University of Sydney Policy Reform Project

Research Paper for NSW Council of Social Service: Summarising data on gambling, homelessness and economic disadvantage in NSW

Authors: Avish Sharma, Aladdin Ko, Jesse Gray, Longye Tian and Claudio Campi.

Adviser: Dr Anna Boucher, Associate Professor.

Contact: Mr James Hall, Coordinator, University of Sydney Policy Reform Project, <james.l.hall@sydney.edu.au>.

Table of Contents

Executive Summary

This paper summarises available data on the number of electronic gaming machines (EGMs), levels of homelessness, and levels of economic disadvantage in Local Government Areas (LGAs) of NSW.

Initially, summarising the available data proved difficult. This is because some LGAs have insufficient data on the number of EGMs and levels of homelessness. In trying to address this issue, we synthesised and rationalised available datasets (such as those provided by NSW Liquor and Gaming and the Australian Bureau of Statistics (ABS)) to ensure comparability. Such alterations are described in the Methodology section of the paper.

The section labelled 'Descriptive Statistics' outlines the statistics requested. The LGAs with the highest number of EGMs and their respective levels of homelessness are:

- Canterbury-Bankstown, with 5067 EGMs and 2582 homeless people.
- Central Coast, with 4634 EGMs and 1031 homeless people.
- Fairfield, with 3857 EGMs and 2226 homeless people
- Sydney, with 3729 EGMs and 5061 homeless people
- Newcastle, with 3015 EGMs and 797 homeless people.

The LGAs with the lowest number of EGMs and their respective levels of homelessness are:

- Kyogle, with 81 EGMs and 34 homeless people.
- Cabonne with 104 EGMs and 24 homeless people
- Warrumbungle with 141 EGMs and 15 homeless people
- Greater Hume & Lockhart, with 148 EGMs and 24 homeless people
- Inverell, with 170 EGMs and 42 homeless people

There is strong correlation (0.75) between the number of EGMs and the level of homelessness in LGAs. However, once population size of all LGAs is accounted for, there is no longer a strong correlation, and EGMs are only weakly correlated with most socio-economic variables. Hence, explanatory factors other than EGMs appear to be driving

homelessness. Interestingly, there is a negative correlation between EGMs and the SEIFA [Socio-economic Index for Areas], an aggregate of the measures of advantage and disadvantage; indicating that EGMs are less likely to be present in communities that are more economically advantaged. Furthermore, EGMs could still be part of the causal pathway towards homelessness – qualitative data may be better equipped to explore this second point. Detailed statistics are provided in the 'Descriptive Statistics' section.

Note that when wanting to capture the relationship between EGMs and homelessness/other economic indicators, such correlation calculations may actually not be very informative. This is because correlation does not account for additional factors that may influence the number of EGMs and levels of homelessness/disadvantage in an LGA. For this reason, this paper includes some basic regression analysis in the section labelled 'Regression'. The difference between regression and correlation is that regression controls for other factors that may influence levels of homelessness (such as percentage of people unemployed, income levels, gender etc), and hence captures the specific effect the number of EGMs has on levels of homelessness/disadvantage. The regression found that on average, an additional electronic gaming machine in an LGA is expected to increase the number of homeless people in that LGA by 0.64 people, holding all else equal.

Suggestions for future data collection:

Note that these regression results are not predictive and are not supposed to be used for inference. In order to accurately capture the effect of EGMs on homelessness, the regression would need to include many more control variables. Rather, this paper includes these regressions to serve as a guide as to what future research and analysis should focus on. To this extent, the paper makes the following recommendations:

1. **Common Reporting Standards** – the NSW Government and the Federal Government should use common reporting and statistical standards between key agencies to make it easier to collate data, draw insights and prevent possible distortion of data.

2. **More Aggregate Research** – the NSW Department of Communities and Justice should conduct more aggregate studies on the relationship between gambling, homelessness and economic disadvantage at the LGA level. A greater body of research will prevent model misspecifications and spurious correlations from being advanced; enabling the interplay of gambling, homelessness and economic disadvantage to be studied more accurately.

3. **More Data** – the NSW Department of Communities and Justice should collect more data on the homelessness population. This will prevent gaps from being present in LGA-level data and allow homelessness to be studied more precisely.

4. **More Qualitative Research** – more qualitative studies are needed to analyse the relationship between gambling and homelessness. Although our analysis showed that the number of EGMs had a negligible effect on homelessness once other measures of socio-economic disadvantage are controlled for, it is very much likely that EGMs could be part of the causal pathways that lead people to homelessness and this would be best analysed through qualitative rather than quantitative means.

1. Background Information (Review of Literature)

The regulation of gambling policies in Australia has been historically considered as a State prerogative, as each government can decide how to implement such policies and manage the revenue resulting from such activities. However, due to growing concerns on the adverse social effects of gambling and the development of new online and interactive technologies, the Commonwealth has shown increased interest in the matter. For this reason, the early 2000s saw several bills and legislations aimed at curbing the spread of online interactive gambling, such as the *Interactive Gambling Act 2001* (Cth) (Australasian Gaming Council 2017). Regulation around gambling has been further tightened with the introduction of ad hoc bodies, such as Liquor and Gambling NSW in the latter state, and an Independent Liquor and Gambling Authority (ILGA).

Gambling is recognized as a concerning public health and social issue in Australia. For example, Liquor and Gaming NSW reports that 53% of adults has gambled in the past 12 months, while 7.2% are considered to be at moderate-risk and 1% are classified as problem gamblers (Liquor and Gaming NSW).

Problem gambling is defined as the difficulty in limiting money and time resources devoted on gambling activities, and it is usually conceived of as a spectrum with varying level of intensity, meaning that it is harder to unambiguously determine its severity (AGC 2016). In fact, individuals may shift from healthy gambling to problem gambling and vice versa; moreover, people affected by problem gambling exhibit a wide variety of characteristics. Some frameworks have been introduced to better define gambling problems, such as the Victorian Gambling Screen (VGS) and the Problem Gambling Severity Index (PGSI), which ask targeted questions related to behaviours and beliefs of the individuals tested. Usually, respondents will report how often they engage in selected activities, and a severity score will be assigned accordingly. It is difficult to compare surveys that try to assess the exact rates of problem gambling, as methodologies and timeframes vary; nonetheless, it is recognized that problem gambling might affect from 1% up to 3% of the adult population in Australia (AGC 2016). Moreover, while gambling issues can affect individuals from any background and socioeconomic status, single unemployed young males have been shown to be the most vulnerable demographic. A report specific to the NSW context estimates that 0.8% of adults are problem gamblers, with a greater incidence for males and Aboriginals and Torres Strait Islanders (Sproston et al. 2012).

