Mauritius time series

Obiwenobi

2024-12-06

1 Introduction

This document aims to define whether macroeconomics variables in Mauritius are endogenous. The dataset used to produce these figures and tables is merge from various sources. It gathers data from 1964 to 2018.

2 Data summary

The dataset **dbase** includes the following variables:

```
##  [1] "year"         "birth"        "death"        "infmortality" "stillbirth"
##  [6] "marriage"     "divorce"      "gdp"          "g"            "inf"
```

- **birth**: The number of live births in a year per 1,000 mid-year population.
- **death**: The number of deaths in a year per 1,000 mid-year population.
- **infmortality**: The number of infant deaths in a year per 1,000 live births during the year.
- **g**: Annual growth rate
- **inf**: Annual inflation rate,Percent, Not Seasonally Adjusted
- **gdp**: Gross domestic product, billions of US $
- **marriage**: The number of persons married in a year per 1,000 mid-year population.
- **divorce**: The number of persons divorced in a year per 1,000 mid-year population.

How many observations do we have ? The function **nrow()** returns the number of rows and which is in the number of observations in our situation.

```
nrow(dbase)
```

```
## [1] 55
```

Let's look at the descriptive statistics of our variables.

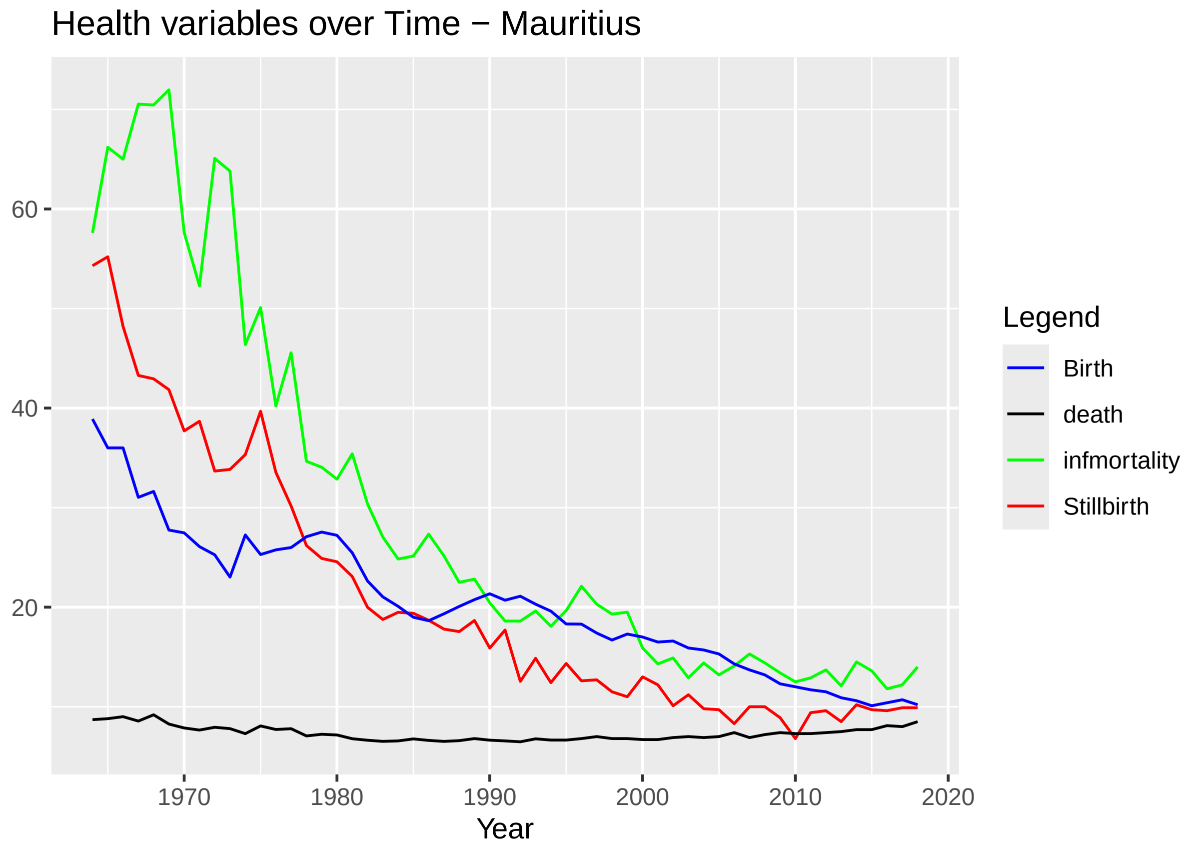
```
summary(dbase)
```

```
##       year          birth           death        infmortality     stillbirth
##  Min.   :1964   Min.   :10.10   Min.   :6.476   Min.   :11.80   Min.   : 6.80
##  1st Qu.:1978   1st Qu.:15.50   1st Qu.:6.770   1st Qu.:14.40   1st Qu.:10.05
##  Median :1991   Median :19.60   Median :7.162   Median :20.44   Median :15.89
##  Mean   :1991   Mean   :20.29   Mean   :7.321   Mean   :29.48   Mean   :20.72
```

```
##  3rd Qu.:2004   3rd Qu.:25.61   3rd Qu.:7.752   3rd Qu.:37.81   3rd Qu.:28.18
##  Max.   :2018   Max.   :38.90   Max.   :9.188   Max.   :71.97   Max.   :55.20
##     marriage         divorce            gdp                g
##  Min.   : 9.947   Min.   :0.2169   Min.   : 0.3163   Min.   :-0.79486
##  1st Qu.:15.741   1st Qu.:0.4878   1st Qu.: 3.2169   1st Qu.:-0.33537
##  Median :18.400   Median :1.3889   Median : 6.4193   Median :-0.05803
##  Mean   :17.353   Mean   :1.4790   Mean   : 7.3360   Mean   : 0.73759
##  3rd Qu.:20.057   3rd Qu.:2.1000   3rd Qu.: 9.0364   3rd Qu.: 0.35264
##  Max.   :23.012   Max.   :3.8000   Max.   :41.9995   Max.   :16.65818
##       inf
##  Min.   : 0.3163
##  1st Qu.: 3.2169
##  Median : 6.4193
##  Mean   : 7.3360
##  3rd Qu.: 9.0364
##  Max.   :41.9995
```

