Mauritius time series

Obiwenobi

2024-12-06

1 Introduction

This document aims to define whether macroeconomic variables in Mauritius are endogenous. The data set used to produce these figures and tables is the intersection of several sources. It gathers data from 1964 to 2018.

2 Data summary

The data set **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, here, is in the number of observations.

```r
nrow(dbase)
```

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

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

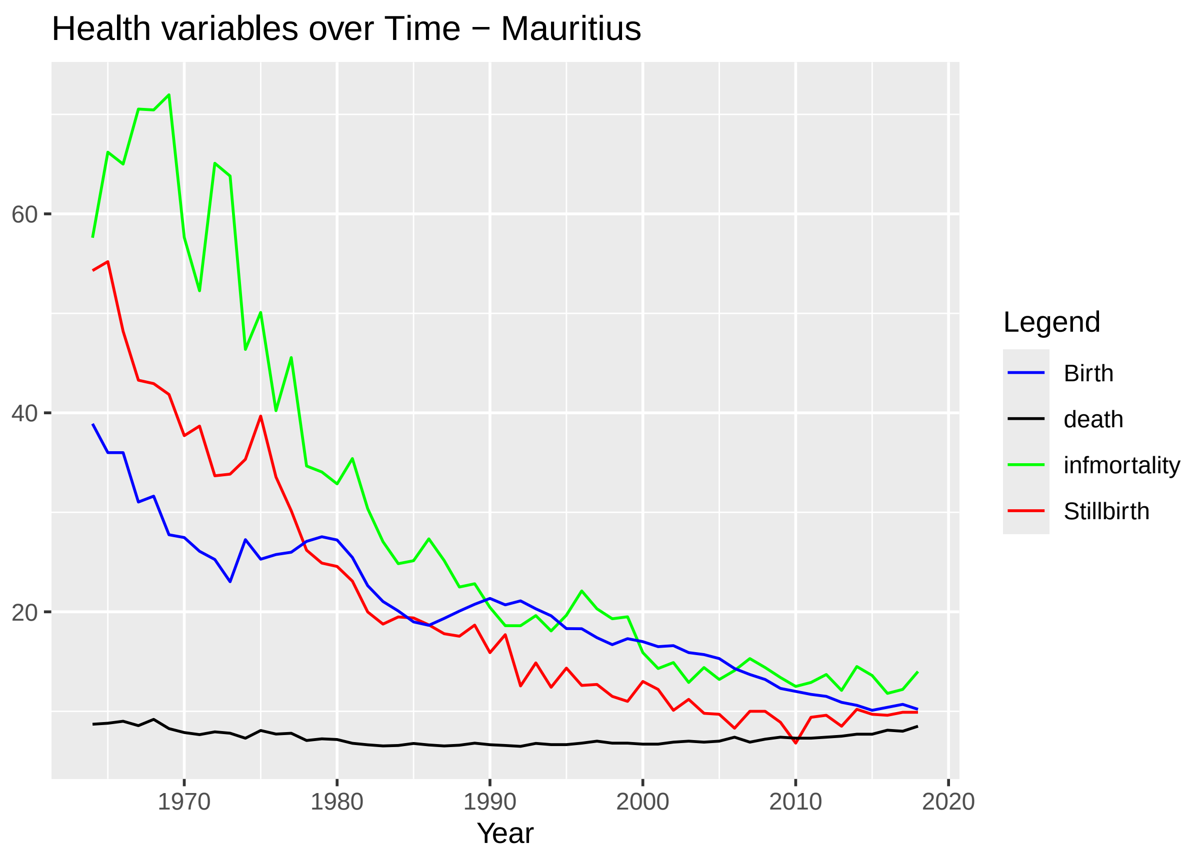
```r
summary(dbase)
```

```
##       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.

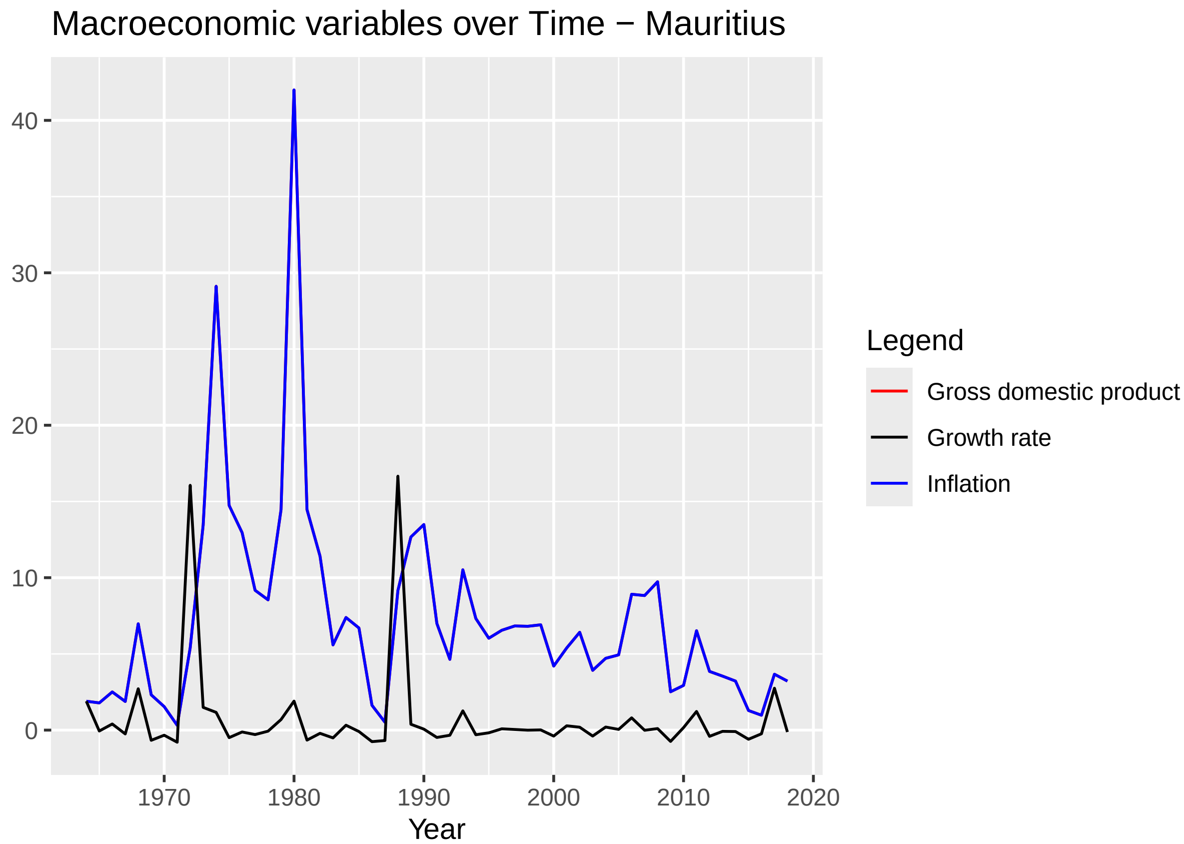
```
dbase %>%
  ggplot() +
  geom_line(aes(y = stillbirth, x = year, color = "Stillbirth")) +
  geom_line(aes(y = death, x = year, color = "death")) +
  geom_line(aes(y = infmortality, x = year, color = "infmortality")) +
  geom_line(aes(y = birth, x = year, color = "Birth")) +
  xlab('Year') +
  ylab('') +
  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.

