**COMPETITION 65691 (Econ 27): January 27, 2020**

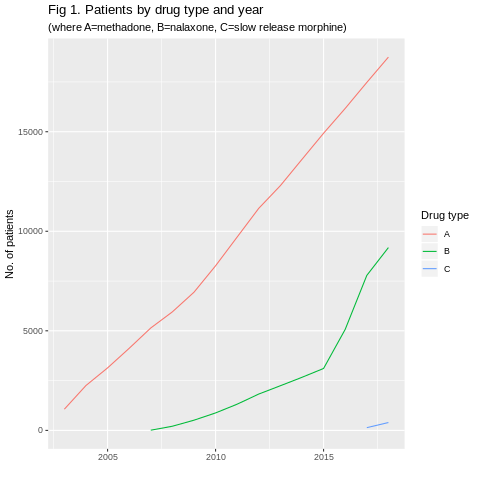
**PART 1** (relevant code in R can be found in the file ‘written assignment 1.ipynb’ )

1. How many clients have participated in the OAT program? 44,200.
2. What is the client count, by drug, and year?

Table 1. Clients by drug type, and year

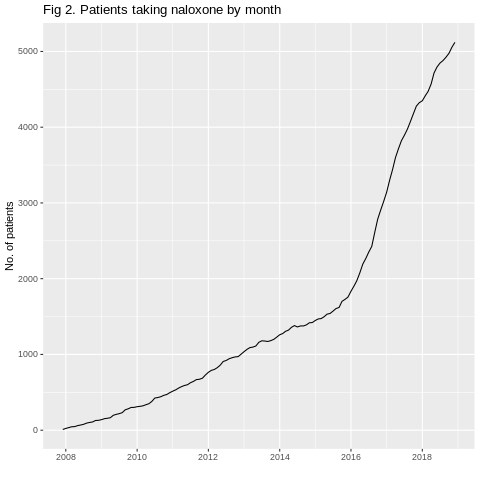
(where A=methadone, B=nalaxone, C=morphine)

|  |  |  |  |
| --- | --- | --- | --- |
| Year | A | B | C |
| 2003 | 1064 | na | na |
| 2004 | 2253 | na | na |
| 2005 | 3139 | na | na |
| 2006 | 4122 | na | na |
| 2007 | 5143 | 9 | na |
| 2008 | 5952 | 208 | na |
| 2009 | 6940 | 512 | na |
| 2010 | 8267 | 877 | na |
| 2011 | 9712 | 1317 | na |
| 2012 | 11153 | 1825 | na |
| 2013 | 12294 | 2243 | na |
| 2014 | 13613 | 2666 | na |
| 2015 | 14924 | 3110 | na |
| 2016 | 16169 | 5069 | na |
| 2017 | 17476 | 7784 | 137 |
| 2018 | 18748 | 9181 | 395 |



1. Did the policy changes in October 2015 and July 2016 affect the client counts?

* Visually, both policies appear to increase the number of clients (Fig 2). However, only the October 2015 policy appears to be significant.
* To test the Oct 2015 significance, I created the monthly ‘difference’ variable[[1]](#footnote-2) of the number of patients taking naloxone. The ‘difference’ variable was divided into two: (1) number of clients before Oct 2015, and (2) clients after Oct 2015 but before July 2016. The increase in clients taking naloxone following the 2015 policy has a significant t-statistic (t = -5.5707, df = 8.2738, p-value = 0.0004687)
* To test the July 2016 policy, the ‘difference’ variable was again divided into two: (1) number of clients taking naloxone after Oct 2015 but before July 2016 , and (2) clients taking naloxone after July 2016. The increase in clients taking naloxone following the 2015 policy is not significant (t = -1.8715, df = 20.104, p-value = 0.0759)

****

**PART 2** (relevant R code can be found in file ‘written assignment 2.ipynb’ )

**BACKGROUND:** Fundamentals of time series data

The sequential nature of time series data may have the following implications on modelling and forecasting[[2]](#footnote-3). Observations from previous (lagged) time periods may help predict observations of subsequent periods. Such time series data are said to violate the property of stationairty[[3]](#footnote-4). These data may contain the following properties, which can be identified and used to coerce them into stationarity for modelling:

* Autoregression is when observations of previous time periods help predict observations of subsequent periods. For example, observation y(t- p) from p lagged periods may predict future y(t). A projection model that fits the data is an autoregressive (**AR**) model of order **p**.
* Moving average is when the shocks from residuals of previous time periods help predict observations of future periods. Here, residual e(t-q) from lagged period q may predict future e(t) and y(t). A model that fits this data is a moving average (**MA**) model is of order **q**.
* ARIMA models put together both AR and MA properties.

**ASSUMPTIONS:**

Variables that may predict community A’s data include: the health authority’s and community A’s client numbers and service volume data respectively, and the province’s total population .

