



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

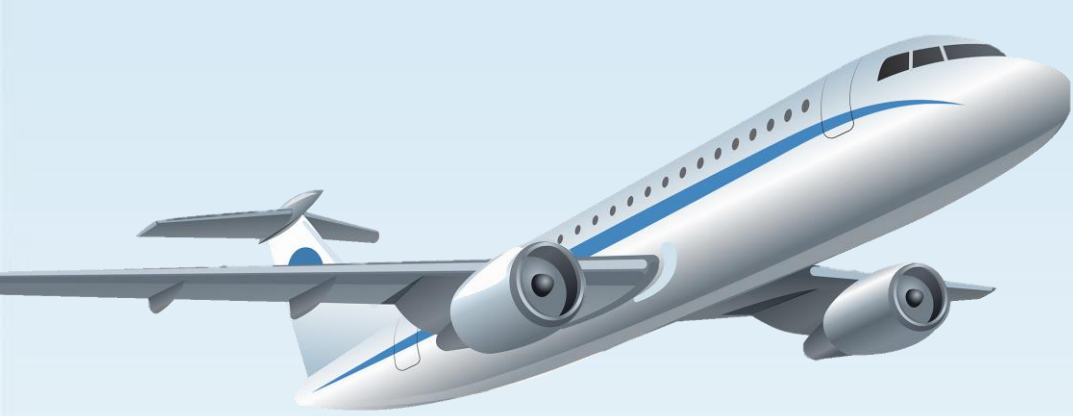


DIPARTIMENTO
MATEMATICA

Air traffic through the years

A time-series analysis on three European countries

CAMILO BETANCOURT NIETO
INÊS CAEIRO JESUS
OKSANA ABRAMOVA



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
MATEMATICA

Air traffic through the years

A time-series analysis on three European countries

CAMILO BETANCOURT NIETO
INÊS CAEIRO JESUS
OKSANA ABRAMOVA

What we will see



Introduction



Evolution of air traffic throughout the years



What might **influence** air traffic?



Statistical modeling



Discussion



Introduction

As **climate change effects** become more evident, public interest in air traffic has raised.

European airlines look to flight path reform to help cut emissions

An EU plan to streamline control of the region's airspace to boost fuel efficiency has stalled

Transport and mobility

Modified 19 Jan 2024

Image © Stefan Bakker, My City/EEA

euronews. My Europe World Business Sport Green Next Health Travel Culture Video

Image © Stefan Bakker, My City/EEA

Home > Topics > In-depth topics > Transport and mobility

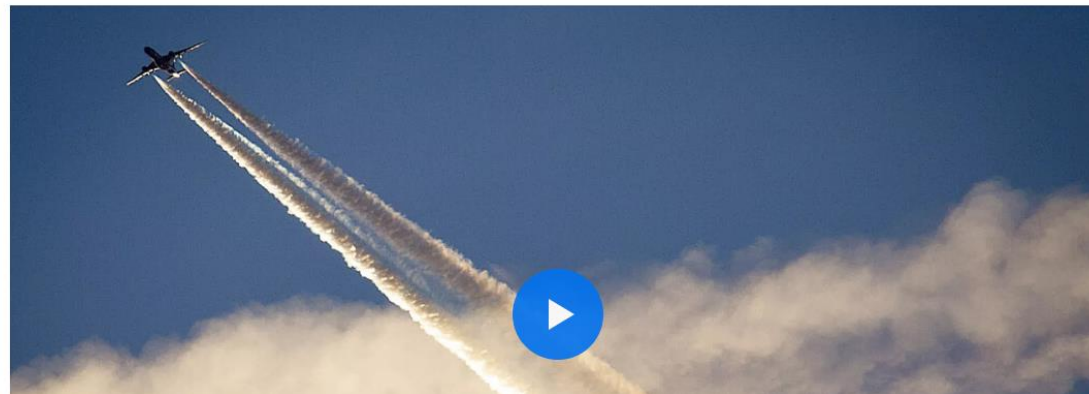
Transport is a vital sector but our current mobility

in de
our h

THE CONVERSATION
Academic rigour, journalistic flair

Home > News > World

Air traffic is booming again and environment activists aren't happy



Reducing air travel by small amounts each year could level off the climate impact

Published: November 5, 2021 3:12pm CET

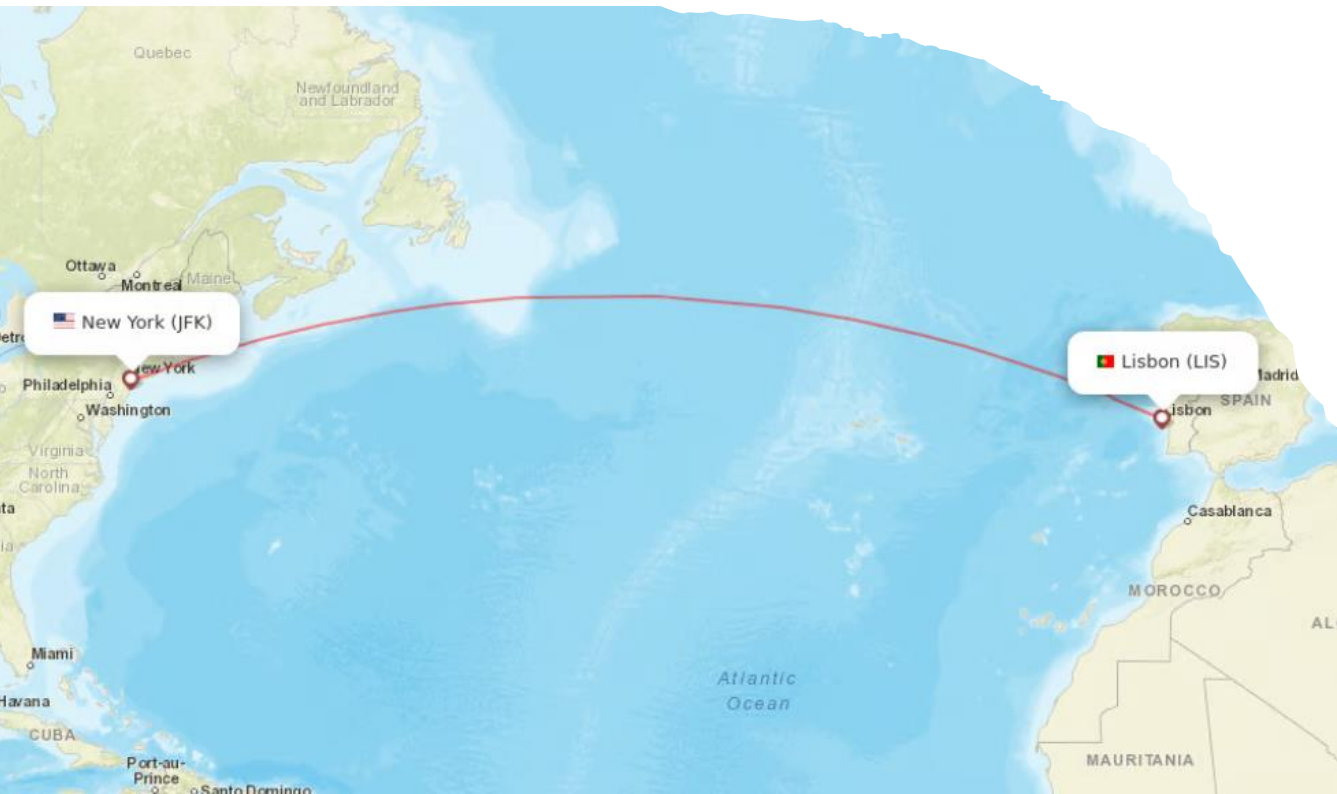


NEWS EDITORIAL FEATURES MEDIA THE CANARY SUPPORT

Port has gone rogue
air traffic polluters to
pay for their CO₂ emissions
Schiphol is taking a stand

Introduction

The **aviation sector** creates **13.9% of the emissions** from transport, making it the second biggest source of transport GHG emissions after road transport.



If global aviation were a country, it would rank in the **top 10** emitters.

Someone **flying from Lisbon to New York and back** generates roughly the same level of emissions as the average person in the EU does by **heating their home for a whole year**.

Source: [European Commission](#)

Introduction

*“To achieve climate neutrality, the European Green Deal sets out the need to reduce transport emissions by 90% by 2050 (compared to 1990-levels). **The aviation sector will have to contribute to the reduction.**”*

Source: [European Commission](#)

Thus, it is helpful to understand the evolution of air traffic through time.

To restrict the scope of our work, we focus on the **domestic** air traffic of three European countries:



Spain



Italy



Germany

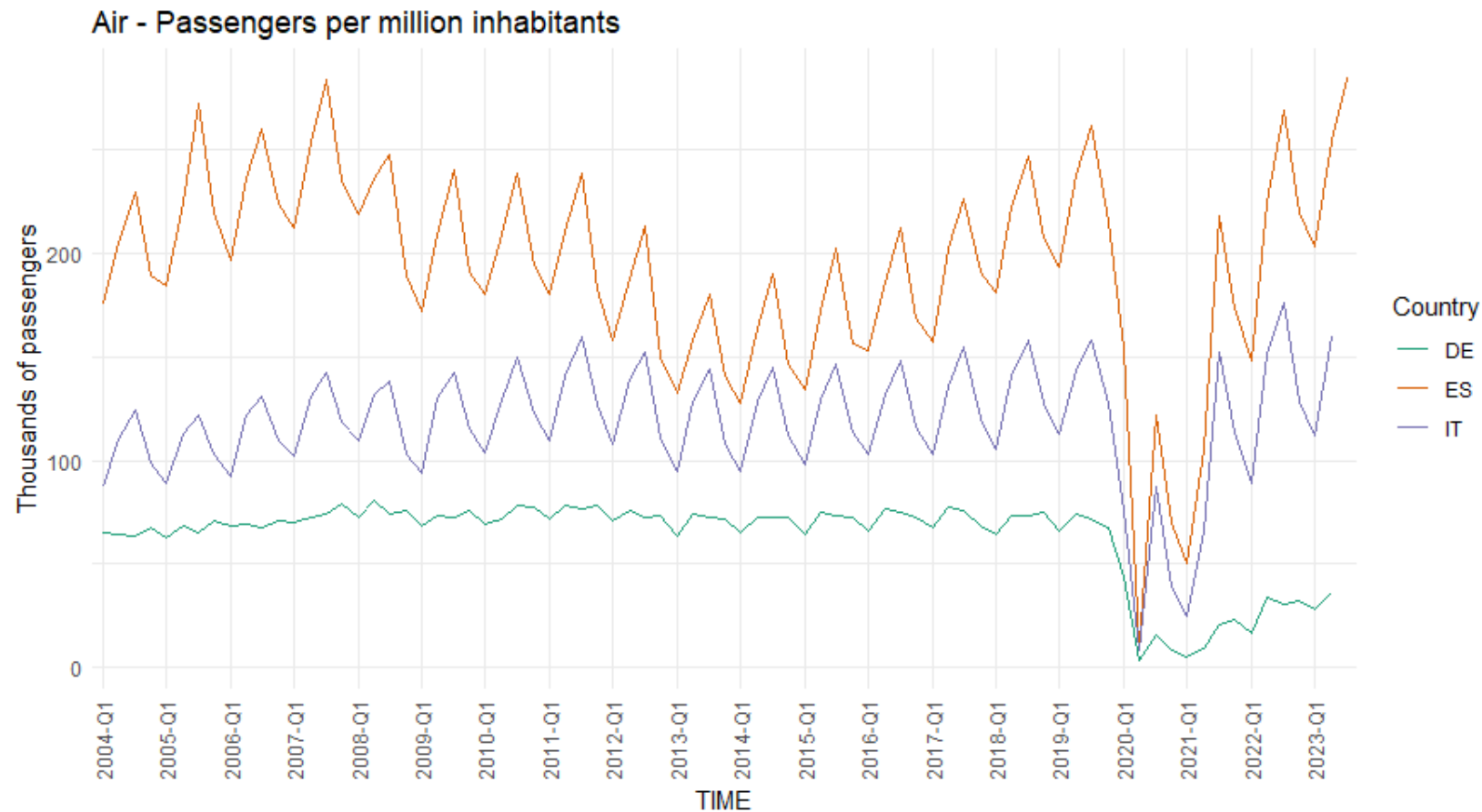
Evolution of air traffic throughout the years

Our data:

- Transportation measurements taken from [Eurostat](#), the Statistical Office of the European Union.
- We used **quarterly data** (to match with our explanatory variables).
- We focus on **internal transportation** only (within the country).
- Complete information from **2002-Q1** to **2023-Q2** for the three countries.
- We focus on passengers carried per million inhabitants* of the country.

* We used the annual demographic data from [Eurostat](#). When doing this project, we did not have information from 2023, so we forecasted it by applying an exponential smoothing method with a damped trend.

