Protocol: The effect of changing to white road lighting on crime in the UK city of Leeds.

Drafted by Paul Marchant 21 June 2023

Leeds City Council (LCC) relit its roads, changing from orange (low pressure sodium) lamps to white light, under the UK Government's Private Finance Initiative mechanism. The aim of this research is to determine what effect this change had on the number of crimes in the City reported to the police. This work follows the study on Road Traffic Collisions (RTCs) in Leeds (Marchant and Norman 2022). It also follows on from the study of newly installed lighting and its effect on crime in the boroughs of London (Marchant 2011).

The crime data to be used is crimes reported to and recorded by the West Yorkshire Police (WYP). This data has been supplied by the WYP, already aggregated to the Mid Super Output Area (MSOA) level. It goes from the start of 2005 to the end of 2015. There are 840897 crimes recorded in file, allocated to 24 offence types, e.g., Burglary dwelling & Burglary non dwelling. The lighting data (inventory) for the same time period has been supplied by LCC. This was also used in the Leeds RTC study mentioned above.

It is the intention to put the crime data, after any necessary cleaning, into the appropriate format, that is weekly counts of crime by MSOA, (the crime weekly rates). Cleaning will involve, e.g., removing cases which are not known to be from Leeds, or which have erroneous dates. Details of the cleaning and the processing will be reported.

The crime file given has the reported earliest date and time that the crime could have occurred and also the latest date and time that it could have occurred. From the two dates it is possible to calculate the times of sunrise and sunset which then allows the periods when road lamps are on to be known. If the earliest date and time that the crime could have occurred and the latest possible date and time are not too temporally widely spaced, it is possible to calculate whether the crime occurred in daylight or darkness. If the temporal spacing is too great however, the darkness_daylight status is of course unknown. The computer code for generating whether the crime occurred in darkness, or daylight, or whether it is not possible to tell, will be made available.

The best estimate of the date and time of the crime will be taken as the mid-point between the earliest and latest times. The crime will be allocated to the appropriate week on the basis of this midpoint.

The Outcome Variable will use the number of crimes of all types that occurred in that week in that MSOA. The file of individual crimes will be aggregated into weeks within MSOAs.

The new aggregated file will be joined to the existing lighting data of weekly increase in new lamps, by MSOA, to analyse. This runs to Sept 2013 when Part Night Lighting (PNL) started in some areas.

It is the aim of the study to use the subset (probably the majority) of crimes where the darkness_daylight status is known, in order to see how the ratio of the number of crimes in darkness to the number in daylight is related to the implementation of the lighting scheme. (as was done in the study on RTCs). Examining the ratio of darkness to daylight crimes will help to act against confounding, that is to act against other changes in an area which might influence the level of crime both day and night (such as, more policing both day and night).

The effect of the new lighting on the ratio can also be estimated by examining the effect on the count of darkness crimes and count of daylight crimes separately, then combining these to obtain the estimate of the ratio. This was done in the paper for RTCs.

However, the effect of the lighting change on the sum (count) of crime for which the darkness_daylight status is known can also be determined. This can be compared to the effect on the sum of crime for every lighting status, daylight, darkness and unknown. The count models for the mean crime weekly rate are therefore: 1) darkness, 2) daylight, 3) the sum of darkness and daylight and 4) the sum of darkness, daylight and unknown. However, because of multiplicity of analyses, the focus will be on where darkness or daylight is known.

The modelling will follow that of the Leeds RTC study; multilevel ('mixed') with MSOA at Level 2 and week at level 1.

The link function for directly examining the ratio of the number of crimes in darkness to that in daylight is:

$\text{logit}(\pi_{ij}) = \log(\pi_{ij}/(1-\pi_{ij}))$

where i indicates the level 1 index and j indicates the level 2 index and π_{ij} = the proportion of crimes occurring in darkness = $\mu\text{dark}_{ij} / (\mu\text{dark}_{ij} + \mu\text{daylight}_{ij})$. Where $\mu\_\_\_{ij}$ is the mean number of crimes occurring, in darkness or daylight, in week i and MSOA j .

so $1 - \pi_{ij}$ = the proportion of crimes occurring in daylight = $\mu\text{daylight}_{ij} / (\mu\text{dark}_{ij} + \mu\text{daylight}_{ij})$.

