Prediction and analysis of the residential buildings permits

Schnieder project

Justine Mao, Nikita Gusarov, Tom Vieillescazes
Under direction of Beatrice Roussillon and Mathias Verdiere

Université Grenoble Alpes

10 january 2020

Introduction

Objectives

- Analyse the factors influencing construction permits number.
- Elaborate a predictive model for the construction permits to be issued.

Presentation plan

- Litterature study presentation
- Database construction
- Data presentation
- Data analysis
- Modelisation
  - Simple OLS models
  - Time series models
  - Alternative models

Université Grenoble Alpes

Schneider Electric

Litterature

Sources

Nous avons commencé notre étude de la littérature par la recherche des articles méthodologiques et théoriques en lien avec le sujet traité.

Les moteurs de recherche scientifique les plus utilisés ont été :

- Science Direct
- Google Scholar
- Elsevier
- SAGE journals

Variables of interest

Number of building permits postponed by one quarter (French Building Federation, 2017)	+
10-year OAT rate lagged by two quarters (French Building Federation, 2017)	-
Unemployment rate lagged by two quarters (French Building Federation, 2017)	-
Deadline for housing outstandings in quarter sales shifted by one quarter (French Building Federation, 2017)	-
Political measure (French Building Federation, 2017)	+ or -
Age of the population (Lindh et al, 2008) - Individus moins de 20 ans - Individuals between 20 and 59 years old (National Institute for Demographic Studies, 2019) - Individuals over the age of 75	 ++ + -
Real estate prices (Essay et al, 2015)	-
GDP (Lindh et al, 2002)	+

Table 1: Variables of interest

Database

Data sources

We construct our database using macroeconomical variables figuring in different french and international databases :

- Eurostat
- Sit@del2
- INSEE
- OCDE

Data already studied

The variables that we have already presented are :

- Number of the building permits,
- Surface authorised for construction (not used in analysis anymore),
- GDP,
- Long term interest rate,
- Households spendings,
- Households investments.

Changes and advances

We have extracted as well the following variables :

- Population part by age group :
  - less than 20 years old,
  - from 20 to 59 years old,
  - from 60 years and more.
- Proxy variables serving to capture policies changes :
  - elective cycles dummy variables,
  - political party appartainance.
- Real estate housing prices,
- Household confidence index.

Data

Descriptive statistics

Statistic	Mean	St. Dev.	Min	Max
Year	2,006.000	7.360	1,994	2,018
GDP	1,956.328	219.289	1,545.800	2,285.900
Interest rate	3.903	1.867	0.468	7.535
Household spendings	1,018.228	122.541	801.300	1,186.200
Household investments	111.336	12.820	87.300	134.800
Number of permits	381,173.600	70,800.620	275,711	526,592
Population $<$ 20 y.o.	25.007	0.686	24.003	26.370
Population 20-59 y.o.	52.887	1.389	49.991	54.176
Population $>$ 60 y.o.	22.105	1.964	19.882	26.006

Table 2: Descriptivve statistics for annual data

Data transformation

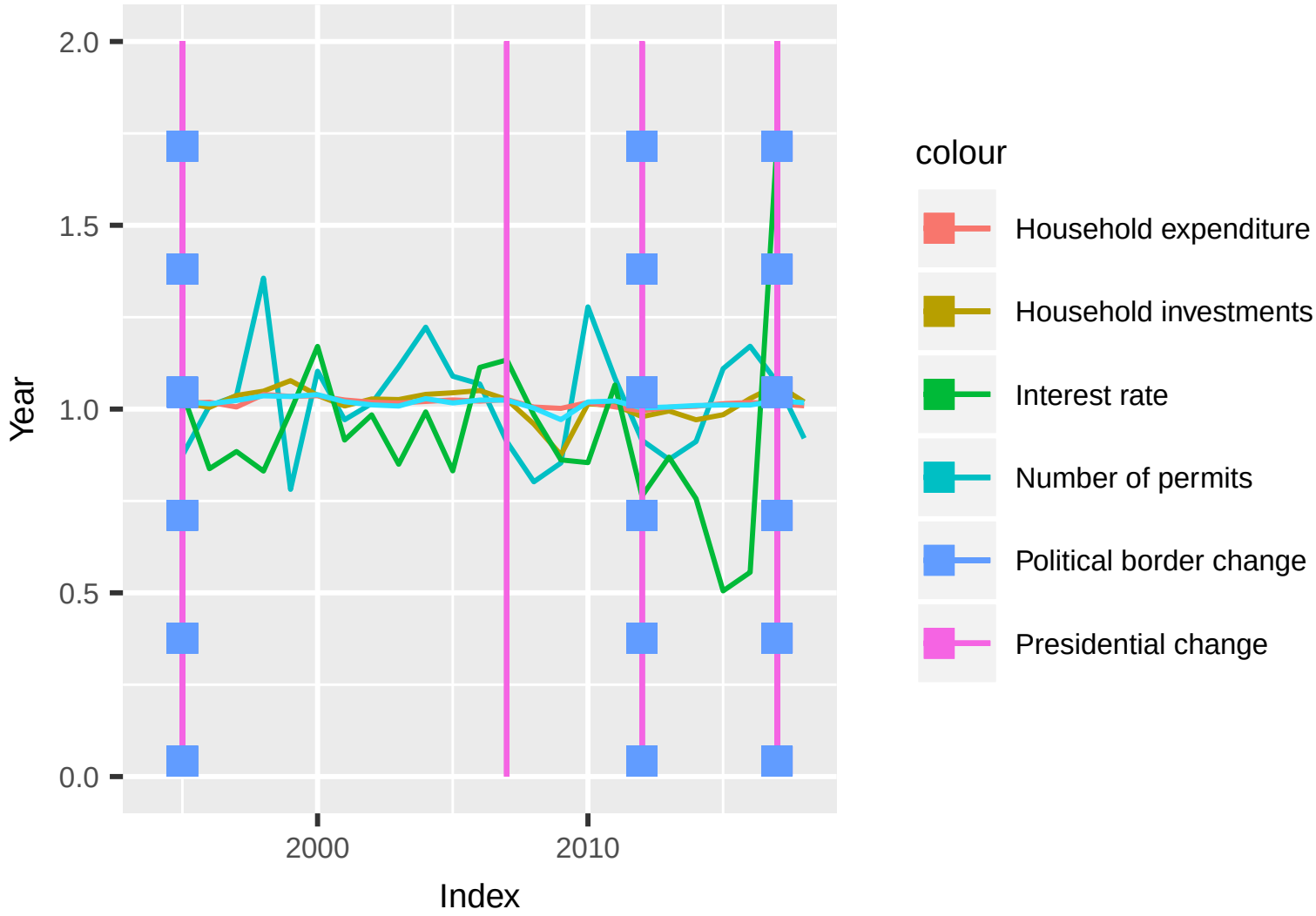
In our work we prefer to operate over indiced variables.