How did the variables evolve during the years ? First, the health data.

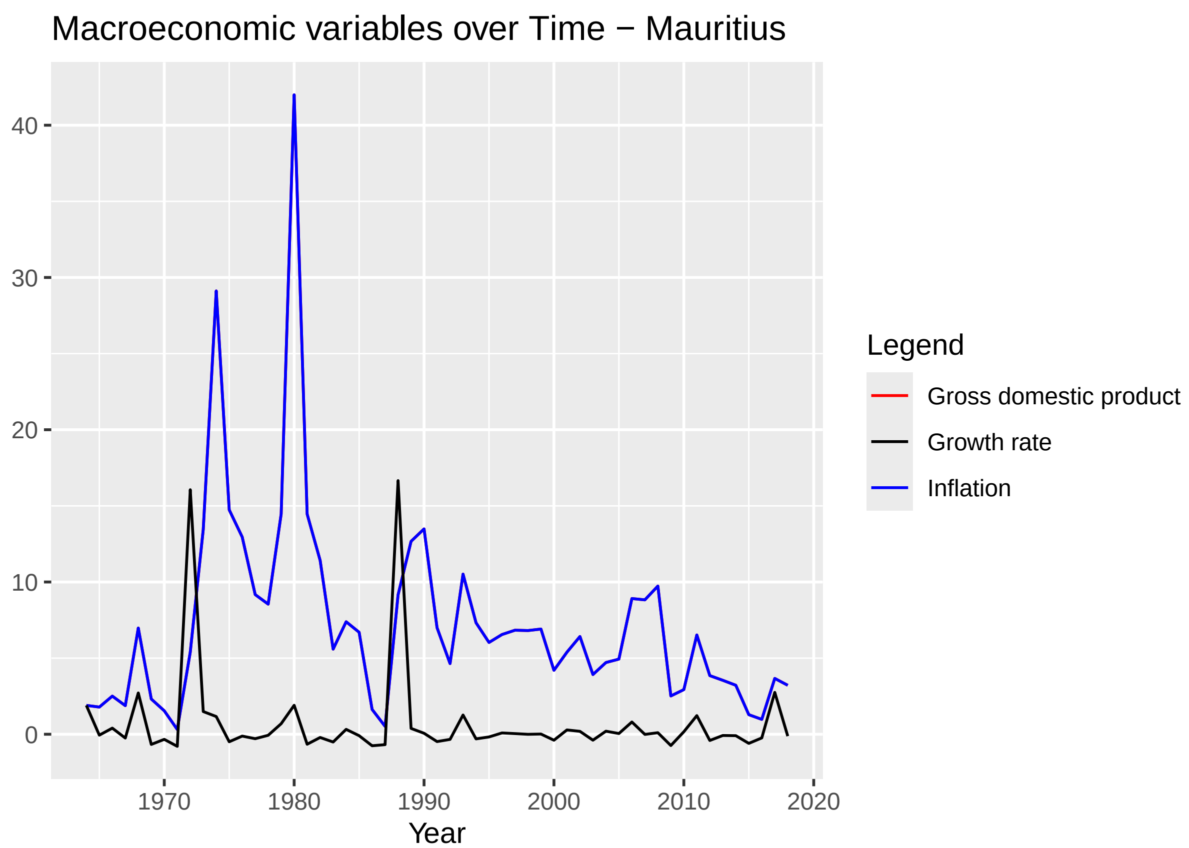
```r
dbase %>%
  ggplot() +
  geom_line(aes(y = stillbirth, x = year, color = "Stillbirth")) +
  geom_line(aes(y = death, x = year, color = "death")) +
  geom_line(aes(y = infmortality, x = year, color = "infmortality")) +
  geom_line(aes(y = birth, x = year, color = "Birth")) +
  xlab('Year') +
  ylab('') +
  labs(title = "Health variables over Time - Mauritius", color = "Legend") +
  scale_color_manual(values = c("Stillbirth" = "red", "Birth" = "blue","death"="black",
                                "infmortality"="green"))
```