**METHODOLOGY:**

1. The following observations are used to build historic variables, from 2012-2019: the health authority’s and community’s client numbers and service volumes respectively, and the provincial population [[4]](#footnote-5) .
2. The variables are used to estimate the following ratios for 2012-2019:

* Health authority’s clients as a share of total provincial population (a ratio I termed as ratio ‘**ha\_p’** in my Excel spreadsheet)
* Community A’s clients as a share of the health authority’s clients (termed as ratio‘**a\_ha’**)
* Service volume per client in Community A (termed as ratio ‘**a\_n**’)

1. I estimate AR and MA orders for the above ratios. These orders help fit a model that projects the ratios into to 2020-2022. Ising the 2020-2022 ratios, I project client numbers and service volumes for 2020-2022 as follows:

* Provincial population (t) \* ha\_p ratio (t) = health authority clients (t).
* Health authority’s clients (t) \* a\_ha ratio (t) = community A’s clients (t)
* Community A’s clients (t) \* a\_n ratio (t) = community A’s service volume

**RESULTS:**

What follows is an illustrative example for estimating one of the ratios (ha\_p) for 2012-2019. As well, this estimatesAR and MA orders; and the ratios for 2020-2022 (code is in file ‘part2.ipynb’).

* Between 2012-2019, the ha\_p ratio increased in a linear fashion, with AR of order 1 (p =1) and MA of order 2 (q=2), shown in Fig1 (left).
* On Fig 1 (right), I fit p and q to a projection for 2020-2022 (p=1, d=0, q=2). Here, d =0, because ‘ha\_p’ increased in a linear fashion, with no difference ‘d’ variable needed [[5]](#footnote-6).
* As described in the methodology, together with the remaining ratios, I project community A’s client numbers and service volumes (Table 1). The Excel spreadsheet contains the remaining ratios, AR and MA orders used for projecting client numbers and service volumes ( Table 1.)

Fig 1. AR and MA orders (left); fitted model (right) for ‘ha\_p’

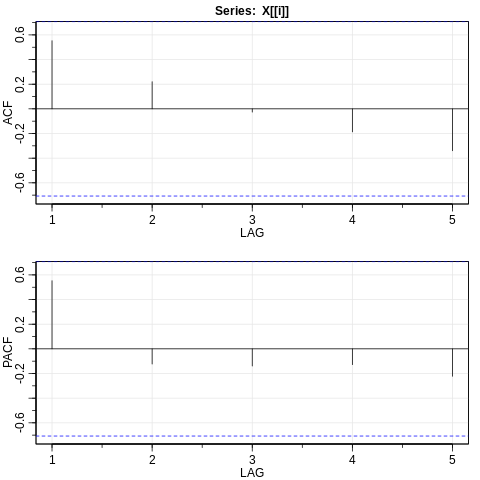
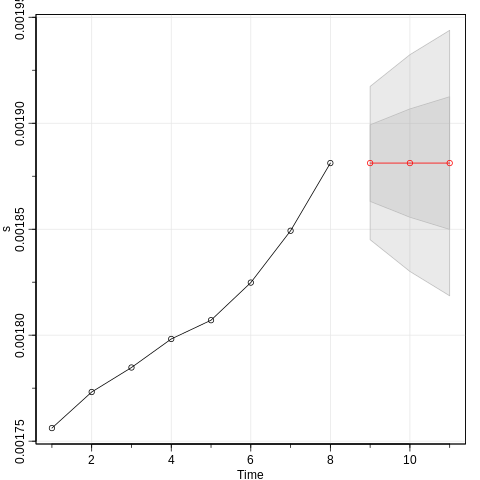


Table 1. Forecast clients and service hours for community A

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Clients | Service vol. | Grwth % a | Funding grth%b |
| 2020 | 1839 | 30,090 | 2 | 2 |
| 2021 | 1861 | 30,318 | 1 | 2 |
| 2022 | 1879 | 30,558 | 1 | 2 |

‘a’ -growth rate is the same for clients and service volume; ‘b’-health authority funding growth

**CONCLUSION**

* Community A’s number of clients and service hours are predicted using their own and the health authority’s lagged data, and provincial populations.
* Community A’s projected growth rate in client numbers and service hours is between 1-2 per cent, which is less than the health authority’s estimates (2%.)

**LIMITATIONS**

* The discrepancy with the health authority’s growth estimates maybe partly due to errors in my R code. Errors in my code prevent me from estimating AR and MA orders for the ‘difference’ of ratios/variables .
* Errors also prevent me from estimating the residuals, related to the AR and MA orders. Having the residuals would lead to a better model fit.

1. A difference variable is estimated by subtracting one lagged observation. [↑](#footnote-ref-2)
2. Econometrics by Erasmus University, Rotterdam accessed on Coursera. [↑](#footnote-ref-3)
3. Stationary time series variables have constant mean, variance and covariance with residuals over time. [↑](#footnote-ref-4)
4. Obtained from ‘B.C. Stats population projection’ <<https://bcstats.shinyapps.io/popProjApp/>> [↑](#footnote-ref-5)
5. For example, a difference of 1 is estimated by subtracting one lagged observation. [↑](#footnote-ref-6)