Solar energy consumption in Spain

Laia Porcar
Philippe Robert
Luis M. López

1.
Intro

Types of solar energy, Importance of boosting solar energy consumption, Brief history of solar energy in Spain.

Interesting facts

Spain is one of the countries in Europe with most solar hours.

Solar power is the most abundant energy source on earth.

Solar power plants can last 40 years or more.

China is the world leader in solar energy… by a lot. (absolute value)

Solar is the fastest energy source to deploy and doesn't produce any noise pollution.

In a single hour, the amount of power from the sun that strikes the Earth is more than the entire world consumes in one year.

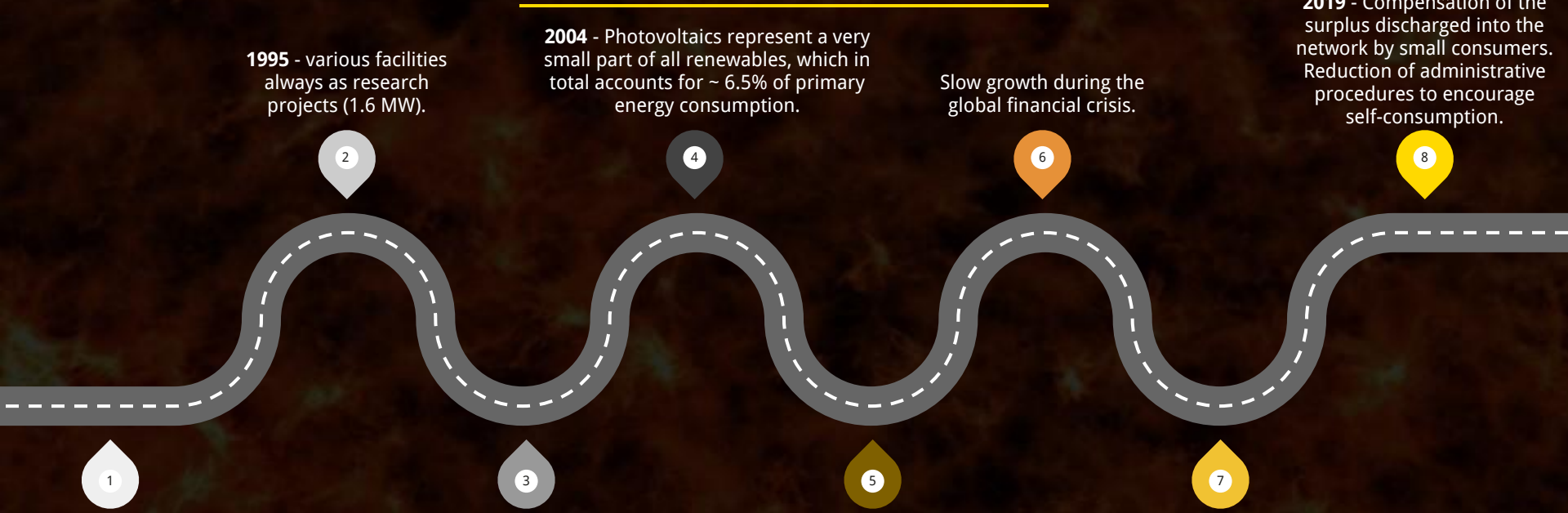
Why promote the use of solar energy in Spain?

Great amount of sunshine hours.

European commitments to install renewable energies.

Convenience of reducing the great dependence on external energy and increasing energy autonomy.

Brief history of solar energy in Spain



1. **1984** - Iberdrola installed the first P.V. plant connected to the grid (100 kWp).
2. **1995** - various facilities always as research projects (1.6 MW).
3. Some incentives to promote use.
4. **2004** - Photovoltaics represent a very small part of all renewables, which in total accounts for ~ 6.5% of primary energy consumption.
5. **2007** - set premiums and fixed regulated rates (high profitability, investments)
6. Slow growth during the global financial crisis.
7. **2015** - approval of the "sun tax" for users with self-consumption facilities. For maintenance of the conventional electrical network
8. **2019** - Compensation of the surplus discharged into the network by small consumers. Reduction of administrative procedures to encourage self-consumption.

What types of solar energy are there?

Photovoltaic

Photovoltaic modules transform solar radiation into electrical energy.

Thermal

Solar collectors convert solar radiation into heat.

The sun's energy is harnessed directly by means of materials and construction solutions.

Passive

Combination of solar energy with other renewable energies (mainly wind).

Hybrid

2.
The data

Source and structure of our data, Questions we aim to answer.

Our data

- Diverse datasets from Our World in Data, all merged into one dataframe of 57 rows and 16 columns:
  - Year.
  - PV price.
  - Installed capacity.
  - Share from solar.
  - Per capita solar energy consumption in Spain, USA, Germany, China, Australia and India.
  - Per capita wind, nuclear, oil, gas and coal energy consumption in Spain.
- Only useful from year 1989.

Questions we aim to answer

- How will the solar energy consumption in Spain behave in the next 3-4 years?
- How does the data reflect the history of solar energy in Spain?
- What influence does Spain have in other countries? What countries influence Spain?
- What has the introduction of solar energy in Spain represented for other energies?