Statistic	Mean	St. Dev.
Year	12.000	7.360
GDP	1.017	0.014
Interest rate	0.936	0.238
Household spendings	1.017	0.011
Household investments	1.014	0.041
Number of permits	1.022	0.148
Population $<$ 20 y.o.	0.996	0.002
Population 20-59 y.o.	0.997	0.004
Population $>$ 60 y.o.	1.011	0.007

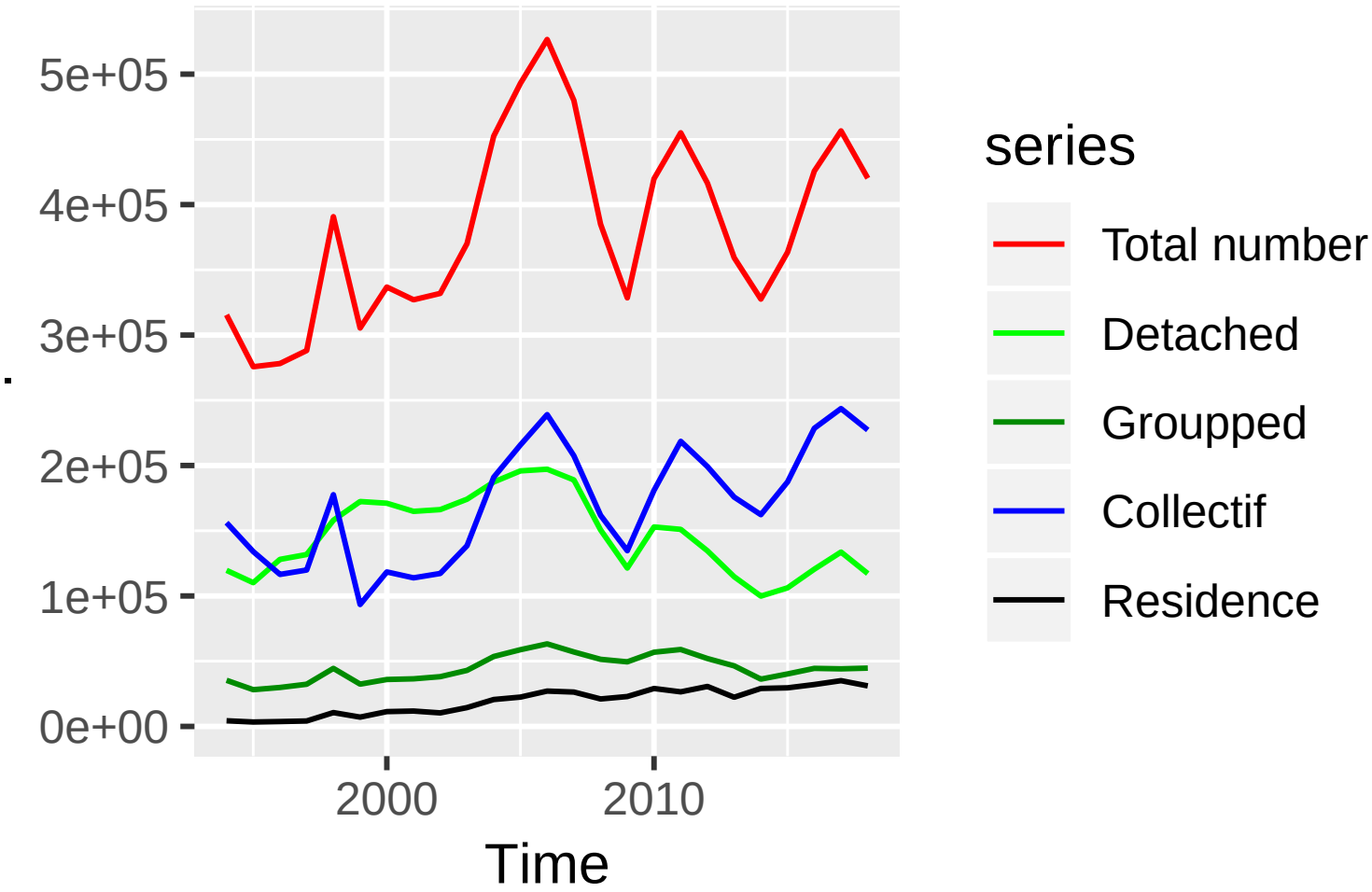
Table 3: Descriptive statistics for transformed data