This leads to:

$\text{logit}(\pi_{ij}) = \log(\mu\text{dark}_{ij} / \mu\text{daylight}_{ij})$ that is the logarithm of the ratio, the mean number of crimes in darkness divided by the mean number of crimes in daylight, in week i and MSOA j.

This ratio we can call the daylight adjusted darkness crime weekly rate.

For the count models for the mean crime weekly rate the link function is the logarithm $= \log(\mu_{ij}) = \log(\text{mean crime weekly rate})$.

Crime occurrence is anything but statistically independent, for one thing crimes are clustered within criminals. Therefore, we must expect to encounter overdispersion both in estimating the ratio and the counts.

The linear combination predicting the level of the appropriate crime response variable will be as given in the RTC study as:

$= \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_{Mk}\text{Month}_k + \beta_{HI}\text{PubHol}_I + \beta_W(L_{ij} - \langle L_{ij}\rangle_j) + \beta_B(\langle L_{ij}\rangle_j - \langle\langle L_{ij}\rangle\rangle)$

L_{ij} = the number of new lamps operating at time i, in area j

where $\langle L_{ij}\rangle_j$ denotes the mean over i, in area j, <<>> the mean of the area means, t = time in appropriate units that the midweek is from the start of the series.

The β_0 term is the intercept coefficient, which is expected to be 'random', as different areas will have different levels of crime, $\beta_1 t + \beta_2 t^2 + \ldots$ represents the time trend .The β_w term represents the effect of deviation of the number of lamps from its mean <> in the area, giving the within-area effect of lighting change. This coefficient will enable the effect of changing lighting to be seen. The β_B term represents the between area term, the effect of the deviation of the mean number of lamps in each area from the mean of area means <<>> of the number of lamps over all areas. The advantage of this centring of the between area term is that the intercept will give the value of the mean crime level for a typical area and time in the series, rather than one with a zero mean number of lamps.

Instead of using the number of lamps installed, an alternative is to use the fraction of the final number installed in the given area in order to see what difference this makes.

Exponentiating to give the mean ratio or the mean count from the logarithmic form, will give a factor $\exp(\beta_w(L_{ij} - \langle L_{ij}\rangle_j))$, when comparing the effect, on the ratio or mean count, at the end, i =T, to that at the beginning, i =1, of the period of relighting. That is, this is the factor by which the mean ratio or mean count has been multiplied, due to the change in the amount of lighting and will be approximately $1+ \beta_w(L_T - L_1)$ if $|\beta_w(L_T - L_1)| << 1$.

The overdispersion in the underlying variation will mean that Extra Binomial variation will need to be examined in estimating the daylight adjusted crime weekly rate and Overdispersed Poisson or Negative Binomial variation in any measure of mean count. Also using Gaussian models may be suitable if count rates are sufficiently high.

Offsetting by the period of daylight or darkness will reduce the effect of the month indicators to just picking up extra effects of seasonality through, such as, different traffic flows in different months. Having the appropriate time exposure to darkness or daylight accurate down to about a minute will give more precision to seasonality, especially of importance when the length of daylight and daylight are rapidly changing, as is the case around an equinox.

Whether using offsets or not, the key effect, of the lighting change, will still be the factor by which the number of events, or daylight adjusted darkness number of events, is multiplied, from start to finish.

Maximum Likelihood Estimation is thought to be adequate for the modelling in this study as the amount of data is large.

References

Marchant PR and Norman PD (2022), To determine if changing to white light street lamps improves road safety: A multilevel longitudinal analysis of road traffic collisions during the relighting of Leeds, a UK city, *Applied Spatial Analysis and Policy* 15(5):1-26, DOI: 10.1007/s12061-022-09468-w

Marchant PR (2011), Have new street lighting schemes reduced crime in London?" *Radical Statistic*s, 104, pp39-48 - http://www.radstats.org.uk/no104/Marchant2_104.pdf .