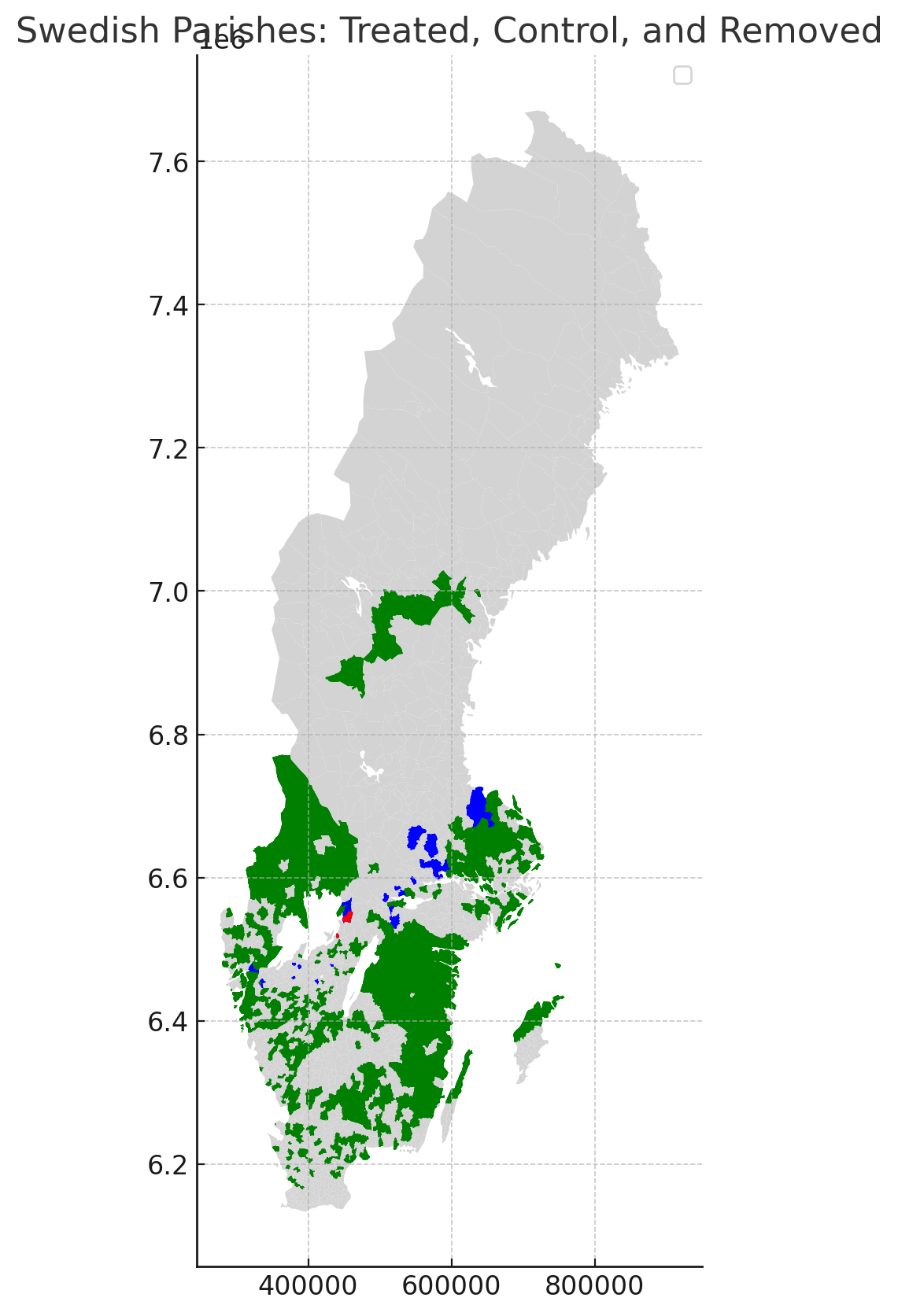
Shc5 is farmers and fishermen, shc6 is low skilled workers. The figure shows avg. across 1890 and 1900.