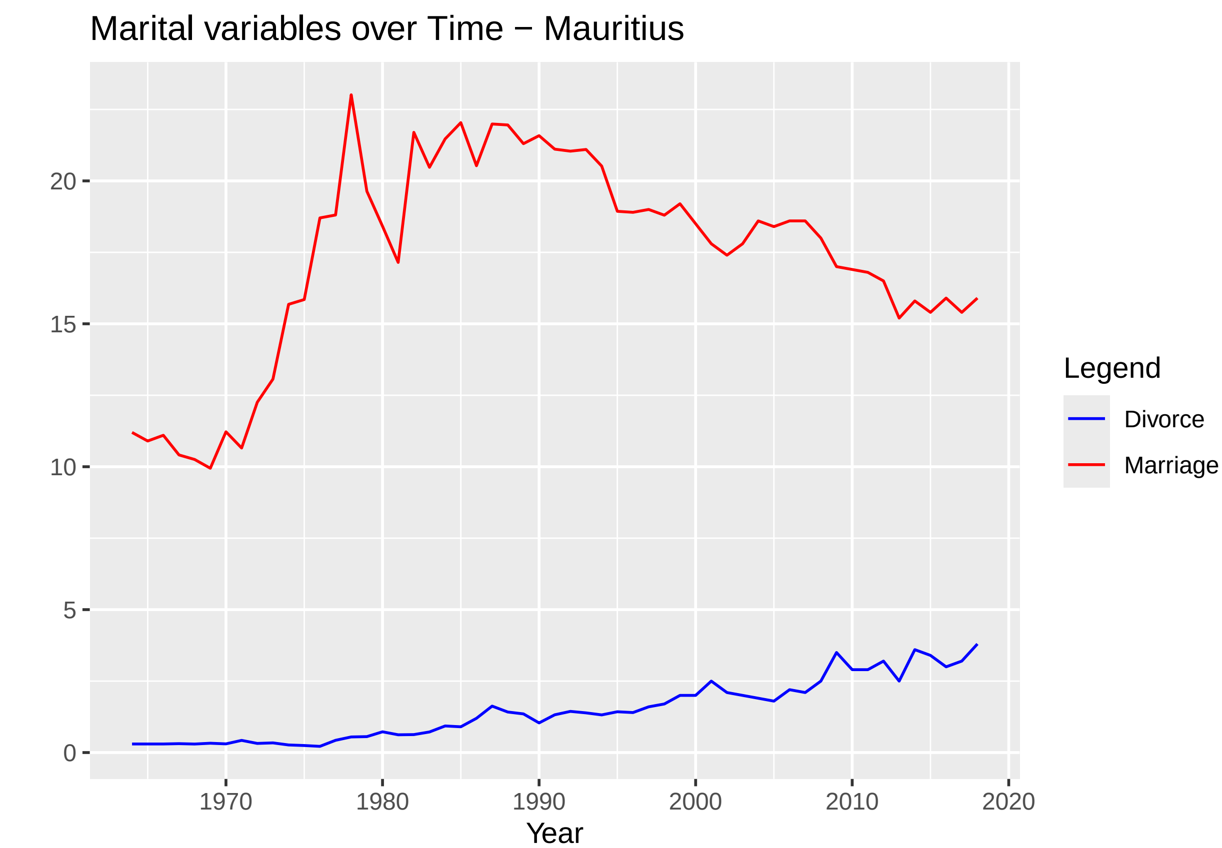
```r
dbase%>%
ggplot() +
  geom_line(aes(y =gdp , x = year, color = "Gross domestic product")) +
  geom_line(aes(y =inf , x = year, color = "Inflation")) +
  geom_line(aes(y =g , x = year, color = "Growth rate"))+
  xlab('Year') +
  ylab('') +
  labs(title = "Macroeconomic variables over Time - Mauritius",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



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

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 the previous plot did. At first glance, there does not seem to be any linear correlation between both variables.

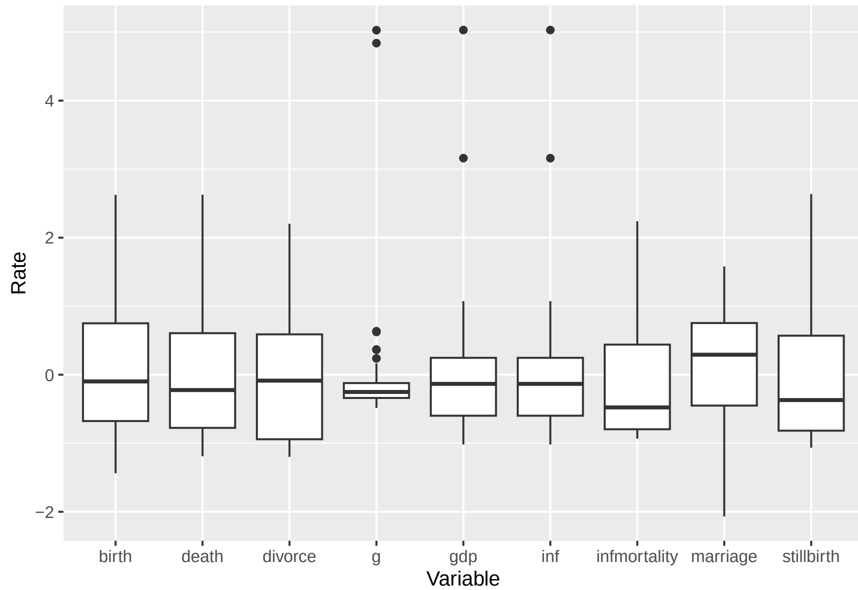
In the previous figure, we did plot inflation which features a sharp increase around **1980**. At the same period, there is an inverse trend to inflation in marriage. Can we infer whether there is a relationship between marriage and divorce ? As discussed previously, the low frequency of data might hamper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

Finally, before embarking in the inference journey, let's take a look at 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 box plot 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. Unsurprisingly, we see that there are some outliers as well for inflation and GDP. As our sample is small, it wouldn't be surprising if these outliers hamper our inference on 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 approaches.

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 marriage. One might suggest the following model:

$$marriage_t = \beta_1 divorce_t + \beta_2 gdp_t + \beta_3 inf_t + \beta_4 g_t + \beta_5 birth_t + \beta_6 death_t + \varepsilon_t$$