We can see that every rate is decreasing and might probably be correlated. Evaluating this effect through regression would probably be misleading as all those variable are likely to be endogenous. Methods such as Vector Auto-Regressive models suit well to identify such relationships but require high frequency data such as monthly data. The data set only provides yearly data which does not ensure the estimations to be converging.

Before we investigate this, let's consider macroeconomic data.

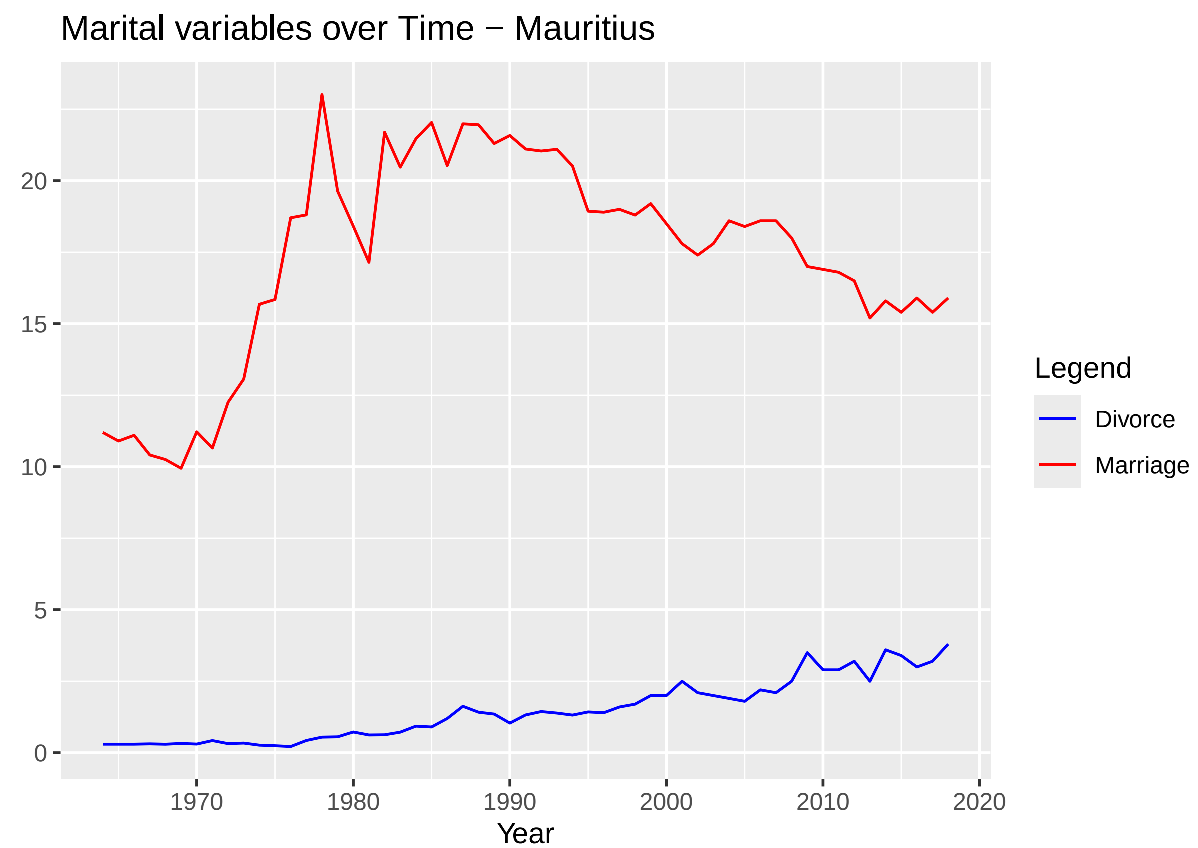
```
dbase%>%
ggplot() +
  geom_line(aes(y =gdp , x = year, color = "Gross domestic product")) +
  geom_line(aes(y =inf , x = year, color = "Inflation")) +
  geom_line(aes(y =g , x = year, color = "Growth rate"))+
  xlab('Year') +
  ylab('') +
  labs(title = "Macroeconomic variables over Time - Mauritius", color = "Legend") +
  scale_color_manual(values = c("Gross domestic product" = "red",
                                "Inflation" = "blue","Growth rate"="black"))
```