All of the other variables are not statistically different between the treatment and control groups at 300km, (also 250 and 350, it’s not very sensitive).



So we can drop control group parishes with high values of shc5 and low values of shc6 in order to get comparable results.

That is what we do here, dropping 40 parishes so that we have a control group that is comparable to the treatment group in 1890 and 1900.



Here, the control parishes are in green, blue are treated parishes are treated and red are control that we remove.

# What is the impact on the number of parishes?

Treated in theory: 198

Treated in 1930 census: 123

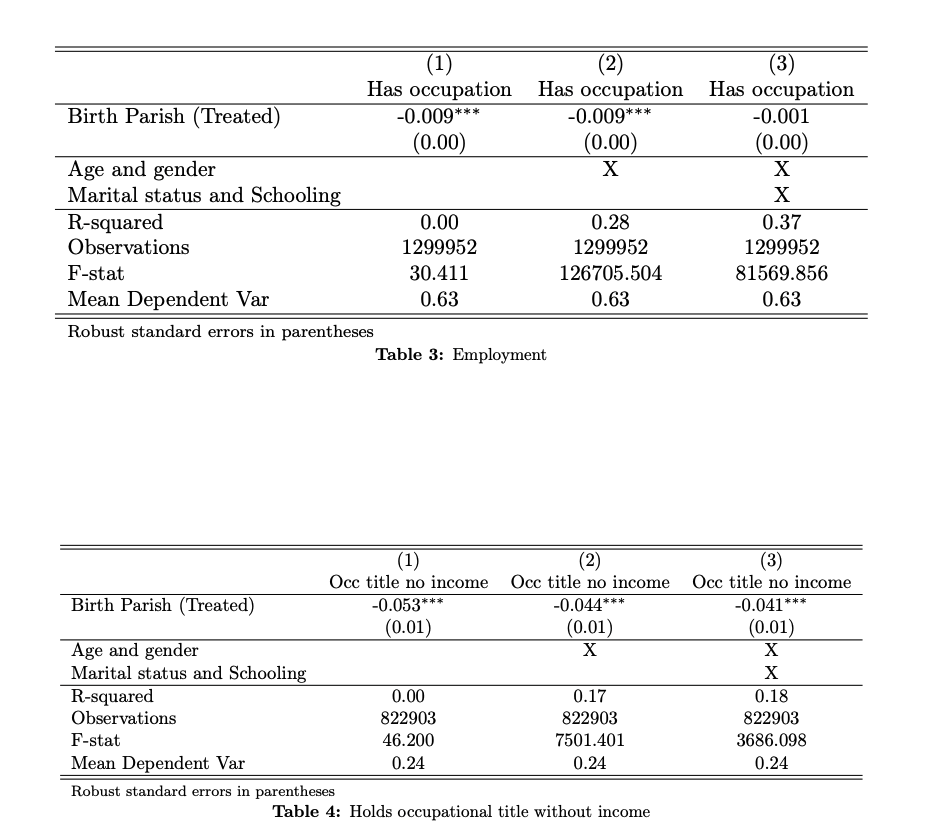
Treated in final sample: 123

Control in theory (at 300km): +- 1400

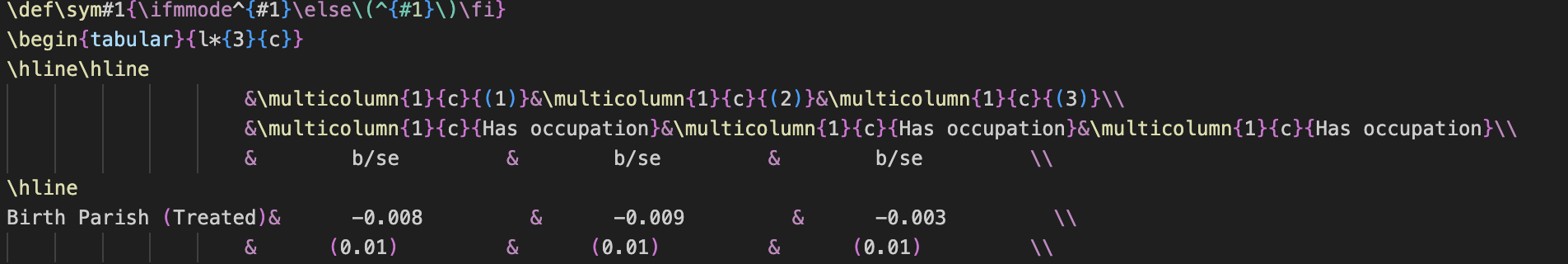
Control in 1930 census (at 300km): 937

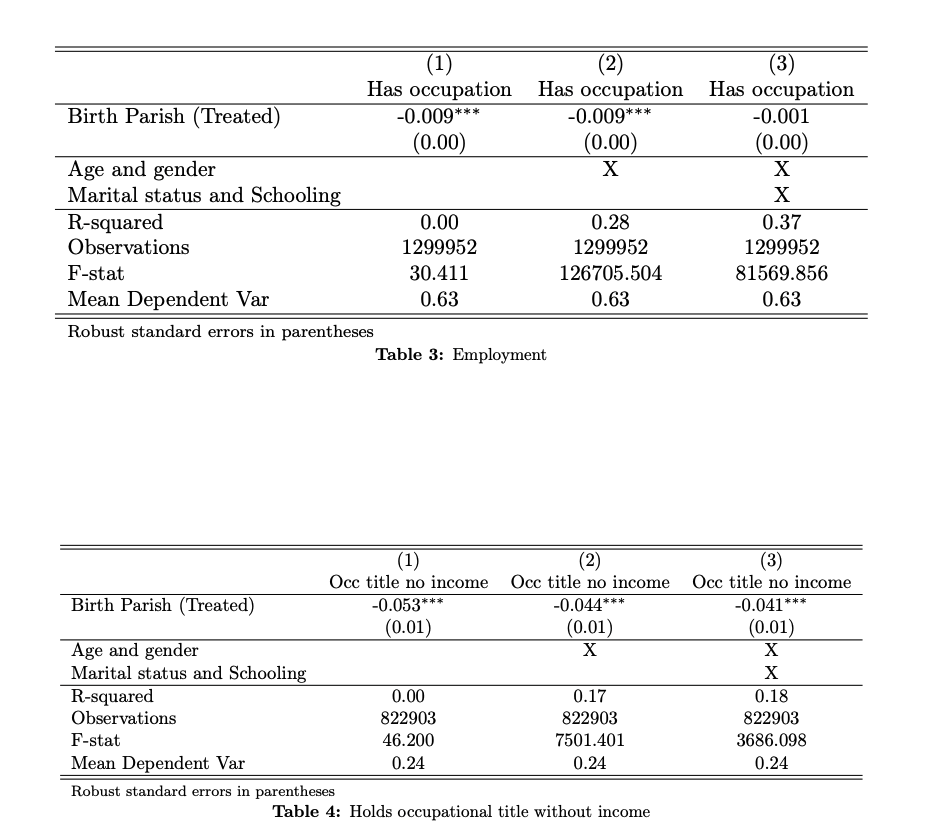
Control in final sample: (at 300km): 897