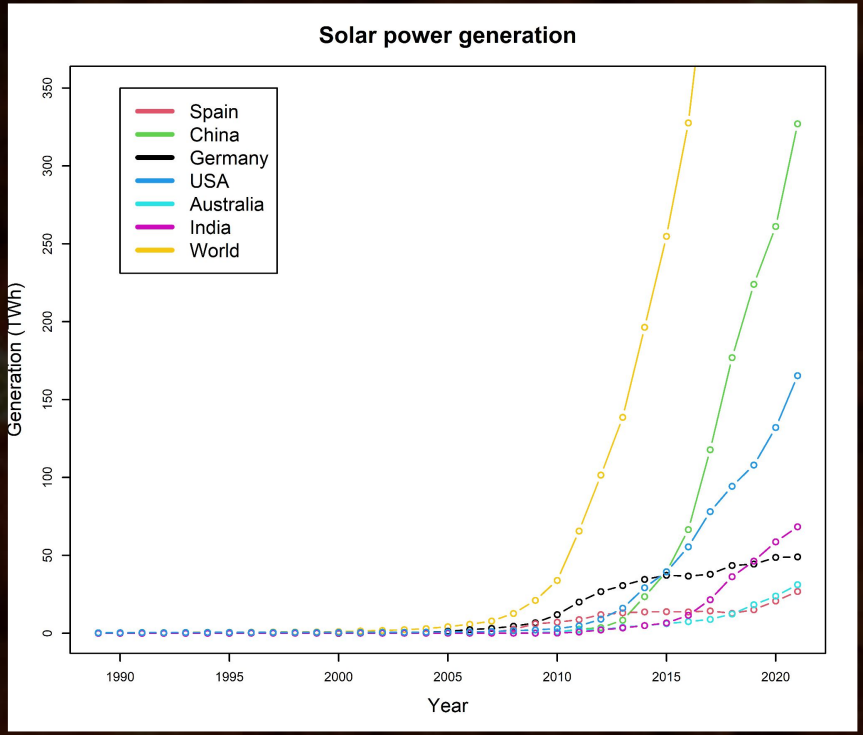
Our process

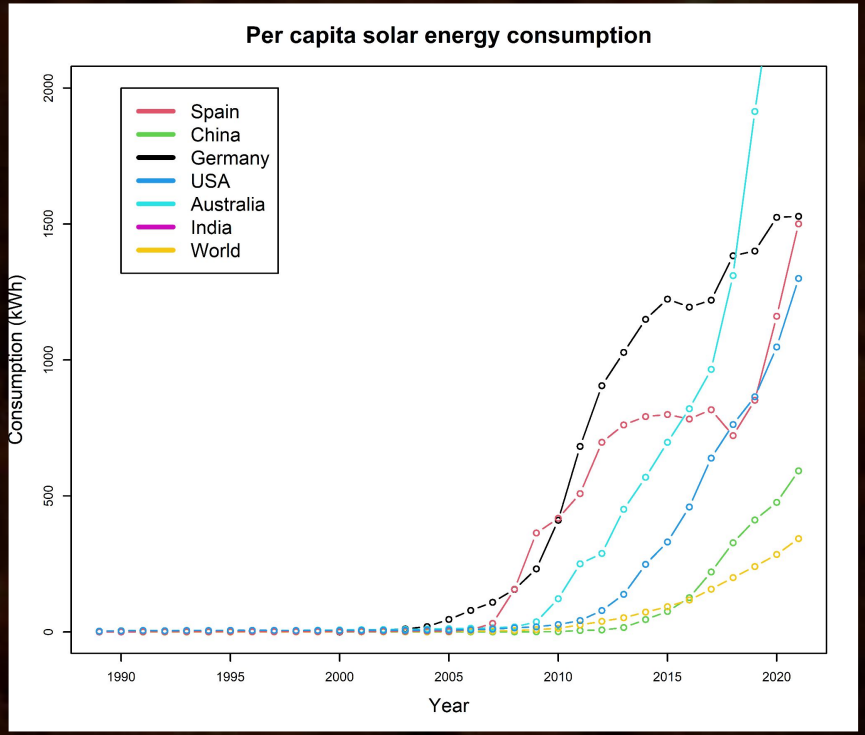

EDA
Univariate diffusion models
Competition between energies
Influence on/of other countries
GAM and gradient boosting

3.
Exploratory data analysis

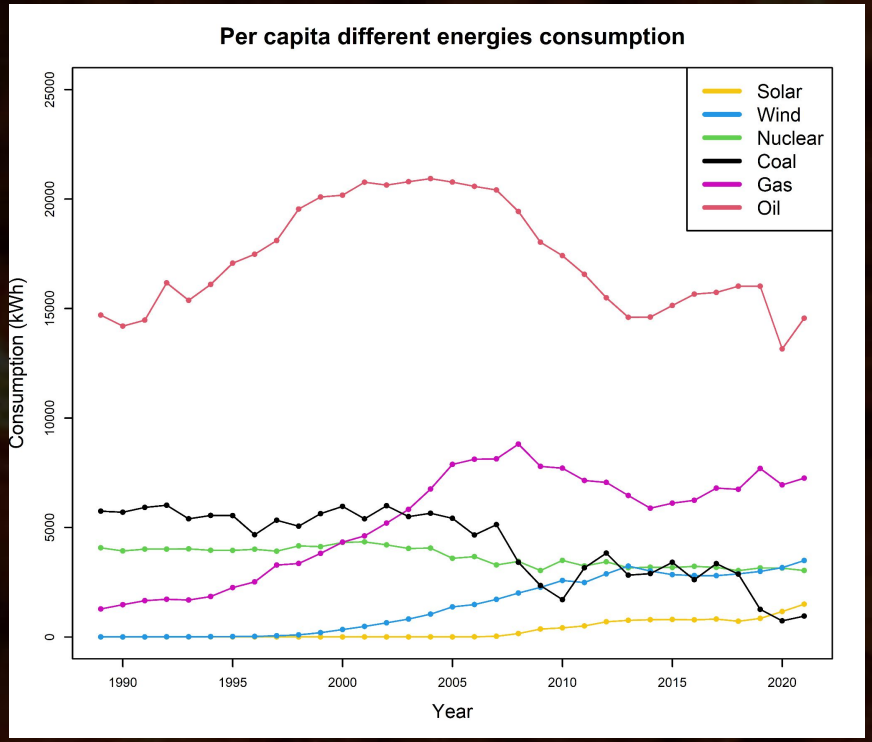
Solar power generation, Per capita solar energy consumption, PV capacity, PV price.

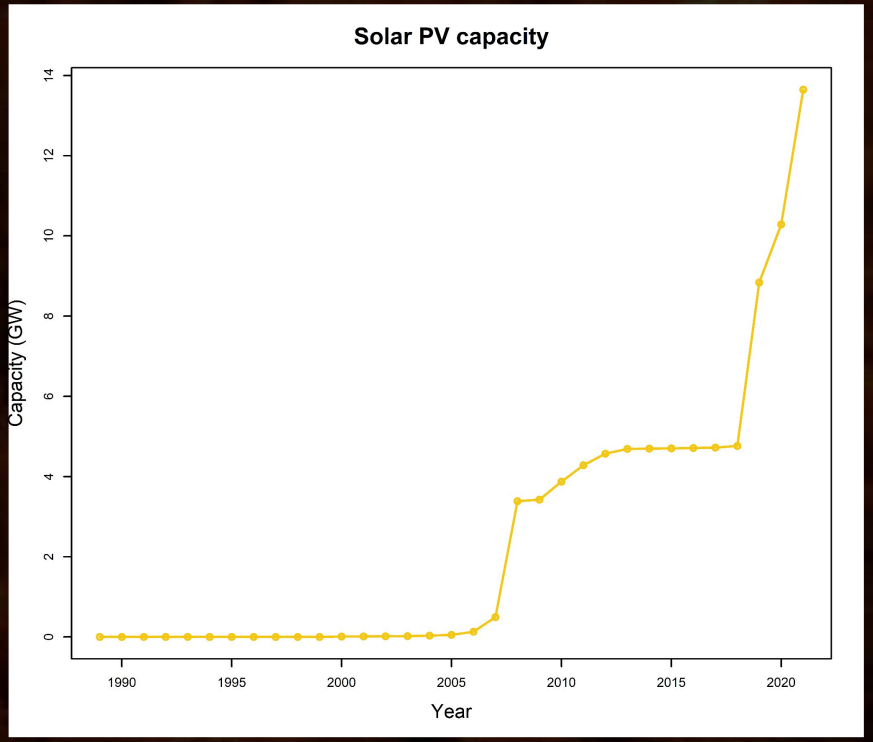
Some EDA


Solar power generation
Spain
China
Germany
USA
Australia
India
World
Generation (TWh)
0
50
100
150
200
250
300
350
1990
1995
2000
2005
2010
2015
2020
Year

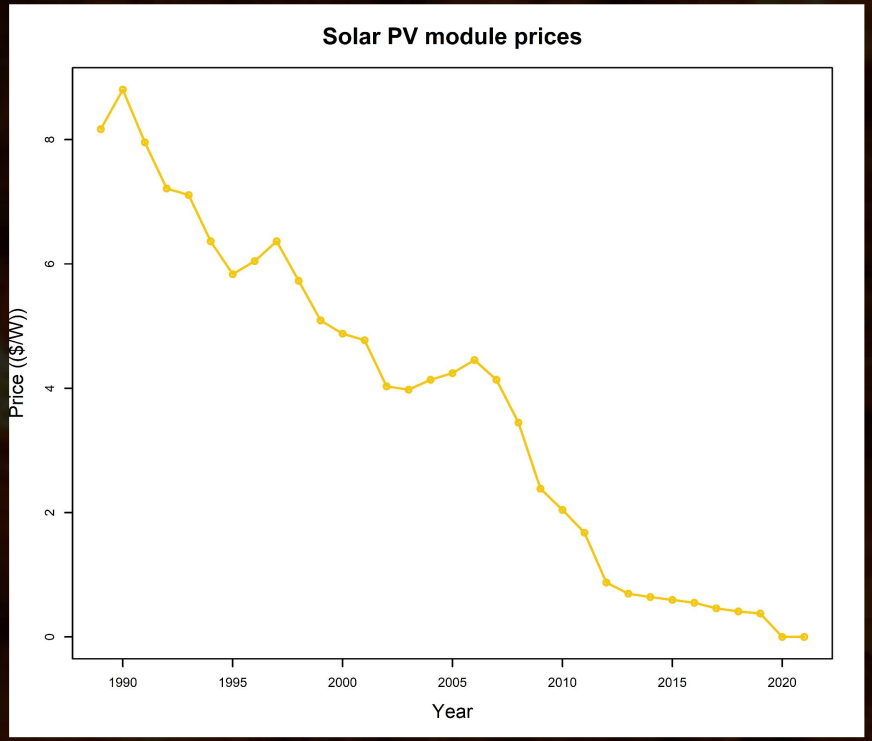

Per capita solar energy consumption
Spain
China
Germany
USA
Australia
India
World
Consumption (kWh)
0
500
1000
1500
2000
1990
1995
2000
2005
2010
2015
2020
Year