Evolution of air traffic throughout the years

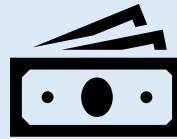


What might influence air traffic?

From the publicly available data, we considered several factors that could possibly explain the air traffic for these countries throughout the years:



COVID-19
(restrictions)



Economic activity
(Gross Domestic Product)



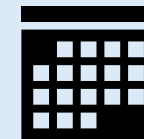
Ticket prices



Interaction/competition
(with rail transport)



Tourism



Time dependencies
(trend, seasonality, etc.)

COVID restrictions

- Data from the Oxford COVID-19 Government Response Tracker ([OxCGRT](#)).
- We used the record of restrictions on internal movement between cities/regions.
- Ordinal variable with three levels*:

0	No measures
1	Recommend not to travel between regions/cities
2	Internal movement restrictions in place

* It was monthly data that we grouped for the quarter and rounded to the closest value to keep it as an ordinal variable.



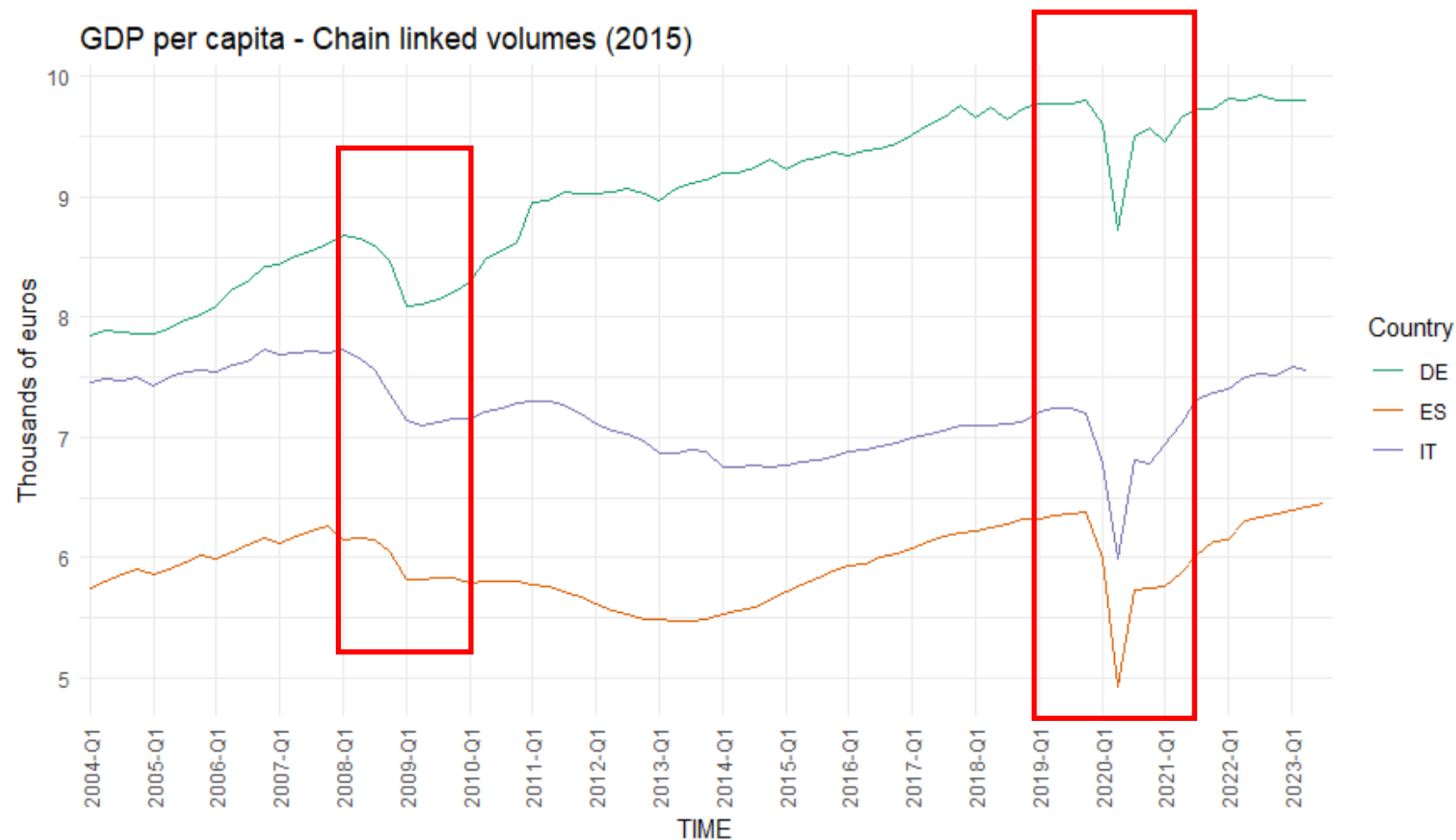
Economic activity

- COVID-19 is not only relevant because of the governments' policies on movement between regions.
- People were consuming less and earning less money during this time. The world was stalled for a while.
- We think people are less likely to fly under this conditions.
- In fact, extending this reasoning, economic activity (not only in presence of COVID) could be a factor to explain the willingness to travel.
- Gross Domestic Product (per capita) can capture these fluctuations in economic activity.

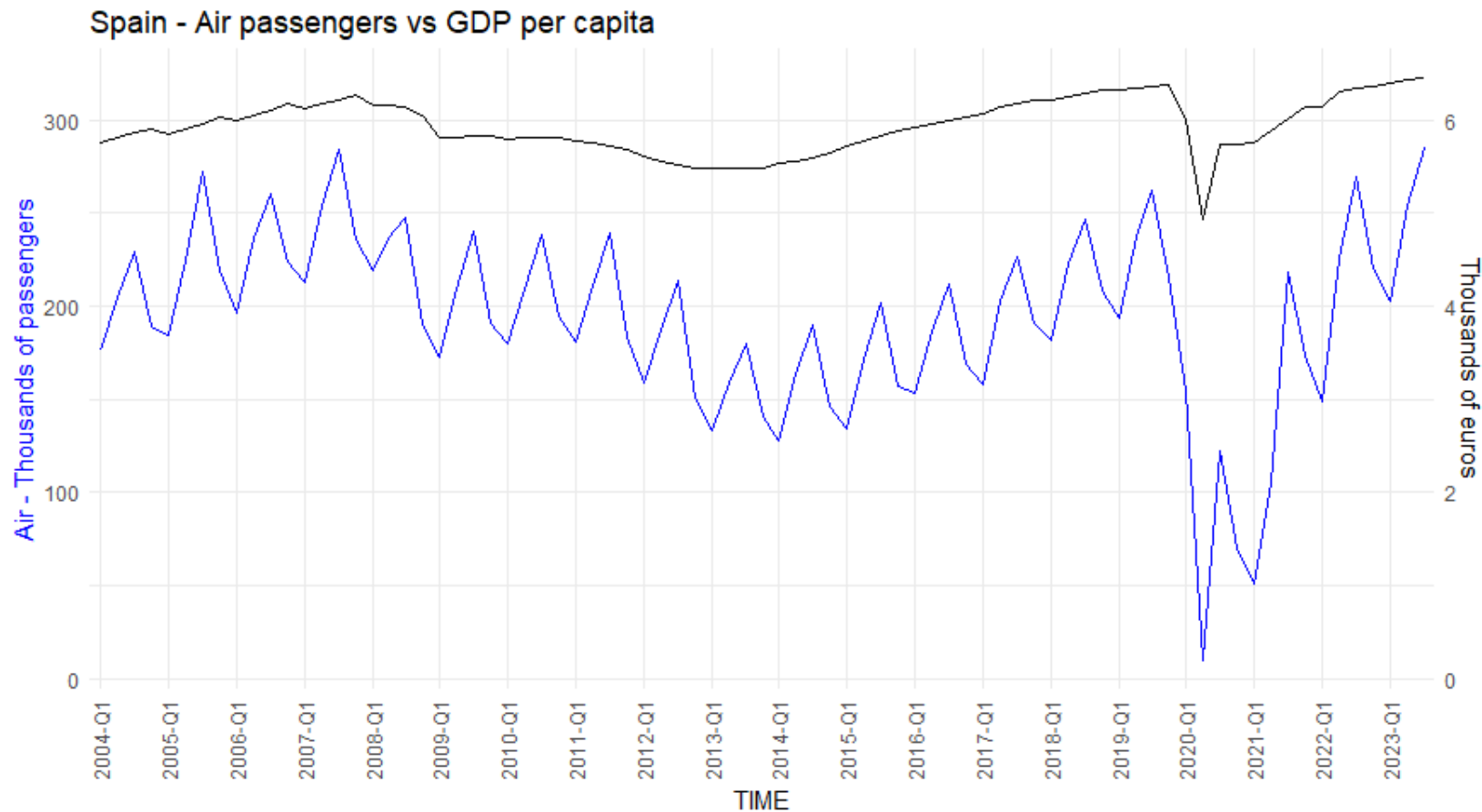


Economic activity

- GDP data from [Eurostat](#) at market prices.
- Was already [adjusted](#) for inflation with 2015 as reference.
- Already included calendar and seasonal adjustments (more on this later).
- We used the GDP per capita to control for the population.



