Submission
No 734

SYDNEY'S NIGHT TIME ECONOMY

Organisation: Centre for Translational Data Science, University of Sydney

Date Received: 2 July 2019

THE UNIVERSITY OF SYDNEY

Sally Cripps Prof
Director
Centre for Translational Data Science

2 July 2019

The Centre for Translational Data Science at The University of Sydney has undertaken a data driven study of the impact of the Sydney lockout laws on non-domestic assaults in Kings Cross and surrounding geographical areas from January 2005 – December 2017. The attached technical report has been written by Nicholas James, Dr Roman Marchant and Professor Sally Cripps.

Our findings deviate from prior analysis done on the topic, in particular work done by the Bureau of Crime Statistics and Research (BOCSAR) who suggest that the lockout laws had a material impact in reducing crimes in the Sydney Central Business District (CBD). Our research indicates that the lockout laws had no impact on reducing non-domestic assaults in the CBD. The contrasting insights between BOCSAR's analysis and our own are due to:

1. Different techniques in data processing
2. The use of different statistical methodologies.

We look forward to hearing from the parliamentary committee. In particular, we hope to collaborate and provide further data driven analysis on the topic of the Sydney lockout laws as the review process unfolds.

Sincerely yours,

Sally Cripps

Sally Cripps

Centre for Translational Data Science
Room 545, Level 5
Computer Science Building J12
The University of Sydney
NSW 2006 Australia

[redacted]
sydney.edu.au

ABN 15 211 513 464
CRICOS 00026A

Technical Report
A Case Study on the Sydney Lockout Laws

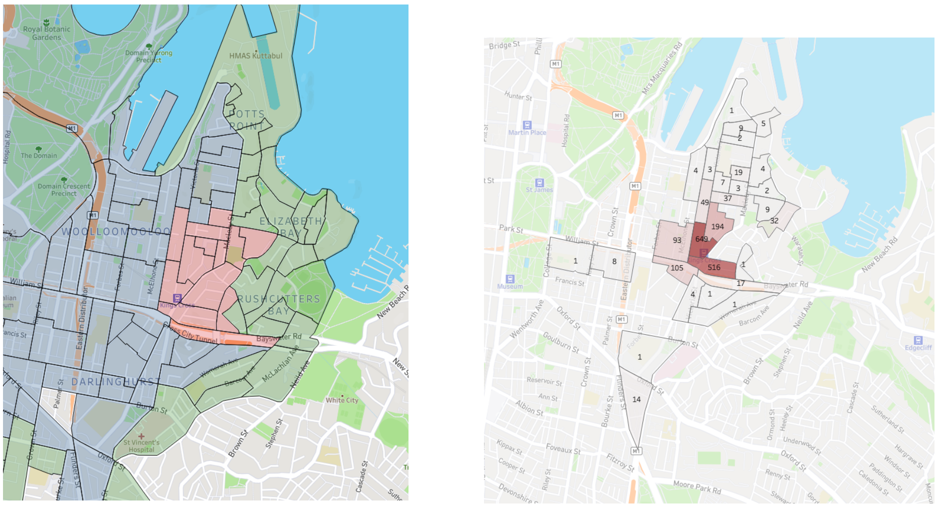
Nicholas James, Roman Marchant and Sally Cripps

Centre for Translational Data Science
The University of Sydney
Sydney, Australia
{nicholas.james, roman.marchant, sally.cripps}@sydney.edu.au

Abstract

Objectives The objective of this report is to provide an evidence based evaluation of the 2014 Liquor Amendment Act (the Sydney lockout laws), with the intention of informing future policy decisions regarding these laws.
Method We follow the *Bureau of Crime Statistics and Research* (BOSCAR) and define the measure of success of the laws to be the reduction in *Non-Domestic Assaults* (NDA) in the areas of New South Wales which were directly targeted by the legislation and the surrounding geographical areas. The analysis in this report consists of multiple stages for analysing time series of non-domestic assaults. The methodology consists of state of the art methods for analysing time series in both the time and frequency domains, modelled within a Bayesian, nonparametric framework. First, we apply adaptive spectral density estimation for change point identification using the AdaptSPEC algorithm. Once structural breaks are detected, temporal trends are estimated using log Gaussian Cox Process regression.
Results We analyse time series of non-domestic assaults for each geographic region and make inference regarding the impact of the legislation. Results show that the impact of the laws was not consistent across the areas of implementation. While violence in Kings Cross was reduced by the introduction of the lockout laws, the same outcome was not observed in the Sydney Central Business District. Results also suggest that there has not been a displacement of crime towards proximal and distal displacement areas.
Comparison with BOCSAR's Findings The results in this report vary to those presented by BOCSAR in their previous analysis, mainly regarding the lockout laws' impact on the Sydney *Central Business District* (CBD). As requested by BOCSAR, we compare our findings with theirs and comment on any difference in conclusions regarding the impact of the lockout laws. We elaborate on the differences and present the reasons behind these discrepancies, which we classify into two categories:

1. Differences in Data Processing:
   We note that the major discrepancy between the two analyses was the manner in which BOCSAR allocated crimes to geographic areas, which we now discuss. Figure 1a presents a map of the areas used by BOCSAR in their latest report Donnelly et al. (2017). This map shows that the area defined as the CBD includes many locations not usually associated with the definition of the CBD, including Woolloomooloo, Darlinghurst and Potts Point. We argue that the areas should be redefined to limit the extension of the CBD to better understand the dynamics of crime in the different



(a) Map of Kings Cross Area, New South Wales, Australia. Blue - *Central Business District* (CBD); Red - *Kings Cross* (KC); Green - *Proximal Displacement Areas* (PDA)

(b) Overall counts of criminal incidents which have been simultaneously assigned to Kings Cross and CBD by Donnelly et al. (2017).

Figure 1: Map of Areas and Comparison

areas. The technique used by BOCSAR led to the allocation of approximately 1900 crimes which occurred in Kings Cross being counted among those that occurred in the CBD as well as Kings Cross. Figure 1b shows the spatial distribution and number of crimes that were counted as both occurring within Kings Cross and the CBD. We would like to make the following three comments about this data processing procedure.

(a) If there is ambiguity about a boundary location, often referred to as a soft boundary, then there are statistical techniques which attempt to address this issue, such as weighting the crime to both areas. For example a weight of 50% to both areas if there is genuinely no knowledge about the boundary.

(b) Figure 1b shows that for the predominant number of crimes assigned to both the CBD and Kings Cross, there is no ambiguity as to the crime's location.

(c) If there is uncertainty in allocating the exact location of certain crimes, it is good practice to do the analysis multiple ways. For example, allocating all the duplicated crimes first to the CBD and then to Kings Cross. If the inference regarding the impact of the lockout laws is the same irrespective of how the crimes were attributed to a location, then from a policy point of view, it is not particularly relevant. However, our analysis shows that allocation of a crime to an area influences the conclusions around the efficacy of the lockout laws. If we use the same allocation scheme as BOCSAR we detect a change point in early 2014, but not otherwise.

2. Differences in Statistical Methods
The technique we have used to analyse the data has been peer reviewed and published in the Journal of the American Statistical Association (JASA)[1], where the properties of the methodology were rigorously tested on simulated and real data. It has been used and cited by many authors since its publication in 2012. The code is publicly available https://cran.r-project.org/web/packages/BayesSpec/BayesSpec.pdf.

[1]https://www.tandfonline.com/doi/full/10.1080/01621459.2012.716340

There are many statistical models which could be used to analyse the effectiveness of the lockout laws. We do not dispute the technique or analysis in Donnelly et al. (2017), however we have taken a different approach. Our approach differs from that of Donnelly et al. (2017) in four ways:

(a) The technique does not assume that the time series are stationary. Instead the number and timing of any structural breaks in the time series are considered to be random variables, and estimated from the data.

(b) The technique does not assume a parametric model for the data within each segment.

(c) We use daily data from January 2005 to December 2017 whereas Donnelly et al. (2017) used monthly data from 2009-2016.

(d) Inference regarding the quantities of interest are made in a Bayesian framework, while those done in Donnelly et al. (2017) are made in a frequentist setting.

Conclusions We have applied a fully probabilistic approach, allowing for identification of structural breaks and adaptive spectral density estimation across time series of non-domestic assaults. This report provides evidence of the impact of past legislation, and outlines differences between our analysis and prior work done by BOCSAR. Further work is required to better understand this complex landscape. We hope this study and its findings will serve as an additional resource in augmenting future policy decisions regarding the lockout laws.

