Supplemental material:

1. Lasso-regularized Quasi-Poisson Fit

Given a set of observations $\{y_i\}_{i=1}^{I}$ and the associated covariates $X_{i,1}, X_{i,2}, \dots, X_{i,N}$., for *i=1...I*, the maximum likelihood-based estimator for a Poisson distribution tries to find the coefficients $\{w_n\}_{n=0}^{N}$ that maximize the following log-likelihood:

$$logL\left(\{w_n\}_{n=1}^{N}\right) = \sum_{i=1}^{I}\left(y_i(w_0 + w_1 X_{i,1} + \cdots + w_N X_{i,N}) - exp(w_0 + w_1 X_{i,1} + \cdots + w_N X_{i,N})\right) - log(y_i!)$$

where ! denotes the factorial function and w_n is the coefficient associated with the nth covariate. Denoting the coefficients that maximize the log-likelihood as $\{\hat{w}_n\}_{n=1}^{N}$, the dependent variable is simply estimated as $\hat{y}_i = (\hat{w}_0 + \hat{w}_1 X_{i,1} + \cdots + \hat{w}_N X_{i,N})$. Notice that when maximizing the log-likelihood, the summands $log(y_i!)$ are constant and do not affect the outcome of the optimization, hence they are typically removed.

In the context of Poisson regression, it is often the case that the interest is in analyzing "rate data", that is scenarios where the dependent variable is a count of events divided by some measure of that unit's exposure. This requires dividing the original y_i by the exposure level, denoted by s_i, giving rise to the following updated log-likelihood

$$logL'\left(\{w_n\}_{n=1}^{N}\right) = \sum_{i=1}^{I}\left(y_i(w_0 + w_1 X_{i,1} + \cdots + w_N X_{i,N}) - s_i exp(w_0 + w_1 X_{i,1} + \cdots + w_N X_{i,N})\right)$$

where, as we did before, constant terms that do not affect the outcome of the optimization have been removed.

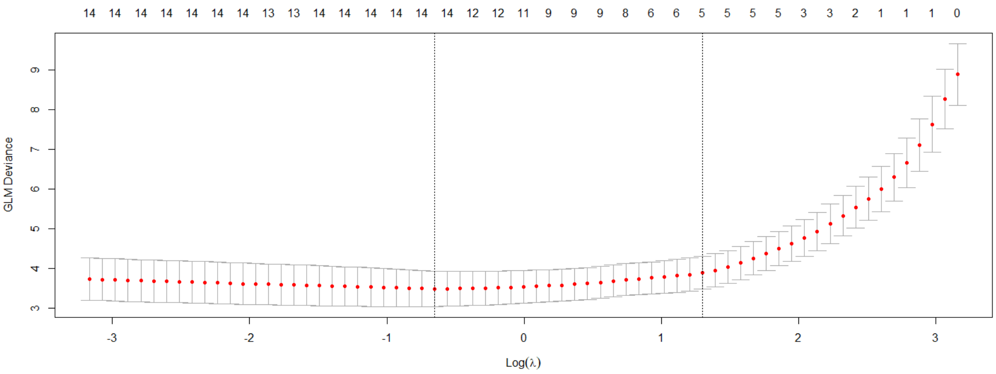
A well-known problem when fitting the log-likelihood above is that all the *N* coefficients will be, in general, non-zero which can be an issue from a point of view of interpretation and statistical accuracy. Mathematically, Lasso addresses this problem "a priori" regularizing the original log-likelihood with a term that penalizes the norm-one of the model coefficients

$$\{\hat{w}_n\}_{n=1}^{N} = argmax_{\{w_n\}_{n=1}^{N}} \ logL'\left(\{w_n\}_{n=1}^{N}\right) - \lambda \sum_{n=1}^{N} |w_n|$$

In the formula above, λ is a (non-negative) tuning parameter that controls the strength of the norm-one penalty $|w_1| + |w_2| + \cdots + |w_N|$, that is, the amount of shrinkage. The higher the λ, the larger the penalty associated with the coefficients $\{\hat{w}_n\}_{n=1}^{N}$. This implies that having non-zero coefficients incurs a larger cost and, hence, the solution generated by Lasso naturally sets many of those coefficients to zero (meaning that the associated covariates are not selected to explain the dependent variable). Indeed, if λ is very large, none of the variables is selected and, if $\lambda=0$, all of them are selected.