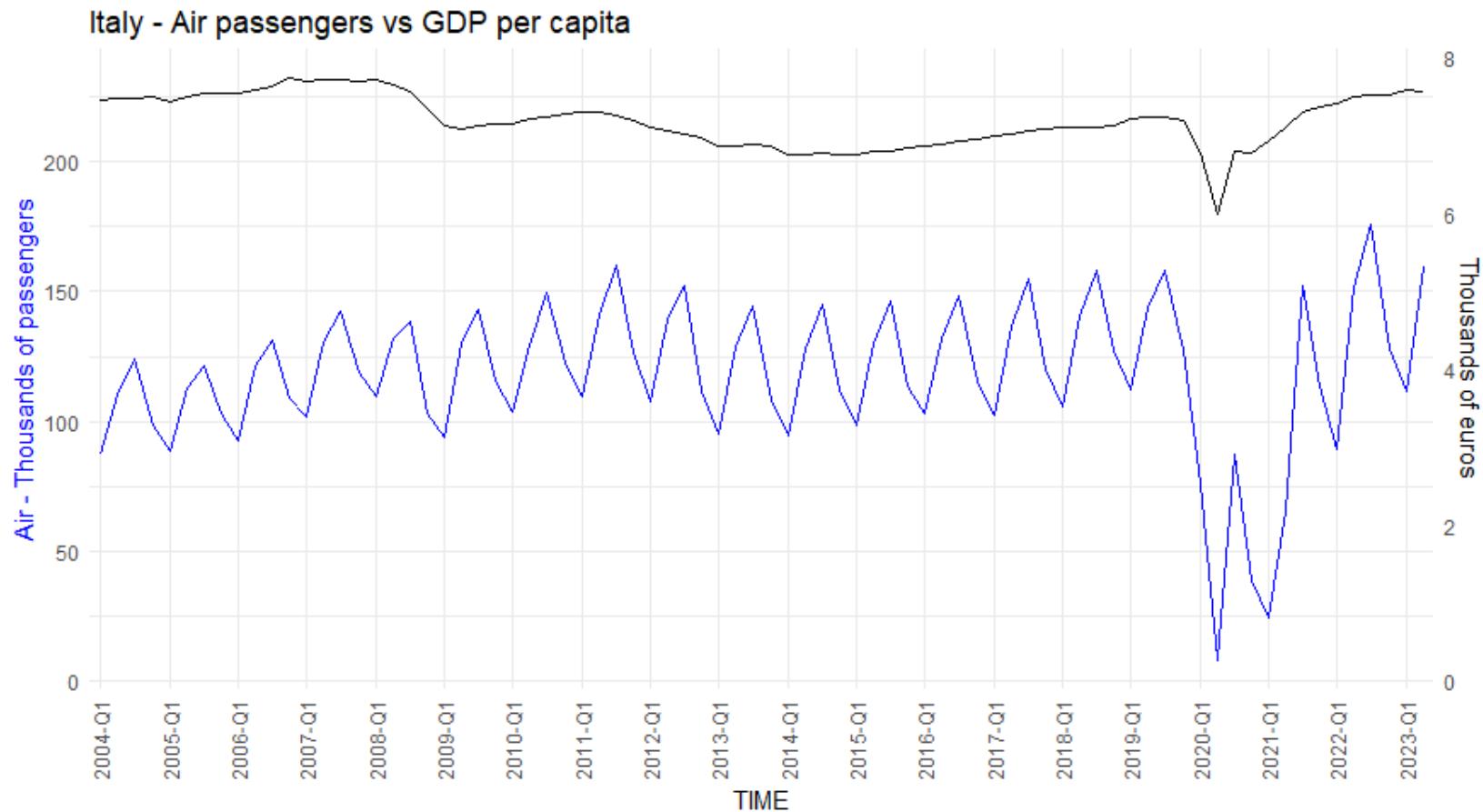
GDP per capita - Spain



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



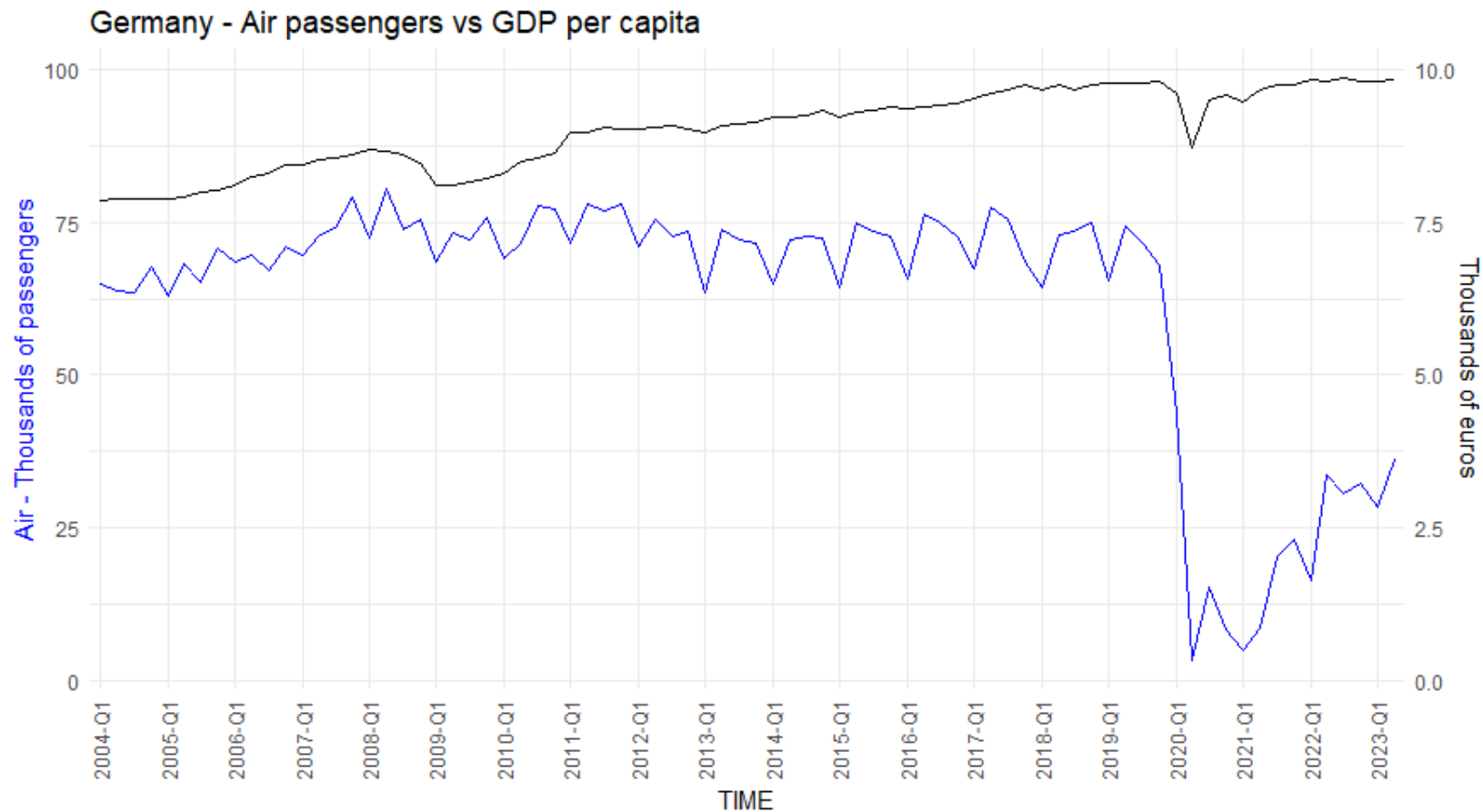
GDP per capita - Italy



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



GDP per capita - Germany

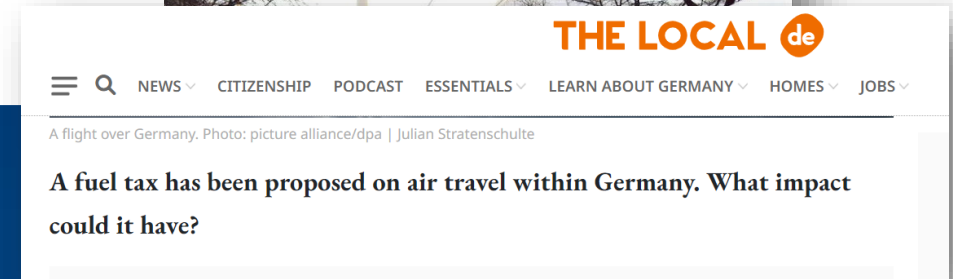
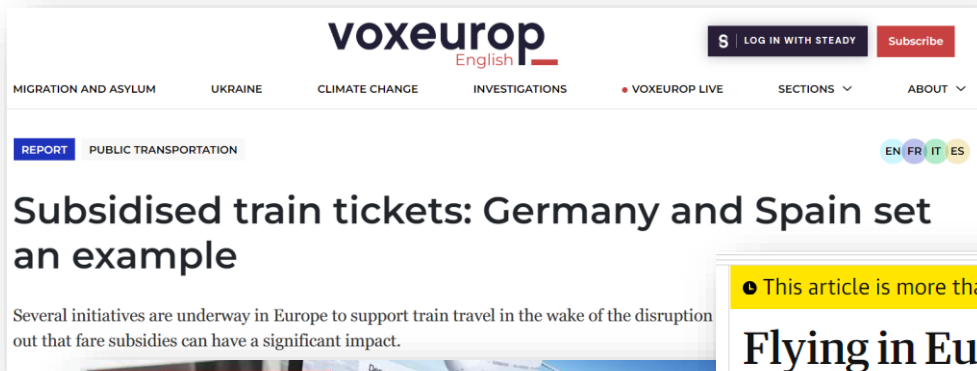


* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



Prices

- How do people react with respect to the price of their mode of transportation?



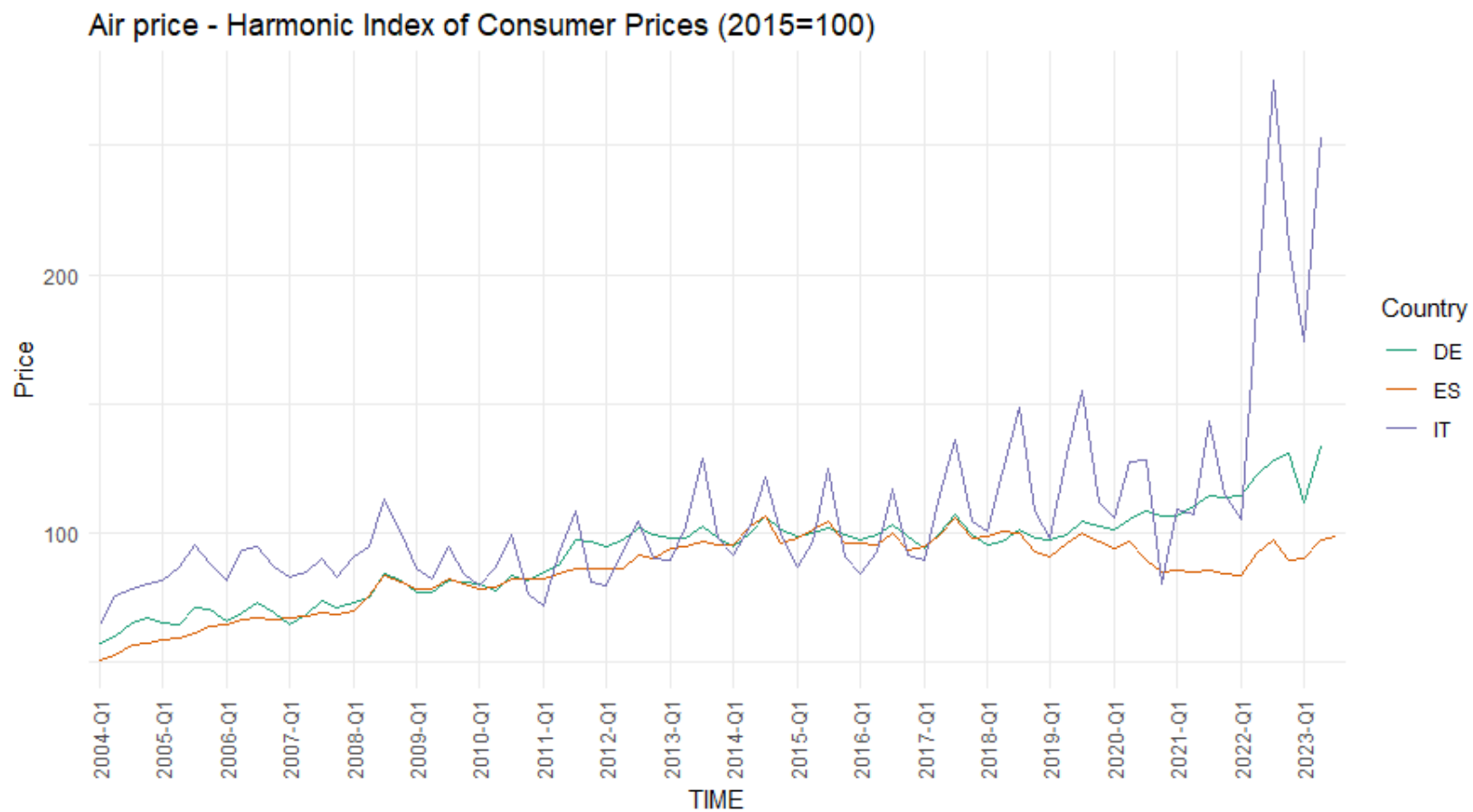
Prices

- Harmonised index of consumer prices (HICP) from [Eurostat](#).
- HICP measures **consumer price inflation**, harmonized to compare between countries.
- We took the prices for:
 - Passenger transport by air*
 - Passenger transport by railway
- Monthly data, grouped as the average per quarter.
- Unit of measure: Index, 2015=100

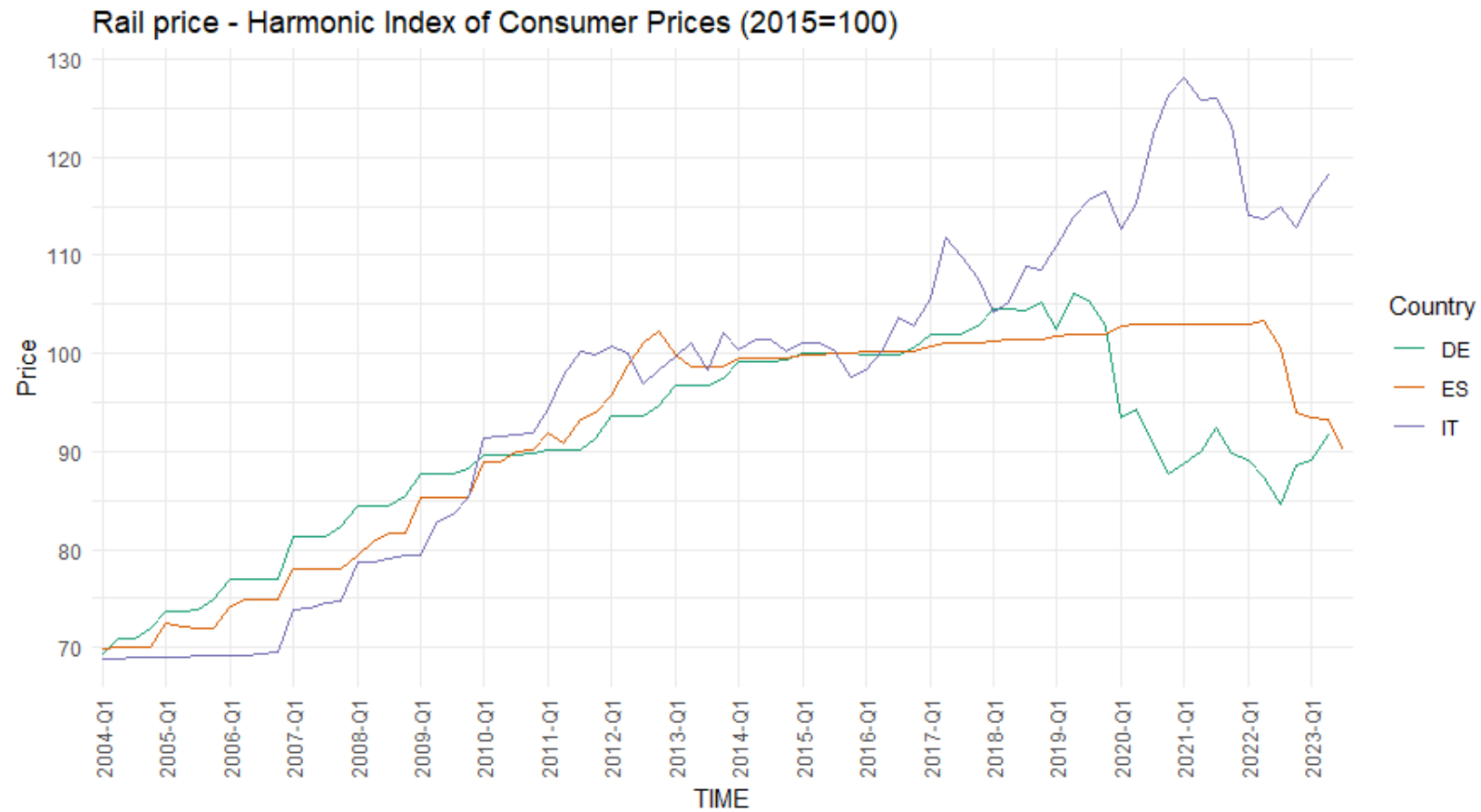
* There is also a measurement of domestic flights prices. Unfortunately, it is too recent and there are many missing values.