1 Introduction

Following a run of late night assaults and the catalytic deaths of Thomas Kelly and Daniel Christie, the Sydney lockout laws, introduced in February 2014, aimed to reduce alcohol-related assaults. There is significant literature evidencing the relationship between the concentration of liquor outlets and alcohol-related social harm (Chrikritzhs et al., 2007, Escobedo and Ortiz, 2002, Gruenewald et al., 2006, 2002, Jewell and Brown, 1995, LaScala et al., 2000, Lipton and Gruenewald, 2002, Zhu et al., 2004). Within Australia, Stevenson et al. (1999) demonstrated that there was a strong positive correlation between overall alcohol sales in an area and its incidence of assault in Sydney and New South Wales. Burgess and Moffat (2011) used spatial methods to demonstrate that zones surrounding licensed premises had a higher assault rate than zones surrounding commercial premises. With the intention of reducing alcohol-related violence and improving public safety, the New South Wales Government introduced the Liquor Amendment Act (henceforth referred to as the lockout laws) in February 2014, which placed several restrictions on licensed premises:

1. 1:30 am lockouts at clubs, hotels etc. in Sydney *Central Business District* (CBD) and *Kings Cross* (KC).
2. Ceasing alcohol service in venues at 3am.
3. Freezing new liquor licences in Sydney CBD and KC.
4. Banning takeaway alcohol sales after 10:00pm across New South Wales.
5. Extending temporary and long-term banning orders, intended to exclude known criminals from most licensed establishments in KC and Sydney CBD.

This report inspects the dynamics of crime in the areas of interest, which mainly correspond to the Sydney *Central Business District* (CBD), *Kings Cross* (KC), *Proximal Displacement Areas* (PDAs) and *Distal Displacement Areas* (DDAs). These areas are shown in Figure 2 and were previously defined by Donnelly et al. (2017) in a study by the NSW *Bureau of Crime Statistics and Research* (BOCSAR). Despite these areas appearing to be disjoint (Donnelly et al., 2017), there is considerable overlap, which is further investigated in Section 2.4. Donnelly et al. (2017) identified that the reforms resulted in significant reductions of non-domestic assaults in the CBD and KC areas, with 13% and 49% declines respectively. The authors also suggest a displacement of crime towards PDAs and DDAs, where crime increased by 12% and 17%.

The objective of this report is to analyse non-domestic assault time series for respective geographic areas of interest. The time series are generated by aggregating daily non-domestic assault counts for each of the geographic areas defined by Donnelly et al. (2017). The daily aggregated counts of reported non-domestic assaults are shown in Figure 3, while the monthly counts are shown in Figure 4. This analysis utilises state of the art statistical learning methodology for change point detection and spectral density estimation; the AdaptSPEC method Rosen et al. (2012). There are several advantages of using a fully probabilistic nonparametric Bayesian framework for tackling such a complex problem:

1. Change point detection - The method utilised in this report does not make any assumptions regarding the number or location of change points in the time series. The Bayesian framework enables uncertainty quantification when estimating both the number and location of change points.
2. Adaptive spectral density estimation - The AdaptSPEC method produces an estimate of the time-varying spectral density. This allows for inference regarding the evolution of periodicities over time.

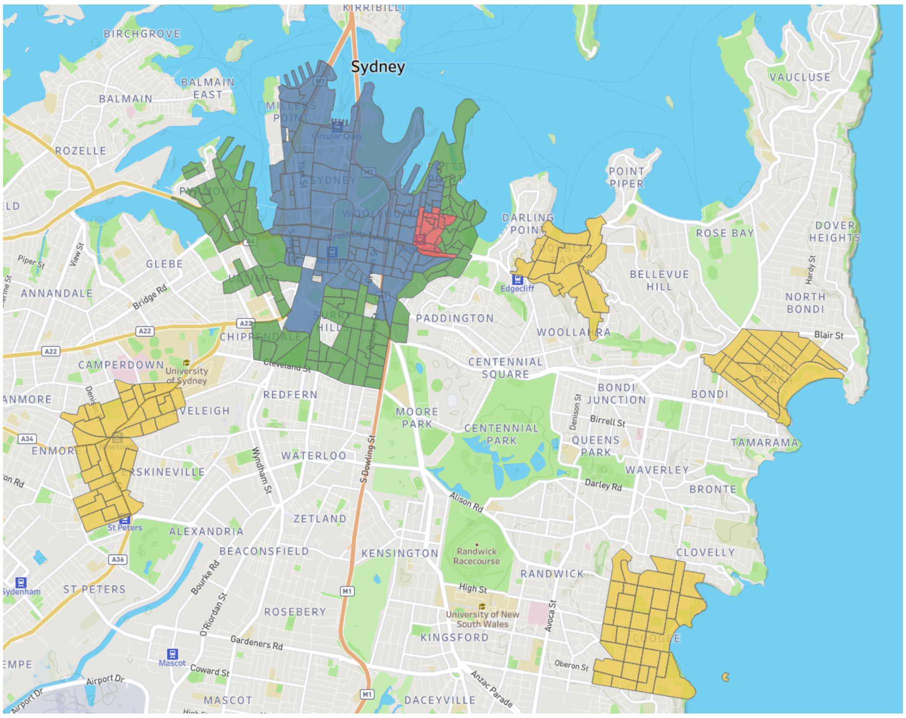


Figure 2: Map of Sydney, New South Wales, Australia. Highlighting areas of analysis: Blue - *Central Business District* (CBD); Red - *Kings Cross* (KC); Green - *Proximal Displacement Areas* (PDA); Yellow - *Distal Displacement Areas* (DDA)

3. Flexible estimation of temporal dynamics using a non-parametric model - Within each segment, we place a log Gaussian Cox Process. This model makes no assumptions about the functional form of the trend, and provides a posterior probability distribution over the space of functions that better represent the underlying dynamics of non-domestic assaults.

The main contribution of this report is providing a probabilistic, Bayesian treatment for modelling non-domestic assaults across specific areas of New South Wales to assess the effectiveness of the lockout laws across space and over time.

This report is structured as follows. Section 2 presents the results of applying the methodology on non-domestic assaults in the regions seen in Figure 2. Section 3 highlights the conclusions of this report and the future work we propose to better understand the impact of the laws. Finally, the Appendix provides further technical discussion and the theoretical background on the probabilistic methods of time series analysis applied in this paper.

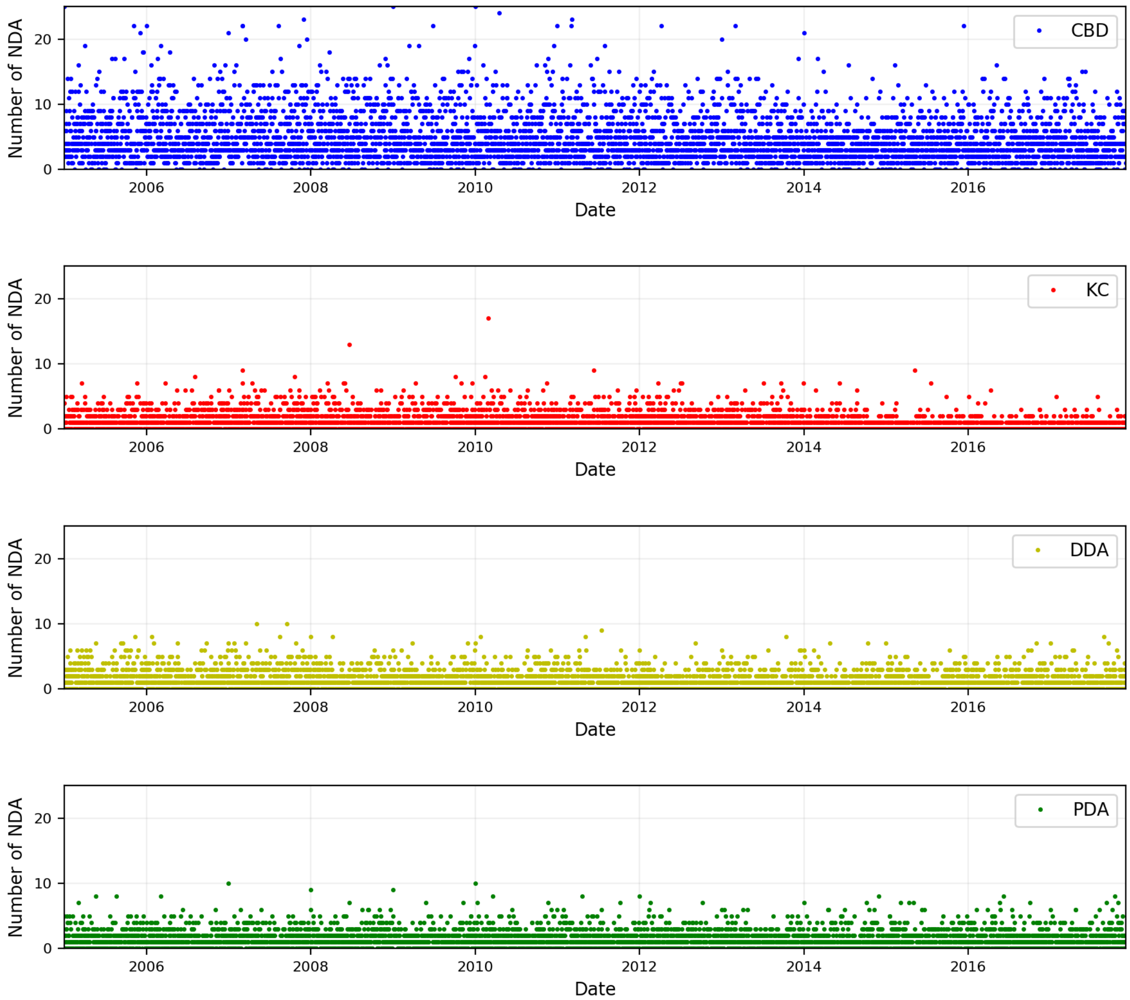


Figure 3: Daily counts of NDA over the areas defined in Figure 2. The colours indicate the respective region corresponding to Figure 2

