Regression analyses thus far

In this section we compare six models fit to raw IBD estimates (using a ordinary linear model, OLM) and dichotomized IBD estimates (using a generalised linear model, GLM). This six models are,

- Model 1: ~ geo_dist + time_dist
- Model 2: ~ geo_dist * time_dist
- Model 3: ~ geo_dist + time_dist + Tumaco + Guapi + Buenaventura + Quibdo + Tado
- Model 4: ~ geo_dist * time_dist + Tumaco + Guapi + Buenaventura + Quibdo + Tado
- Model 5: ~ geo_dist + time_dist + Within
- Model 6: ~ geo_dist + time_dist * Within

GLMs

The glm models without interaction terms (1,3 and 5) are preferable to those with (2,4,6). Of those without interaction terms, model 3 has the most favourable AIC. In terms of binned residules, model 1 is poor; models 3 and 5 are comparable, but many unexplained anomalies remain (Figure 1). If we fit to across site comparisons only using model 1 (models 3 and 5 irrelvant when intra site comparisons are removed), regression coefficients remain significant and negative (AIC is not comparable since data are different); there are more binned residual outliers, however. We thus henceforth discount models fit to partial data.

OLM

All the lm models (1 to 6) seem to be viable candidates, with more as-expected coefficient signs and significant coefficients than those generated under the GLMs. According to the residual squared error, M3 seems to be the best model. It is far from fitting the linear model assumptions, however (Figure 2).

Conclusion

To conclude, model 3 appears to be the best of all linear models and GLMs (regression coefficients, Table 1). Neither fit the data perfectly however (the violations of the linear model seem worse, but this is arguably because of a lack of plots for the GLM). I suspect the addition of travel distance, will improve model fit.

Table 1: Regression coefficient estimates with p-values in parentheses.

	Generalised linear model	Ordinary linear model
(Intercept)	-2.86e+00 (0.00e+00)	8.58e-02 (4.21e-275)
geo_dist	-9.21e-04 (2.52e-05)	-8.42e-05 (2.64e-34)
time_dist	-3.29e-03 (8.85e-48)	-8.97e-05 (2.17e-35)
TumacoTRUE	1.37e+00 (1.37e-77)	9.15e-02 (1.03e-196)
GuapiTRUE	2.00e+00 (1.28e-130)	2.16e-01 (0.00e+00)
BuenaventuraTRUE	9.35e-01 (4.17e-13)	2.78e-02 (9.59e-06)
QuibdoTRUE	6.28e-01 (1.09e-08)	9.64e-02 (3.44e-88)
TadoTRUE	6.99e-01 (5.05e-02)	1.51e-01 (1.09e-13)