Plotting inflation, growth rate and gross domestic product such as the figure above might be misleading. Because the different values have different scales, it is misleading to plot these three time series together. Because of larger values, slight changes in smaller values are unobtrusive. Scatter plot, though they have another purpose, is more suitable to investigate such data relationship.

Finally, let's consider marriage and divorce values.

```
dbase%>%
ggplot() +
  geom_line(aes(y =marriage , x = year, color = "Marriage")) +
  geom_line(aes(y =divorce , x = year, color = "Divorce")) +
  xlab('Year') +
  ylab('') +
  labs(title = "Marital variables over Time - Mauritius", color = "Legend") +
  scale_color_manual(values = c("Marriage" = "red",
                                "Divorce" = "blue"))
```

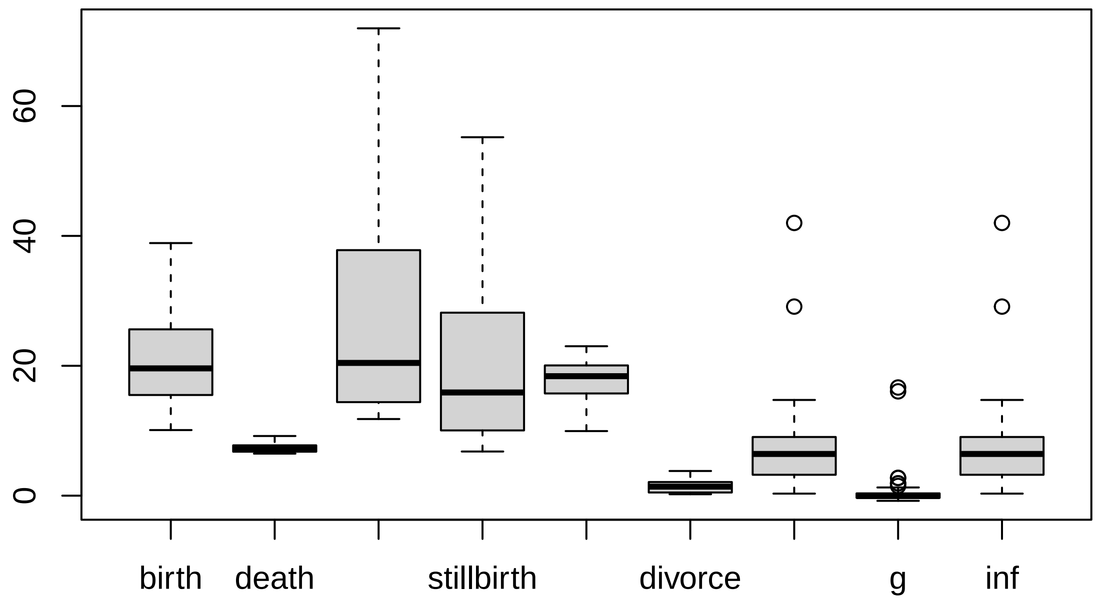


Recall that these variables are measured on the same scale. Interpreting these values does not require the same caution as for the previous plot. At first glance, there does not seem to be any linear correlation between both variables.