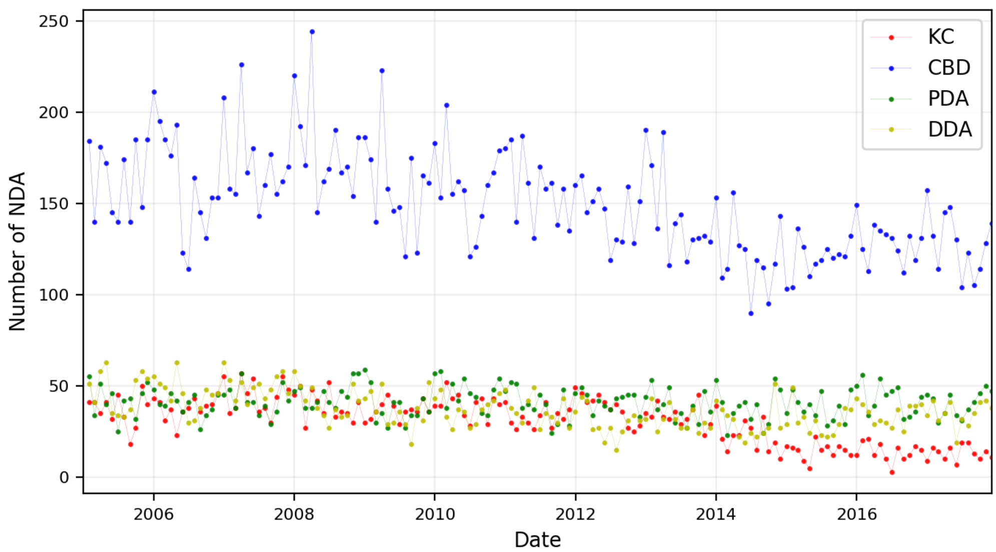


Figure 4: Monthly counts of NDA over the areas defined in Figure 2. The colours indicate the respective region corresponding to Figure 2

2 Case Study: Sydney Lockout Laws

This section presents details on data manipulation, methodology and results of the analysis conducted over criminal incidents in the areas of interest. We have omitted the mathematical details of the analysis, which are further presented in the Appendix, Section 4.

2.1 Data

This report is based on data provided by the NSW *Bureau of Crime Statistics and Research* (BOCSAR). Specifically, we have used the Unit Record Criminal Incident Dataset[2]. This dataset contains 9.9 million historical records since 1995 across New South Wales, Australia, for multiple crime types. Of these, we consider 39,864 criminal incidents corresponding to non-domestic assaults between 1 January 2005 and 30 November 2017 in the areas of CBD, KC, PDA and DDA.

BOCSAR have provided us with geo-referenced criminal incidents, referenced to a *Statistical Area level 1* (SA1) [3]. There are certain limitations with the geo-referencing process BOCSAR uses for assigning incomplete SA1 data. The problem mainly arises from assigning all crimes with uncertain locations to the centroid of the geographical region in the lowest level of the referencing tree (Street, Suburb, State). To overcome this, we requested raw data of addresses and landmarks to conduct our own geo-referencing process, however this request was not approved by BOCSAR on the grounds of identifiability.

Considering the data is aggregated at the SA1 level, the following step in the process is to determine which SA1s belong to which respective geographic area (KC, CBD, PDA or DDA). The map shown in Figure 2, depicts the geographical segmentation, where the black boundaries within each coloured region identify an SA1. All criminal incidents within an SA1 were only assigned to the area which they belonged. Of the total NDA crimes in these areas, 23,148 (58.07%) were assigned to the CBD, 4,767 (11.96%) were assigned to KC, 5,609 (14.07%) were assigned to DDA and 6,340 (15.9%) were assigned to PDA. Of particular interest is to inspect Figure 5, which shows a zoomed in version of the Kings Cross area. It is important to see how the CBD area extends all the way through Woolloomooloo, Darlinghurst and even parts of Potts Point. We argue that the areas should be redefined to limit the extension of the CBD to better understand the dynamics of crime in the different areas.

Once the data for each SA1 is assigned to a region (KC, CBD, PDA or DDA), it is trivial to recover the daily and monthly counts of crimes for each region, which are shown in Figure 3 and Figure 4 respectively.

2.2 Methodology

The method consists of applying the following processing steps to extract and analyse the data.

1. Assign SA1 codes to each area (KC, CBD, PDA, DDA).
2. Aggregate incidents by day for each geographic area.
3. Find breakpoints and stationary segments in each time series (details in Section 4.2).
4. Aggregate incidents by month for each geographic area.
5. Fit a non-parametric model, and produce an estimate of the temporal trend for each stationary segment using Section 4.3.

[2]Dataset reference code nm1816581, year 2018

[3]For more information on statistical areas, defined by the Australian Bureau of Statistics, visit https://www.abs.gov.au/

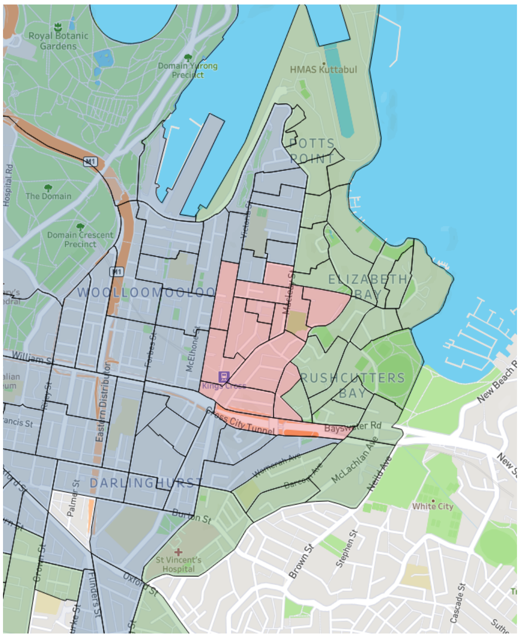


Figure 5: Map of Kings Cross Area, New South Wales, Australia. Blue - *Central Business District* (CBD); Red - *Kings Cross* (KC); Green - *Proximal Displacement Areas* (PDA)

Area Name	Number of Breaks	Dates of Breaks
Central Business District (CBD)	1	August 2011
Kings Cross (KC)	2	January 2014; April 2016
Proximal Displacement Areas (PDA)	0	-
Distal Displacement Areas (DDA)	1	April 2008

Table 1: Number and date of structural breaks found in the time series for each area.

2.3 Results

This section highlights the main results for the spectral and temporal analysis. For a deeper and more technical description of the results, we refer the reader to the Appendix, Section 4.2.

First, the individual time series for each region are processed using the AdaptSPEC algorithm (Section 4.2 and Rosen et al. (2012)). AdaptSPEC provides the modal estimate of structural breaks within a time series. The details of these breaks are shown in Table 1 and in Figure 6 as dashed vertical lines, coloured depending on the region.

The key observation regarding the breaks detailed on Table 1 and Figure 6 is that Kings Cross is the only area for which there was a structural break close to the dates the lockout laws' introduction. For all other areas, including the CBD, PDA and DDA, there are no structural breaks close to March 2014. This suggests that the lockout laws did not have a noticeable impact in these areas.

Using the structural breaks, it is possible to recover stationary segments within the overall time series. Each segment is modelled using a Log Gaussian Cox Process, as described in Section 4.3 of the Appendix.

The trend in NDA in Kings Cross is characterised by a sharp reduction in crime at the time of the lockout laws, followed by a further slow decline after the lockout laws' implementation. Prior to the lockout laws the monthly number of NDA in Kings Cross was relatively consistent, with a trend ranging between 30-35 NDA per month. At the time of the lockout laws' enactment in February 2014 (which was almost coincidental with the identification of the first change point), the monthly NDA reduced to $\sim$ 20 NDA per month. The trend in monthly NDA slowly declined until April 2016 (the second change point) where it reached a stable rate of $\sim$ 10-15 NDA per month. Our analysis demonstrates that the lockout laws had a meaningful immediate impact on reducing NDA in Kings Cross, followed by a a sustained slow decline in NDA per month.