We run the analysis for different values of λ, each of them giving rise to a different number of active variables. We then select the optimum value of λ (and, correspondingly, the optimal number of active variables) using a 10-fold cross validation approach. Under this approach, we implement the following six steps: 1) splitting the I data points into 10 groups of size I/10; ii) fixing the value of λ; iii) fitting the Lasso model leaving one group out and using as input the data in 9 of the groups, iv) assessing the fitness of the obtained coefficients by predicting the outputs for the data points in the only group that was not used in step iii; v) repeating steps iii and iv this for the other 9 groups and average the error obtained in the 10 trials; and vi) repeating steps ii, iii, iv and v for all the values of λ. The results obtained are shown in the following figure. The final value of λ is chosen as one between the value of that achieves the smallest error/deviation (green circle and green line) and the one achieving 1 standard error from the smallest deviation (blue circle and blue line).

Cross-Validated Deviance of Lasso-regularized Quasi-Poisson Fit

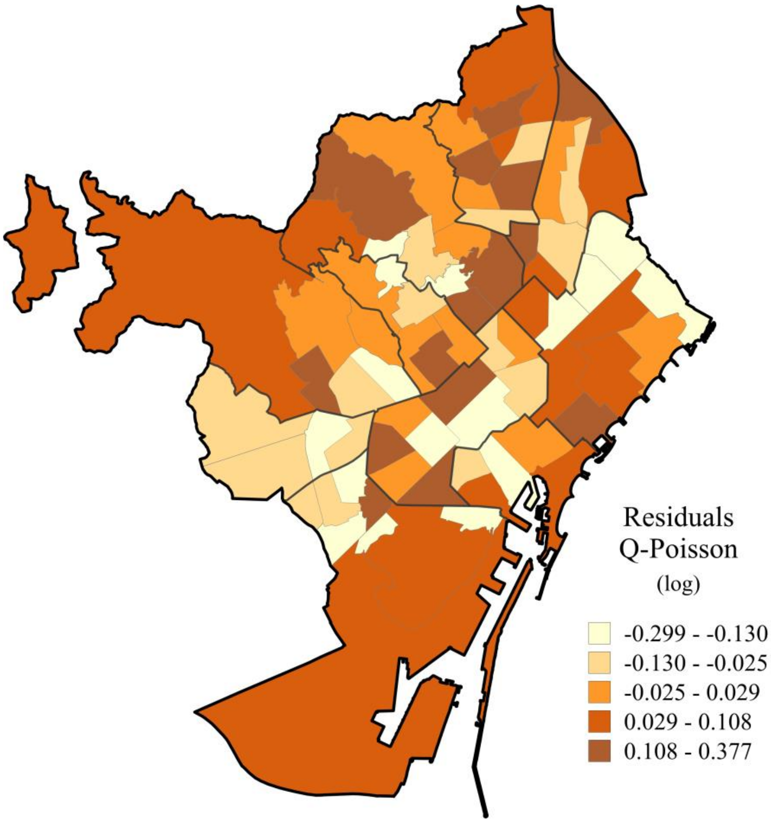


Ranking of the most explanatory variables running a Lasso-regularized Quasi-Poisson analysis

	Covariate
1	HDI high migrants
2	Mobility during lockdown
3	Percentage 70+
4	Post-secondary education3
5	Population density (urban)
6	Nursing homes
7	Health workers
8	Old-age sex ratio
9	Persons per dwelling
10	Life expectancy
11	Percentage 0-15
12	HDI low migrants
13	Vehicles per inhabitant
14	One person households
15	Income
16	Population de-registration

2. Spatial autocorrelation of residuals

Spatial distribution of residuals – Quasi-Poisson model (log)



Spatial autocorrelation of residuals. Moran's I (1st order Queen contiguity)