Graphical representation

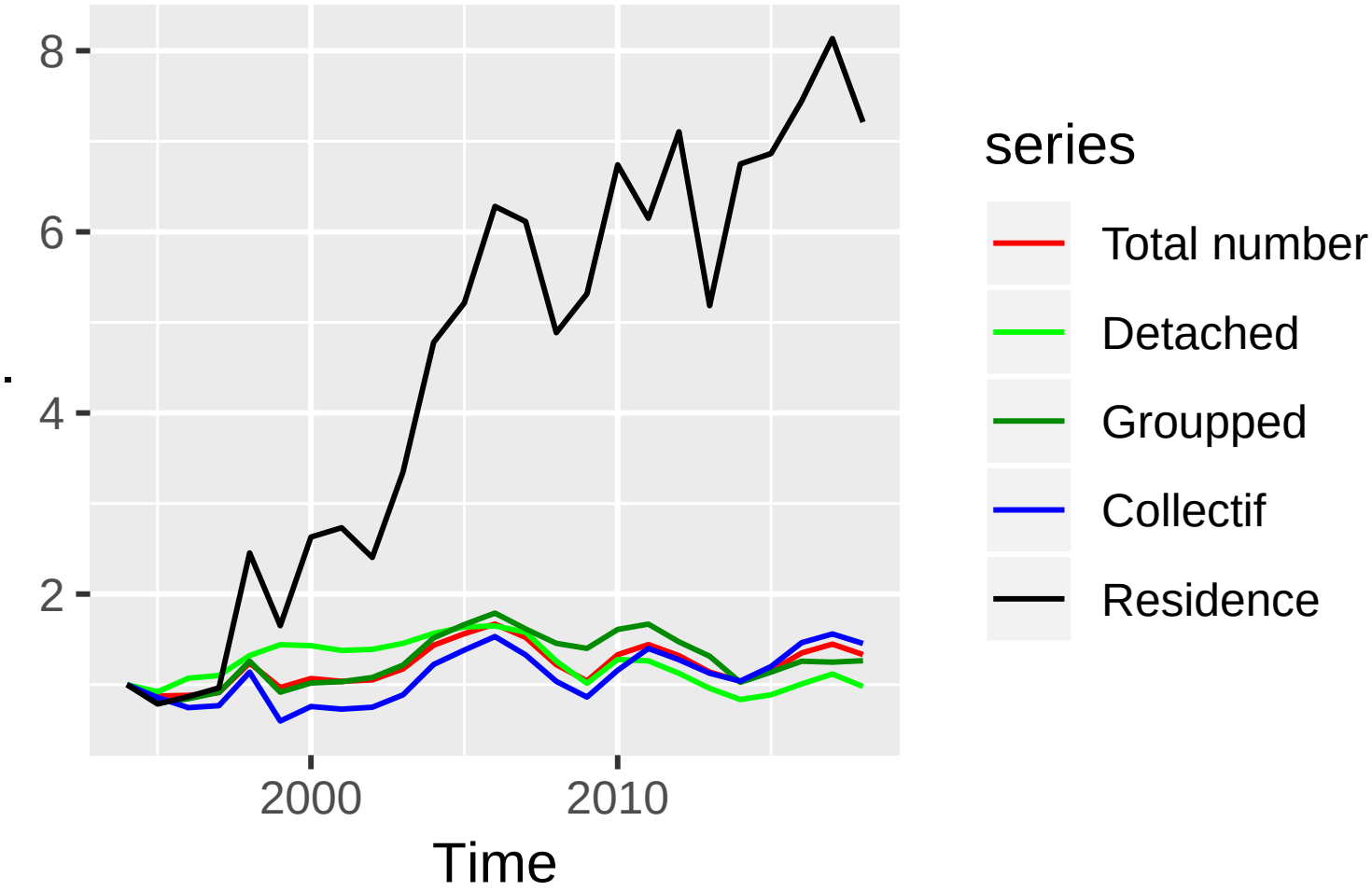
Political changes



Different housing types



Transformed data for different housing types



Data analysis

Analysis

We are going to procede in this section as follows :

- Cross-corelation study,
- Autocorrelation study,
- Partial autocorrelation study.

This analysis will be effectuated over the integrity of variables.

Cross-correlation study

In order to verify stationarity we use the following tests :

- augmented Dickey-Fuller test (ADF), which makes it possible to test the hypothesis of the non-stationarity of a time series;
- Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) to check stationarity on trend (KPSS-T) or level (KPSS-L).

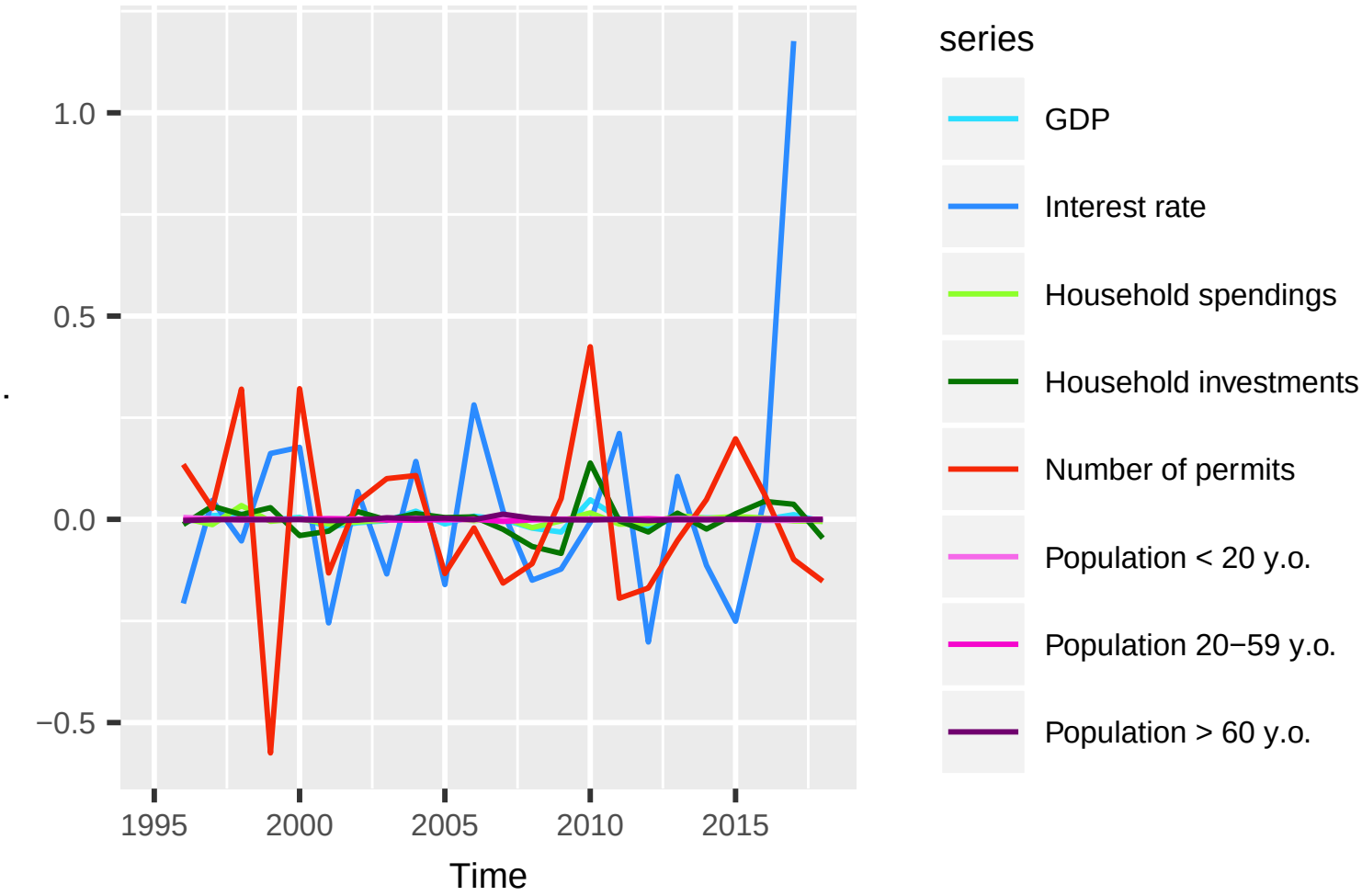
These two tests allow us to verify the hypothesis of the stationarity of the series studied and, if necessary, correct it by applying a transformation on the series in question.