Several forms of gambling are allowed in Australia, from the aforementioned Interactive Gaming to Racing, Lotteries and Electronic Game Machines. The latter are of particular concern, as reports (AGC 2016) show how EGMs are the primary gambling activity of those individuals who suffer from problem gambling. Throughout Australia, EGMs are only available at licensed venues, which are usually clubs or casinos, and their type is highly regulated by each state or territory. All jurisdictions determine a maximum of EGMs allowed on their territory; for example, NSW caps them at 99,000 with 1,500 at The Star casino.

Gambling represents a prime concern due to its adverse financial effects on affected individuals, placing them at greater risk of hardship, bankruptcy and homelessness. Moreover, populations that are affected by higher rates of problem gambling are also more likely to report other concurring issues that are likely to exacerbate financial struggle. As reported by Nower et al. (2014), several studies have shown that problem gamblers are often subject to personality, mood and other psychiatric disorders. These rates are also more prevalent among the homeless population, which suggests patterns of co-morbidity and underlying health characteristics that make individuals more likely to be both compulsive gamblers and homeless. Moreover, each of the two populations has shown higher than average rates for the other issue, meaning that homeless people are more likely to be problematic gamblers and vice versa. Holdsworth and Tiyce (2012) recognize that homelessness is a multidimensional and complex issue which is determined by a plethora of inter-connected underlying factors, some structural like unemployment, poverty, social exclusion, and other psychological like childhood trauma, mental illness, domestic violence. Nonetheless, most of these are also associated with problem gamblers who often voice despair and powerlessness when coping with such ‘complex needs’ that are difficult to tackle in isolation. It must also be noted that these factors tend to compound on each other and over time, raising the barriers to resolution (Holdworth and Tiyce 2012; Matheson et al 2014).

When looking at the specific link between homelessness and gambling, it is hard to pinpoint a specific pattern of causation due to the highlighted complexity of these issues; nonetheless, such relationship has been often analysed and recognized as crucial. For example, Sharman et al. (2015) starts from the definition of problem gambling as an Impulse Control Disorder to show how individuals facing financial struggle will engage in risky behaviour to escape poverty. Therefore, the homeless are more likely to decide to gamble in a desperate attempt to regain economic stability. Holdsworth and Tiyce (2012), on the other hand, specifically interviewed homeless individuals and enquired about their gambling behaviours, highlighting additional reasons that correlate these two realities. For instance, gambling could be a way to temporarily escape worries associated with homelessness, providing limited psychological relief and a sense of belonging. On the other direction of this relationship, gambling might push otherwise stable individuals into homelessness on channels other than the obvious erosion of disposable income. For example, addiction to gambling might decrease trust within a household, placing pressure on a marital relationship and eliminating mutual support, which might eventually cause one partner to move out.

2. Summary of Methodology

Combining data from varying sources provided many challenges in obtaining the statistical measures and insights that are detailed throughout the rest of the report. Although a full methodology covering the data organisation and cleaning process can be found in the appendix to this report, it should be noted that where data was unavailable or could not be neatly organised together, groups of LGAs, rather than individual LGAs became the smallest unit of analysis. This point should be kept in mind when reading and extrapolating any insights presented in this report. Other challenges in managing data are further detailed in the appendix.

3. Descriptive Statistics

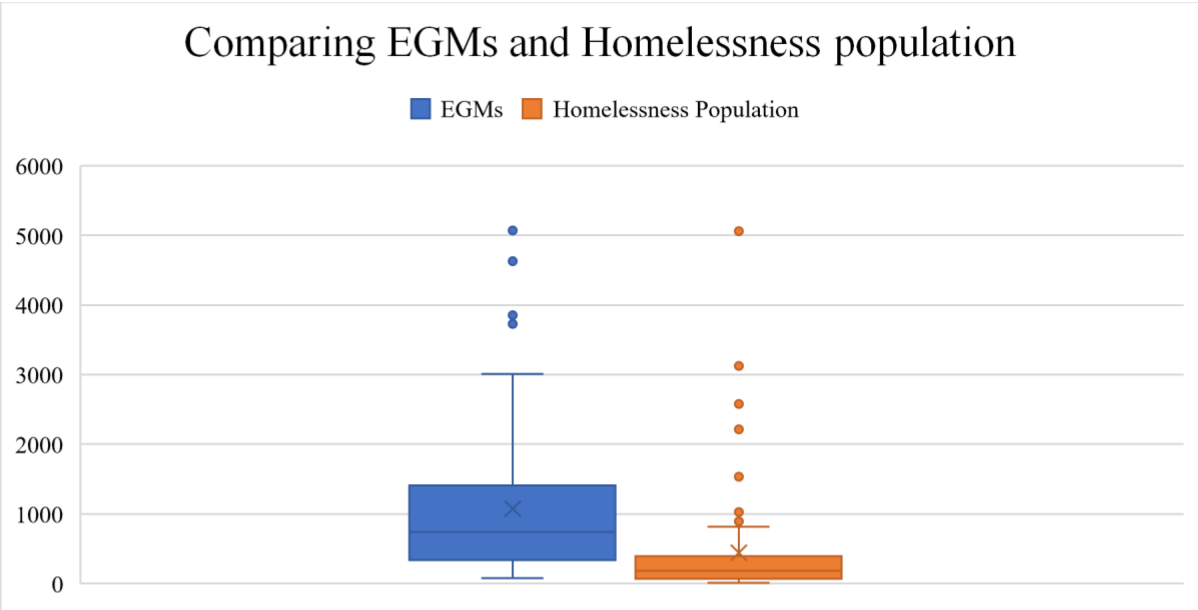
Based on 2016 data, Table 1 summarises various descriptive statistics relating to homelessness, EGMs and economic disadvantage, across LGAs in NSW.

Table 1: Descriptive statistics for homeless population, homelessness rate (per 10,000 people), number of EGMs, and EGMs ate (per 10,000 people), and proportion of economically disadvantaged residents.