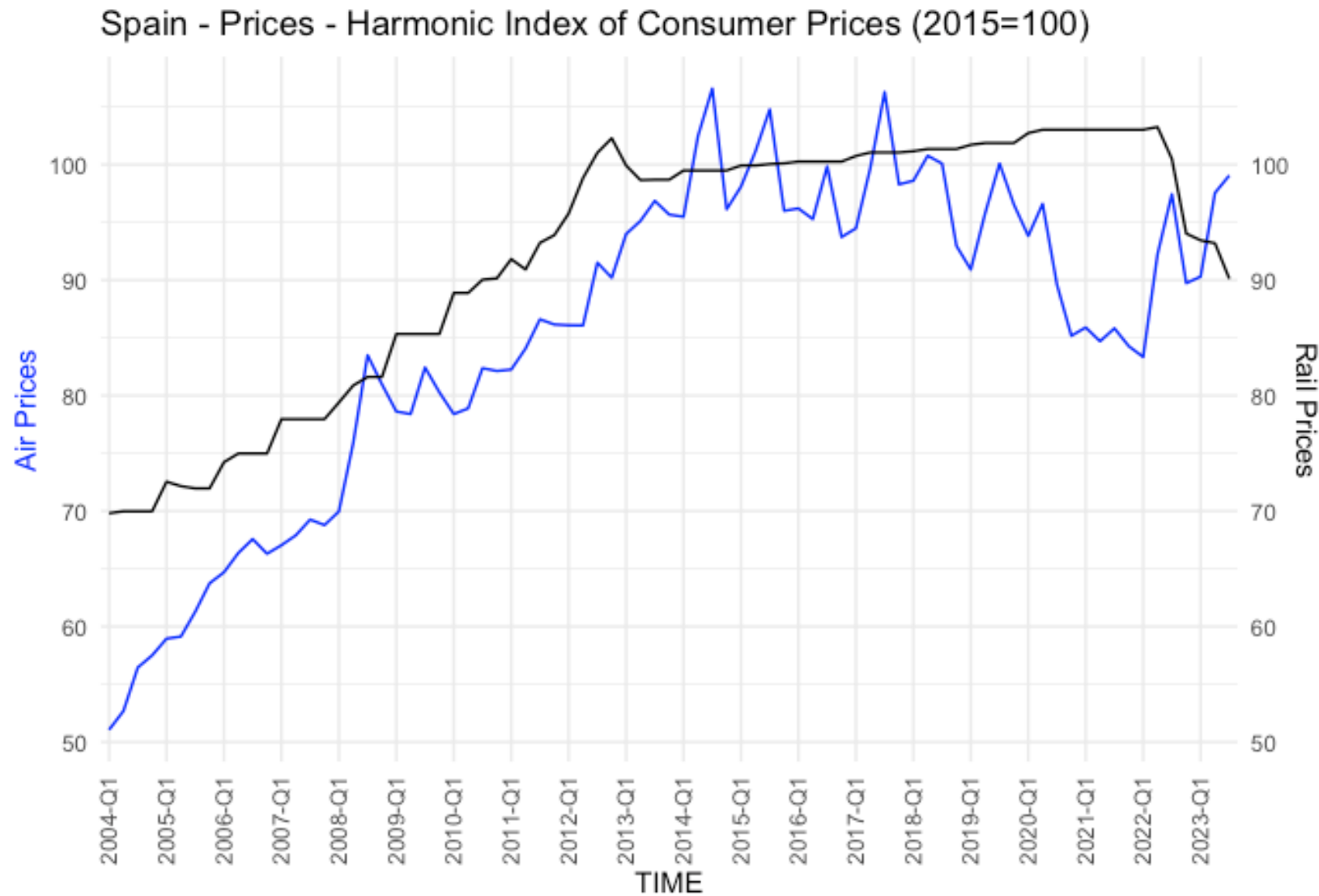
Air prices



Rail prices

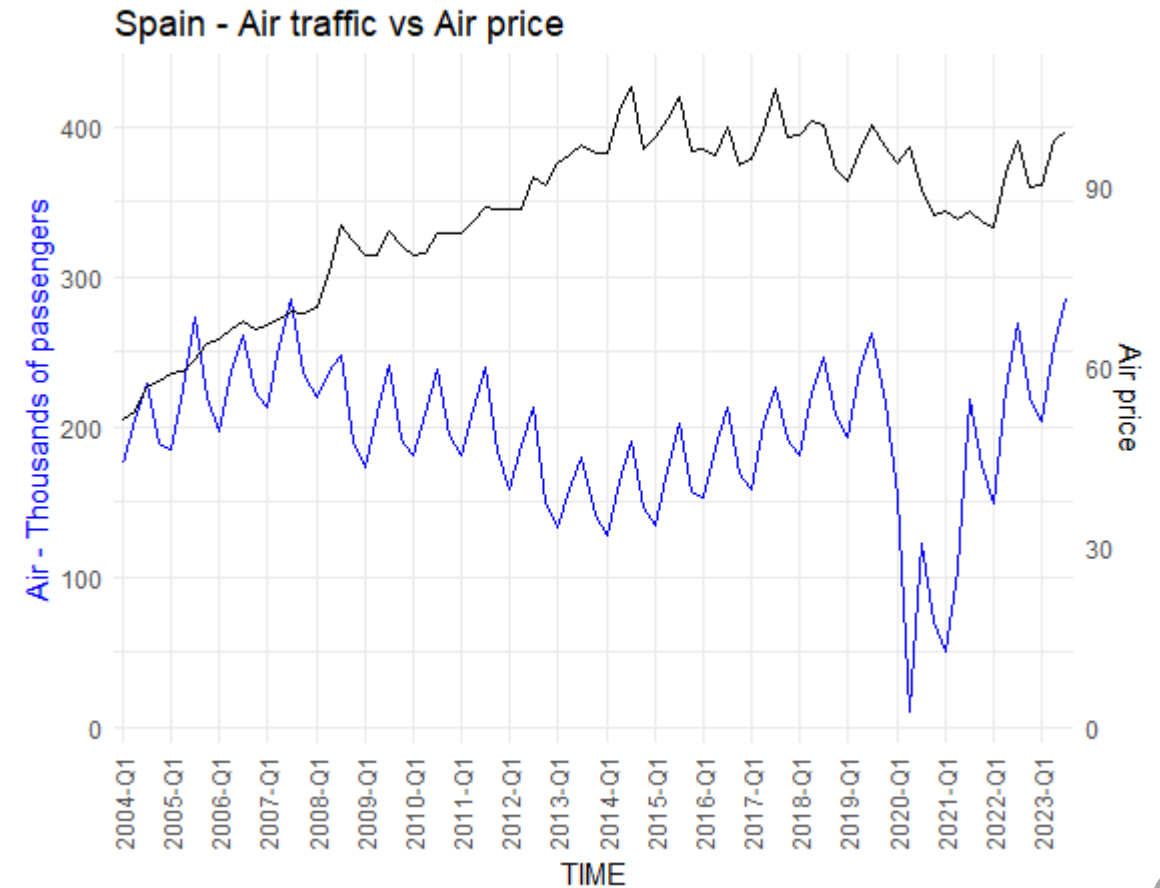
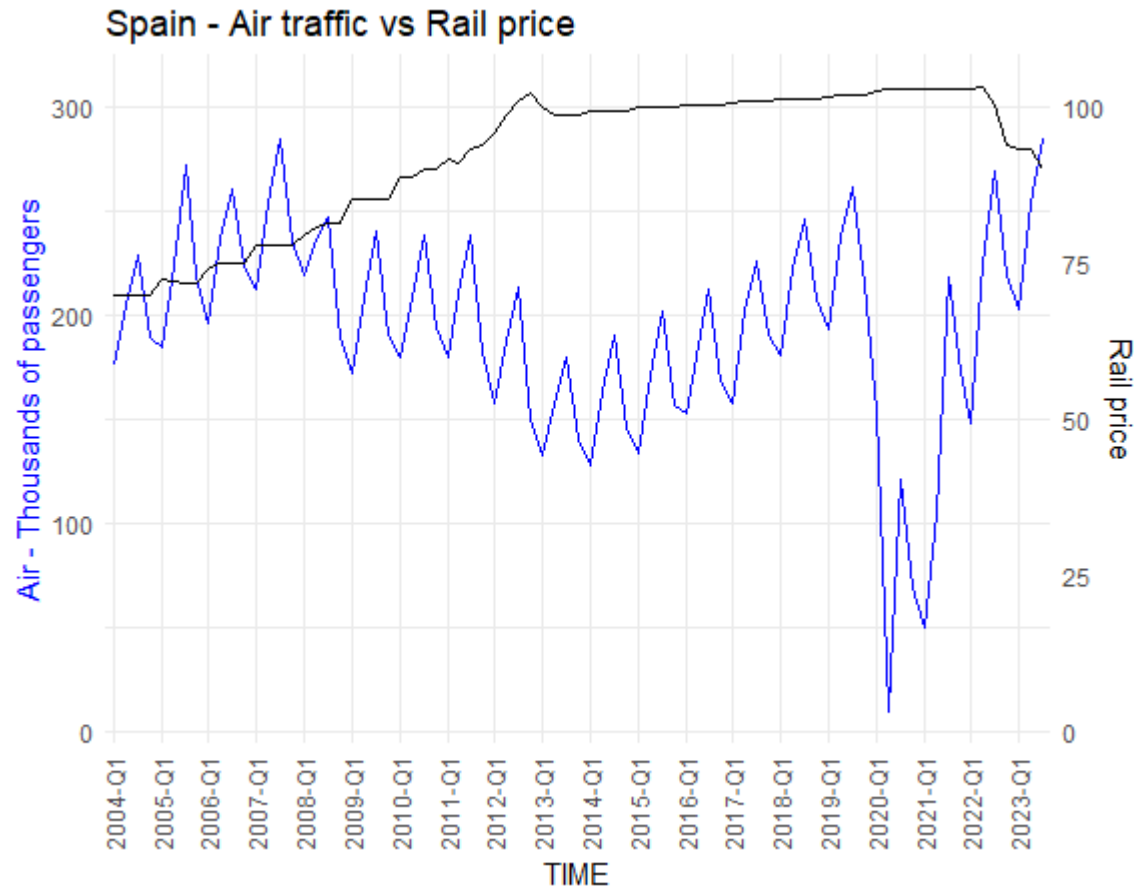


Transportation prices – Spain



Air traffic vs Ticket prices – Spain

- Not a clear pattern with respect to prices.

