

## Exercises 8: Spatial models

### A spatial probit model

In “katrina.csv” you have data on 673 businesses in downtown New Orleans. Your goal is to model how quickly these businesses re-opened in the wake of Hurricane Katrina, as a function both of known regressors and unknown spatial variation. The data is from LeSage et. al. (2011).<sup>1</sup> The variable codes were taken straight from the journal website<sup>2</sup> and are as follows.

*long* longitude coordinate of store  
*lat* latitude coordinate of store  
*flooddepth* flood depth (measured in feet)  
*logmedinc* log median income  
*smallsize* binary variable for “small size” firms  
*largesize* binary variable for “large size” firms  
*lowstatuscustomers* binary; low socio-economic status of clientele  
*highstatuscustomers* binary; high socio-economic status of clientele  
*owntypesoleproprietor* binary; indicates “sole proprietor” ownership type  
*owntypenationalchain* binary; indicates “national chain” ownership type  
*y1* reopening status in the very short period 0-3 months; 1=reopened, 0=not  
*y2* reopening status in the period 0-6 months; 1=reopened, 0=not  
*y3* reopening status in the period 0-12 months; 1=reopened, 0=not  
*street* which of three streets the shop is on

If you install the R library “ggmap” and its dependencies, then you’ll be able to visualize the data on a Google map:

```
qmpplot(long, lat, data = katrina, matype="roadmap", source="google")
```

Fit a probit model for re-opening status, choosing one of 3, 6, or 12 months as the threshold defining your binary outcome (or model them all separately). Incorporate relevant regressors in your model, as well as spatial correlation. Think carefully about the spatial structure. You’ll notice that the shops fall on three separate streets; thus you might treat spatial covariation within a single street as independent of spatial covariation across streets. Within a street, you might think about Gaussian processes to model a spatial “intensity of re-opening” function. You could also consider autoregressive models, with “nearest neighbor” effects. If you’d like a reference on Bayesian spatial modeling, consider the classic paper by Diggle et. al. on model-based geostatistics.<sup>3</sup>

<sup>1</sup> J. P. LeSage, R. K. Pace, N. Lam, R. Campanella and X. Liu (2011), New Orleans business recovery in the aftermath of Hurricane Katrina, Journal of the Royal Statistical Society A, 174, 1007–27

<sup>2</sup> <http://www.blackwellpublishing.com/rss/Volumes/Av174p4.htm>

<sup>3</sup> <http://www.isds.duke.edu/~fei/samsi/Readings/DiggTawnMoye1988.pdf>