Modeling the relationship in this manner relies on the critical assumption that there is no endogeneity between the variables. There are statistical tests available to verify the validity of this assumption.

Furthermore, this approach presumes that the relationship is contemporaneous, which may not be suitable for all models. For example, consider the relationship between mortality and economic growth. These variables may not affect one another simultaneously within the same period.

One could hypothesize that a higher marriage amount may lead to future economic growth, as it strengthens households. Conversely, a negative economic growth shock directly impacts the broader economy, which in turn affects the individuals' ability to plan, causing an immediate decline in the marriage amount.

Thus, we will model marriage as a dependent variable of other macroeconomic variables.

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

Let's define our macroeconomic variable as the following vector:

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

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 η_t 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 referred to as the **VAR Impulse Response Function** (**VAR IRF**)

3.1 Linear contemporaneous additive model

Let's model marriage as dependent of other macroeconomic variables as:

$$marriage_t = \beta_1 divorce_t + \beta_2 gdp_t + \beta_3 inf_t + \beta_4 g_t + \beta_5 birth_t + \beta_6 death_t + \varepsilon_t$$

We assume that this model satisfies the Gauss-Markov theorem and that we may obtain the BLUE[^2] estimator. The `R` code to estimate this model is [^2]: Best Linear Unbiased Estimator. See here.