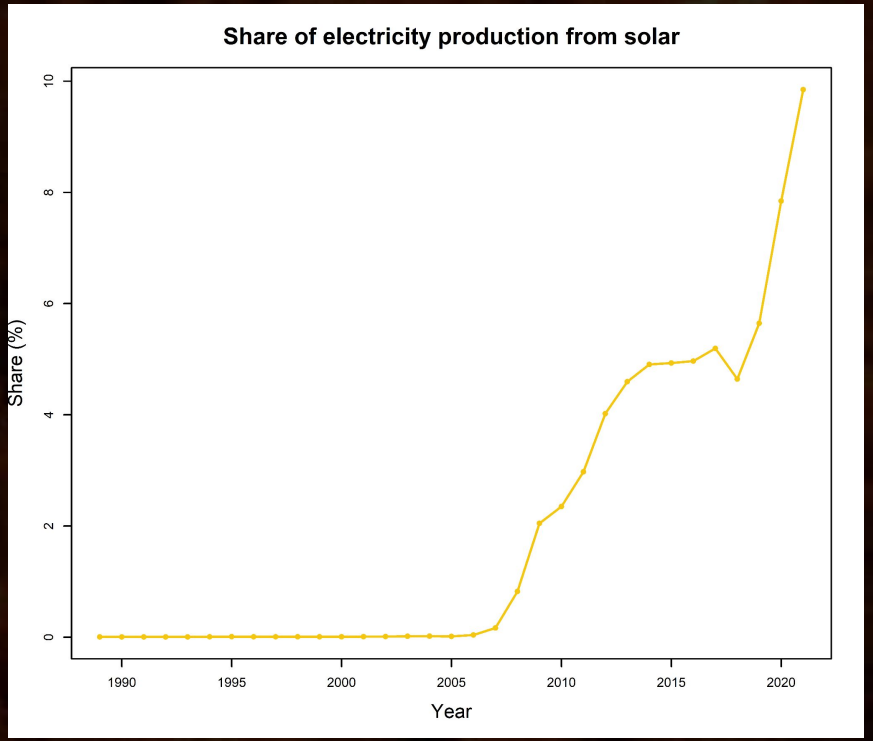
Some EDA


Per capita different energies consumption
Solar
Wind
Nuclear
Coal
Gas
Oil
Consumption (kWh)
0
5000
10000
15000
20000
25000
1990
1995
2000
2005
2010
2015
2020
Year


Solar PV capacity
Capacity (GW)
0
2
4
6
8
10
12
14
1990
1995
2000
2005
2010
2015
2020
Year

Some EDA


Solar PV module prices
Price (($/W))
0
2
4
6
8
1990
1995
2000
2005
2010
2015
2020
Year


Share of electricity production from solar
Share (%)
0
2
4
6
8
10
1990
1995
2000
2005
2010
2015
2020
Year

4.
Univariate diffusion models

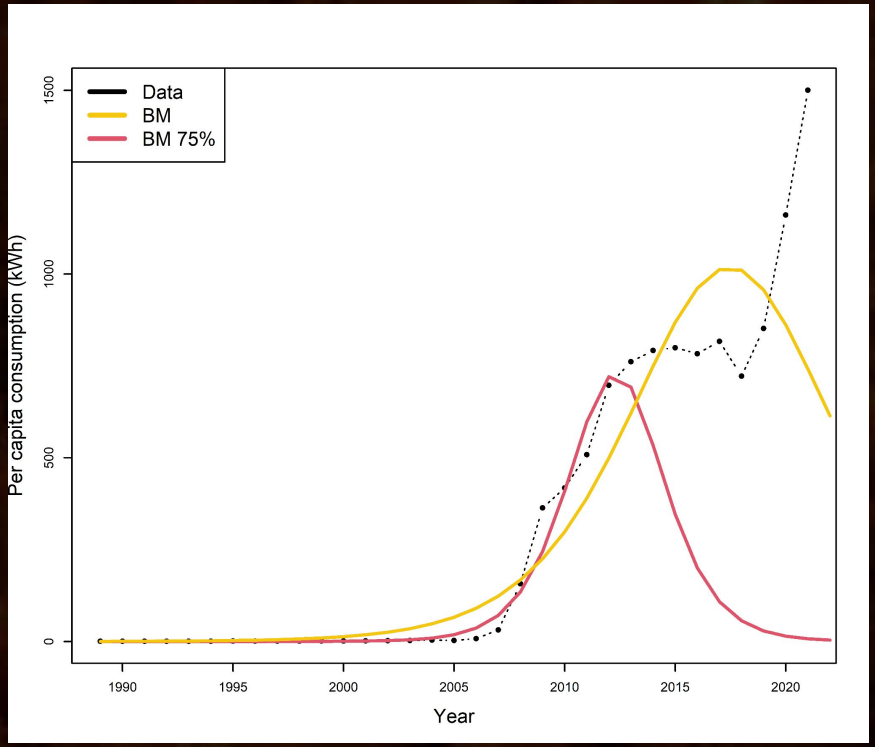
Bass model, Generalized bass model and Guseo-Guidolin model

Bass model

Full BM:

- $m = (1.24 \pm 0.08) \times 10^4$
- $p = (2.4 \pm 1.0) \times 10^{-5}$
- $q = 0.33 \pm 0.02$
- $R^2 = 0.9955479$
- $RSS = 1.7 \times 10^6$

Process governed by imitation.



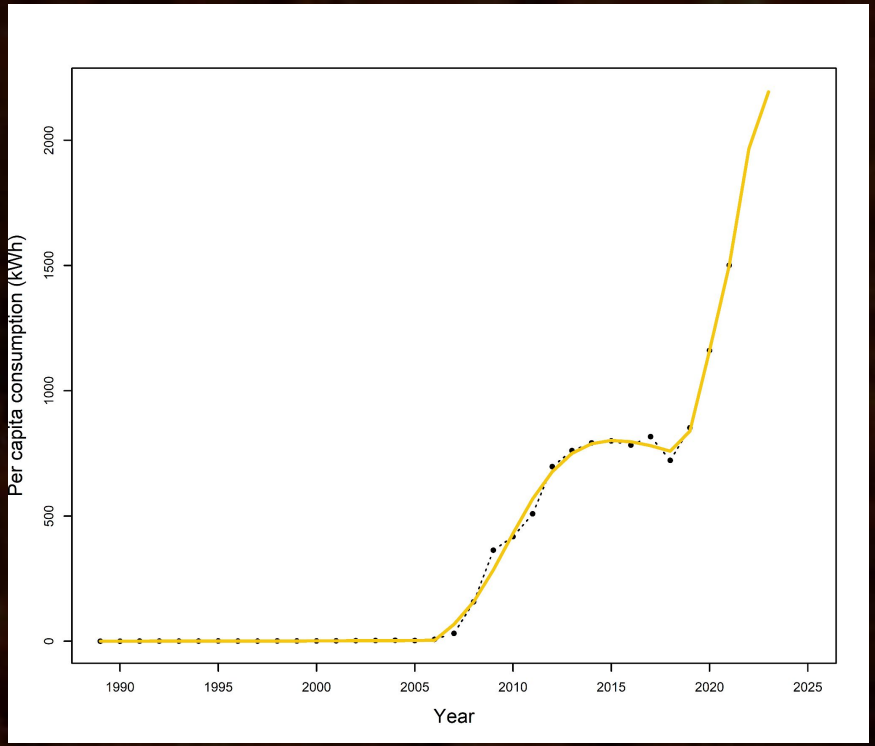
GBM with two exponential shocks

Exponential shocks:

- 2007: set fixed premiums and regulated taxes.
- 2019: elimination of the sun tax.