Descriptive Statistics	Homeless Population	Homelessness Rate (per 10,000 people)	EGMs	EGMs Rate (per 10,000 people)	Proportion of Economically Disadvantaged Residents (%)
Mean	443.29	39.42	1075.53	141.32	13.66
Median	185	28.89	739	126.20	13.79
Standard Deviation	769.32	33.49	1038.17	99.83	3.73
Sample Variance	591858.31	1121.64	1077796.44	9965.65	13.91
Kurtosis	16.99	12.36	3.46	30.95	0.25
Skewness	3.72	3.05	1.83	5.00	0.33
Range	5052	221.21	4986	793.10	18.55
Minimum	9	6.03	81	43.32	6.76
Maximum	5061	227.24	5067	836.42	25.30
Sum	37680	N/A	91420	N/A	N/A
Count	85	85	85	85	85

According to Figure 1, on average, the number of electronic gaming machines is higher than that of the homelessness population in each LGAs. The box for EGMs is taller than that of the homeless population, which illustrates that the variability of number of EGMs is larger across different LGAs. On the other hand, these two variables share one common feature: both of them have extreme outliers which are data points 1.5 times IQR from the mean.

Figure 1: Boxplot comparing EGMs and Homelessness population



A similar conclusion applies when comparing EGM and Homelessness rate (see Figure 2).

Figure 2: Boxplot comparing EGM and Homelessness Rate

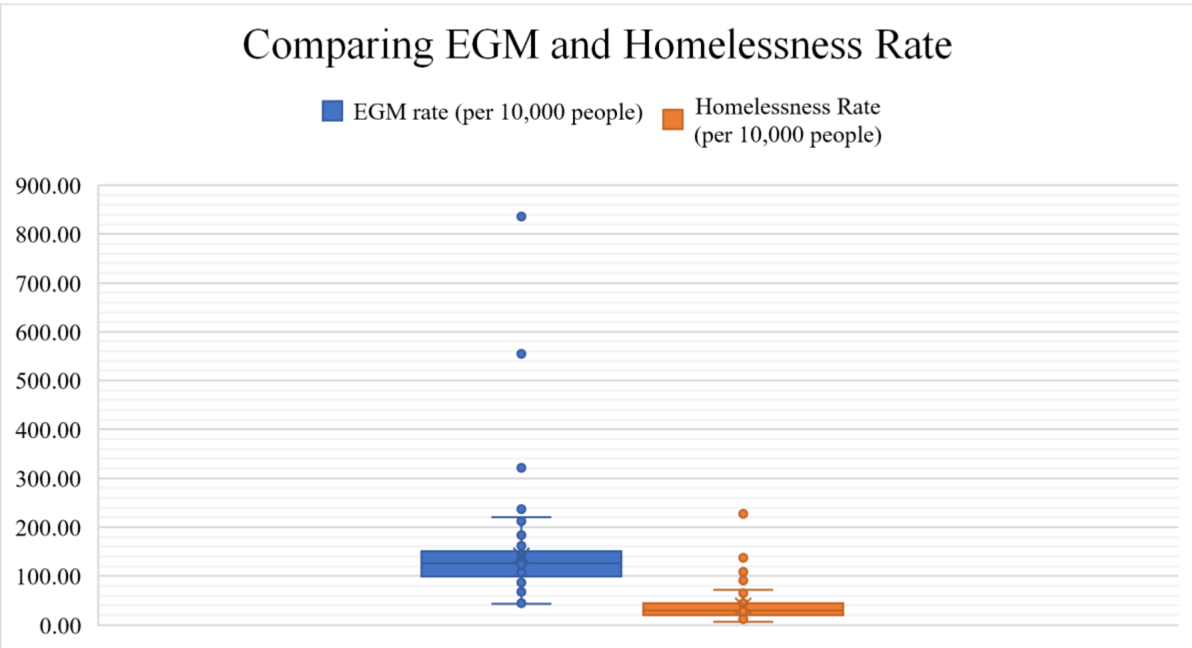


Table 2 shows the descriptive statistics for the proportion of residents in LGAs classified as economically disadvantaged, by various gender and age demographics, across NSW

Table 2: Descriptive statistics for economic disadvantage by gender and age demographics – men, women, children, young people, and older people (all units are percentages).

Descriptive Statistics	Men	Women	Children	Young People	Older people
Mean	11.91	13.38	17.81	12.77	12.19
Median	11.85	13.26	17.79	13.29	12.00
Standard Deviation	3.37	3.49	6.11	4.47	2.93
Sample Variance	11.39	12.17	37.37	20.01	8.58
Kurtosis	0.35	1.28	-0.44	2.51	4.17
Skewness	0.34	0.55	0.31	0.57	1.11
Range	17.23	19.02	27.84	29.15	18.37
Minimum	5.60	7.21	5.52	2.10	6.29
Maximum	22.83	26.23	33.36	31.25	24.66
Count	85	85	85	85	85

Table 3 shows the descriptive statistics for the proportion of residents in LGAs classified as economically disadvantaged, by various age and workforce demographics, across NSW.

Table 3: Descriptive statistics for economic disadvantage by age and employment demographics – people of working age, people employed full time, people employed part time, unemployed people, young people not in labour force, older people not in labour force (all units are percentages).

Descriptive Statistics	People of working age	People employed full time	People employed part time	Unemployed people	Young people not in labour force	Older people not in labour force
Mean	13.00	4.27	6.90	35.03	30.81	13.53
Median	12.48	4.30	6.60	34.31	32.14	13.59
Standard Deviation	3.99	2.00	2.33	9.48	7.40	3.56
Sample Variance	15.96	3.99	5.45	89.94	54.77	12.66
Kurtosis	-0.20	10.61	0.06	4.49	0.71	1.87
Skewness	0.46	2.05	0.65	1.09	0.19	0.72
Range	17.42	14.29	11.77	65.88	39.28	19.95
Minimum	6.19	1.22	2.10	12.50	15.40	6.08
Maximum	23.60	15.50	13.87	78.38	54.68	26.03
Count	85	85	85	85	85	85

The statistics in Table 2 and Table 3 are demonstrated more intuitively in the boxplot Figure 3. It is clear from Figure 3 that young people not in the labour force and people who are unemployed are more likely to be economically disadvantaged across LGAs, on average.