The trend in NDA in the CBD had significantly higher variability over time when compared with other geographic regions. NDA per month in the CBD rose steadily until August 2008 reaching a high of $\sim$ 170 NDA per month. Proceeding this, NDA started to gradually decline. This decline did coincide with the global decline in NDA across New South Wales. The only detected change point in the CBD's time series was located in August 2011 where the trend in monthly NDA dropped from $\sim$ 155 per month to $\sim$ 145 per month. The second segment of the CBD's time series (August 2011 - December 2017), had an initial drop in the trend of NDA per month reaching a low of $\sim$ 125 NDA per month in late 2014, followed by an upswing - reaching a local maximum near the end of 2016. This analysis shows that the Sydney lockout laws did not have a meaningful or sustained impact on NDA per month in the CBD region.

The trend in PDA's NDA per month is largely constant over time. PDA averaged approximately 40 NDA per month between January 2005 and December 2017, and there was no change point detected over the entire time period analysed. This analysis suggests that the lockout laws had no impact on NDA per month in PDA.

The trend in DDA NDA per month increased slowly from January 2005 until their peak in April 2008, where the monthly NDA reached $\sim$ 50. A change point was detected in April 2008, which was accompanied by a sharp drop in NDA per month down to $\sim$ 35. Between April 2008 and the implementation of the lockout laws in February 2014, NDA per month slowly declined reaching a low point of $\sim$ 30 NDA per month. Between the lockout laws' implementation and December 2017, the trend has steadily increased to $\sim$ 40 NDA per month. Although the trend in NDA per month for DDA steadily increased directly after the lockout laws' implementation, no change point was identified during the period Feburary 2014 - December 2017. Accordingly our analysis is unable to support the suggestion that NDA has displaced from Kings Cross and the CBD into other areas such as PDA and DDA.

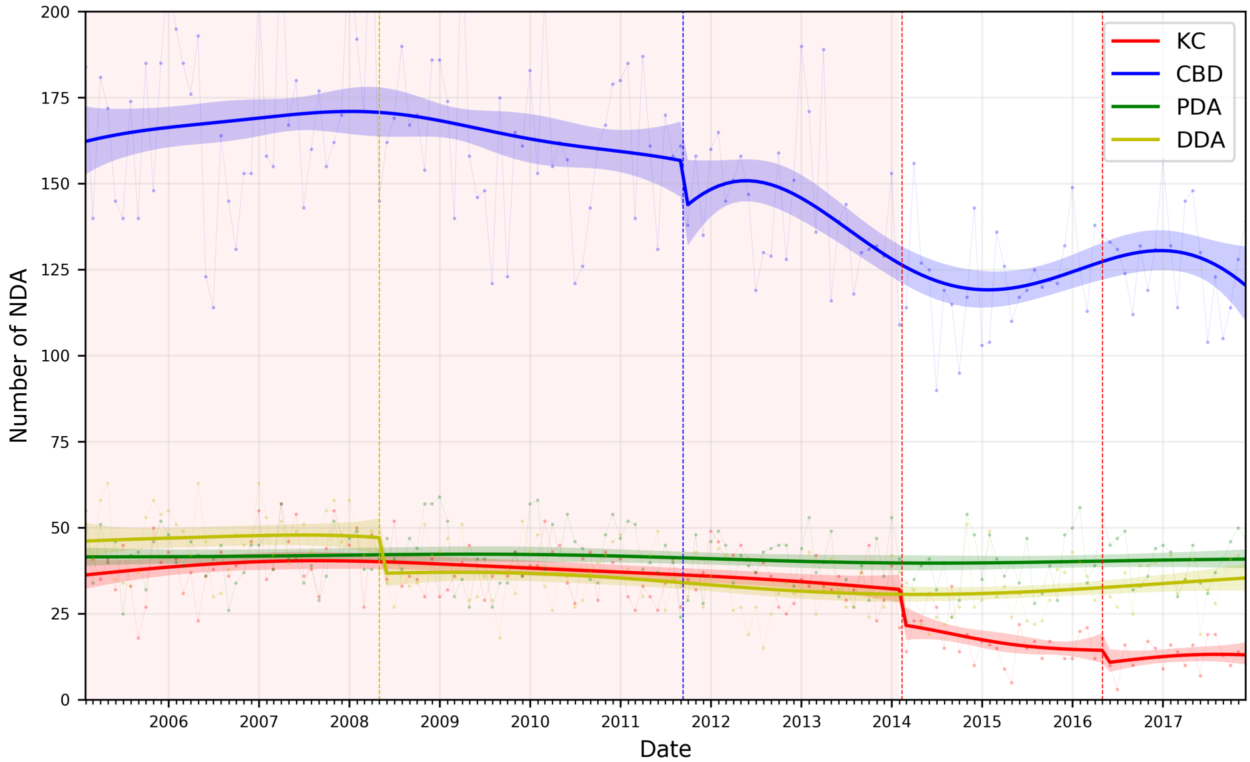


Figure 6: Monthly crime data including trend for each stationary segment and for each area.

2.4 Differences to BOCSAR historical reports

There is an important difference between the results in this report and the results obtained by the *Bureau of Crime Statistics and Research* (BOCSAR) in their previous analysis Donnelly et al. (2017). The main difference is the inference around crime numbers in the CBD area. While BOCSAR has stated that there is a reduction in crime of 13% in the CBD area, we argue that there is no reduction. In fact, as shown in Figure 6, the trend has been oscillating since the introduction of the laws in 2014 and has not declined over time.

Over the past month, we have worked in collaboration with Suzanne Poynton, Acting Director of BOCSAR, who is aware of the differences in methods and results. In order to identify these differences, BOCSAR has provided CTDS the exact data used for the report on Donnelly et al. (2017), which we call *Precinct Dataset* (PD).

The goal is then to understand differences in the respective approaches and the source of varying conclusions around the CBD by fully exploring possible differences between the Precinct and SA1 dataset used in Section 2, which we call *SA1 Dataset* (SA1D).

After thorough investigation, we narrowed down the source of different conclusions in the CBD, which arise from different time series being analysed. Figure 7, shows the individual differences in monthly counts for the CBD area between PD and SA1D. A crucial observation is that BOCSAR's precinct data for the CBD, in red, presents larger counts that our data in the period before the lockout laws, but highly similar counts after the introduction of the laws.

This helped us identify the source of discrepancies: There is a large number of crimes ($\sim$ 1900 incidents) which are assigned simultaneously to the CBD and Kings Cross for Donnelly et al. (2017), due to overlap in these regions. This overlap is not clear and no further explanation is presented in Donnelly et al. (2017). According to BOCSAR,

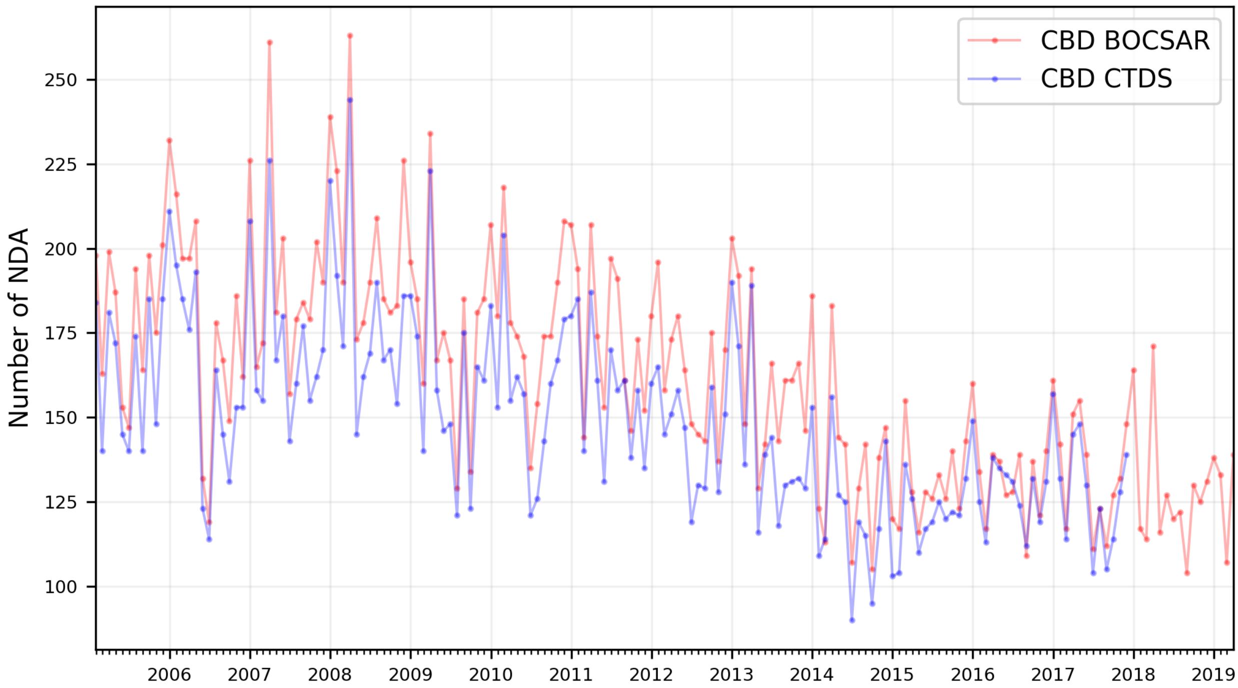


Figure 7: Differences between time series data amongst datasets, for CTDS SA1 assembled dataset and BOCSAR's precinct dataset.

the crimes that are simultaneously taken into account in the CBD and KC arise from uncertain geo-referencing of incidents at the street level. Using the raw data in PD, we show the spatial and temporal distribution of crimes in the overlapped region between the CBD and KC and these are presented in Figure 9 and Figure 8. This evidence indicates that any inference over the CBD is importantly driven by the behaviour of KC NDAs. More precisely, the apparent decline in the CBD, observed by Donnelly et al. (2017), is an artefact of the overlapping region between KC and CBD.