Stationarity tests

	ADF	KPSS.T	KPSS.L
GDP	0.31	0.10	0.10
Interest rate	0.17	0.10	0.10
Household spendings	0.02	0.10	0.06
Household investments	0.42	0.10	0.10
Number of permits	0.01	0.10	0.10
Population $<$ 20 y.o.	0.52	0.10	0.06
Population 20-59 y.o.	0.67	0.10	0.01
Population $>$ 60 y.o.	0.57	0.10	0.03

Table 4: Stationarity tests, p-values

Differenced data

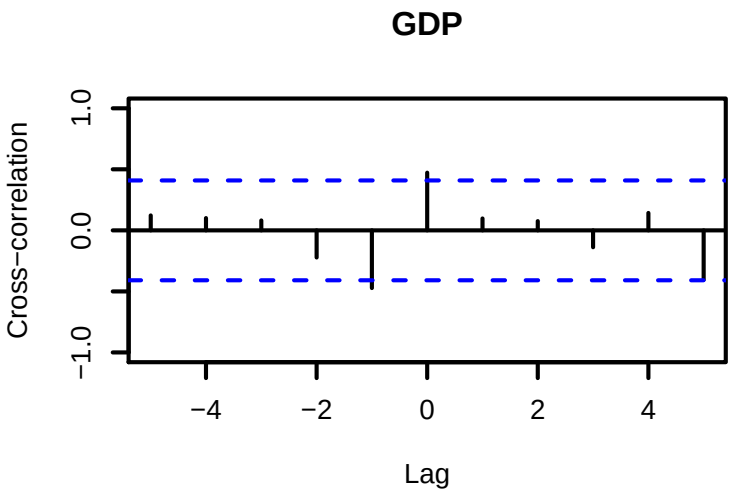


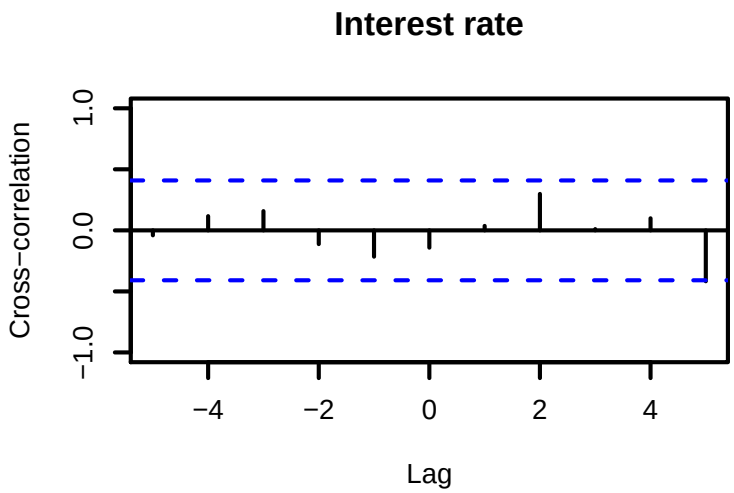
Stationarity reverification

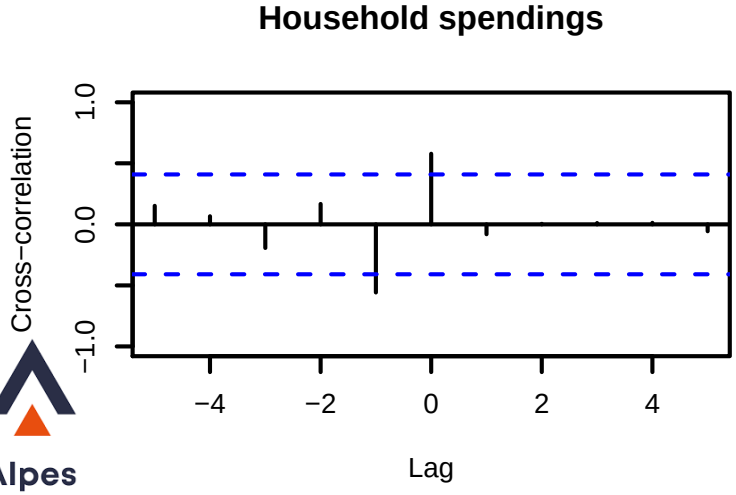
	ADF	KPSS.T	KPSS.L
GDP	0.03	0.10	0.10
Interest rate	0.29	0.10	0.10
Household spendings	0.04	0.10	0.10
Household investments	0.06	0.10	0.10
Number of permits	0.01	0.10	0.10
Population $<$ 20 y.o.	0.27	0.10	0.10
Population 20-59 y.o.	0.57	0.10	0.10
Population $>$ 60 y.o.	0.55	0.08	0.10