Figure 3: Economic disadvantage by various gender, age and employment demographics

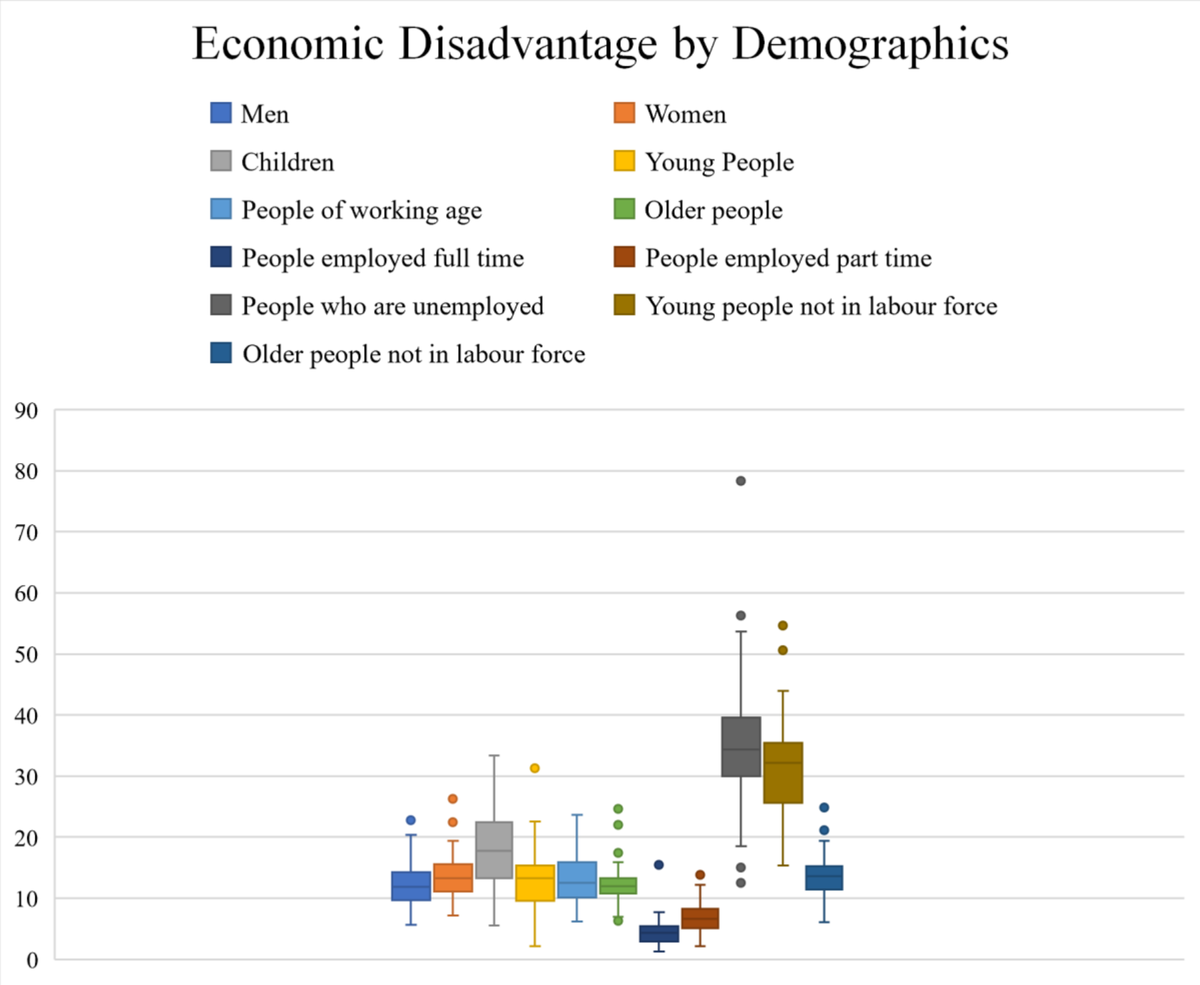


Figure 4 (see next page) describes the relationship between the incidence of EGMs, levels of homelessness and the extent and of each separate demographics of economic disadvantage across NSW Wales through a Pearson correlation coefficient, which measures the statistical relationship between two variables of interest. These data control for population size. A correlation coefficient of 1 implies strong positive relationship while a correlation of -1 implies strong negative relationship. Coefficients closer to 0 tend to signify no relationship between the variables of interest. It must be emphasized that no causal relationship can be inferred from the correlation coefficients.

Figure 4: Correlating the presence of EGMs, homelessness rate, and select socio-economic indicators – insolvencies, domestic violence, household income, labour force, Aboriginality, unemployment rate, SEIF Index (percentile) and economic disadvantage

	EGMRat..	Homele~e	Insolv~s	Domest~e	Househ~e	Labour~e	Aborig~l	Unempl~e	SEIFAI~e	Economic Disadva..
EGM Rate(per 10,000 p..	1.000									
Homelessness Rate(per..	-0.042	1.000								
Insolvencies	-0.025	0.057	1.000							
Domestic Violence	-0.029	0.068	0.999	1.000						
Household Income	-0.033	0.051	0.993	0.992	1.000					
Labour Force	-0.252	0.346	0.032	0.055	0.021	1.000				
Aboriginal	0.151	-0.193	-0.020	-0.023	0.007	-0.508	1.000			
Unemployment Rate	-0.106	0.252	-0.007	0.002	-0.044	0.023	0.244	1.000		
SEIFA Index (Percentile)	-0.288	0.117	-0.012	-0.009	0.001	0.540	-0.667	-0.584	1.000	
Economic Disadvantage..	0.003	0.282	0.011	0.016	-0.027	-0.166	0.282	0.813	-0.641	1.000

Sum of Correlation Coefficients (color) broken down by Variables vs. Variables_.

Correlation Coefficients

-0.667 1.000

The correlation coefficients show that there is very weak negative correlation between the number of EGMs and level of homelessness when measured per 10,000 people. While there is a quite strong correlation between the number of EGMs and level of homelessness (0.75), this is not evident once population size is controlled for (-0.042). This correlation coefficient indicates that there is little to no relationship between the number of EGMs and the number of homeless in an LGA once the population size is taken into consideration. Moreover, EGMs are weakly correlated with most socio-economic indicators (see the first column of Figure 4). An aggregate of these indicators, the SEIFA index, shows slightly stronger correlation. This measures the relative advantage and disadvantage of LGAs through an index. The negative correlation coefficient between LGAs and EGMs indicates that in less disadvantaged areas, there are fewer EGMs.

The 10 LGAs with highest numbers of electronic gaming machines (from highest to lowest) are Canterbury-Bankstown, Central Coast, Fairfield, Sydney, Newcastle, Wollongong, Blacktown, Cumberland, Penrith and Lake Macquarie. The 10 LGAs with lowest numbers of electronic gaming machines (from lowest to highest) are Kyogle, Cabonne, Warrumbungle, Greater Hume with Lockhart, Inverell, Glen Innes Severn, Upper Hunter, Bland with Narrandera, Narrabri and Snowy Valleys.