BOCSAR has agreed to collaborate with CTDS and generate an updated joint analysis of the Lockout Laws, and further improve the geocoding for incidents with missing data in overlapping regions.

3 Conclusion and Future Work

We have analyzed non-domestic assaults in New South Wales using an adaptive spectral analysis technique. The method can determine the number and location of change points along a (possibly non-stationary) time series. There are several insights from our case study. First, the lockout laws have been very effective in Kings Cross - reducing the total number of assaults and alleviating the strong periodic component that corresponded to a weekly assaults trend. There was no change point detected in the CBD around the time of the lockout laws, and the data suggests that the lockout laws have had no effect in reducing crime in this area. There are no change points detected for PDA or DDA that relate to the lockout laws' enactment.

The absence of change in the CBD certainly questions whether the Kings Cross lockout laws have been an effective method of changing peoples' late night violent behaviours, or simply eradicating an area of once popular late night venues.

There is a considerable amount of work required to fully understand and continuously monitor the effect of this legislation amendment. We propose improving the geo-

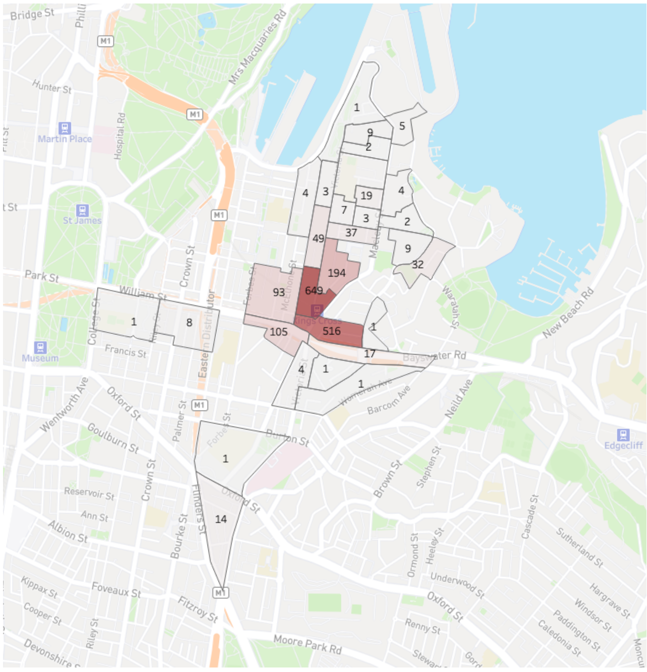


Figure 8: Overall counts of criminal incidents which have been simultaneously assigned to Kings Cross and CBD by Donnelly et al. (2017).

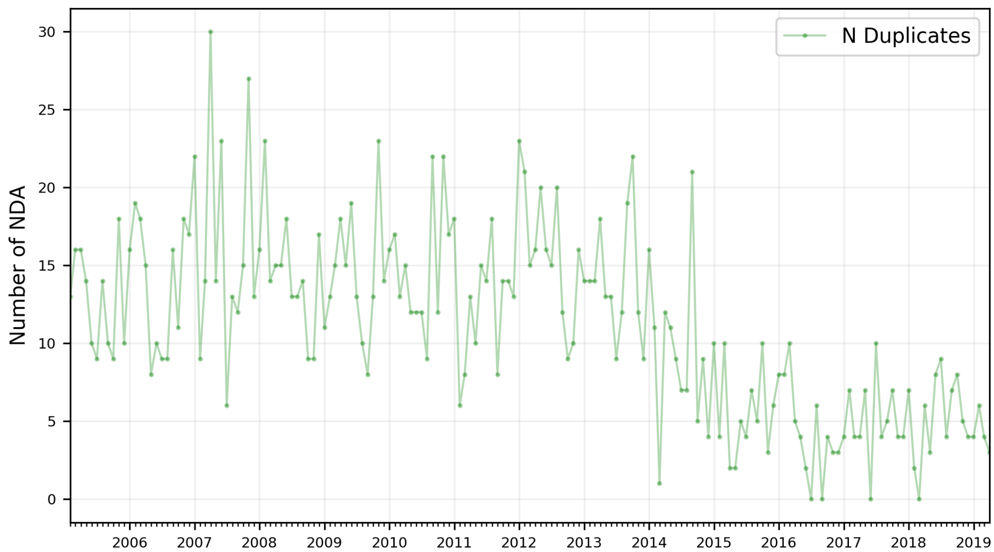


Figure 9: Monthly counts for criminal incidents which have been simultaneously assigned to Kings Cross and CBD by Donnelly et al. (2017).

referencing of crimes that present uncertain locations. We also propose conducting this same analysis on the improved geo-referenced data and with a new geographical definition of Precincts.

Acknowledgment

We would like to thank the NSW *Bureau of Crime Statistics and Research* (BOCSAR) for providing us with access to the Unit Record Criminal Incident Dataset, from which we extracted daily counts of non-domestic assaults. Reference code nm1816581, year 2018. We also thank BOCSAR for ongoing discussions and data/information sharing around their report Donnelly et al. (2017). This was crucial in understanding the differences between the two approaches.

References

Burgess, M. and Moffat, S. (2011). The Associaton Between Alcohol Outlet Density and Assaults on and Around Licensed Premises. *Contemporary Issues in Crime and Justice*, 147.

Chrikritzhs, T., Catalano, P., and Pascal, R. (2007). Predicting alcohol-related harms from licensed outlet density: A feasiblity study. *National Drug Law Enforcement Research Fund Monograph Series*, 28.

Dahlhaus, R. (1997). Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics*, 25(1):1–37.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 101(473):223–239.

Donnelly, N., Poynton, S., and Weatherburn, D. (2017). The effect of lockout and last drinks on non-domestic assaults in Sydney: An update to September 2016. *Contemporary Issues in Crime and Justice*, 201.

Escobedo, L. G. and Ortiz, M. (2002). The relationship between liquor outlet density and injury and violence in New Mexico. *Accident Analysis and Prevention*, 34:689–694.

Gruenewald, P. J., Freisthler, B., Remer, L., LaScala, E. A., and Treno, A. (2006). Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction*, 101:666–677.

Gruenewald, P. J., Johnson, F., and Treno, A. (2002). Outlets, drinking and driving: A multilevel analysis of availability. *Jounral of Studies on Alcohol*, 63:460–468.

Jewell, R. T. and Brown, R. W. (1995). Alcohol availability and alcohol-related motor vehicle accidents. *Applied Economics*, 27(8):759–765.

LaScala, E. A., Gerber, D., and Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis and Prevention*, 32:651–658.

Lipton, R. and Gruenewald, P. J. (2002). The spatial dynamics of violence and alcohol outlets. *Journal of Studies on Alcohol*, 63:187–195.

Rasmussen, C. E. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology.

Rosen, O., Wood, S., and Stoffer, D. S. (2012). Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association*, 107(500):1575–1589.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer Texts in Statistics.

Stevenson, R. J., Lind, B., and Weatherburn, D. (1999). The Relationship between Alcohol Sales and Assault in New South Wales, Australia. *Addiction*, 94(3):397–410.

Whittle, P. (1957). Curve and Periodogram Smoothing. *Journal of the Royal Statistical Society*, 19(1):38–63.

Wood, S., Rosen, O., and Kohn, R. (2011). Bayesian mixtures of autoregressive models. *Journal of Computational and Graphical Statistics*, 20(1):174–195.

Zhu, L., Gorman, D. M., and Horel, S. (2004). Alcohol Outlet Density and Violence: A Geospatial Analysis. *Alcohol and Alcoholism*, 39(4):369–375.

Appendix

4 Theoretical Background

This section introduces the reader to the methodology used in this report by summarising the related theory for change point detection and nonparametric regression over count data. For a general background on time series analysis, we refer the reader to Shumway and Stoffer (2017).