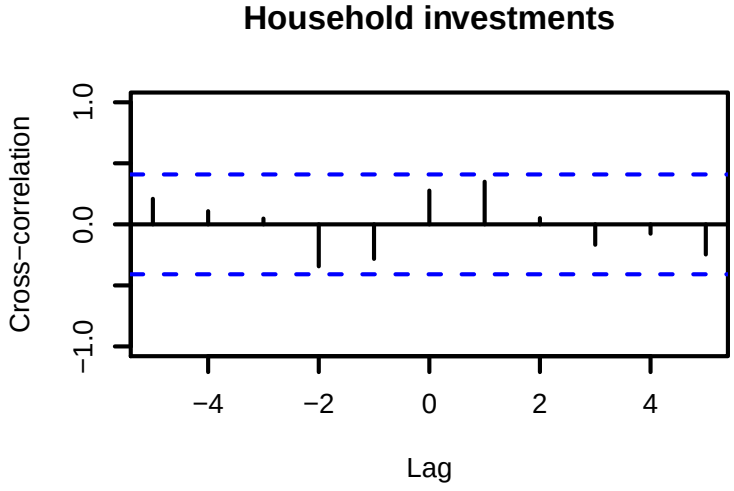
Table 5: Stationarity tests, p-values

Cross-correlation study

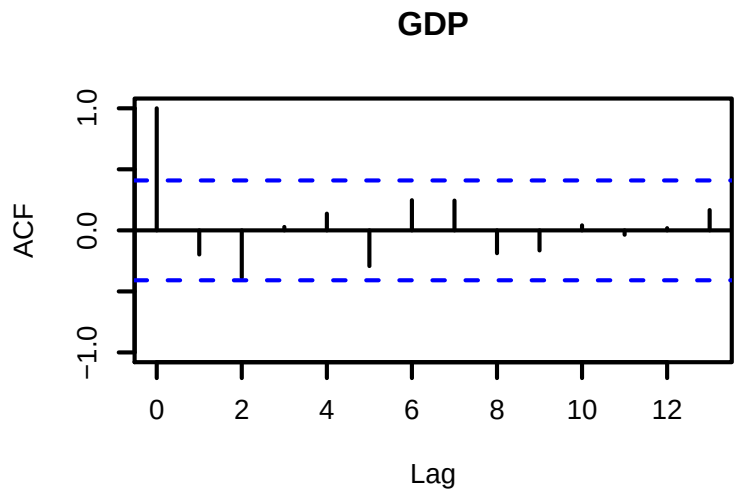


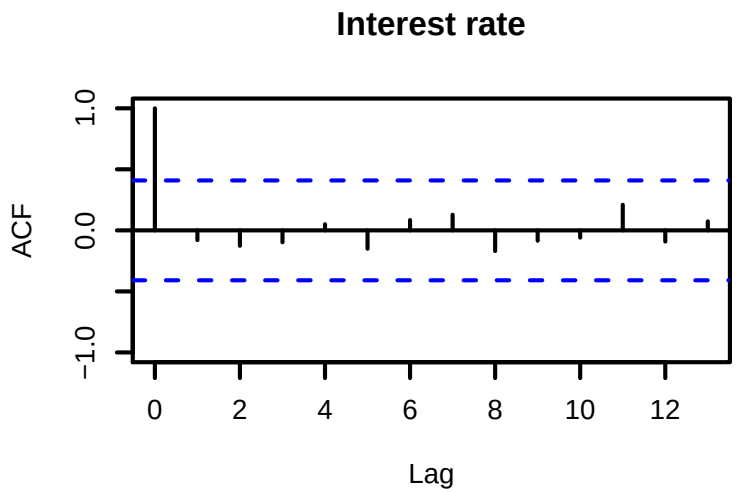


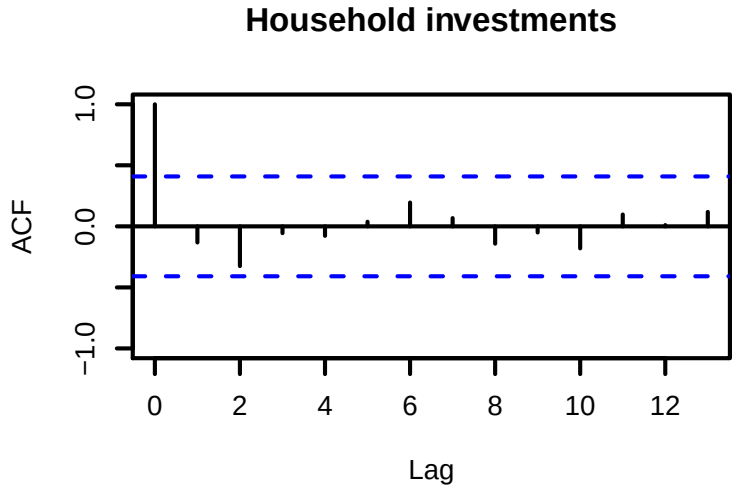




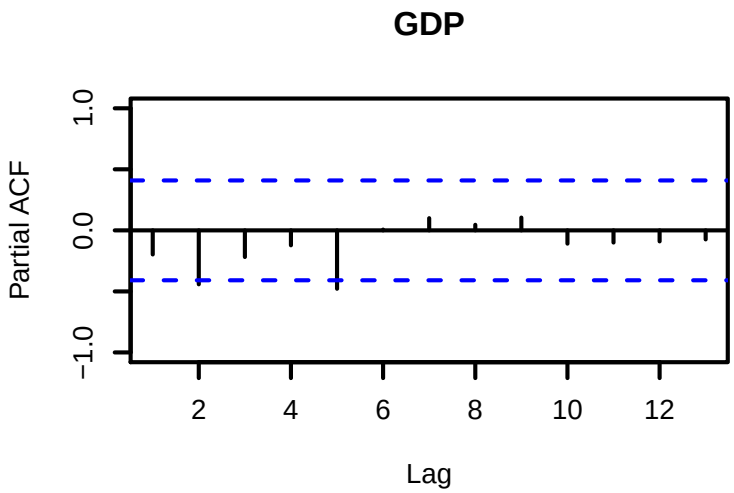
Autocorrelation

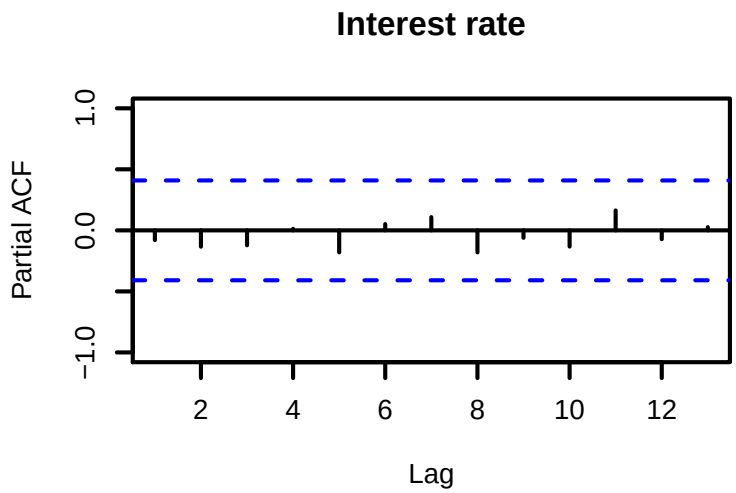


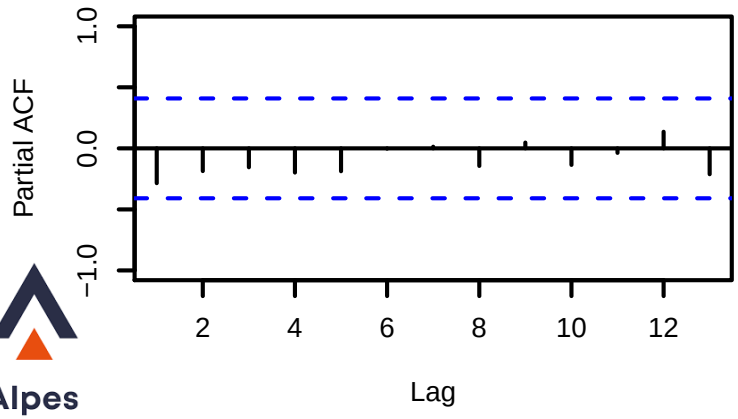


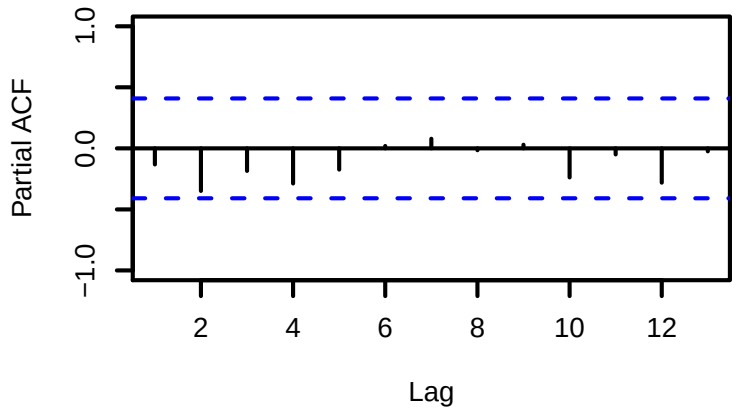


Partial autocorrelation









Modelisation

Different models

- Simple linear models (OLS),
- Time series models (ARIMA),
- Adcanced time series vector models (VAR),
- Alternative linear models, accounting for causality effects.

Model comparison

Model	Advantages	Disadvantages
OLS	Simple to implement	Unable to predict future values without acces to exogenous variables values in future
ARIMA	Relatively simple Allows to predict future values for several periods	Does not take into account supplimentary informa Requires stationary time series data