The graphs below illustrate the numbers of EGMs and homeless people in the 10 LGAs with the highest number of EGMs (see Figure 5), and in the 10 LGAs with the lowest number of EGMs (see Figure 6).

Figure 5: Number of homeless people in the 10 LGAs with the highest numbers of EGMs

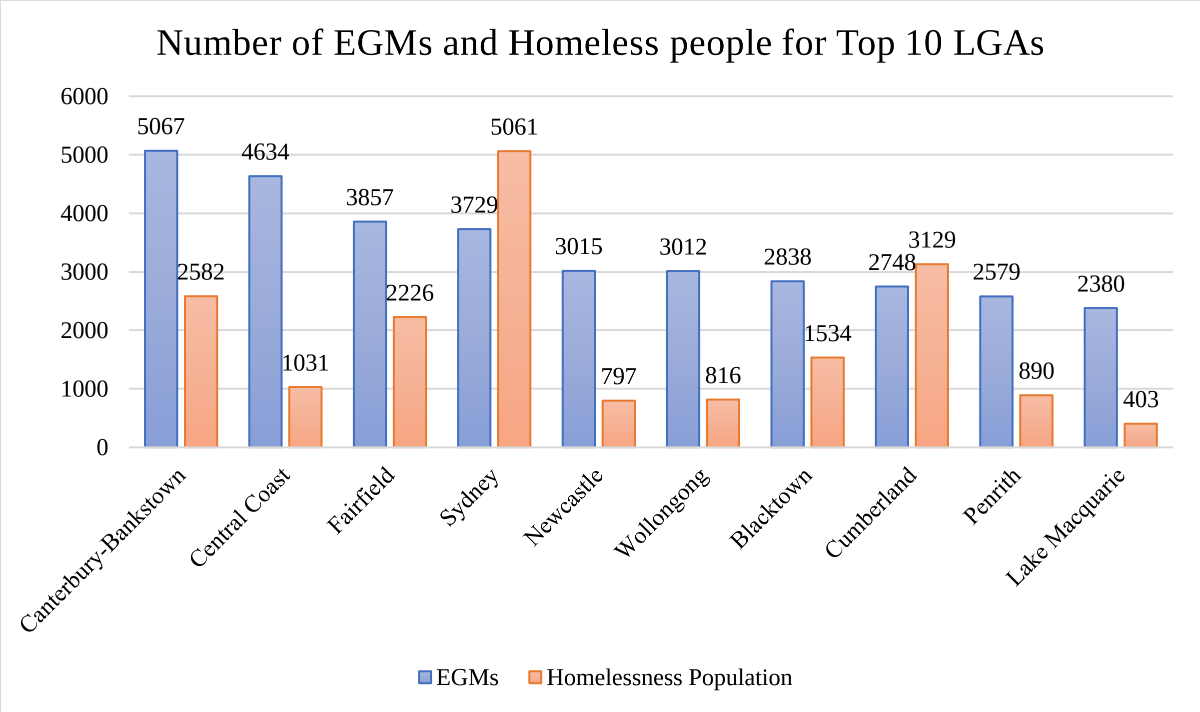
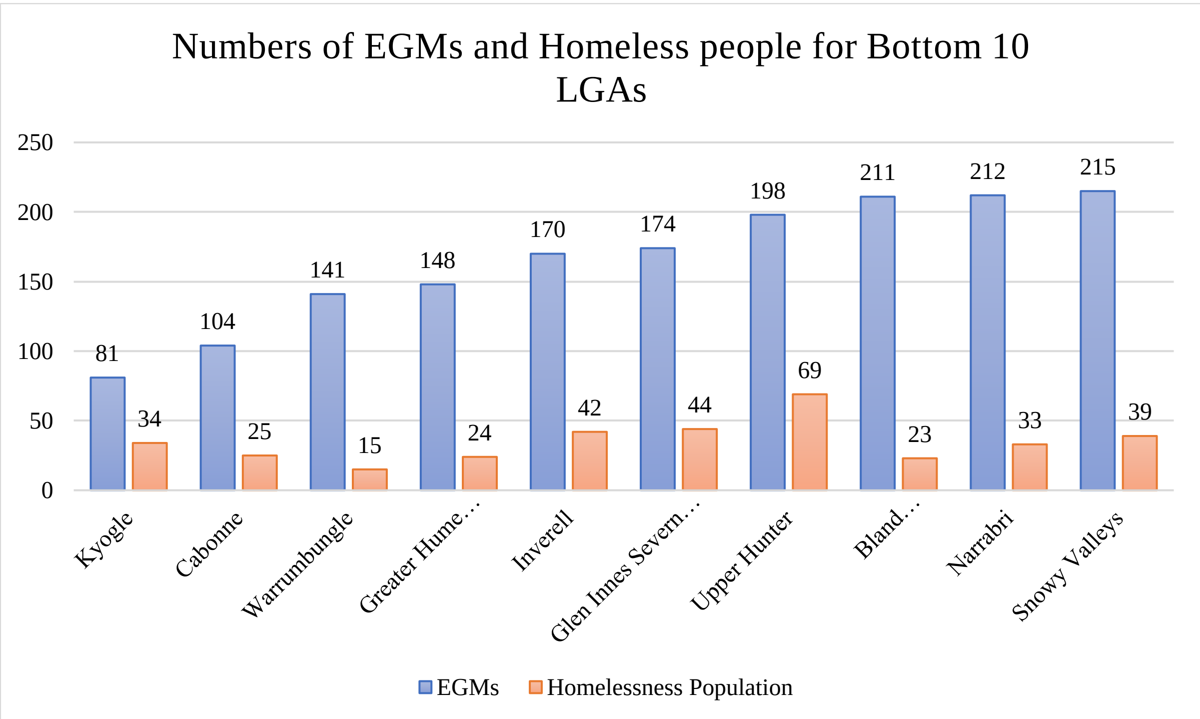


Figure 6: Number of homeless people in the 10 LGAs with lowest numbers of EGMs



4. Regression Analysis

The multiple regression model for homelessness is shown below:

$$HOMELESS = \gamma_0 + \gamma_1 POP + \gamma_2 DV + \gamma_3 AB + \gamma_4 EGM + \gamma_5 INSO + \gamma_6 HouseINC + \gamma_7 LABOUR + \gamma_8 UNEMP + \sigma$$

Where,

- HOMELESS – Number of homelessness population in LGA
- POP – Population in the LGA;
- DV – Number of recorded cases of domestic violence in LGA;
- AB – The percentage of aboriginals in LGA;
- EGM – Number of electronic gaming machines in LGA;
- INSO – Number of personal insolvencies in LGA;
- HouseINC – Average household income in LGA;
- LABOUR – Number of labour force in LGA;
- UNEMP – Unemployment rate in LGA.