Parameters:

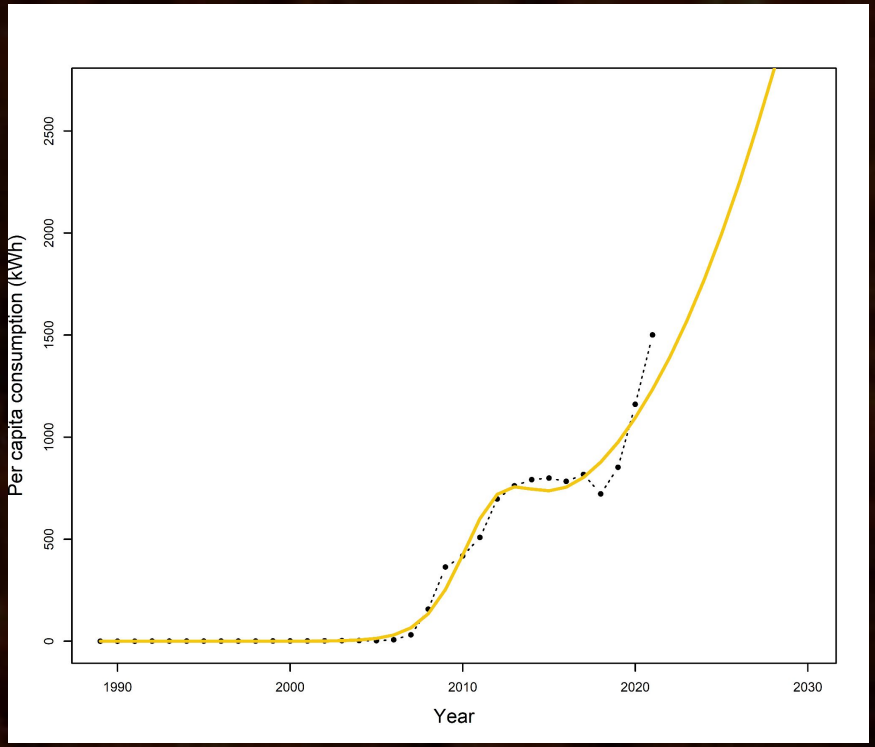
- m = (1.24 ± 0.08) x 10^4
- Negative b1: exponential decaying behavior. Significant.
- Positive b2: exponential increasing behavior. Not significant.
- a1 and a2 strongly significant.
- R^2 = 0.9999759
- RSS = 7 x 10^3



GGM

Parameters:

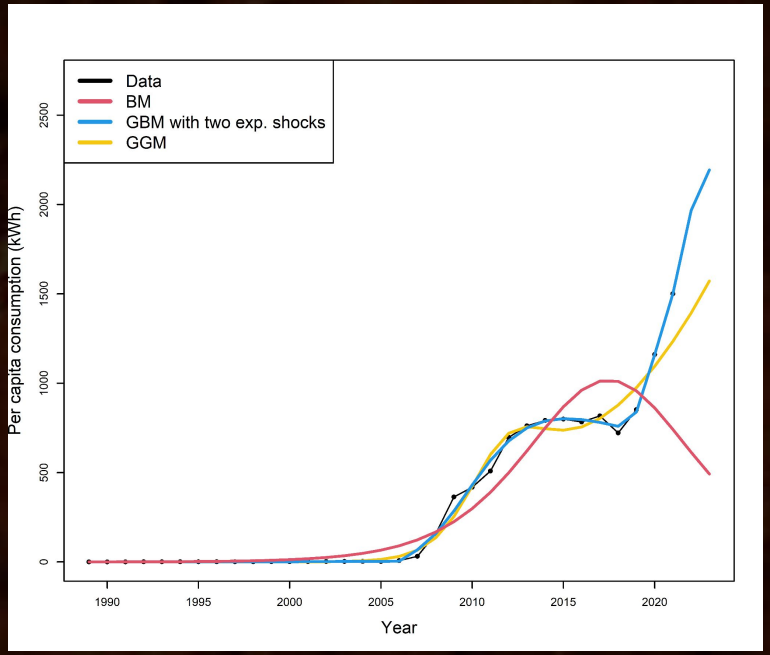
- $R^2 = 0.9996121$
- $RSS = 1.1 \times 10^5$

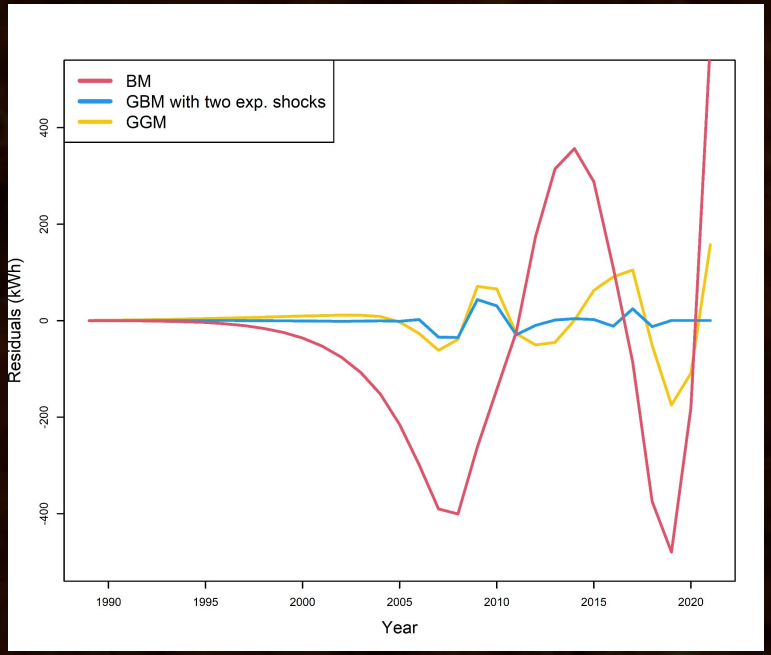


Model comparison

Best is GBM: lowest r-squared, trend well captured, might overfit.

Worst is BM: highest r-squared, trend not captured, too simple.