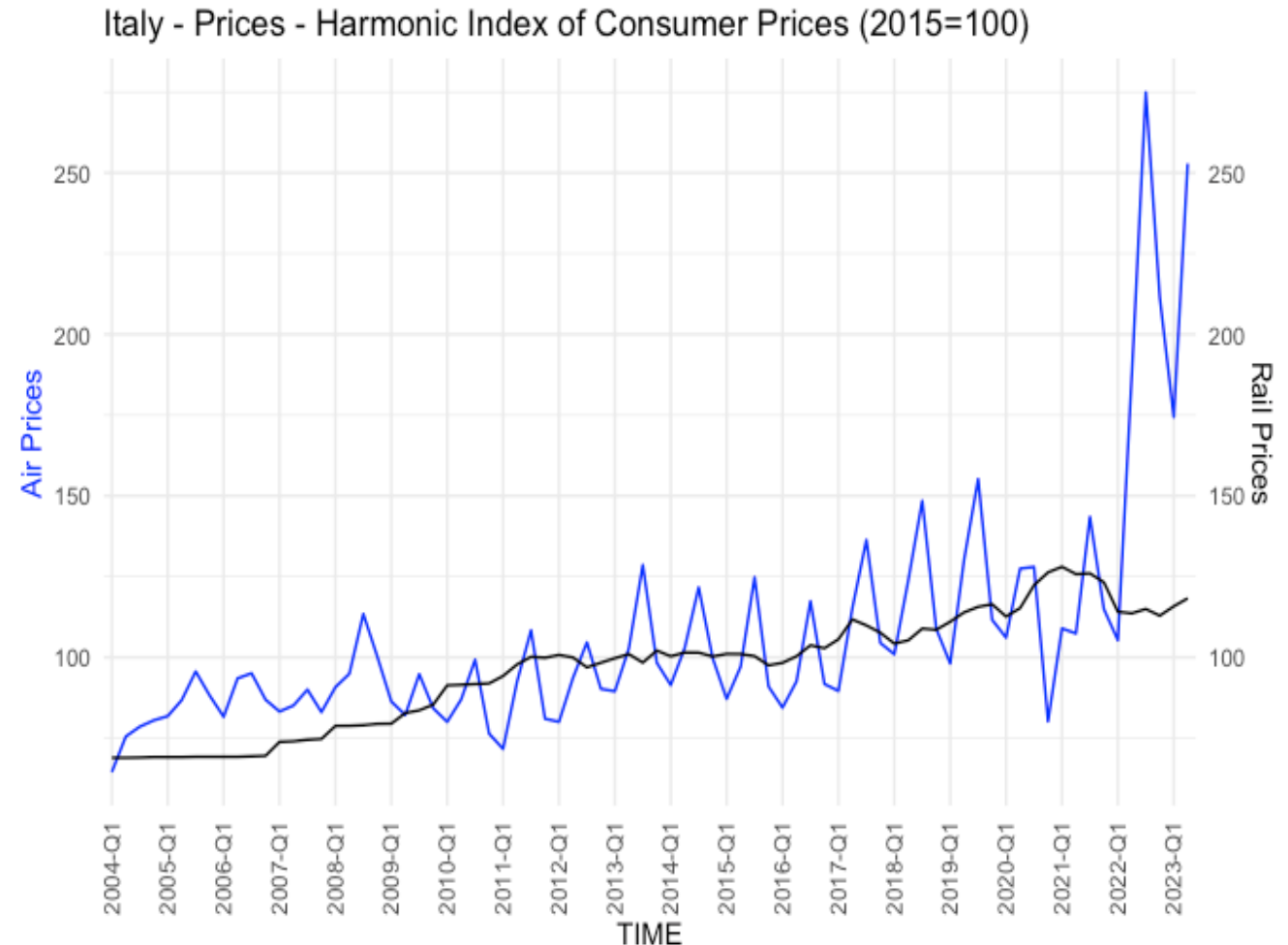


Transportation prices – Italy

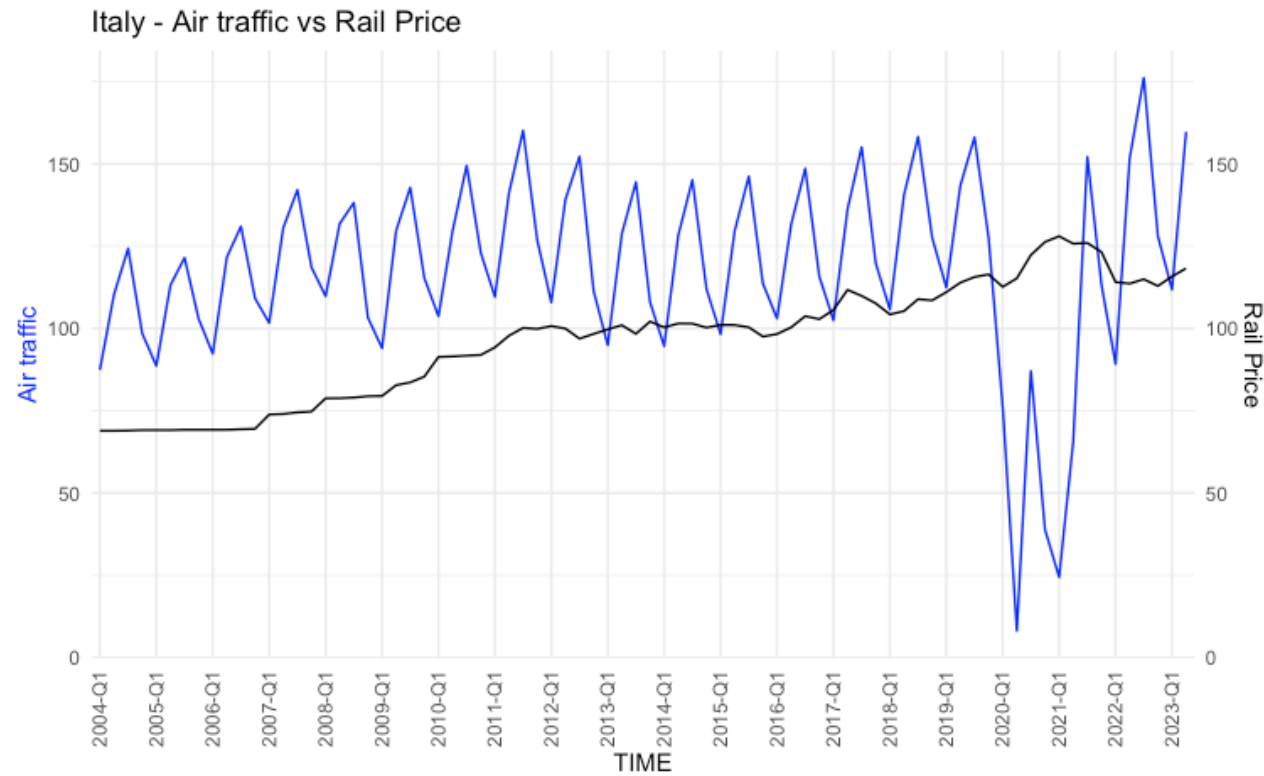
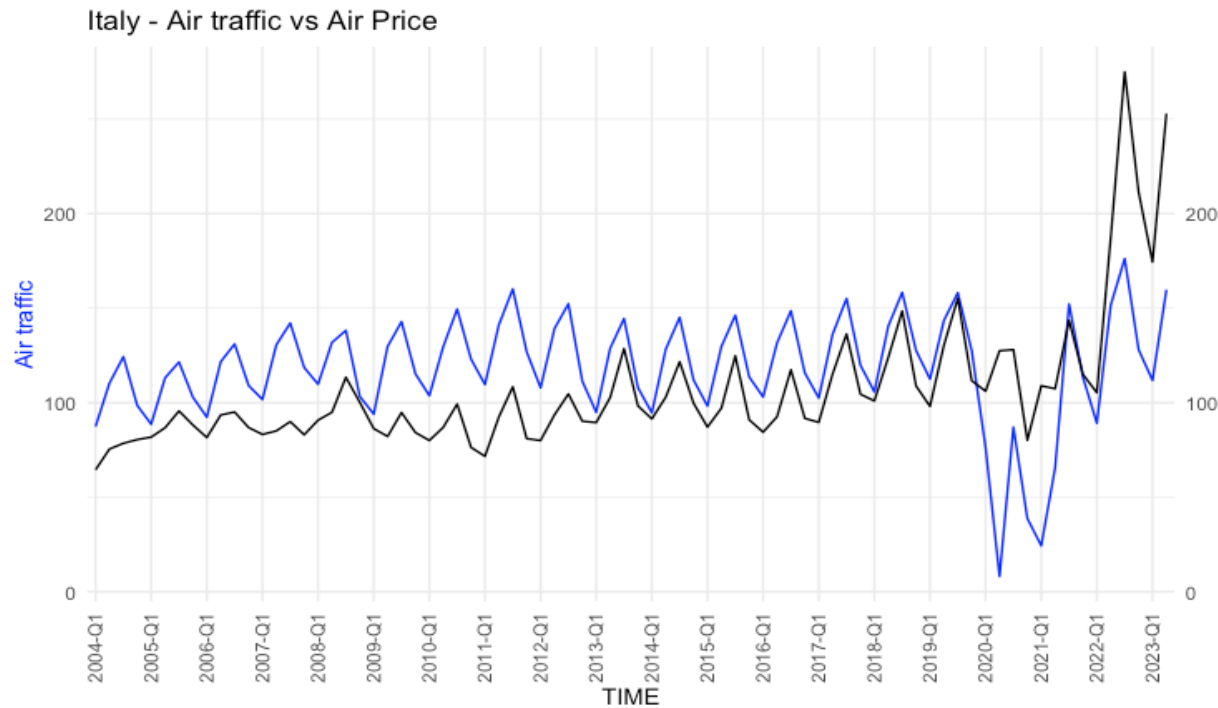
Flight prices from Sardinia and Sicily to Rome were as much as **830 per cent** above average. ([Corriere.it](https://www.corriere.it))

- **Pent-up demand** after COVID crisis
- Inflated **fuel prices**
- **Airlines algorithms** (customer data, dates of booking and travel, competitor's fares, etc.)
- **Scarcity of connections** between mainland and islands

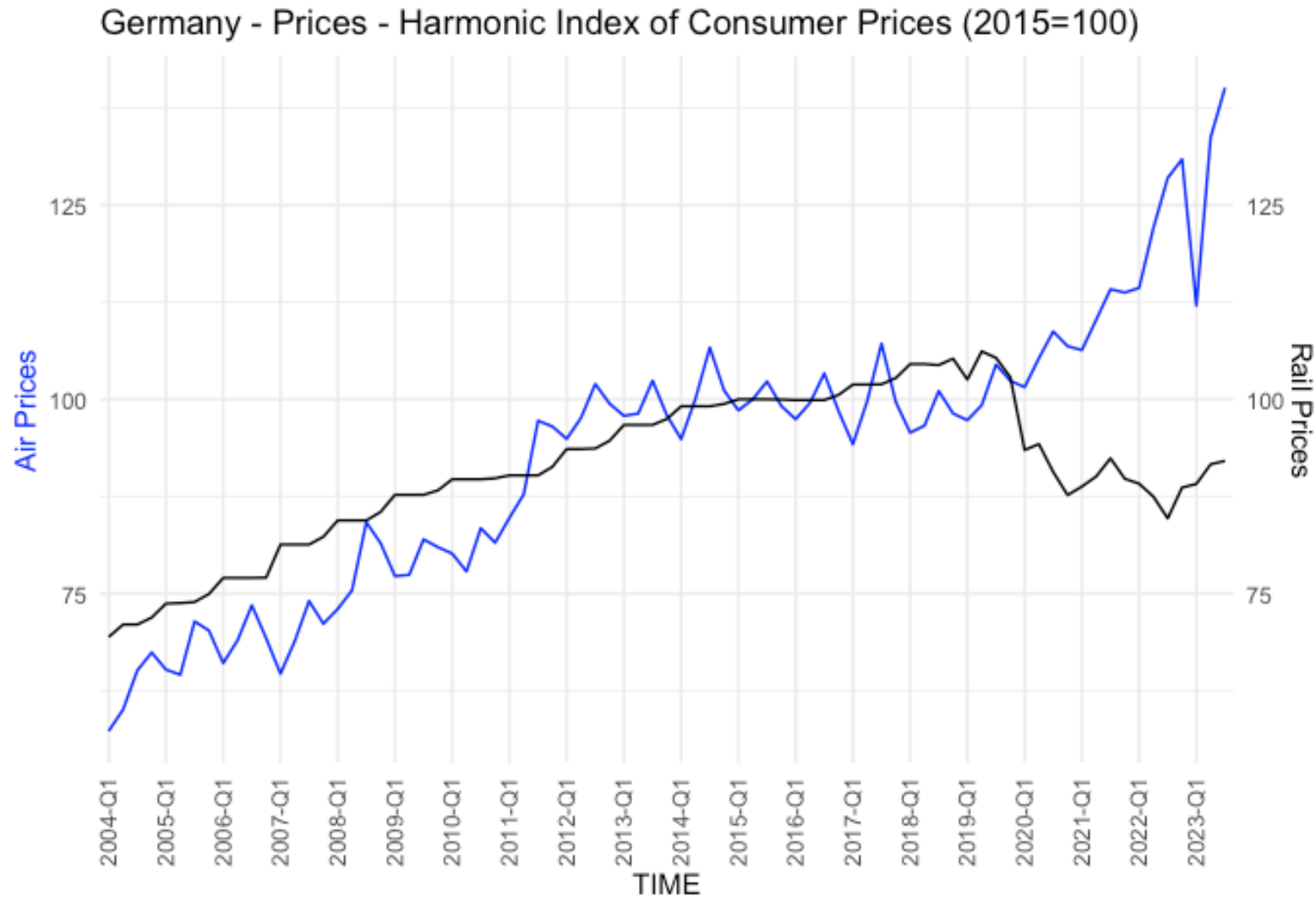
In 2023, Italian government attempts to put a cap on internal flight prices, but quickly retreats.