In the previous figure, we did plot inflation which feature a sharp increase around **1980**. At the same period, there is inverse trend to inflation in marriage. Can we infer whether there is a relationship between marriage and divorce ? As discussed previously, the low frequency of data might hamper inference, especially if we seek to investigate long term effect.[1]

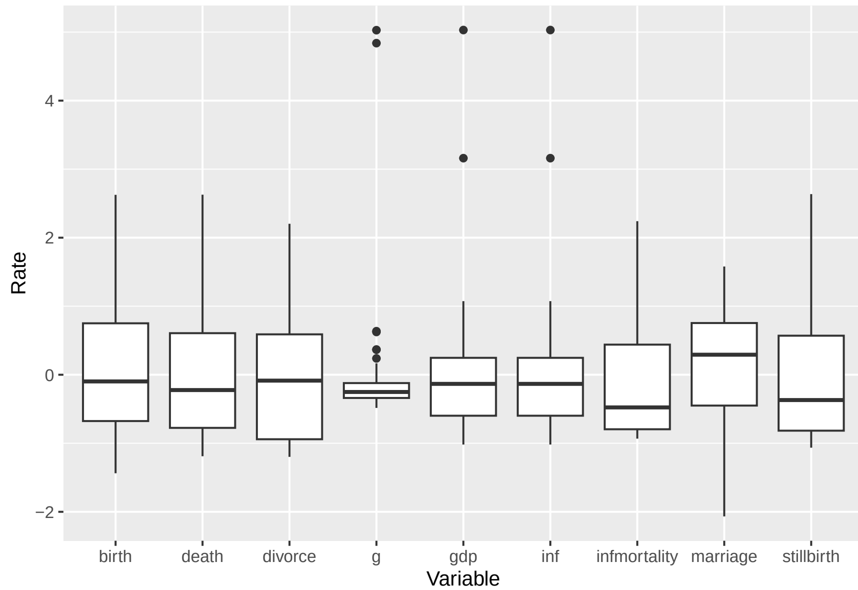
Finally, before embarking in the inference journey, let's take a look to the distributions of our variables.

[1]Indeed, if many lags are used, a substantial share of data is to be put aside. For instance, considering 8 lags implies that the first eight observations are not considered in regressions, which almost amounts to 15% of the observations.



Note that due to different scales, interpreting this boxplot is not straightforward. How about standardizing data before plotting ?

```
ndata <- dbase%>%
  mutate(across(2:10,scale))  %>%
  pivot_longer(cols=2:10,names_to="Variable",values_to="Rate")%>%
  ggplot()+geom_boxplot(aes(x=Variable,y=Rate))
ndata
```



We see that growth features severe outliers. Without surprise, we see that there are some outliers as well for inflation and GDP. As our sample is small, it wouldn't be surprising that these outliers hampers our inference of the relationship between these three variables.

Nonetheless, we see that other variables distribution do not present outliers, which might produce faster converging estimation.

3 Modelling relationships

In this part we will seek to model the relationship between our variables before estimating whether their relationships are statistically significant. We will consider two different approach.

1. **Linear Additive model**: We consider that there is a linear additive relationship between the variables. For instance, we could consider that we have a set of variables that could explain death rates. We can include time *Fixed-Effects* to assess the effect of time. One might suggest the following model:

$$death_t = \alpha_t + divorce_t + gdp_t + inf_t + g_t + birt_t + \varepsilon_t$$

Modelling the relationship as such relies on the assumption that there is no endogeneity between our variables. There exist tests to verify this assumption.

2. **Vector Auto Regressive models (VAR)**: This method aims to define the endogenous relationship between variables considering that their present and lag value have an impact on future values.

Let define our macroeconomic as a the following vector:

$$\mathbf{Y}_t = \begin{bmatrix} birth_t \\ death_t \\ infmortality_t \\ \vdots \\ inf_t \end{bmatrix}$$

We might consider that this vector is dependent on its past values. Therefore we may write our model such as:

$$\mathbf{Y}_t = \mathbf{A}_1\mathbf{Y}_{t-1} + \mathbf{A}_2\mathbf{Y}_{t-2} + \mathbf{A}_3\mathbf{Y}_{t-3} + ... + \mathbf{A}_p\mathbf{Y}_{t-p} + \eta_t$$

Where is the disturbance vector. Then, from this model we can model a shock on one of our variables and estimate the shock effect on future values of the other variables. This is the **Impulse Response** which can be plotted and named as the **VAR Impulse Response Function** (**VAR IRF**)