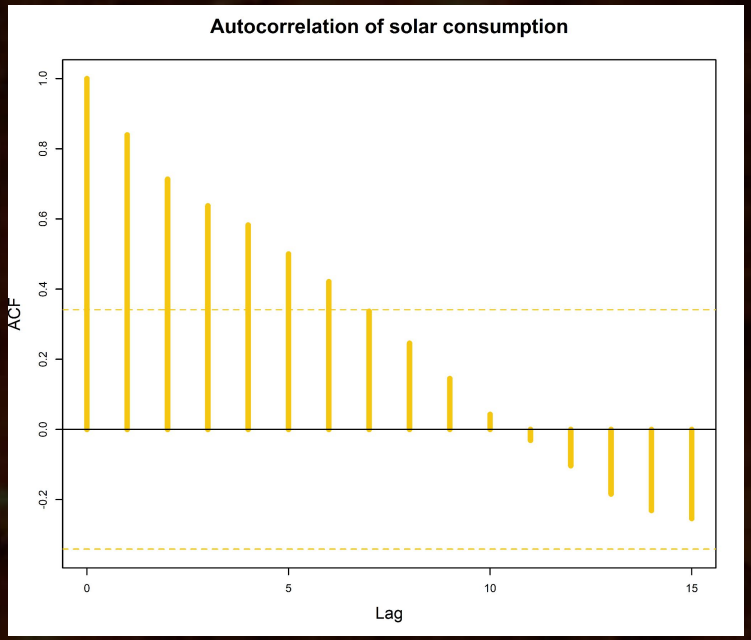


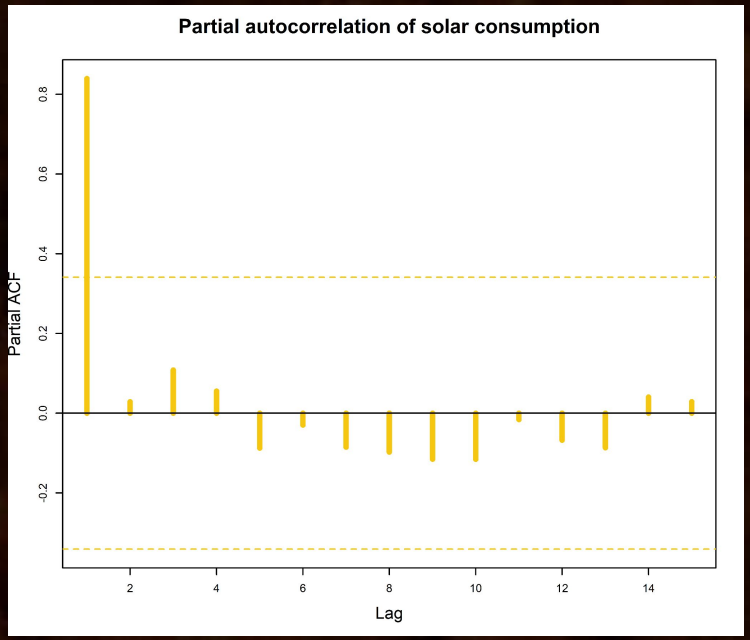
5.
ARIMA model

For Business Plans, Marketing Plans, Project Proposals, Lessons, etc

Autocorrelation of solar per capita consumption

Clear presence of a trend in ACF, AR component of order 1 seems appropriate in PACF



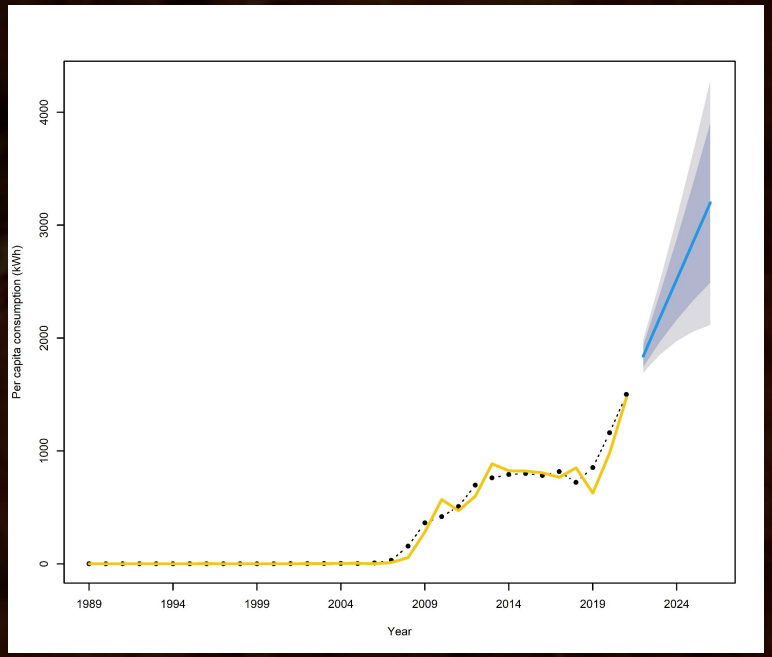


More on ARIMA

Auto ARIMA and forecast.

ARIMA(0,2,0)

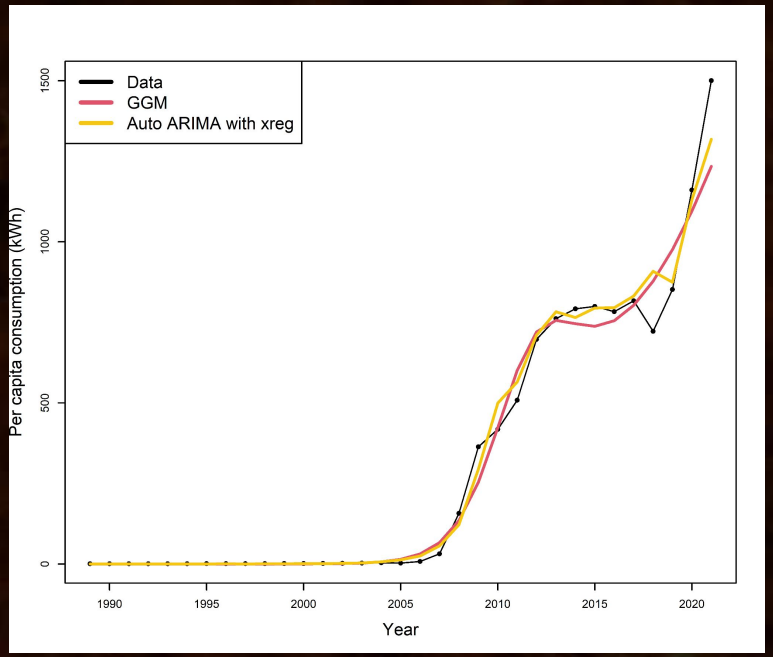
AIC = 357.2, RMSE = 70.1



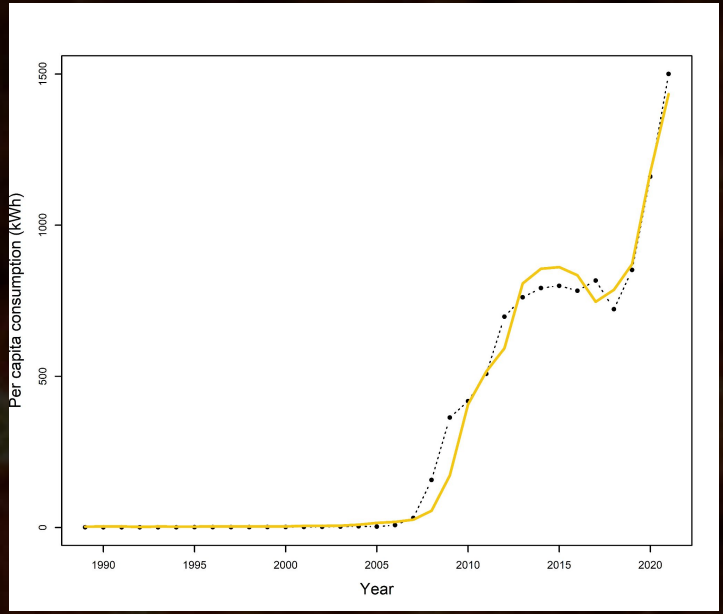
Auto ARIMA with GGM as external regressor.