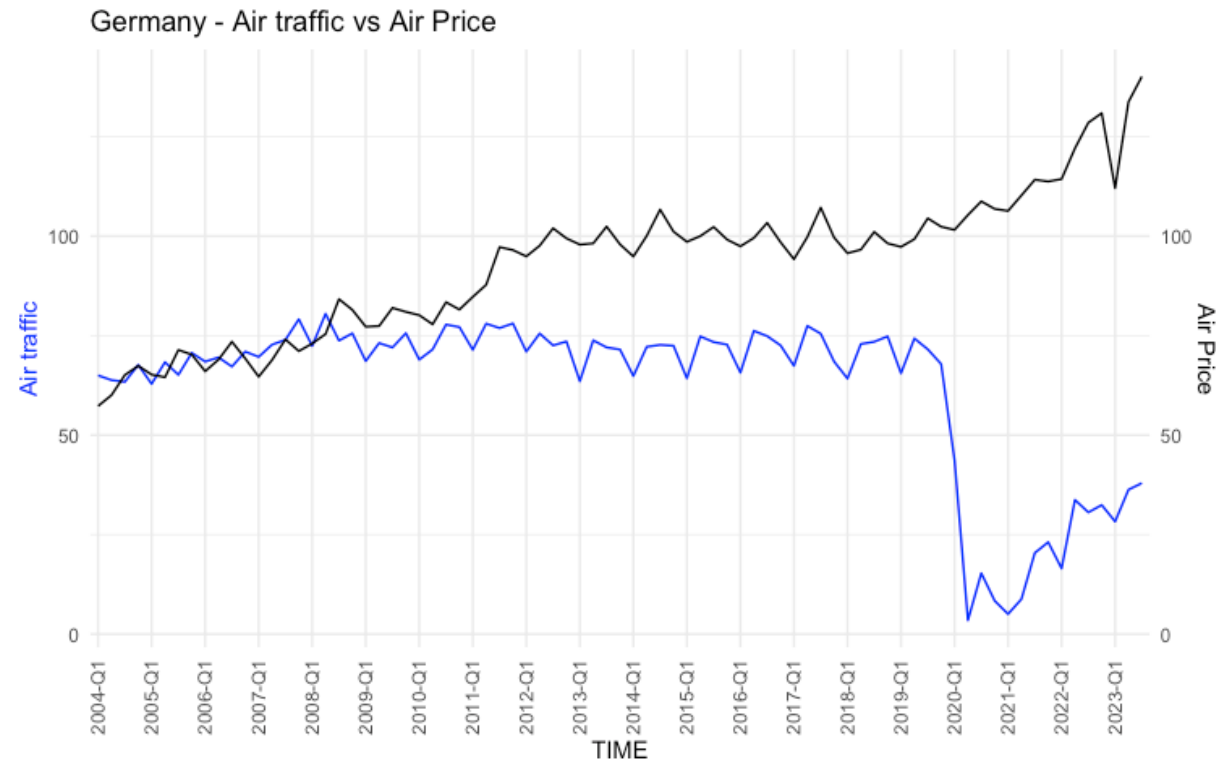
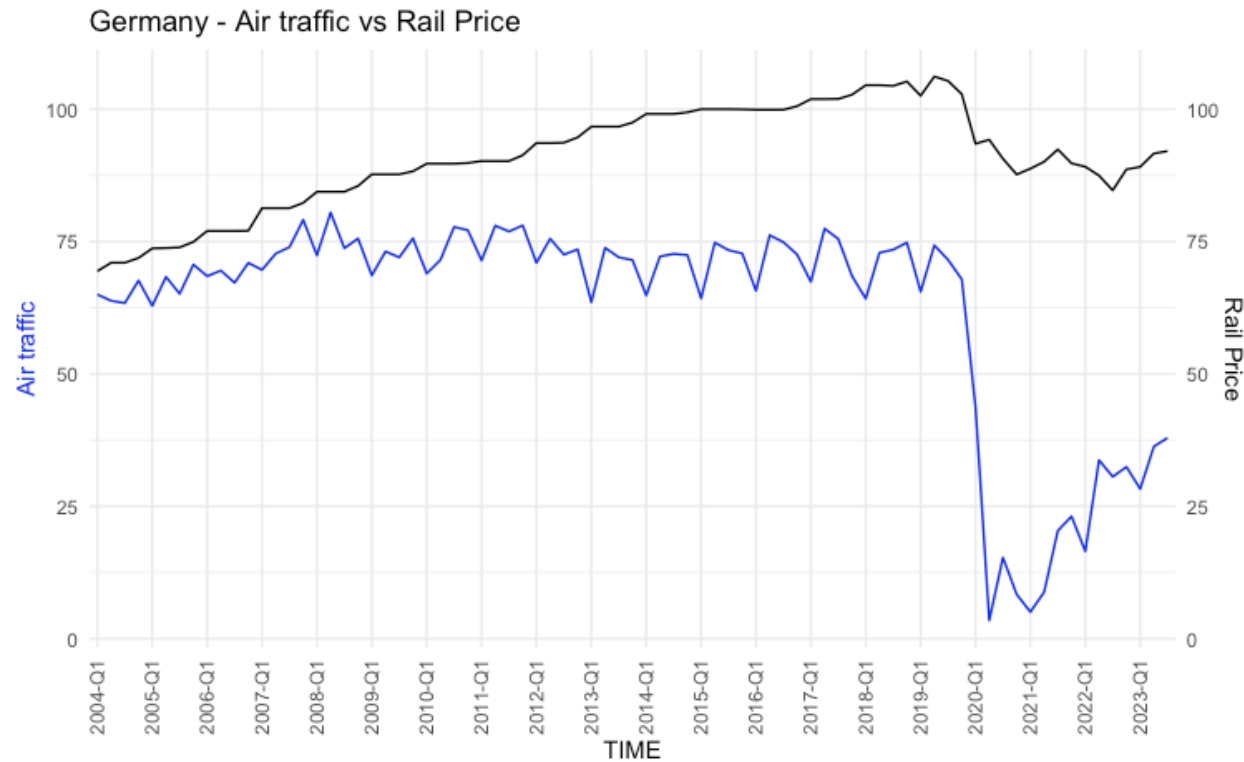
Air traffic vs Ticket prices - Italy



Transportation prices - Germany



Air traffic vs Ticket prices - Germany



Interaction/competition with rail transport

- Again, we used rail transport measurements from Eurostat.
- Only **quarterly** data available*.
- Complete information from **2004-Q1** to **2023-Q2** for the three countries
- We considered two variables:
 1. Thousands of passengers carried
 2. Millions of passenger-kilometers (pkm)**
- We also **normalized** the variables per million inhabitants of the country.

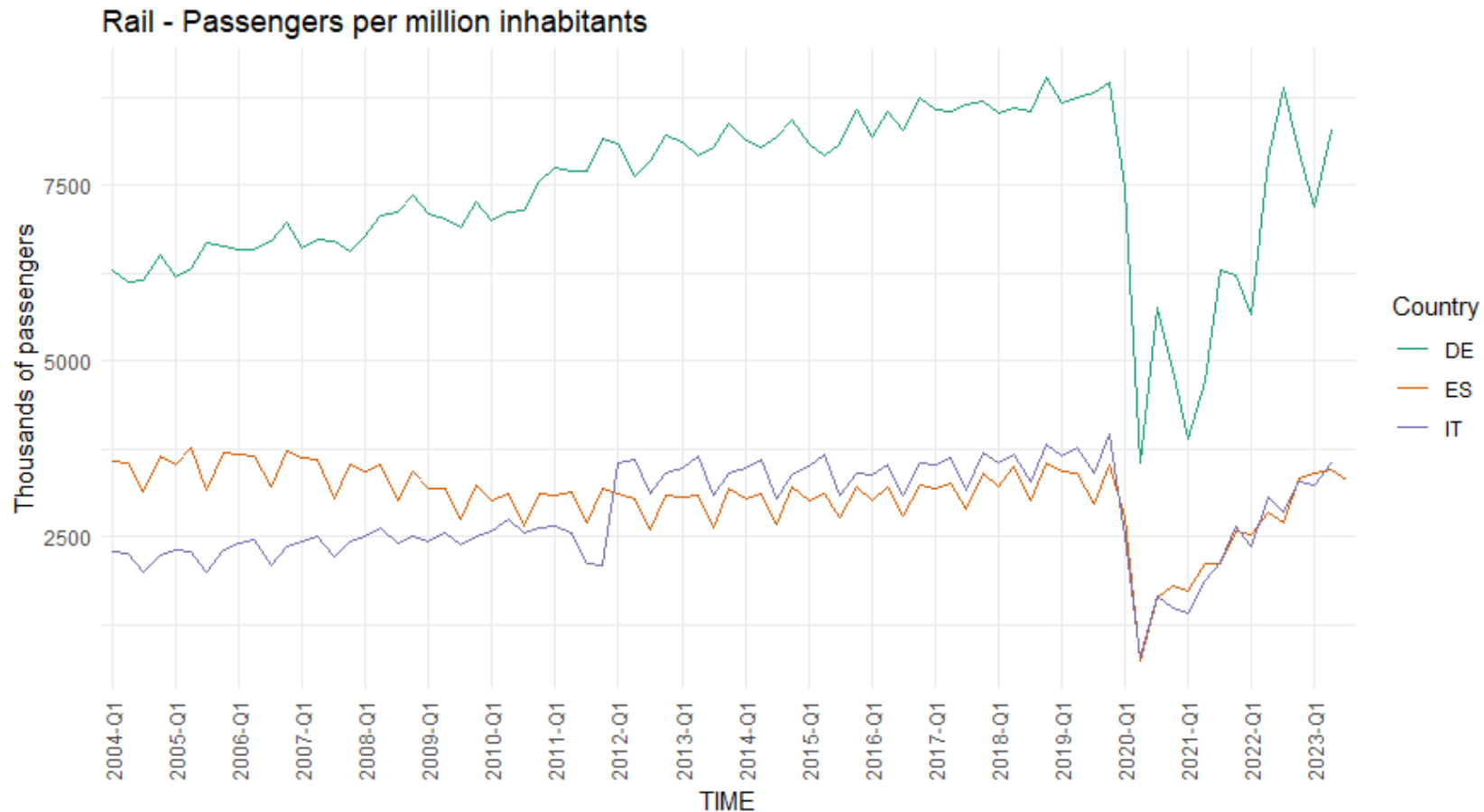
* This is why we used the quarterly data for the air traffic time-series.

** [A passenger-kilometre](#) (pkm) represents the transport of one passenger over one kilometer. It is obtained by multiplying the two measurements.