VAR	Even longer prediction horizon	Multiple restriction on data structure : - Stationarity, - No cointegration.
OLS and causality	Gives predictions for short intervals (1 or 2 periods) No limiting restrictions on the data used	Only short-term predictions are possible

Table 6: Different models' specification

Linear models

OLS estimations

In this part we explore direct links between varibales without wausality imlications.

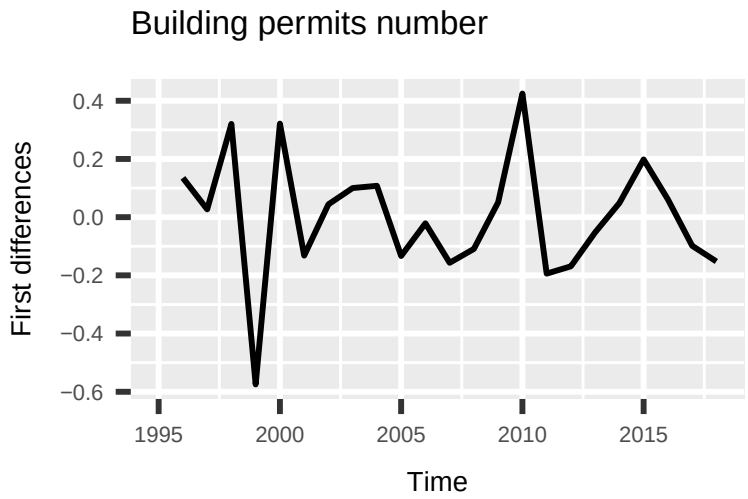
	Model 1	Model 2	Model 3	Model 4	Model 5
GDP	4.38*	5.40*	1.10	2.00	2.98
	(2.10)	(2.22)	(4.89)	(4.94)	(4.93)
Interest rate		−0.17		−0.18	−0.18
		(0.13)		(0.14)	(0.14)
Household spendings			3.13	1.77	3.38
			(3.86)	(4.02)	(4.17)
Household investments			0.56	0.92	0.68
			(1.47)	(1.49)	(1.48)
Time					0.01
					(0.01)
R^2	0.17	0.24	0.20	0.26	0.32
Adj. R^2	0.13	0.16	0.08	0.10	0.12
Num. obs.	24	23	24	23	23
RMSE	0.14	0.14	0.14	0.14	0.14

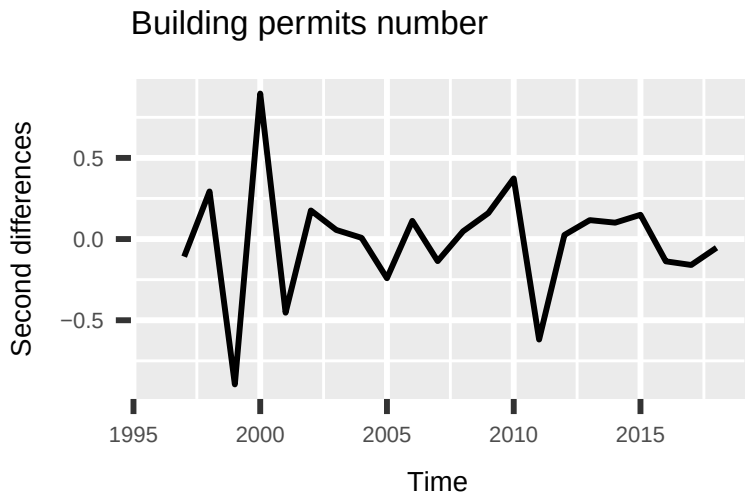
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