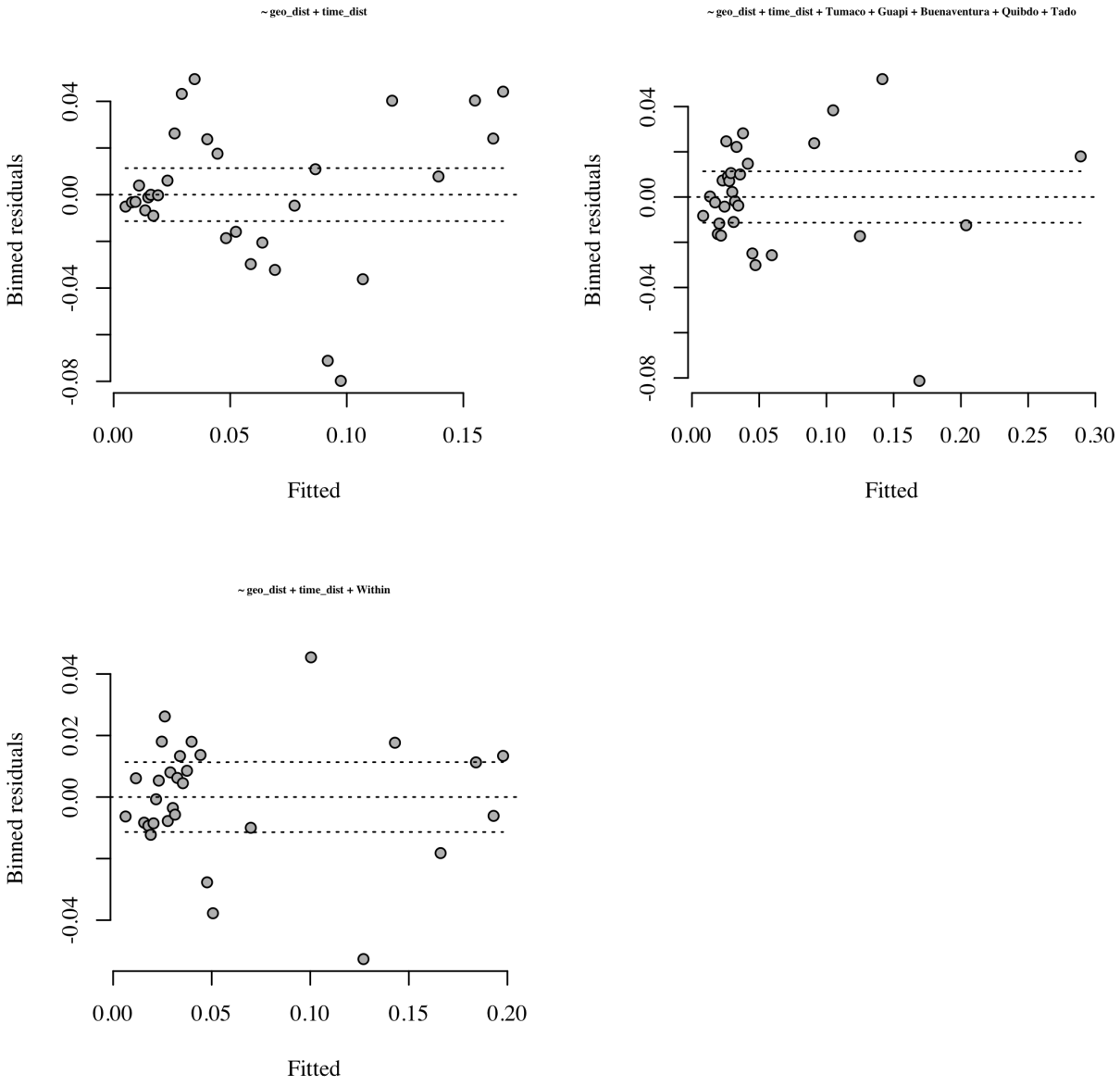

Figure 1: Plots of binned residuals for candidate models fit to dichotomized IBD estimates.

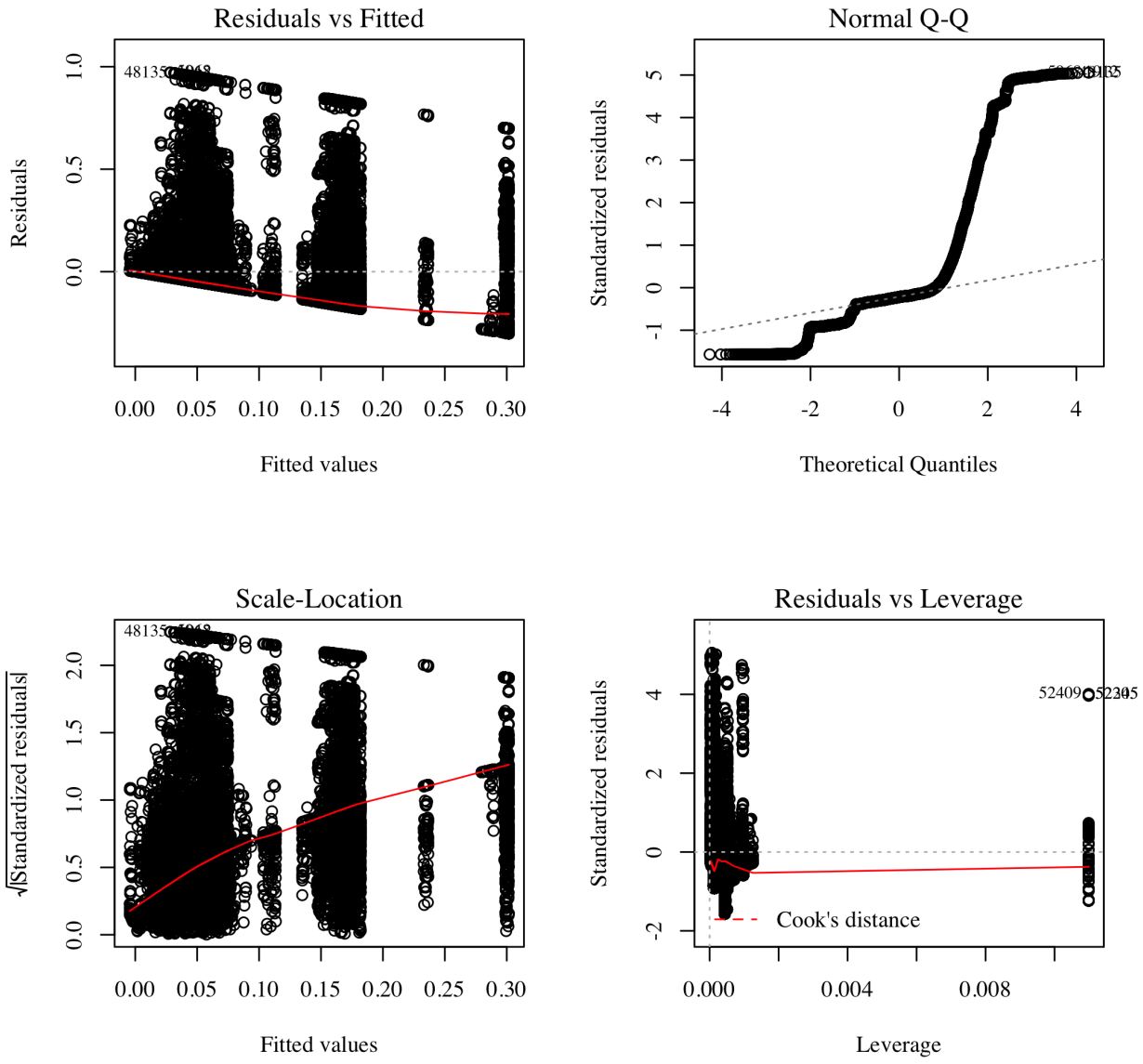

Figure 2: Plots of candidate linear least square models fit.