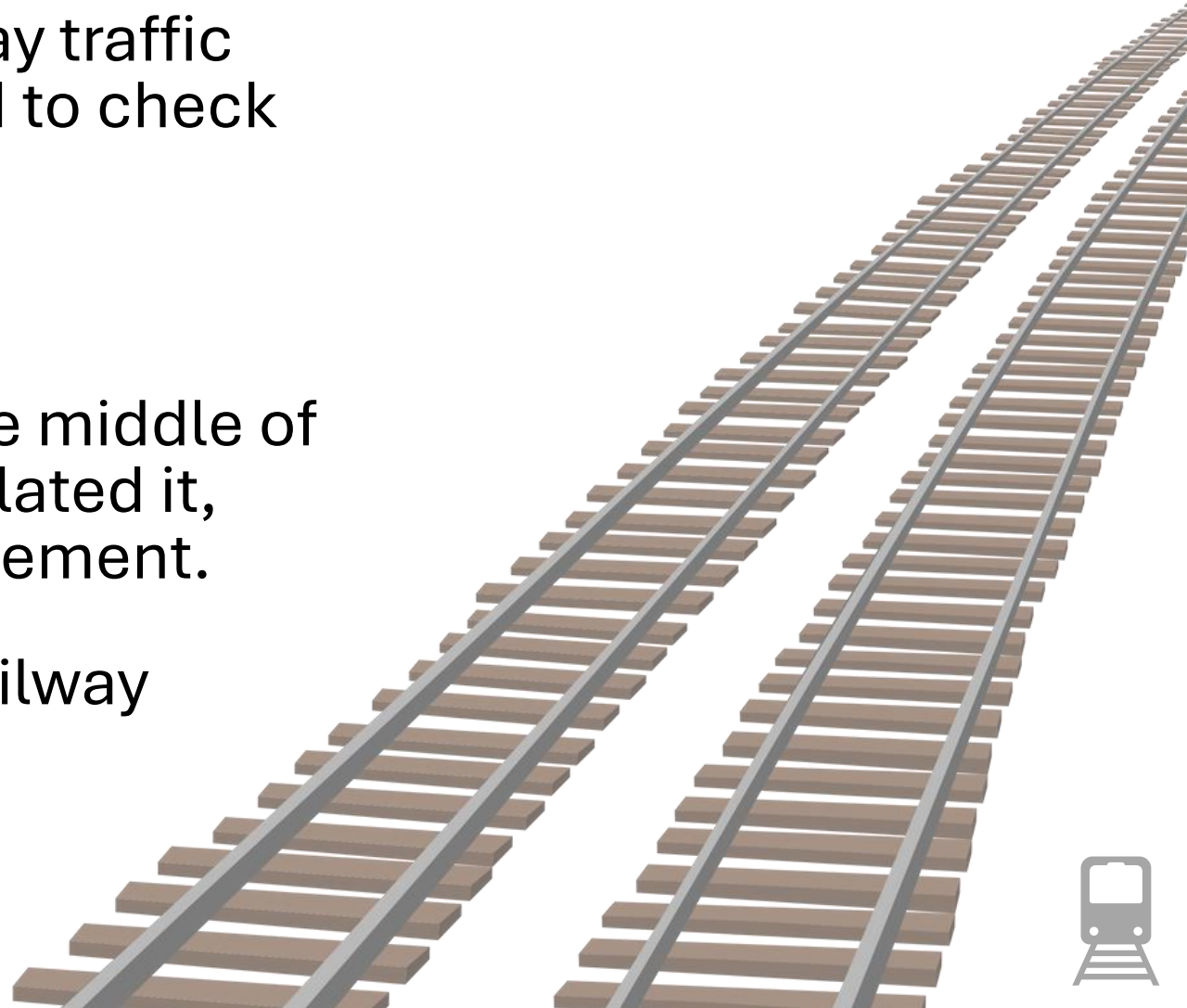
Railway traffic

Contrary to what we saw with the air traffic, Germany has a much higher volume of passengers in comparison to Spain and Italy.



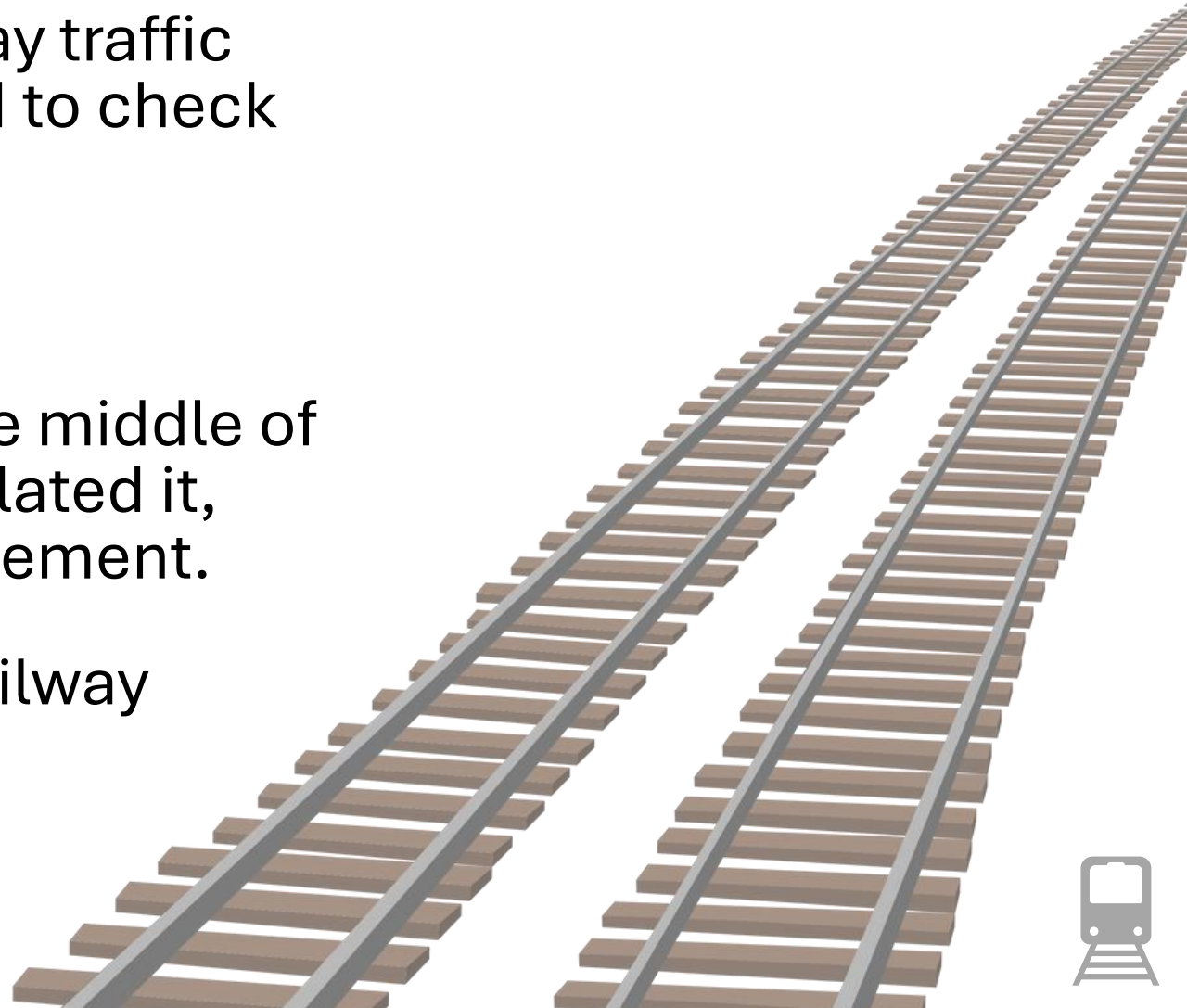
Railway density

- Given the difference of the railway traffic among the countries, we wanted to check the rail densities.
- Also data from Eurostat.
- It had some missing values in the middle of the series, so we linearly interpolated it, since it is a rather stable measurement.
- It is computed as passenger's railway length divided by land area.

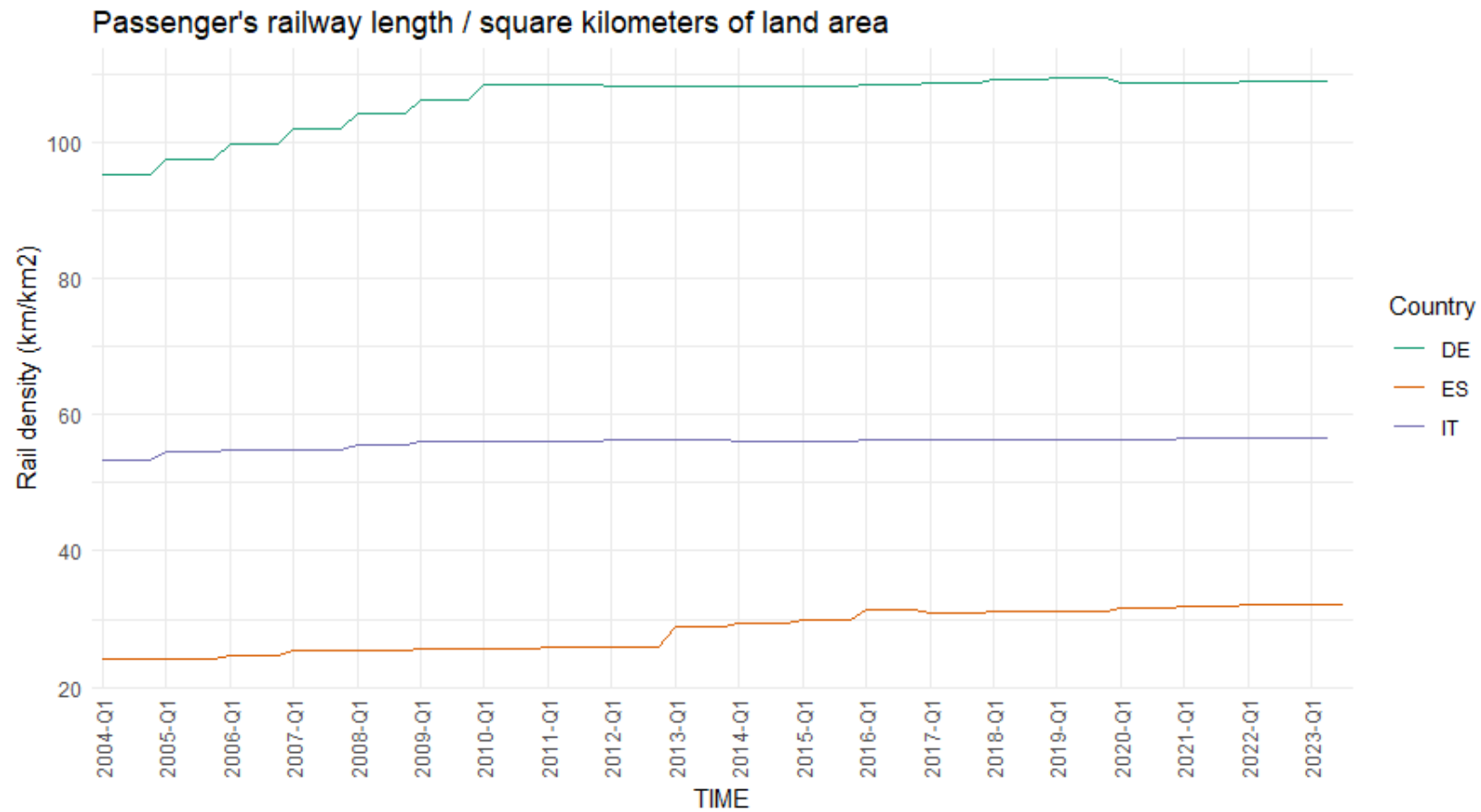


Railway density

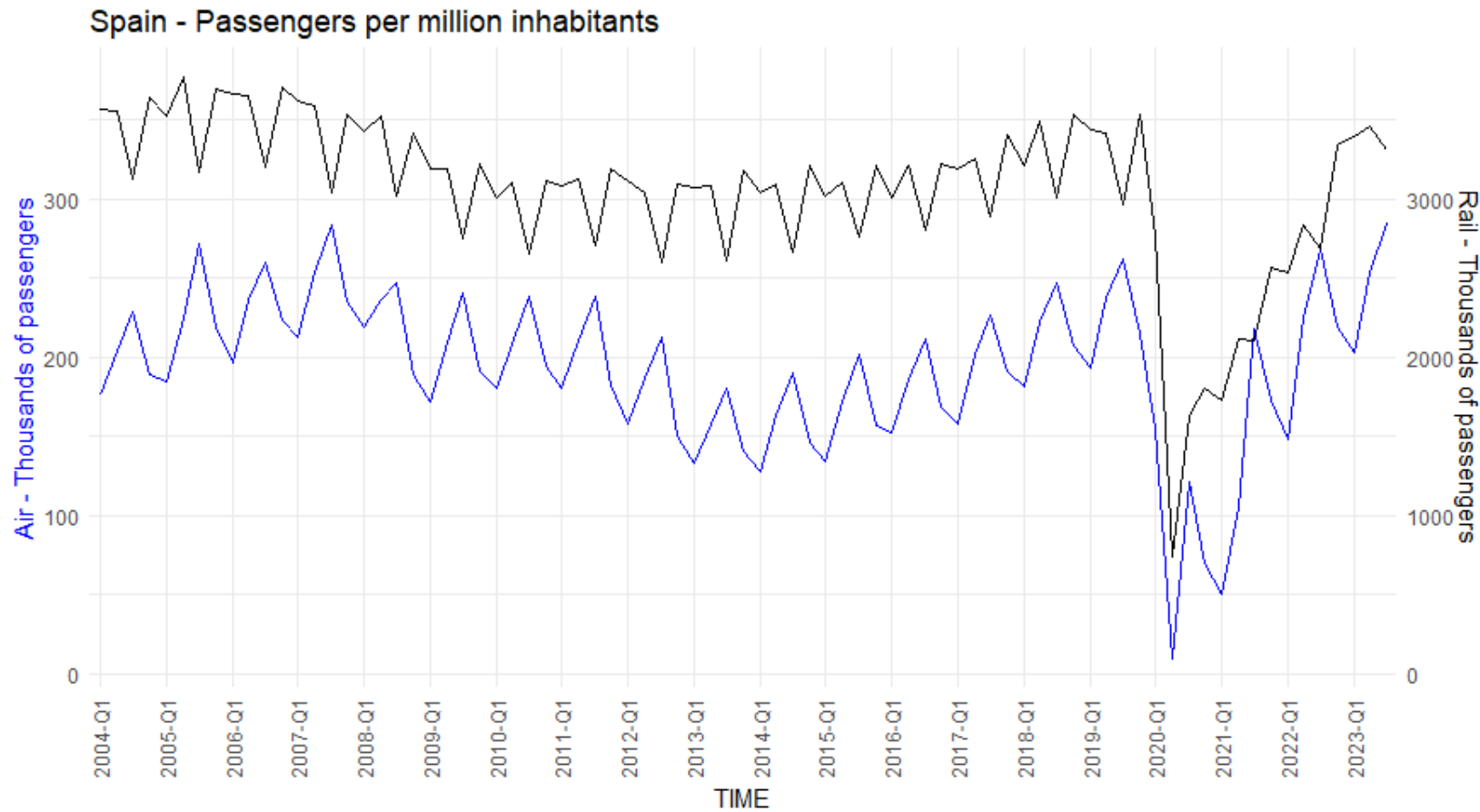
- Given the difference of the railway traffic among the countries, we wanted to check the rail densities.
- Also data from Eurostat.
- It had some missing values in the middle of the series, so we linearly interpolated it, since it is a rather stable measurement.
- It is computed as passenger's railway length divided by land area.