ARIMA(0,0,1)

AIC = 361.01, RMSE = 51.7



More on ARIMA

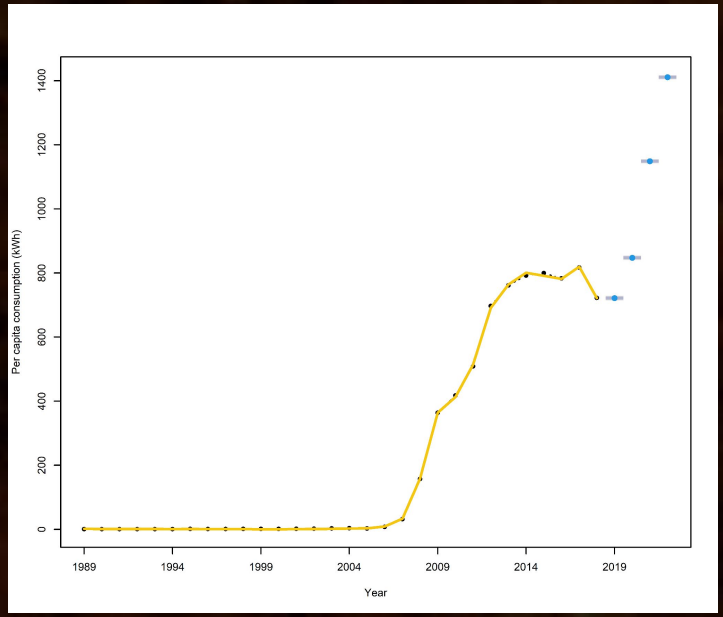


Auto ARIMA (1,0,0) with per capita consumption of countries as external regressors.

AIC = 365.9, RMSE = 50.9

Auto ARIMA (0,0,0) with per capita consumption of countries, share from solar, PV price and capacity as xregs. Forecast.

AIC = 167.1, RMSE = 3.00



6.
Multivariate diffusion models

Competition between resources.

Competition between solar and wind

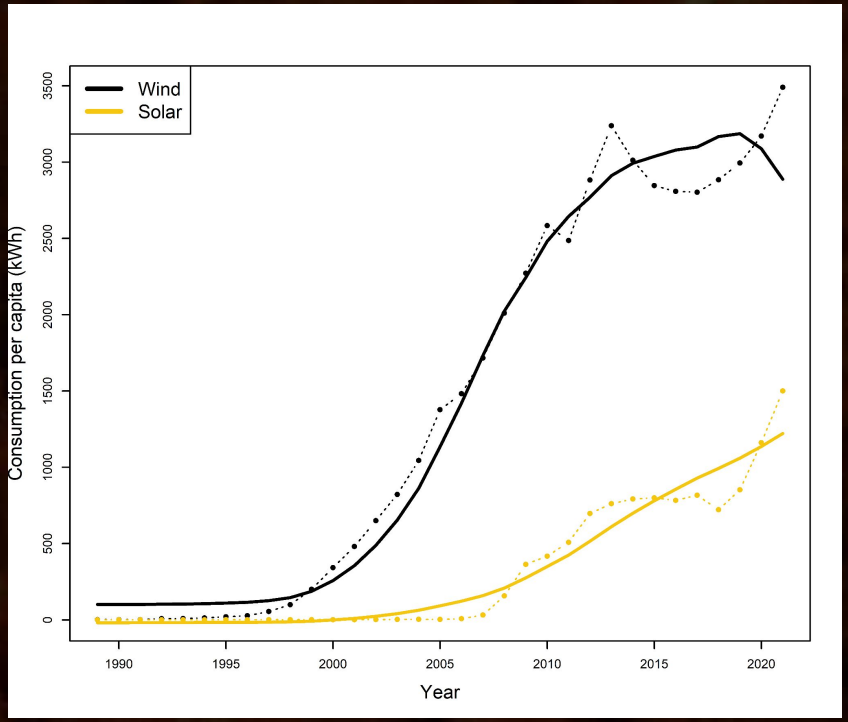
Parameters:

- q_{1c} + δ = 0.21: pos. and signif.
- p_{1c} (0.01) and q2 (0.07) not signif.
- q_{1c} = -0.57: neg and signif.
- q_2 - γ = 0.02: posiv. but not signif.

Within imitation coef. (q_{1c} + δ) indicate growth trend for wind.

Cross imitation coef (q_{1c}) indicates the solar impose competitive pressure and considered as mature technology.

Poor fit for both wind and solar energy: model underestimate growth.



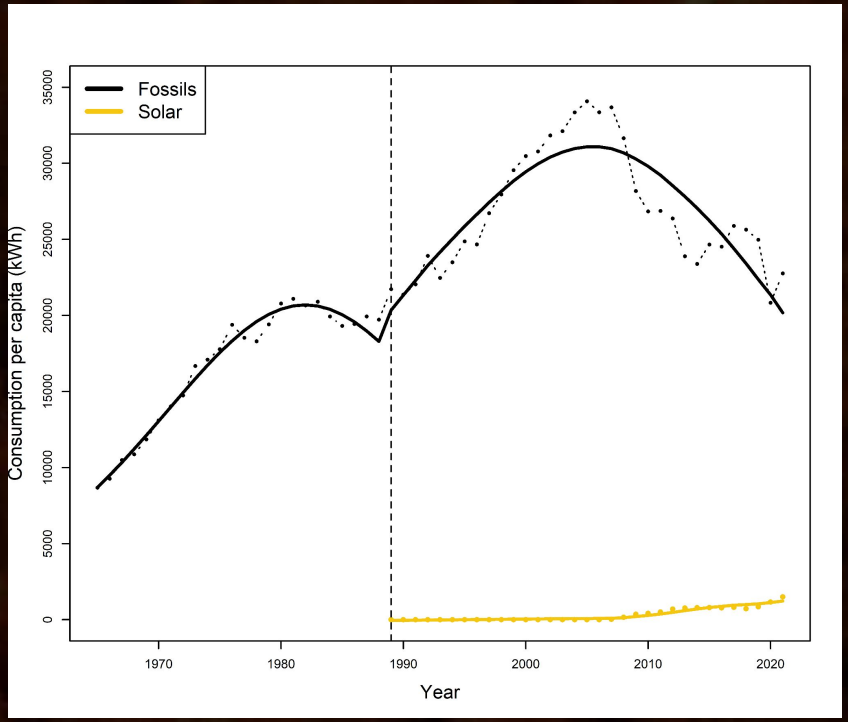
Competition between solar and fossils

Parameters:

- p_{1c} = 0.02: pos. and signif.
- q_{1c} + δ (0.06) and q2 (0.38) not signif.
- q_{1c} = -0.61: neg but not signif.
- q_2 - γ = 0.00: neutral but not signif.

Innovation coefficient (p_{1c}) slightly pos. indicating a small growth trend.

Good fit for both fossils and solar energy.