4.1 Time Series

The objective of quantitative time series analysis is to build mathematical models that best represent the realisation of a temporal stochastic process. In the context of this report, a time series $\boldsymbol{y}$ is defined by a series of integer values, equally spaced in time, i.e.

$$\boldsymbol{y} = \{y_t\}_{t=1,\dots,T} \ , \tag{1}$$

where $y_t \in \mathbb{N}$ represents the number of events within the time interval Δt.

A crucial characteristic of a time series, relevant for this report, is the autocovariance function γ, which is defined by

$$\gamma(s,t) = \mathrm{cov}(y_s, y_t) = \mathbb{E}[(y_s - \mu_s)(y_t - \mu_t)] \ \ \forall s, t \ , \tag{2}$$

where $\mu_\cdot = \mathbb{E}(y_t)$ is the mean function. The autocovariance measures the linear dependence between two points of the time series. When the time series exhibits regular behaviour (that is, the statistical properties do not change over time), one refers to this as a stationary time series, which must satisfy

$$\gamma(t+h, t) = \mathrm{cov}(y_{t+h}, y_t) = \mathrm{cov}(y_h, y_0) = \gamma(h, 0) \equiv \gamma(h) \ . \tag{3}$$

The spectral density $f(\nu)$ of a stationary time series is a function over the frequency domain ν which contains the same information as the autocovariance function and is given by

$$f(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h} \quad -\frac{1}{2} \leq \nu \leq \frac{1}{2} \ . \tag{4}$$

A popular representation of a time series in the frequency domain is achieved by calculating the *Discrete Fourier Transform* (DFT). The DFT of a time series $\{y_t\}$ consisting of T observations, at frequency ν_k is defined as

$$x(\nu_k) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \times (\cos(2\pi\nu_k t) - i \sin(2\pi\nu_k t)) \ ,$$

where $\nu_k = k/T \ \forall k \in \{0, 1, \dots, (T-1)\}$ are the fundamental frequencies. Let the periodogram, a noisy estimate of the true spectral density f at frequency ν_k, $I(\nu_k)$, be the squared modulus of the DFT coefficients

$$I(\nu_k) = |x(\nu_k)\bar{x}(\nu_k)| \ . \tag{5}$$

Whittle (1957) showed that the distribution of $\mathbf{x} = (x(\nu_1) \dots, x(\nu_n))$ is complex normal. Using this property we can rewrite

$$\boldsymbol{x} \sim \prod_{k=1}^{n} \frac{1}{\pi f(\nu_k)} \exp\left(-\frac{I(\nu_k)}{f(\nu_k)}\right) . \tag{6}$$

This representation suggests that $I(\nu_k)$ are i.i.d. with $I(\nu_k) \sim \exp(f(\nu_k))$ and therefore

$$\log(I(\nu_k)) = \log(f(\nu_k)) + \epsilon_k; \quad \epsilon_k \sim \log(\exp(1)). \tag{7}$$

Letting $w(\nu_k) = \log(I(\nu_k))$ and $g(\nu_k) = \log(f(\nu_k))$ we have

$$w(\nu_k) = g(\nu_k) + \epsilon_k. \tag{8}$$

4.2 Change Point Detection

In the context of this report, the main objective is to detect any changes in the structure of the time series. To detect these change points one must remove the assumption of a stationary time series. In this section we present the main principles used for dealing with non-stationary time series.

We now assume that the time series is composed of a series of locally stationary process (as in Dahlhaus (1997)) with an evolutionary spectrum $\boldsymbol{f}(\nu, t)$, which we wish to estimate. This spectrum has an unknown, but finite, number K of piecewise stationary processes, each of length n_s for $s = 1, \dots, K$. A powerful probabilistic estimation method for finding the number of stationary processes and their spectra is called AdaptSPEC and is presented by Rosen et al. (2012).

Given a partition of K segments, we define the partition points to be

$$\boldsymbol{\xi}_K = (\xi_{0,K}, \xi_{1,K} \cdots \xi_{K,K}) \ , \tag{9}$$

with $\xi_{0,K} = 0$ and $\xi_{K,K} = T$. By following Davis et al. (2006), we define

$$A_s = \{t; \xi_{s-1} + 1 < t < \xi_s\} \ . \tag{10}$$

We may therefore rewrite the time series as,

$$\boldsymbol{y} = \sum_{s=1}^{K} y_t^{(s)} \delta(t, A_{s,K}), \tag{11}$$

where, $\delta(t, A_{s,K}) = 1$, if $t \in A_s$ and $\delta(t, A_{s,K}) = 0$ otherwise, and where $y_t^{(s)}$ are independent stationary processes, for $s = 1, \dots, K$, each with spectral density $\boldsymbol{f}_{s,K}(\nu)$.

The joint probability density function of a realization $\boldsymbol{y} = (y_1, \dots, y_T)$ given the individual spectra $\boldsymbol{F}_K = (\boldsymbol{f}_{1,K}(\nu), \dots, \boldsymbol{f}_{K,K}(\nu))$, number of segments K, and partition points $\boldsymbol{\xi}_K$ is

$$p(\boldsymbol{y}|\boldsymbol{F}_K, K, \boldsymbol{\xi}_K) = \prod_{s=1}^{K} p\left(y_{\xi_{(s-1,K)}+1}, \dots, y_{\xi_{(s,K)}} | \boldsymbol{f}_{s,K,K,\xi_{(s,K)}}(\nu)\right). \tag{12}$$

AdaptSPEC is defined within a Bayesian framework. It makes use of prior probability distributions over the unknown values and the time series data to infer a posterior distribution over the number of stationary segments, the partition points and their respective spectral density.

4.2.1 Prior for Spectra

Given a partition defined by K segments, the respective partition points $\boldsymbol{\xi}_K$, and a realization $\boldsymbol{y}^{(s)}$, our goal is to estimate the unknown spectra $\boldsymbol{f}_{s,K}(\nu)$, for $\nu \in (-0.5, 0.5)$. When selecting the prior over $\boldsymbol{f}_{s,K}(\nu)$, we frame the problem of estimating the autocovariance structure of our time series, represented by the power spectrum, as a nonparametric regression problem. In effect, we turn a covariance estimation problem into one of mean estimation, which is more parsimonious and tractable.

To specify the prior, we decompose the unknown function $g_s(\nu_k)$ into its linear and non-linear components so that $g_s(\nu_k) = \alpha_{s0} + h_s(\nu_k)$. A Gaussian Process (Rasmussen and Williams, 2006) prior is placed over the unknown function $h_s(\nu_k)$. The Gaussian Process is a suitably flexible prior. In the case of this specific model, we assume

$$h_s(\nu_k) = \tau_s W(\nu_k), \tag{13}$$

or equivalently,

$$\boldsymbol{h}_s = (h_s(\nu_1), \ldots, h_s(\nu_{n_s})) \sim \mathcal{N}\left(0, \tau_s^2 \Omega\right), \tag{14}$$

where $W(.)$ is a Wiener process, τ_s^2 is a 'smoothing' parameter and the i^{th}, j^{th} element of Ω, are given by $\omega_{ij} = \text{cov}(h_s(\nu_i), h_s(\nu_j)) = \min(\nu_i, \nu_j)$.

For computational convenience we write $\boldsymbol{h}_s$ as a linear combination of basis functions. We perform an eigenvalue decomposition on $\Omega = QDQ'$ and only keep basis functions corresponding to the 30 largest eigenvalues for computational speed. That is, we let $X = QD^{1/2}$ be the design matrix and $\boldsymbol{\beta}_s \sim (0, \tau_s^2 I_{n_s})$ be the vector of regression coefficients, so that $\boldsymbol{h}_s = X\boldsymbol{\beta}_s$ has the required distribution.

4.2.2 Prior for Partition

The partition is defined by the number of locally stationary segments K and the respective partition points, $\boldsymbol{\xi}_K$, given K which separate any candidate number of segments. The prior we place over the partition $\Pr(K, \xi_S) = \Pr(\xi_s|K)\Pr(K)\ \xi_{s,K}$ is as follows;

$$\Pr(K) = \frac{1}{S}, \tag{15}$$

where S is the the upper limit for the number of segments. Given K we decompose the prior on ξ_K into a sequence of discrete uniform prior distributions, so that

$$\Pr(\xi_K \mid K) = \prod_{s=1}^{K-1} \Pr(\xi_{s,K}|\xi_{s-1}, K)\ , \tag{16}$$

where $\Pr(\xi_{j,m} = t \mid m) = 1/p_{s,K}$, for $s = 1, \ldots, K-1$, $p_{s,K}$ is the number of available locations for our partition point $\xi_{s,K}$ and is equal to $T - \xi_{s-1,K} - (K - s + 1)t_{\min} + 1$. The quantity $t_{\min}$ is a user chosen number which represents the minimum number of observations assumed to be sufficient for the Whittle Likelihood approximation to be valid.