```
lm(marriage ~ g + gdp + inf +
     divorce + death + birth, data=dbase) %>% summary()
```

```
##
## Call:
## lm(formula = marriage ~ g + gdp + inf + divorce + death + birth,
##     data = dbase)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3584 -1.0004 -0.2189  1.2675  4.8626
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.686e+01  3.136e+00  14.946  < 2e-16 ***
## g           -3.535e-03  8.375e-02  -0.042    0.967
## gdp          5.231e+03  9.391e+03   0.557    0.580
## inf         -5.231e+03  9.391e+03  -0.557    0.580
## divorce      4.562e-01  6.695e-01   0.681    0.499
## death       -4.285e+00  4.963e-01  -8.633 2.48e-11 ***
## birth        3.714e-02  1.071e-01   0.347    0.730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.905 on 48 degrees of freedom
```

```
## Multiple R-squared:  0.7486, Adjusted R-squared:  0.7172
## F-statistic: 23.82 on 6 and 48 DF,  p-value: 7.598e-13
```

Our results suggest that growth rate does not impact marriage contemporaneously, while mortality rate effect is significant. This model features a high `Multiple R-squared=0.7486` which can be interpreted as the predictive accuracy. One could aggregate new data to test the efficiency of this model. Finally, other variables do not seem to affect significantly marriages in Mauritius on the period 1964-2018.

##Vector Auto Regressive model

Now, we will explore the effect of a macroeconomic shock on marriage over time. Recall that our model is :

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

3.1.1 Baseline regression

The code to produce a VAR estimation in `R` is:

```
var1 <- VAR(dbase[,c("g","inf","divorce","gdp","death","birth","marriage")], p = 4, type = "none",
          season = NULL, exogen = NULL, lag.max = 5,
          ic = c("AIC", "HQ", "SC", "FPE"))
```

This computes an estimation for every $\mathbf{A}_p$ matrix. Let's look at the estimation for the growth equation.

```
var1$varresult$g %>%summary()
```

```
##
## Call:
## lm(formula = y ~ -1 + ., data = datamat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5240 -0.8640 -0.0606  0.8645  5.4883
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## g.l1        -4.131e-01  2.194e-01  -1.883   0.0793 .
## inf.l1      -1.513e+04  2.382e+04  -0.635   0.5350
## divorce.l1  -3.169e+00  2.291e+00  -1.383   0.1868
## gdp.l1       1.513e+04  2.382e+04   0.635   0.5350
## death.l1     3.296e-01  2.508e+00   0.131   0.8972
## birth.l1     4.804e-01  9.473e-01   0.507   0.6195
## marriage.l1  1.594e+00  7.364e-01   2.165   0.0469 *
## g.l2        -3.727e-01  2.176e-01  -1.713   0.1073
## inf.l2      -3.758e+04  2.504e+04  -1.501   0.1542
## divorce.l2   2.237e+00  2.237e+00   1.000   0.3332
## gdp.l2       3.758e+04  2.504e+04   1.501   0.1542
## death.l2    -5.865e+00  2.616e+00  -2.242   0.0405 *
## birth.l2    -9.927e-01  9.334e-01  -1.064   0.3044
## marriage.l2 -1.557e+00  6.763e-01  -2.302   0.0361 *
## g.l3        -1.698e-01  2.467e-01  -0.688   0.5017
## inf.l3      -5.795e+03  2.541e+04  -0.228   0.8227
## divorce.l3  -8.303e-01  2.341e+00  -0.355   0.7278
```