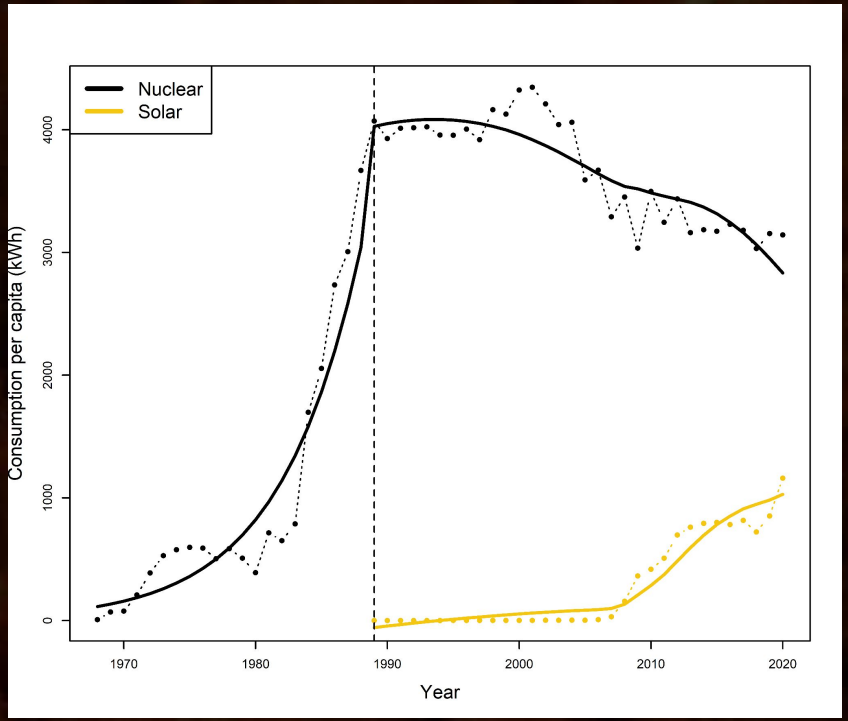
Competition between solar and nuclear

Parameters:

- p_{1c} = 0.02: pos. and signif.
- q_{1c} + δ = 0.03 pos. not signif.
- q_2 = 0.39: pos. and signif.
- q_{1c} = 0.30: pos. but not signif.
- q_2 - γ = 0.00: neutral and signif.

Within imitation coef. (q_2) bigger than innovation coef. (p_{1c}) indicating solar has higher growth.

Cross imitation coef. (q_2 - γ) neutral, might indicate that nuclear has no relationship with the growth of solar.



Summary

- Innovation coefficient (p_2) always non-significant, which was expected since the data show a flat trend until the presence of incentive measures (2007, 2019).
- Solar is the only one with continuous growing trend.
- Model doesn't obtain an adequate significance of parameters to validate a dynamic relationship.
- Unrestricted UCRCD for the three competitions.
- Full competition between the dynamic relationship of solar vs fossil and solar vs nuclear, while we might infer a full collaboration for solar vs wind.

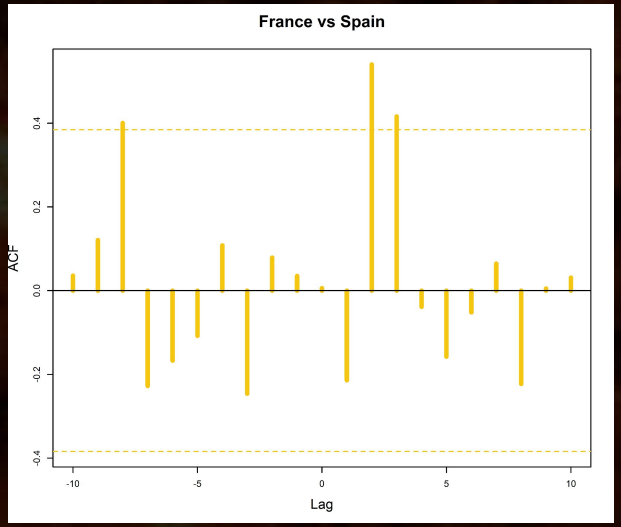
7.
Influence on/of other countries

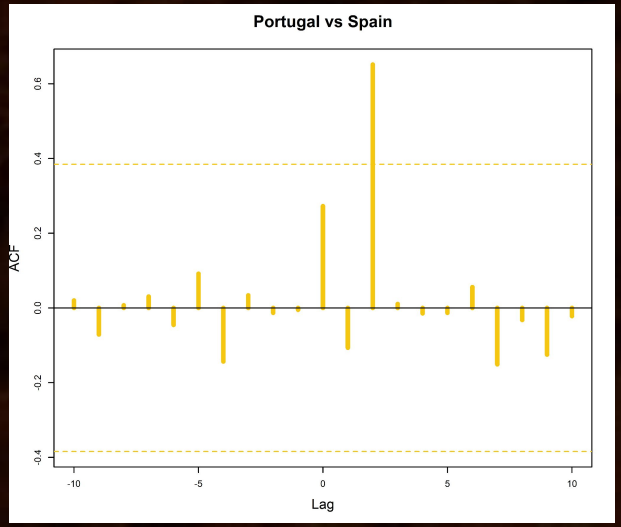
Leading and lagging countries

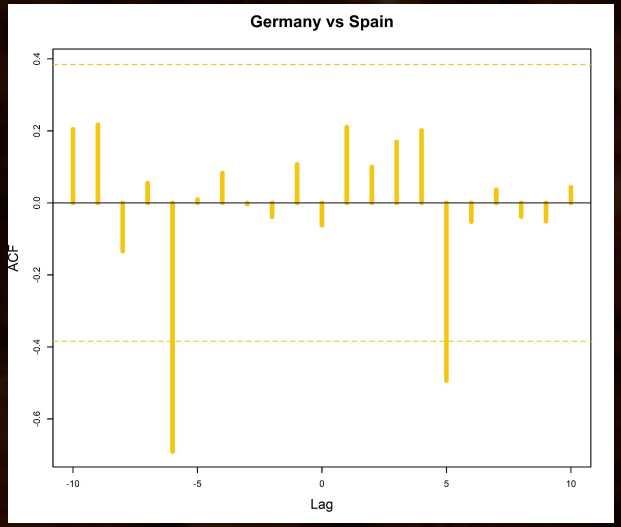
What countries do we consider?

Influence on/of other countries

- Spain leads France and Portugal (with positive correlation).
- Germany leads Spain (with negative correlation)







8.
Price of PV cells and capacity

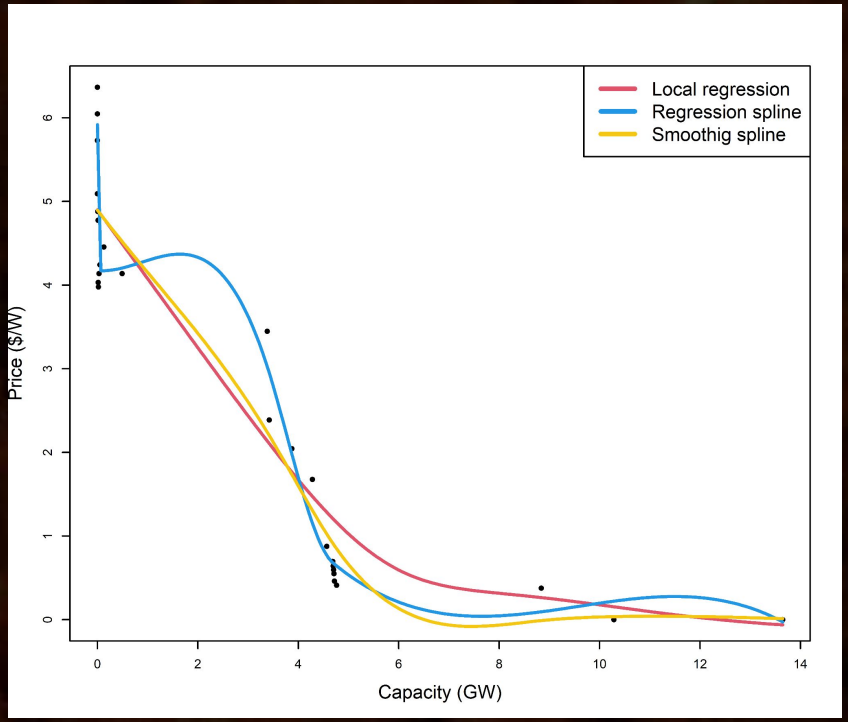
Leading and lagging countries

Modelling price of PV cells as a function of their capacity

Parameters:

- Local regression: h = 3, ngrid = 200.
- Regression spline (cubic): 5 knots.
- Smoothing spline: Lambda = 0.001.