We have included 8 explanatory variables in our model to demonstrate the variability of homelessness across different LGAs. More specifically, population and percentage of aboriginals' variables are chosen to ensure that our further analysis are controlled for population and racial diversities. In addition, to control the economic variability and financial status, we introduce number of personal insolvencies, average household income, labour force size and unemployment rate in each LGA as control variables. Moreover, according from the Australian Human Rights Commission, domestic violence is one of the major causes of homelessness in Australia. Hence, using number of recorded cases of domestic violence in each LGA as a regressor, we ensure the effect of domestic violence on homelessness in controlled.

The variable of interest is the number of electronic gaming machines, and we will conduct further analysis on its coefficient and standard errors. All the demographic statistics are collected from the ABS (ABS 2018). Data of electronic gaming machines is from Liquor and Gaming NSW (Liquor and Gaming NSW 2020). Data for number of domestic violence is gathered in NSW Bureau of Crime Statistics and Research (BOCSAR 2020).

(a) Estimation

The estimation of Model 2 is summarised in Table 4 below.

Table 4: Estimation of Model 2

	Coefficient	**Standard error**	**Robust Standard Error**
Intercept	-757.1959	241.4188	237.5264
POP	-0.0212	0.0051	0.0113
DV	1.1697	0.2993	0.4376
AB	-28.5432	14.1106	12.4387
EGM	0.635	0.1003	0.1742
INSO	-5.4395	1.4174	2.3753
HouseINC	0.0115	0.0984	0.0794
LABOUR	0.0414	0.0086	0.0189
UNEMP	127.0466	33.8152	37.7383

According to the estimated model, it demonstrates that, on average, an additional electronic gaming machine is expected to increase the number of homeless people in that LGA by 0.64 people, holding all else equal. Another model using SEIFA index to replace economic variables is estimated and shows similar results (See more in Appendix). However, this model has a better fit than the other and enables further analysis for specific joint significance tests if needed.

(b) Test for Statistical significance of EGM

The variable of interest is EGM; hence, we conduct a hypothesis test for its statistical significance.

$$H_0: \gamma_4 = 0$$
$$H_1: \gamma_4 \neq 0$$

The test statistic for EGM's coefficient is 3.64 using the heteroskedasticity-robust errors (see section III). We will reject the null hypothesis in favour of alternative hypothesis if the test statistic is below the critical value. And the critical value, for the t distribution with 79 degrees of freedom at 1% significance level, is 2.64. Since $|t|>c$, we reject the null in favour of the alternative at 1% significance level. In conclusion, the number of electronic gaming machines has a statistically significant effect at the 1% level on homelessness in LGAs, ceteris paribus. Moreover, since we use the heteroskedasticity-robust standard errors for testing, we ensure the coefficient is unbiased and consistent.

(c) Check for Heteroskedasticity

After conducting Breusch-Pagan test, we reject the null hypothesis and conclude that there is heteroskedasticity in our data. Therefore, we choose to use the Huber-White standard errors to surmount this issue.

5. Recommendations

After conducting this statistical exercise, this paper recommends the following to ensure a clearer analysis between homelessness, gambling and economic disadvantage can be established. The key recommendations of this paper are as follows:

1. **Common Reporting Standards** – the NSW Government and the Federal Government should use common reporting and statistical standards between key agencies to make it easier to collate data, draw insights and prevent possible distortion of data.

2. **More Aggregate Research –** the NSW Department of Communities and Justice should conduct more aggregate studies between the relationship between gambling, homelessness and economic disadvantage at the LGA level. A greater body of research will prevent model misspecifications and spurious correlations from being advanced; enabling the interplay of gambling, homelessness and economic disadvantage to be studied more accurately.

3. **More Data –**the NSW Department of Communities and Justice should collect more data on the homelessness population. This will prevent gaps from being present in LGA-level data and allow homelessness to be studied more precisely.

4. **More Qualitative Research –** more qualitative studies are needed to analyse the relationship between gambling and homelessness. Although our analysis showed that the number of EGMs had a negligible effect on homelessness once other measures of socio-economic disadvantage are controlled for, it is very much likely that EGMs could be part of the causal pathways that lead people to homelessness and this would be best analysed through qualitative rather than quantitative means.

References

Australasian Gaming Council (2016), A Guide to Australasia's Gambling Industries Chapter Ten: Problem Gambling Prevalence. Available at: <https://www.austgamingcouncil.org.au/system/files/AGCPublications/AGC_DB_10_F2017_0.pdf>

Australasian Gaming Council (2017), A Guide to Australasia's Gambling Industries Chapter One: The Australian Gambling Environment. Available at: <https://www.austgamingcouncil.org.au/system/files/AGCPublications/AGCDB_1_2016-17F.pdf>

Australian Bureau of Statistics, 2018, Economy and Industry, LGAs 2011-2017, cat. no. 1014.0, viewed 20 April 2020, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02012-17?OpenDocument>

Australian Bureau of Statistics, 2018, Education and Employment, LGAs 2011-2017, cat. no. 1014.0, viewed 20 April 2020, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02012-17?OpenDocument>

Australian Bureau of Statistics, 2018, Income, LGAs 2011-2017, cat. no. 1014.0, viewed 20 April 2020, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02012-17?OpenDocument>

Australian Bureau of Statistics, 2018, Population and people, LGAs 2011-2017, cat. no. 1014.0, viewed 20 April 2020, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02012-17?OpenDocument>

Holdsworth, L. and Tiyce, M. (2012), Untangling the Complex Needs of People Experiencing Gambling Problems and Homelessness, International Journal of Mental Health Addiction, Vol.11, pp.186-198

Liquor and Gaming NSW, 2020, Gaming Machine Data, viewed 20 April 2020, <https://www.liquorandgaming.nsw.gov.au/resources/gaming-machine-data>.