```
## gdp.l3       5.795e+03  2.541e+04   0.228   0.8227
## death.l3    -2.642e+00  2.765e+00  -0.956   0.3544
## birth.l3     3.573e-01  1.005e+00   0.356   0.7271
## marriage.l3  6.996e-01  7.740e-01   0.904   0.3803
## g.l4         9.885e-02  2.070e-01   0.477   0.6399
## inf.l4      -2.638e+04  2.402e+04  -1.098   0.2894
## divorce.l4  -1.115e+00  2.221e+00  -0.502   0.6230
## gdp.l4       2.638e+04  2.402e+04   1.098   0.2894
## death.l4     6.322e+00  2.568e+00   2.462   0.0264 *
## birth.l4     1.417e+00  7.355e-01   1.927   0.0731 .
## marriage.l4 -4.428e-01  7.300e-01  -0.607   0.5532
## g.l5        -5.525e-03  1.880e-01  -0.029   0.9769
## inf.l5      -1.753e+04  2.592e+04  -0.676   0.5091
## divorce.l5  -4.379e+00  2.271e+00  -1.928   0.0730 .
## gdp.l5       1.753e+04  2.592e+04   0.676   0.5091
## death.l5     5.342e+00  2.406e+00   2.220   0.0422 *
## birth.l5    -2.069e+00  7.221e-01  -2.865   0.0118 *
## marriage.l5  3.397e-01  5.786e-01   0.587   0.5659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.757 on 15 degrees of freedom
## Multiple R-squared:  0.7964, Adjusted R-squared:  0.3214
## F-statistic: 1.676 on 35 and 15 DF,  p-value: 0.1425
```

We can observe multiple things but the output is not user-friendly. Each variable has five lagged values. The output is quite large and cannot be easily interpreted. We could filter results for interesting estimation, *i.e., the significant estimates.*

```r
resu_g <- coef(var1)$g
colnames(resu_g) <-  c("est","std","tval","pval")
resu_g_sig <- as.data.frame(resu_g) %>%
  filter(pval<0.1)

resu_g_sig
```

```
##                    est       std      tval       pval
## g.l1        -0.4131046 0.2194223 -1.882692 0.07928492
## marriage.l1  1.5943692 0.7364456  2.164952 0.04692461
## death.l2    -5.8654863 2.6156062 -2.242496 0.04046688
## marriage.l2 -1.5568785 0.6762763 -2.302134 0.03607520
## death.l4     6.3222543 2.5677191  2.462206 0.02640029
## birth.l4     1.4174293 0.7354584  1.927273 0.07310250
## divorce.l5  -4.3785857 2.2710269 -1.928020 0.07300272
## death.l5     5.3418516 2.4060859  2.220142 0.04223809
## birth.l5    -2.0686895 0.7220744 -2.864926 0.01180449
```

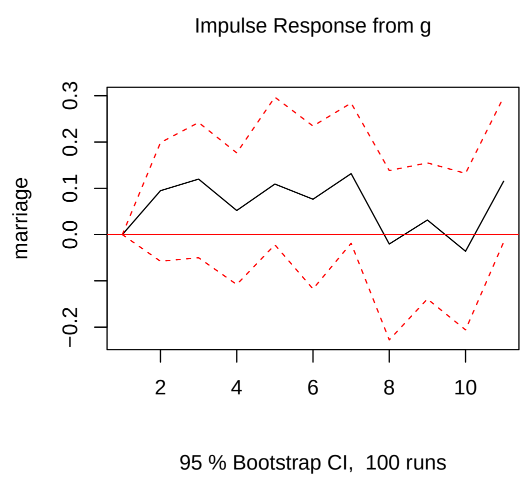
This displays the variables which have a significant effect on marriage. We can already draw few conclusions here. Firstly, we found previously in our contemporaneous model that death had a negative impact on marriage. This is still observed but only with lagged values of mortality rate. Our prior finding could be explained by the auto regressive property of mortality: past mortality could affect contemporaneous mortality and contemporaneous marriage would be affected by past mortality through the intermediary effect of present mortality. Furthermore, we observe that prior growth reduces the marriage rate. While we hypothesized that growth would induce households to forecast positive economic outcomes, it seems that growth repels marriage. Finally, there are long-term effects of macroeconomic variables on marriage, such as the effect of divorce, birth and death.

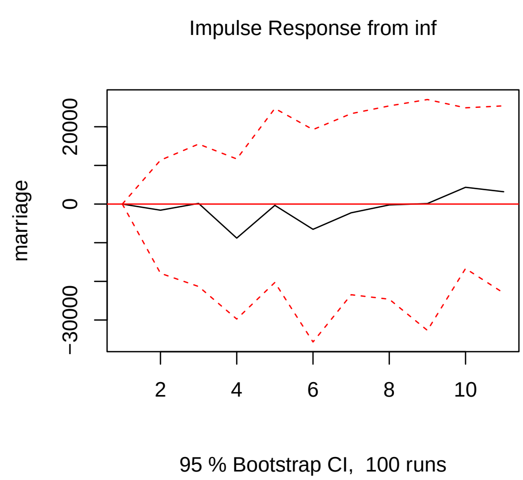
3.1.2 Impulse Response Function

Rather than interpreting the coefficient tables, we can draw results from the VAR model regression to produce the **IRF**. As explained previously, this method simulates a shock on one of our endogenous variable and estimates the effect on other variables. This method is more suitable for interpretation, especially to compare the effect across periods.

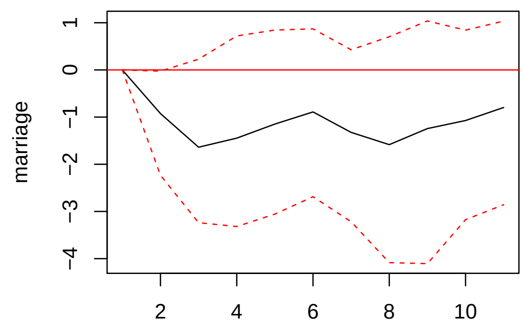
```
irf_est <- irf(var1,response="marriage",ortho=FALSE)
```