Regression spline seems to be the best, but might tend to overfit.



9.
Non parametric models

Multilinear model, Generalized additive model

Linear model

Train-test split of 70-30%

Full linear model

- AIC = 118.0
- F-stat = 3.3 x 10^4
- Most statistically significant:
  - Share from solar (positive).
  - USA, Australia (negative).
  - China (positive).
- MSE: 572

Stepwise regression

- AIC = 110.5
- F-stat = 5.7 x 10^4
- Most statistically significant:
  - Share from solar.
  - USA, Australia (negative)
  - Germany, China (positive).
  - Nuclear (positive, small).
- MSE: 599

GAM

Using loess

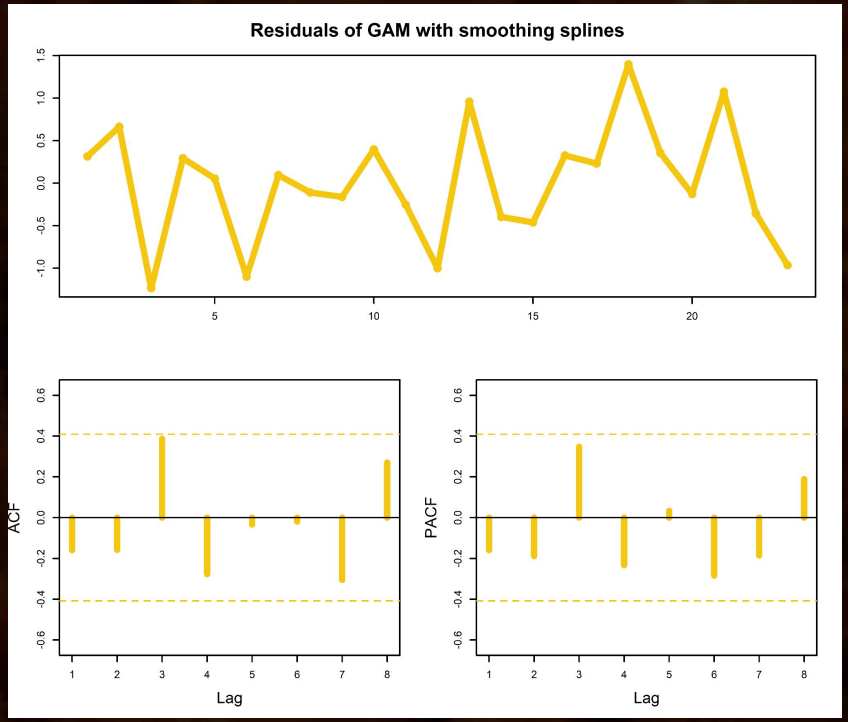
- AIC = 98.3
- Most significant parametric:
  - Nuclear
  - Share from solar
- Most significant non-parametric:
  - Coal, df = 2
- MSE: 464

Using smoothing spline

- AIC = 81.0
- Most significant parametric:
  - Nuclear
  - Share from solar
- Most significant non-parametric:
  - Wind, df = 4
  - Coal, df = 4
- MSE: 464

GAM - Residuals

- No clear pattern in their distribution.
- No autocorrelation
  - Durbin-Watson test: 2.14.



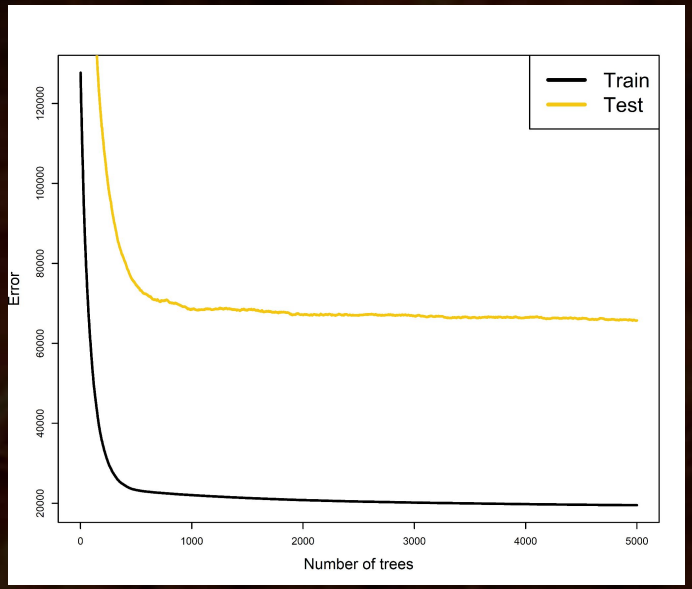
10.
Tree based approaches

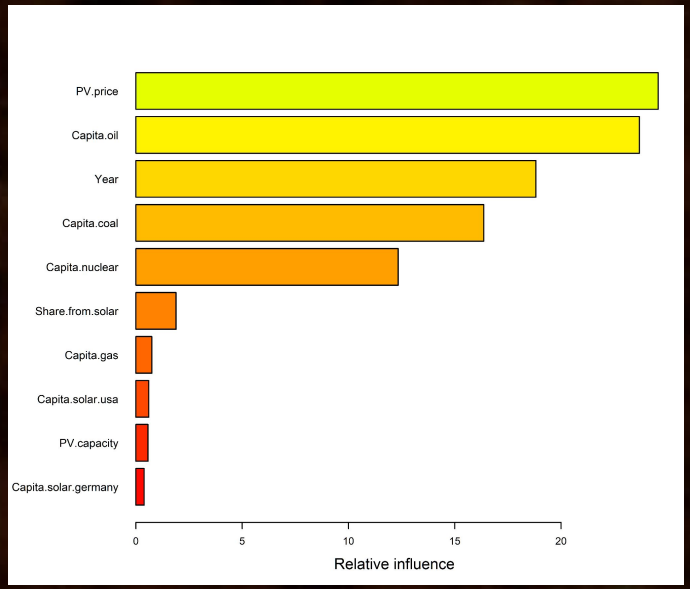
Gradient boosting, Regression tree, Bagging,
Random forest.

Gradient boosting

Given the small amount of data, bagging fraction had to be quite big.

Minimum MSE = 65884.





Bagging and random forest

Tree

- Max depth = 11
- MSE = 13441

Bagging

- 250 trees
- MSE = 30127

Random forest

- 200 trees
- MSE = 31003

Gradient boosting

MSE = 65884 !!

Summary and conclusions

	AIC	MSE (kWh²)
Full linear model	118.0	572
Stepwise regression	110.5	599
GAM - loess	98.2	464
GAM - splines	81.0	464
Gradient boosting	-	65884
Tree	-	13441
Random forest	-	31003
Bagging	-	30127

- Tree based methods have the worst results because of the small amount of data.
- GAM with splines is our most powerful prediction model: lowest AIC and MSE.
- GBM has the best fit, GGM is preferred as xreg.
- Fueled by investment, the growth of solar most impacts the nuclear.

Summary and conclusions

- We expect a big rise in solar energy consumption for the next 4 years.
- The behavior of the data has been corroborated with the history of solar energy in Spain.
- Spain has positive influence on Portugal and France.
- Nuclear energy is the most impacted by solar energy presence.
- Tree based methods have the worst results because of the small amount of data.
- GAM with splines is our most powerful prediction model: lowest AIC and MSE.
- GBM has the best fit, GGM is preferred as xreg.
- Fueled by investment, the growth of solar most impacts the nuclear.

Thank you for your attention!

Any questions?