The prior in Equation 16 states that, as long as there are at least $t_{\min}$ observations in each of the K segments, the first partition point is equally likely to occur at any point in the time series. The prior on subsequent partition points (change points) is similar and states that conditional on the previous partition point, the next possible partition point is equally likely at any location, again subject to the same constraint regarding the minimum number of observations for each segment.

See for example Wood et al. (2011) and Rosen et al. (2012) for more details regarding the implementation.

4.3 Non-parametric Regression over Count Data

In this work, we place a non-parametric regression model over the count data for each stationary segment. Since count data on the time series follows a Poisson distribution, the hierarchical model used is

$$y_t \sim \text{Poisson}(\lambda(t)) \tag{17}$$

$$\log(\lambda(t)) \sim \text{GP}(m(t), k(t, t')) \tag{18}$$

The covariance function k used in this report is the Squared Exponential. For more information we refer the reader to Rasmussen and Williams (2006).

5 Detailed Results

This section presents further analysis of the individual time series using the concepts in Section 4.

Uncertainty in number and location of change points To determine the most probable number of segments across various regions' time series, the modal number of segments is taken from the posterior distribution as seen in Figure 11. Kings Cross' posterior distribution (Figure 11a), indicates there are either three or four segments, with three being the most probable number. Figure 11b and Figure 11d indicate that the modal number of segments for the CBD and DDA respectively is two. Finally, Figure 11c shows that the modal number of segments for PDA is one. Additionally, AdaptSpec allows for uncertainty quantification regarding the location of each change point in a time series. For Kings Cross, Figure 12a suggests there is uncertainty around the location of the first change point, with some probability mass nearly as late as July 2014. The location of the second change point is much more certain, with a sharp peak in April 2016. For the CBD, the location of the only change point is uncertain, with the modal location in August 2011 as seen in Figure 11b. However, the algorithm identifies a 24 month period (June 2010 - June 2012) as to where the cut point may have taken place. As for the DDA, there is a 4 month period where the cut point may have been located, between January 2008 and April 2008, with the modal location taking place in April 2008. This can be seen in Figure 12c.

Uncertainty in Spectra The Bayesian framework undertaken in this paper allows us to quantify the uncertainty around the power spectrum at any point in time. That is, taking a slice at any point in time, one may see the power spectrum and the 95% confidence intervals as seen in Figure 13. Figure 13 shows the median power spectra and 95% confidence interval for each geography of interest before and after the enactment of the lockout laws. Meaningful changes in the power spectrum before and after the lockout laws' implementation, are indicative of changes in the phenomenology of NDA in candidate areas. Figure 13a and Figure 13e, representing Kings Cross' spectrum before and after the lockout laws respectively, display the largest change in any geography's spectrum. The two most pertinent changes in the Kings Cross spectrum are:

1. A significant drop in the amplitude of the spectrum. The area under the spectrum represents the variance of the time series, and the marked decline is indicative of a reduction in the total variance of NDA.
2. The pronounced weekly and bi-weekly periodic components (visible peaks in the spectrum at frequencies ~ 0.14 and ~ 0.33 seen in Figure 13a prior to the lockout laws are almost completely alleviated after the lockout laws' implementation as seen in Figure 13e. The 95% confidence interval around the spectrum in Figure 13e suggest that there is more uncertainty around this estimate than the pre-lockout laws' spectrum in Figure 13a which has significantly narrower uncertainty bounds.

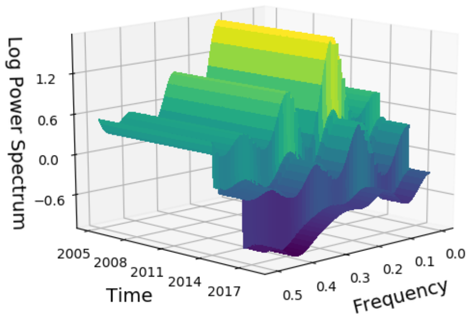
The CBD's only change point is detected in 2011. The most meaningful change between the spectra in Figure 13b and Figure 13f is a drop in the amplitude, representing a drop in the total variance in NDA in 2011 for the CBD. Both spectra in Figure 13b and Figure 13f have narrow uncertainty bounds, suggesting high confidence in this estimate. As there was no change point identified in the PDA time series, there is no change in the spectrum between 2005-2017, hence Figure 13c and Figure 13g are identical. Figures 13d and 13h, representing DDA before and after the lockout laws' implementation, demonstrate a drop in the amplitude between the pre-lockout law and post-lockout law spectra. Both Figures 13d and 13h have narrow bounds around the spectra, suggesting high confidence in the estimation of the spectra which has strong weekly and bi-weekly periodic components.

Kings Cross (KC) There are two change points detected for KC. The first change point is identified in January 2014 and the second is in April 2016. The first change point corresponds to a change in the spectrum that is related to the lockout laws. The power spectrum corresponding to the second segment (January 2014 - April 2016) has a mild drop at the frequency ~ 0.14. This suggests that less of the NDA total variance is comprised of assaults occurring at a weekly periodicity. The second structural break, detected in April 2016, represents a further reduction in the KC NDA variance. The weekly periodicity that had weakened between segments one and two, is almost completely alleviated in the power spectrum belonging to segment 3 (April 2016 - December 2017). This suggests that the weekly periodicity in crimes corresponding to Saturday night violence, is no longer a notable periodicity in KC NDA. The second structural break after the lockout law's implementation is further support that the lockout laws did have a meaningful impact on assault in Kings Cross. Although the trend in NDAs across New South Wales was declining over period where change points were detected, we believe that changes in Kings Cross were idiosyncratic for two reasons. First, the area under the power spectrum (corresponding to variance) between segments is meaningfully different. Second, the frequency components exhibiting most of the power changed materially between change points, highlighting that the phenomenology of NDAs in KC has evolved over time. Our method highlights a significant drop in the total variance of KC NDA, and almost a complete eradication of the weekly periodicity in NDA - which most likely correspond to the Sydney Lockout Laws.

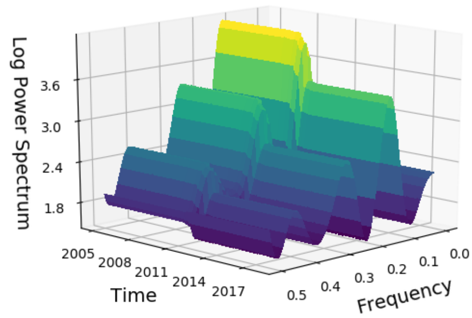
Central Business District (CBD) There is one change point detected for the CBD. The change point is identified in August 2011. This change point is more than two and a half years prior to the implementation of the lockout laws. Our analysis suggests that the lockout laws had no impact on NDAs in the CBD. There are two key insights that drive us to this conclusion. First, there is no change point detected at the time of the lockout laws or after their implementation. Second, the power spectrum for both segments in the CBD's time-varying spectrum (January 2005 - August 2011 and August 2011 - December 2017) are almost identical, with a minor reduction in total variance when transitioning from segment one to segment two. The existence and consistency of marked weekly, bi-weekly and every second day periodicities (frequencies ~ 0.14, 0.33, 0.45) between segments highlights that there has been limited change in the nature of NDAs over time and no evidence to suggest that the lockout laws have impacted NDA in the CBD. The structural break identified in 2011 may be related to the global decline in NDAs across New South Wales, which were in the midst of their decline during this period.

Proximal Displacement Areas (PDAs) There is no change point detected for PDAs. The spectrum exhibits pronounced weekly, bi-weekly and every second day periodicities. Given that there is no change point detected throughout the entirety of the time period analysed, there is no evidence to suggest that the lockout laws had any influence on PDAs.