```
plot(irf_est)
```



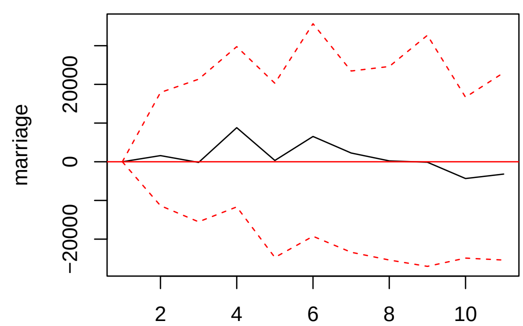


Impulse Response from divorce



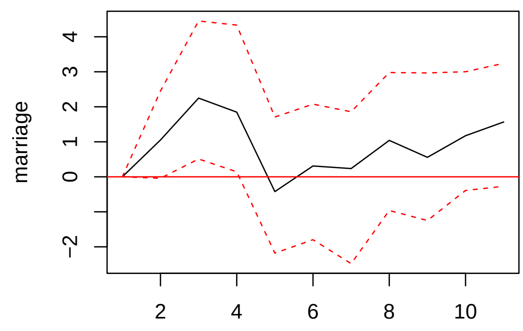
95 % Bootstrap CI, 100 runs

Impulse Response from gdp



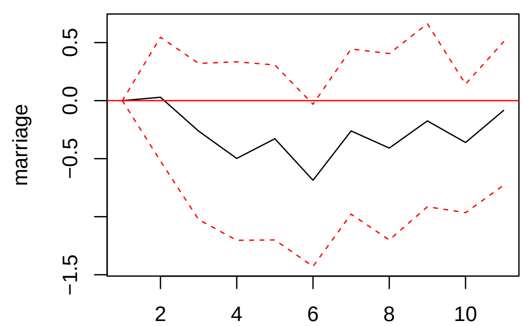
95 % Bootstrap CI, 100 runs

Impulse Response from death



95 % Bootstrap CI, 100 runs

Impulse Response from birth



95 % Bootstrap CI, 100 runs

This models shocks of standard deviation. IRF do not lead to conclude that there are long-term effects of a shock of any variable on marriage. Surprisingly, `growth` and `gdp` do not seem to stimulate marriage.