Matheson, F. et al. (2014), Prevalence of Gambling Problems Among the Clients of a Toronto Homeless Shelter, Journal of Gambling Studies, Vol.30, pp.537-546

Nower, L. et al. (2014), Problem Gambling and Homelessness: Results from an Epidemiologic Study, Journal of Gambling Studies, Vol.31, pp.533-545

NSW Bureau of Crime Statistics and Research, 2020, Domestic violence data for NSW, viewed 20 April 2020, < https://www.bocsar.nsw.gov.au/Pages/bocsar_pages/Domestic-Violence.aspx>.

Sharman, S. et al. (2015), Rates of Problematic Gambling in a British Homeless Sample: A Preliminary Study, Journal of Gambling Studies, Vol.31, pp.525-532

Sproston, K. et al. (2012), Prevalence of Gambling and Problem Gambling in New South Wales, NSW Office of Liquor, Gaming and Racing

Appendix

I. Detailed Methodology

Understanding the data cleaning process is an important step in making sure the interpretations from the analysis of data is valid and that the limits to such analysis are known. Appropriately organising multiple datasets is an imperative procedure in any project and if not done well, can lead to poor analysis and the drawing of incorrect conclusions. Throughout the project, there was several steps taken to alter individual datasets, to ensure data across gambling, homelessness, economic disadvantage and other control variables were consistent. Without this consistency, it would not have been possible to compare descriptive statistics, make correlations or construct regression models, as detailed in this report. This section of the report highlights some noteworthy data cleansing and organisation procedures that were used.

(a) Grouping LGAs

Although NCOSS asked that data from individual LGAs (Local Government Areas) in New South Wales were to be analysed, this was impossible to achieve as certain datasets had incomplete information on some LGAs. At best, if there was no or low information on a certain LGA, this LGA was simply grouped together with another LGA(s). This is what occurred with the data on the number of Electronic Gaming Machines (EGMs), by LGA provided by NSW Liquor and Gaming. In this case, if an LGA had less than 5 hotels/clubs operating in it, data was merged with a neighbouring LGA to ensure the privacy of earnings for individual premises. However, as there were separate datasets for EGMs within clubs and another for EGMs in hotels, the grouping of LGAs was seemingly arbitrary and inconsistent between the two datasets. To avoid this problem, we aggregate groups of LGAs together to report total numbers. Figure 7 provides an illustration of this process.

Figure 7: LGA Grouping Process

Clubs	
LGA	No. of EGMs
A B	X
C	Y

Hotels	
LGA	No. of EGMs
A	W
B C	Z

Total	
LGA	No. of EGMs
A B C	X+Y+W+Z

This created a specific grouping of LGAs and because the data for other variables could be sourced at the individual LGA level, these groups of LGAs became the smallest unit of analysis available. As such, data for other variables had to be grouped in this specific manner, in order to ensure consistency and allow for correlations and regressions to be made.

(b) Accounting for Population

The above method no longer makes it viable to simply compare minimum and maximum values as it would be unfair to compare individual LGAs against groups of LGAs, which themselves would have varying numbers of LGAs within them. Thus, when the data is in absolute value terms, such as the number of EGMs or the homelessness population is considered, comparisons can be made once population sizes are controlled for. This makes logical sense because more populous LGAs or groups of LGAs are likely to have more gambling premises and hence more EGMs. Likewise, more populous LGAs or groups of LGAs are likely to have more homeless individuals.

To address this, we construct the following statistics to control for population size:

- EGM Rate – Number of EGMs in an LGA (or group of LGAs) per 10,000 people. It is obtained by dividing the total number of EGMs in an LGA (or group of LGAs), by the LGA's (or group of LGAs') total population and then multiplying by 10,000.
- Homelessness Rate – Homelessness population in an LGA (or group of LGAs) per 10,000 people. It is obtained by dividing the total homelessness population in an LGA (or group of LGAs), by the LGA's (or group of LGAs') total population and then multiplying by 10,000.

Using these measures, one can then compare where the intensity of EGMs and homelessness is the greatest, across groups of LGAs.

If the statistic at hand is not measured in absolute value terms, but rather is a proportion or must exist within a fixed range of values, a statistic for a group of LGAs was calculated by weighting the observation for each individual LGA by its population and then taking a weighted average as follows:

$$Weighted\ Average = \frac{\sum_{i=1}^{n} Stat_i \times Population_i}{\sum_{i=1}^{n} Population_i}$$

$$where\ i = 1, 2, \ldots n, with\ each\ i\ representing\ data\ from\ an\ LGA,$$

$$n = the\ number\ of\ LGAs\ in\ the\ grouping,$$

$$Stat_i = Statstic\ for\ the\ LGA,$$

$$Population_i = Population\ for\ the\ LGA$$

$$and\ \sum_{i=1}^{n} Population_i = Sum\ of\ the\ Population\ across\ the\ LGAs\ in\ that\ group$$

This ensures the relative populations of the LGAs are preserved when they are grouped together.

(c) Statistical Area Levels (SAx) vs. Local Government Areas (LGAs)