## Does it change the regression results? original regression results

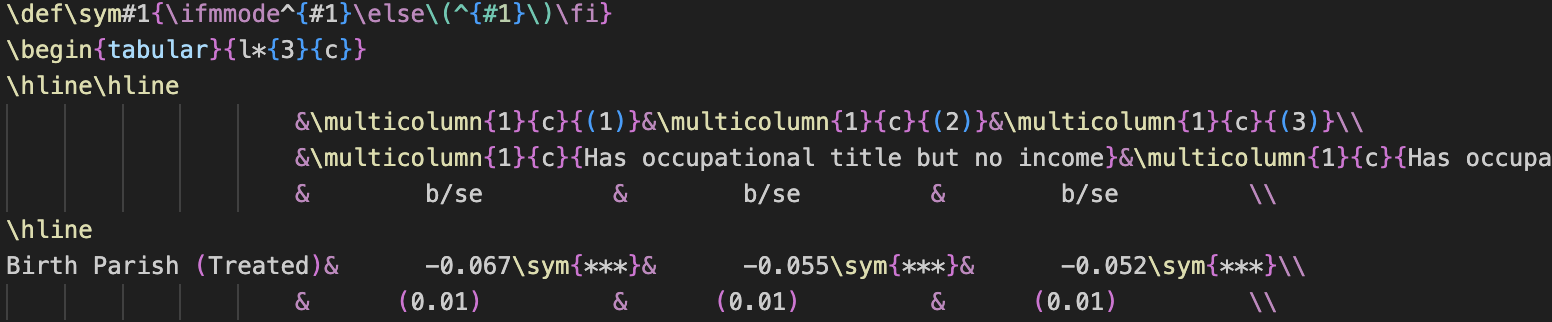


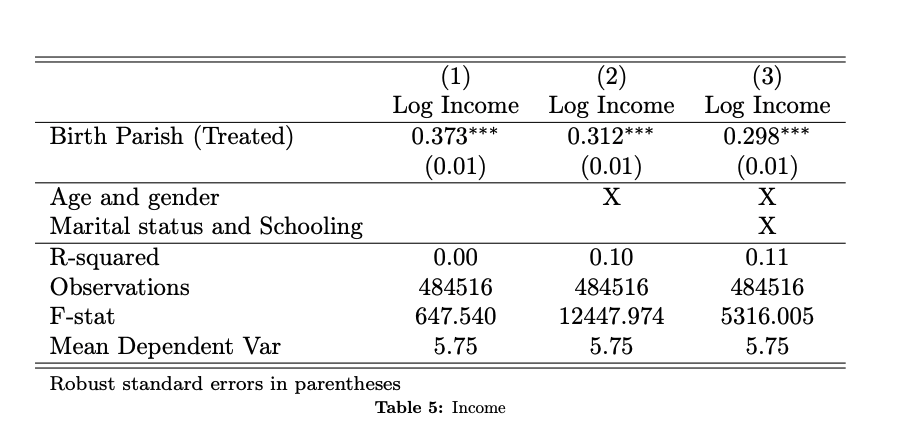
### Compare with table 3



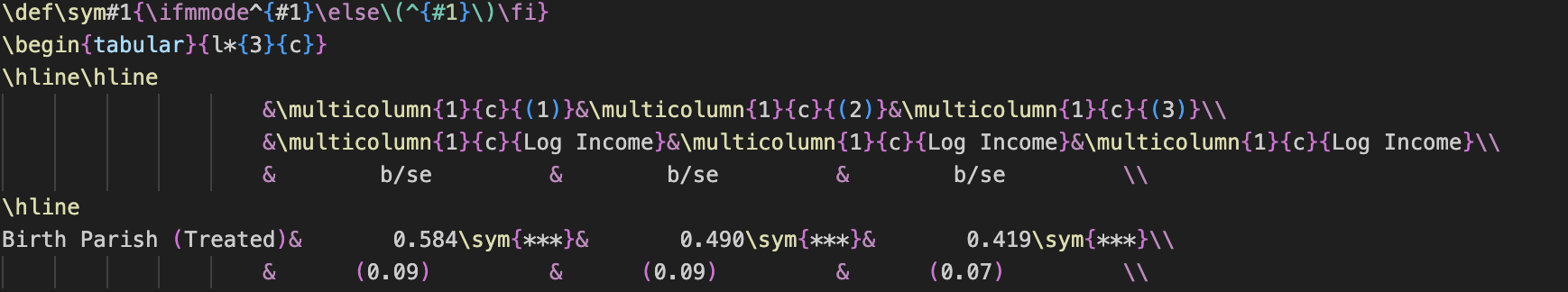


#### Compare with table 4





### Compare with table 5



Observations: 581,851 (more because now we have 300km band instead of 250km).