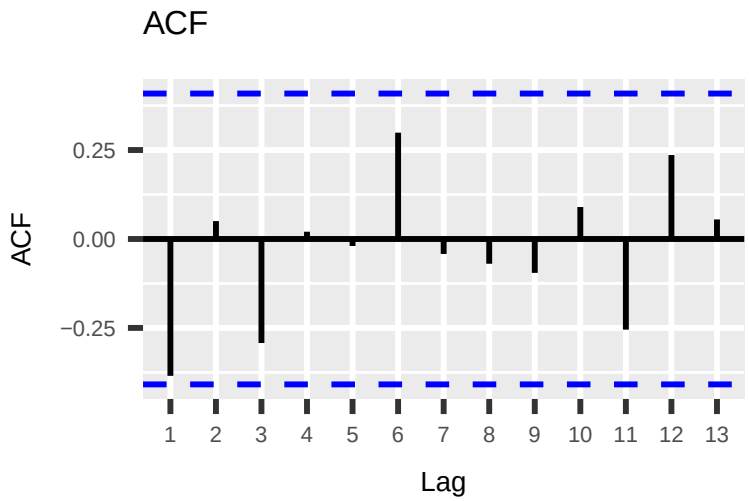
Table 7: OLS models comparison

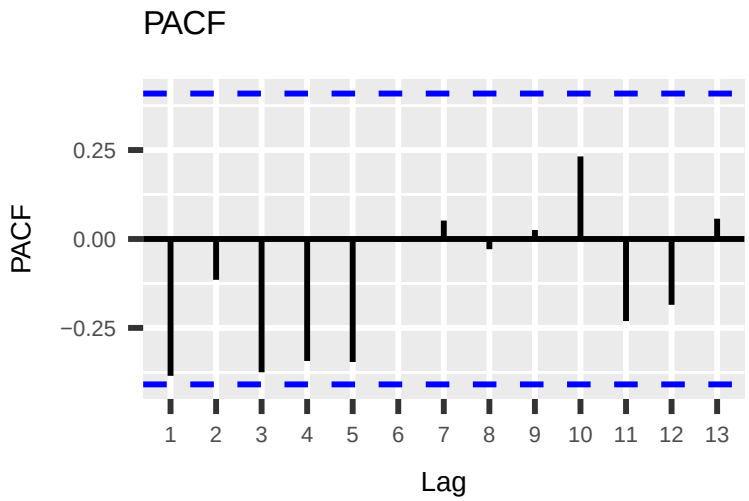
Time series

Time series presentation









ARMA modelisation

	ARMA(1,0)	ARMA(0,1)	ARMA(1,1)	ARMA(1,2)	ARMA(2,2)
AR(1)	−0.38		0.04	−0.96	−0.81
	(0.19)		(0.22)	(0.15)	(0.28)
AR(2)	0.00	−0.00	−0.00	−0.00	−0.00
	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept		−1.00	−1.00	−0.09	−0.12
		(0.12)	(0.12)	(0.24)	(0.23)
MA(1)				−0.91	−0.88
				(0.24)	(0.22)
MA(2)					0.13
					(0.22)
AIC	-4.98	-13.34	-11.38	-9.76	-8.11
Log Likelihood	5.49	9.67	9.69	9.88	10.06