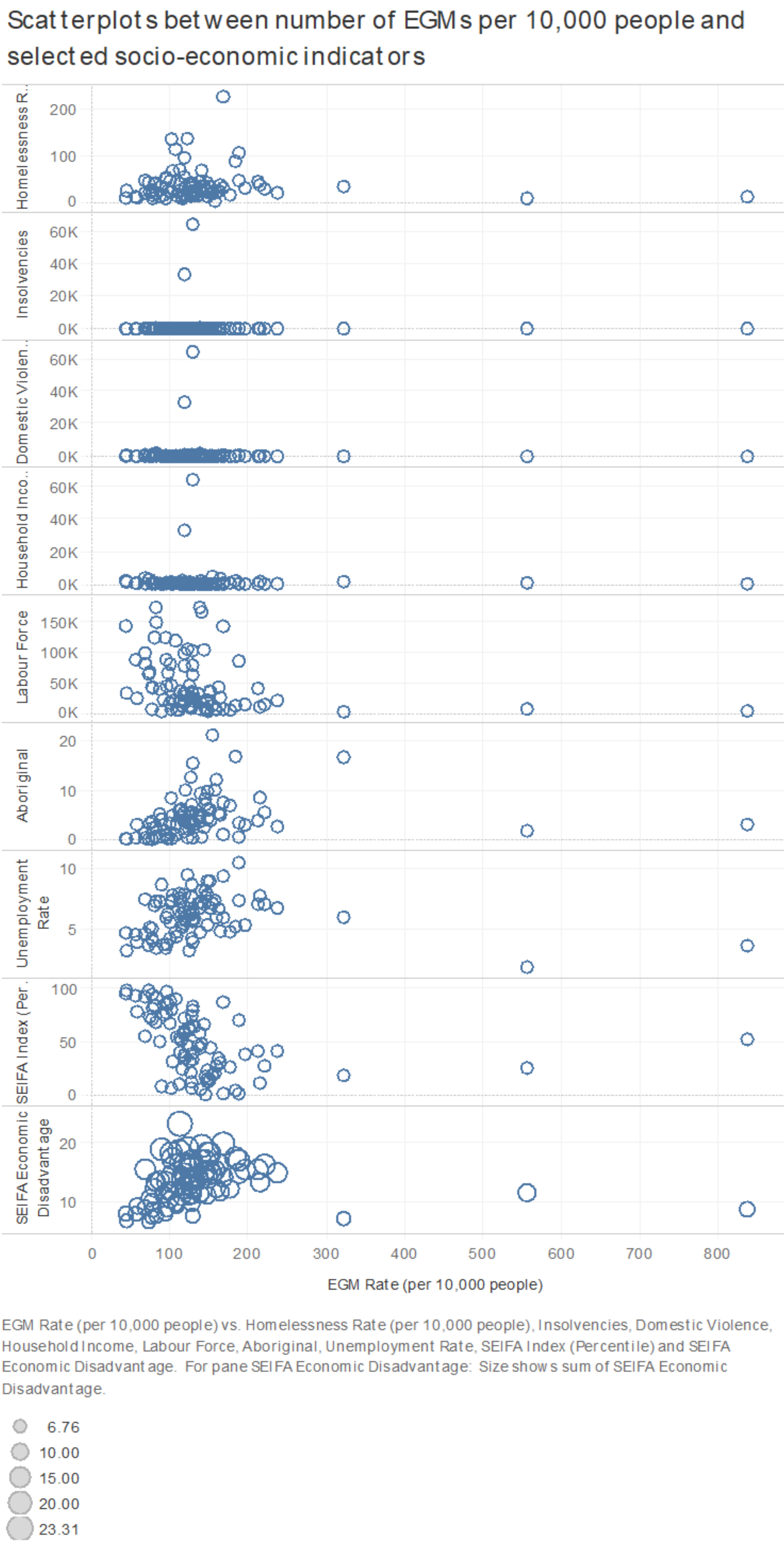
The project also had to deal with differing geographical reporting standards. Specifically, economic disadvantage data from the New South Wales Council of Social Service (NCOSS), was not reported by LGA, but rather by Statistical Areas Level 2 (SA2s). SA2s are part of the Australian Statistical Geographical Standard that the Australian Bureau of Statistics (ABS) uses to release geographically classified statistics. According to the ABS "Statistical Areas Level 2 (SA2) are medium-sized general-purpose areas… [that] represent a community that interacts together socially and economically" (ABS, 2016).

Although the ABS states that "Local Government Area boundaries were considered in the design of the SA2s", there exists no neat relationship between the two standards. Indeed, no dataset which identifies specific SA2s to LGAs was found and as such, the process of identifying each SA2 to an LGA in NSW, was done manually. This involved comparing the boundaries of SA2s and LGAs on an interactive map provided by the ABS, for each of the 576 SA2s in NSW. Nearly every SA2 could be mapped onto a single LGA, but where the SA2 overlapped onto multiple LGAs, the SA2 was mapped onto the LGA with which it shared more of its area with.

As SA2s are smaller in size than LGAs, many SA2s often fall within a single LGA. Thus, a statistic for an LGA can be constructed by taking an average of the individual SA2s data entries that make up the LGA, weighted by each SA2s' population. This is a similar process to that described above for grouping LGAs by the weights of each individual LGA.

II. Scatterplots

Figure 8: Scatterplots of the number of EGMs per 10,000 people and various indicators by LGA



EGM Rate (per 10,000 people) vs. Homelessness Rate (per 10,000 people), Insolvencies, Domestic Violence, Household Income, Labour Force, Aboriginal, Unemployment Rate, SEIFA Index (Percentile) and SEIFA Economic Disadvantage. For pane SEIFA Economic Disadvantage: Size shows sum of SEIFA Economic Disadvantage.

III. Alternative Regression Model

An alternative regression model for homelessness is shown below:

$$HOMELESS = \beta_0 + \beta_1 POP + \beta_2 SEIFA + \beta_3 DV + \beta_4 AB + \beta_5 EGM + \theta$$

Where,

- HOMELESS – Number of homelessness population in LGA;
- POP – Population in the LGA;
- SEIFA – Social-Economics Index for Areas (in percentiles);
- DV – Number of recorded cases of domestic violence in LGA;
- AB – The percentage of Aboriginal people in LGA;
- EGM – Number of electronic gaming machines in LGA;

(a) Estimation

The estimated model 1 is shown below (standard errors are in parenthesis and Huber-White standard errors are in brackets):

$$\begin{array}{lcccccc} HOMELESS = & -143.28 & -\,0.001POP & +\,1.86SEIFA & +\,0.27DV & -\,13.08AB & +\,0.55EGM \\ & (233.4886) & (0.0019) & (2.9949) & (0.2806) & (20.2051) & (0.1287) \\ & [309.7649] & [0.0031] & [4.4846] & [0.2557] & [15.3409] & [0.2762] \end{array}$$
$$R^2 = 0.5668, \quad N = 85$$

According to the estimated model 1, it demonstrates that, on average, an additional electronic gaming machine is expected to increase the number of homeless people in that LGA by 0.55 people, holding all else equal.

(b) Test for Statistical significance of EGM

The variable of interest is EGM; hence, we need to conduct a hypothesis test for its statistical significance.

$$H_0: \beta_5 = 0$$
$$H_1: \beta_5 \neq 0$$

The test statistic for EGM's coefficient is 1.98 using the heteroskedasticity-robust errors (see section III). We will reject the null hypothesis in favor of alternative hypothesis if the test statistic is below the critical value. And the critical value, for the t distribution with 79 degrees of freedom at 5% significance level, is 1.99. Since |t|>c, we reject the null and in favor of the alternative at 5% significance level. In conclusion, the number of electronic gaming machines has a statistically significant effect at 5% level on homelessness in LGA, ceteris paribus.

(c) Check for Heteroskedasticity

After conducting Breusch-Pagan test, we reject the null hypothesis and conclude that there is heteroskedasticity in our data. Therefore, we choose to use the Huber-White standard errors (shown in the brackets of our estimation) to surmount this issue.