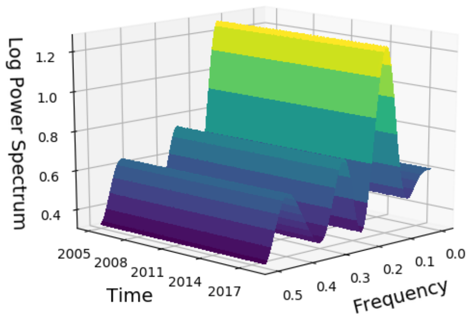
Distal Displacement Areas (DDAs) There is one change point identified for DDAs. The change point is identified in April 2008. The two power spectra in segment one (January 2005 - April 2008) and segment two (April 2008 - December 2017) suggest that the periodicities of NDAs in DDAs were highly similar between the two segments, with the primary points of difference being a reduction in total NDA variance and a slightly less pronounced weekly periodicity (frequency ~ 0.14) in



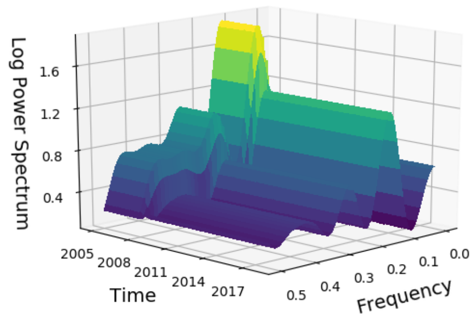
(a) Kings Cross



(b) CBD



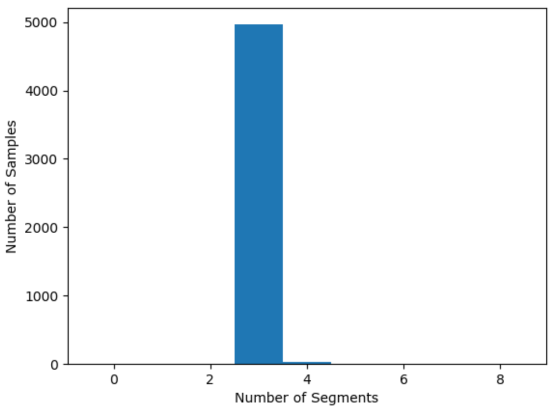
(c) PDA



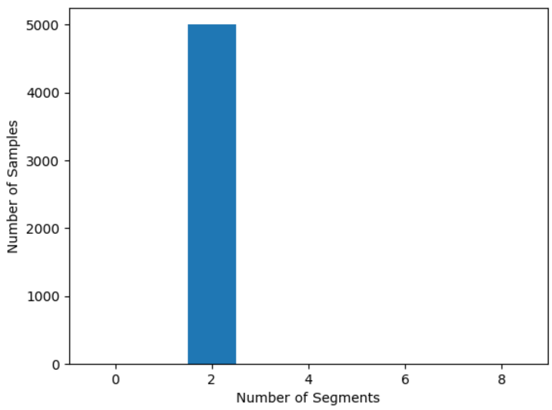
(d) DDA

Figure 10: Time Varying log spectrum for each area.

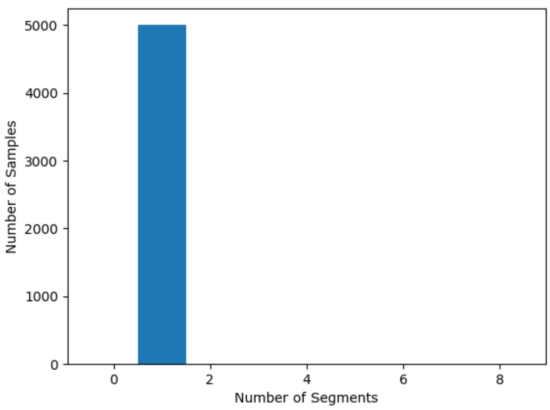
NDAs. The drop in NDAs in DDAs are most likely related to the global decline in NDAs across New South Wales which reached their peak in 2008. However, given that there is no change point detected at the point of the lockout laws or after their implementation, and there is no change in the most significant frequency components when transititioning from segment one to segment two, there is no evidence to suggest that the lockout laws had any impact on NDAs in DDAs.



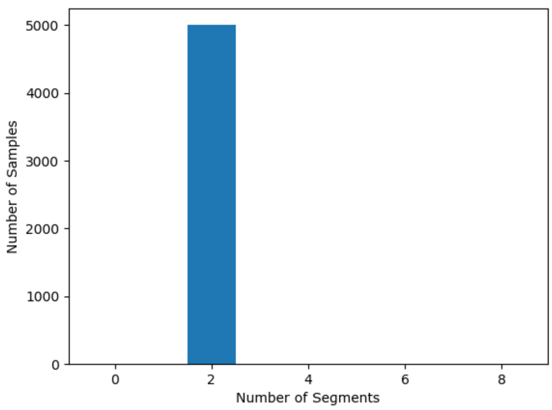
(a) KC posterior number of segments



(b) CBD posterior number of segments

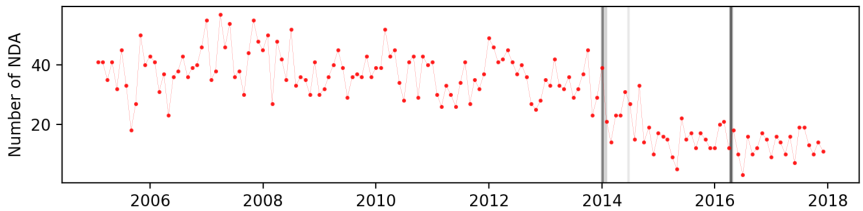


(c) PDA posterior number of segments

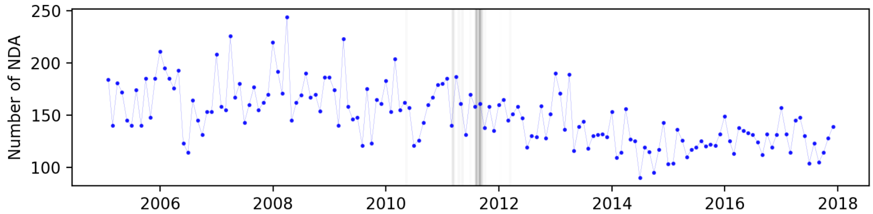


(d) DDA posterior number of segments

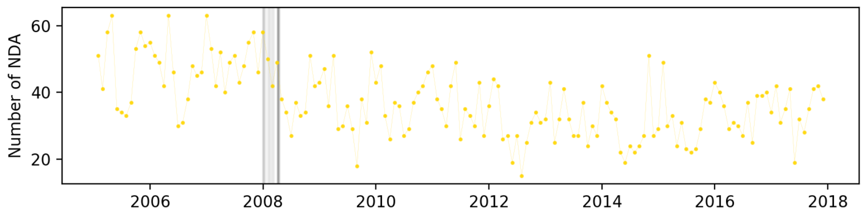
Figure 11: Posterior distribution over number of segments



(a) KC cutpoint uncertainty

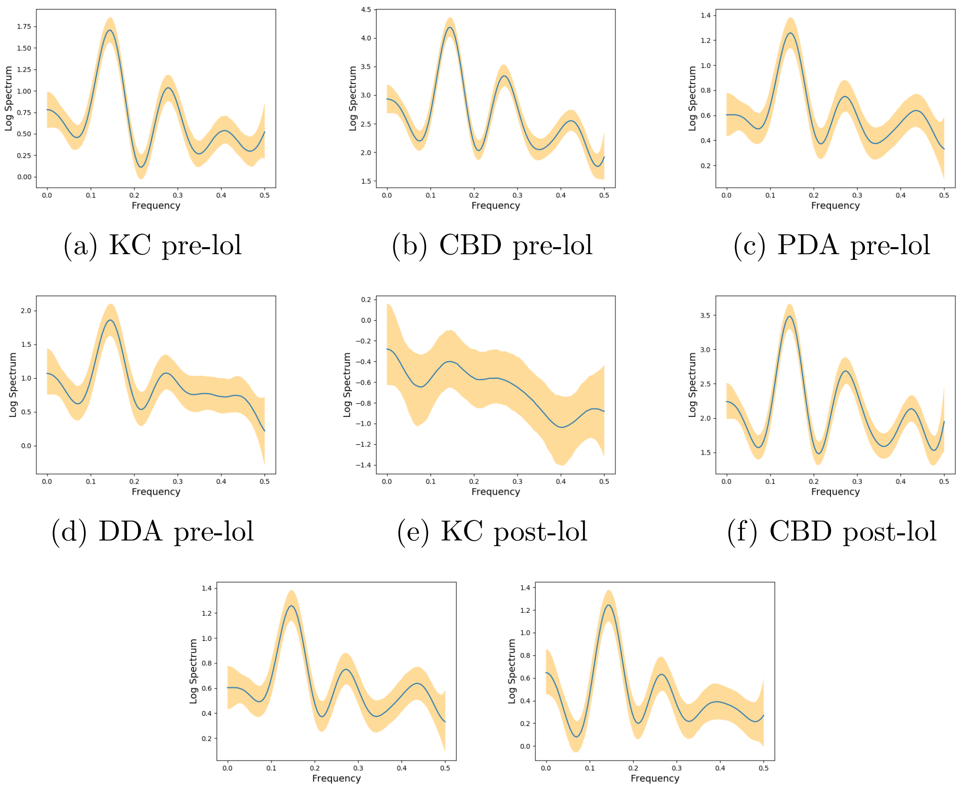


(b) CBD cutpoint uncertainty



(c) DDA cutpoint uncertainty

Figure 12: Cutpoint uncertainty



(a) KC pre-lol (b) CBD pre-lol (c) PDA pre-lol

(d) DDA pre-lol (e) KC post-lol (f) CBD post-lol

(g) PDA post-lol (h) DDA post-lol

Figure 13: Log spectrum and 95 % confidence intervals before and after lockout laws' enactment