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

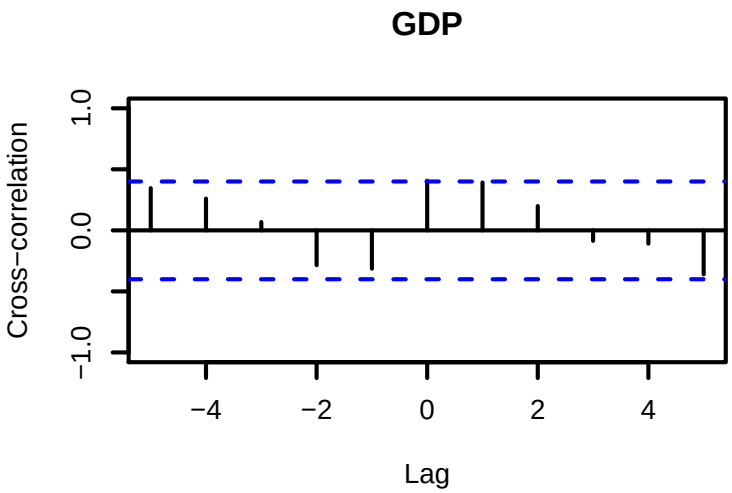
Table 8: Comparaison des modèles ARMA

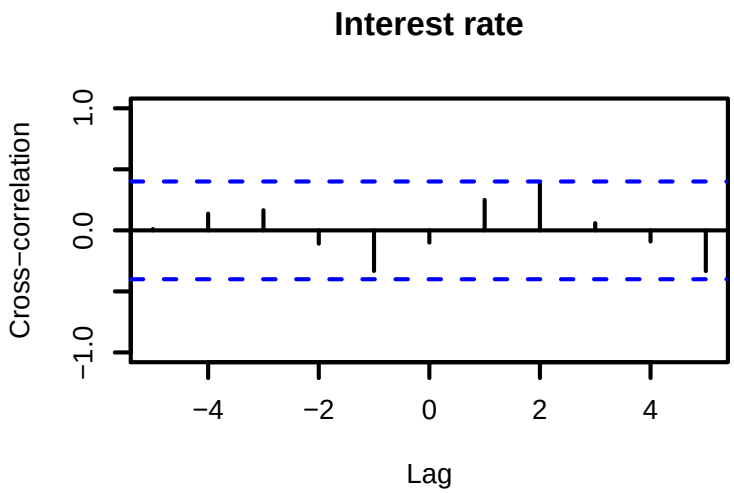
Vector models

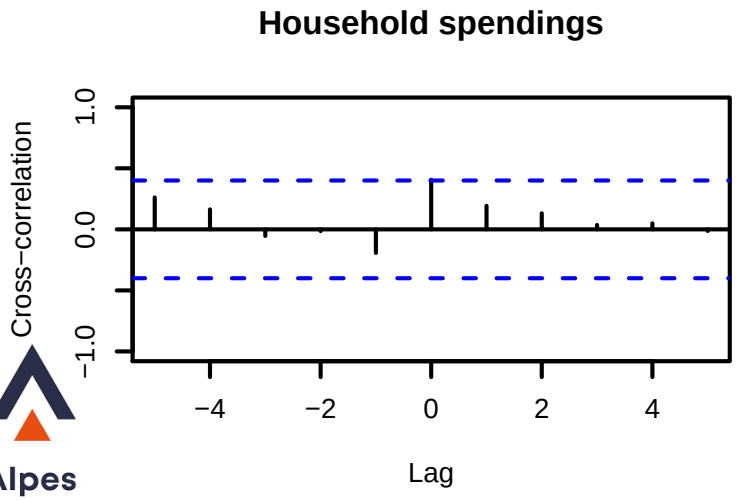
VAR modelisation

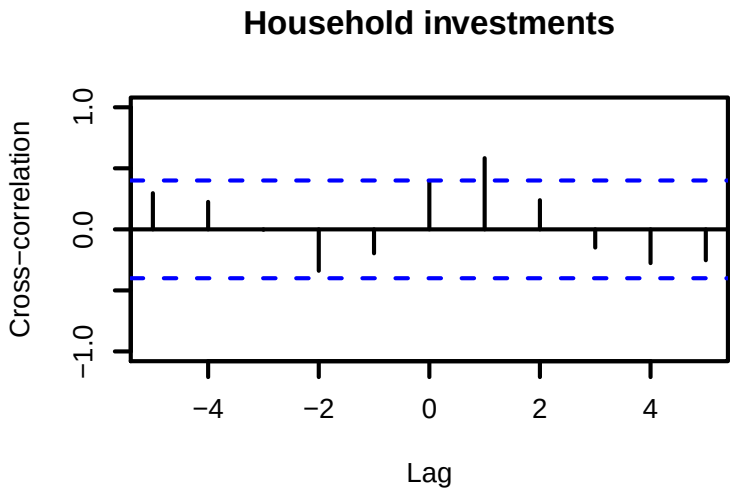
Alternatives

Review of cross-correlations, part 1

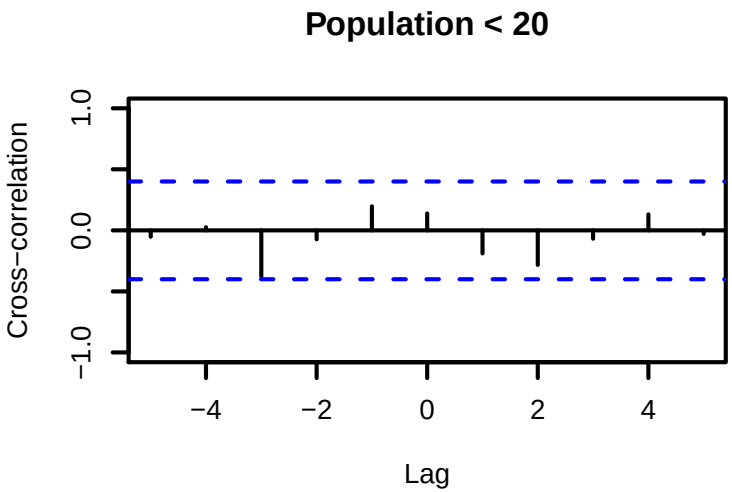


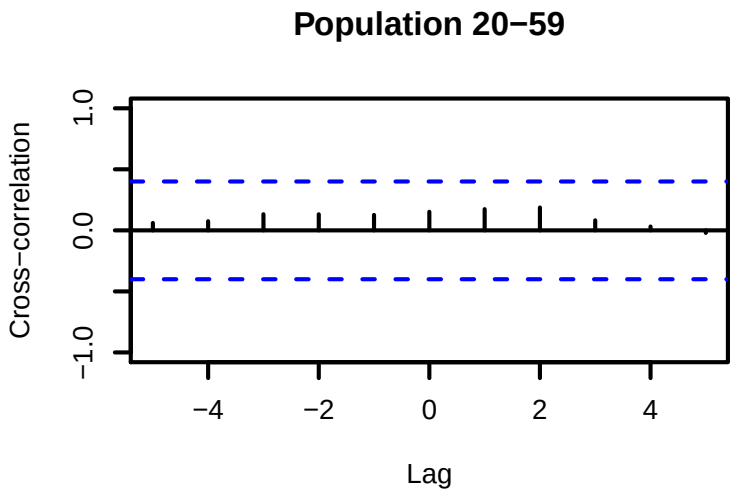


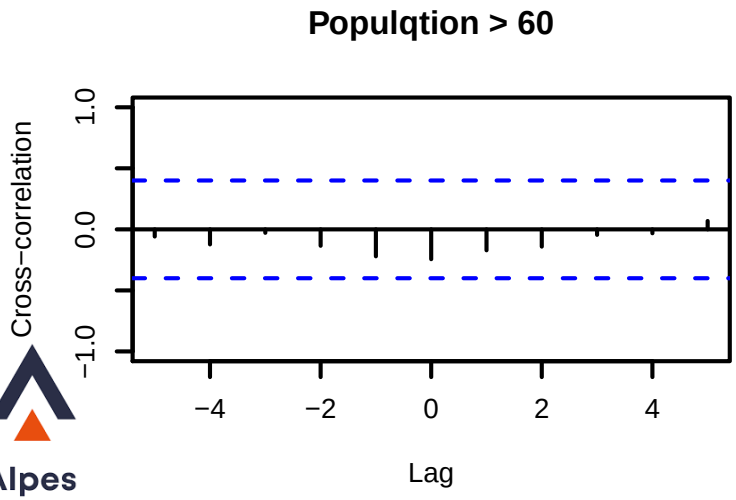




Review of cross-correlations, part 2







Conclusion

Proceedings

- Separate study of different permits types,
- Alternative linear regression implementation (that is similar to ARDL),
- Study of trimestrial data,
- Rapport finalisation,
- Delivrables preparation :
  - Excel automated model,
  - Final rapport,
  - R code organisation,
  - Database documentation.

Université Grenoble Alpes

Schneider Electric