Railway density



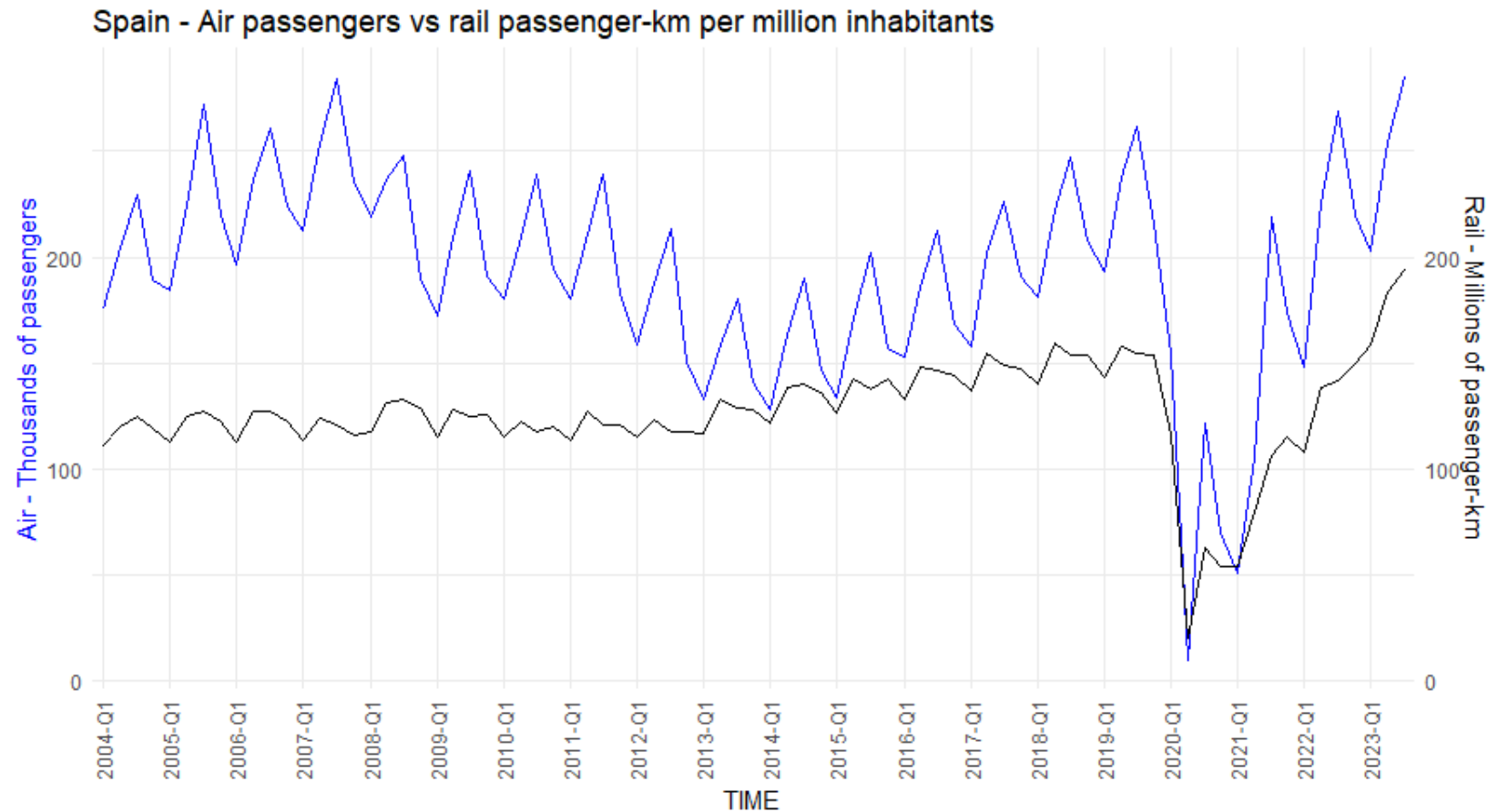
Railway traffic - Spain



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



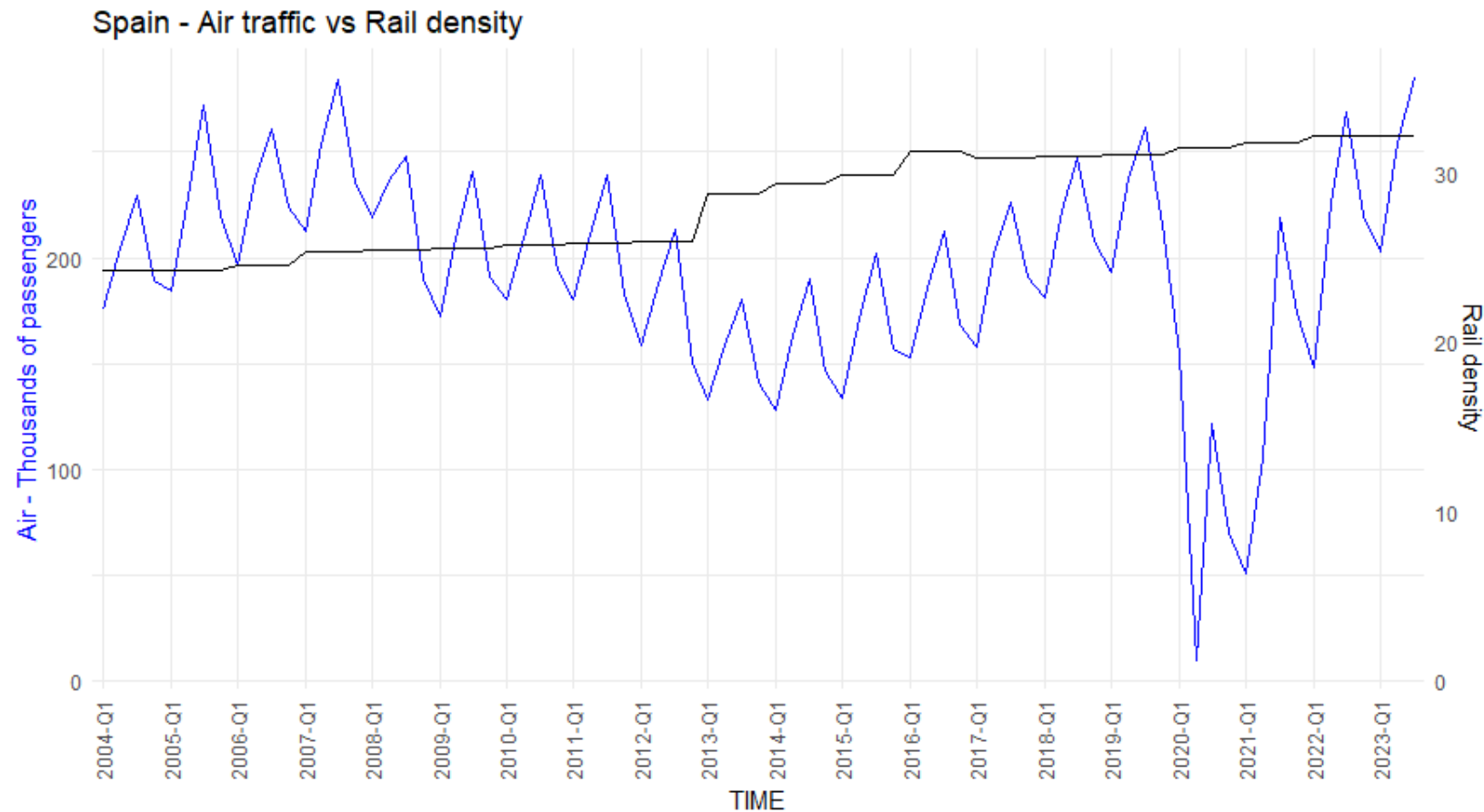
Railway traffic - Spain



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



Rail density - Spain



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.

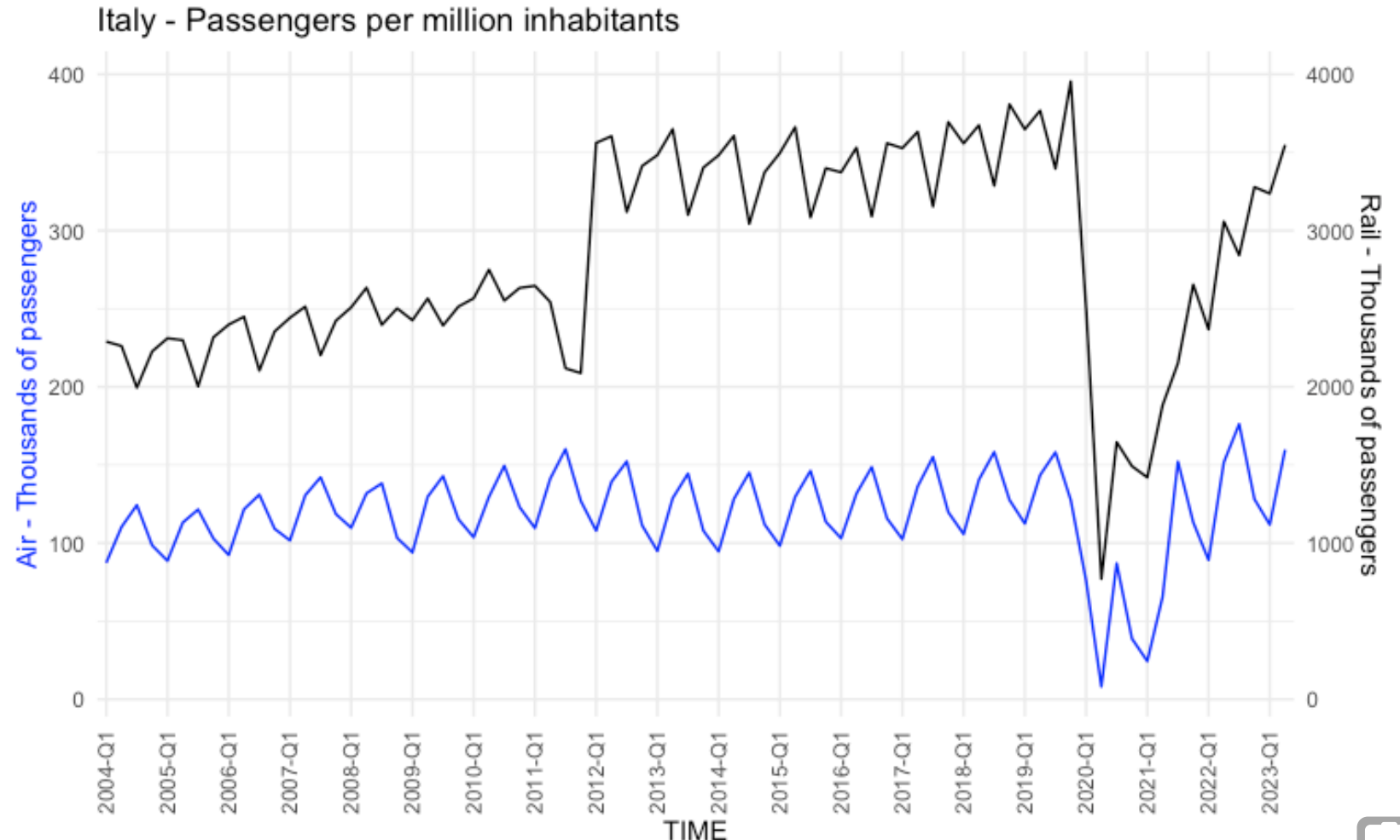


Railway traffic - Italy

Nov 2011: Prime Minister Monti - reform programme and austerity measures to adress **Italy's debt crisis**.

Nov 2011: Trenitalia Frecciarossa introduces trains with four levels of service, together with a reduction of prices

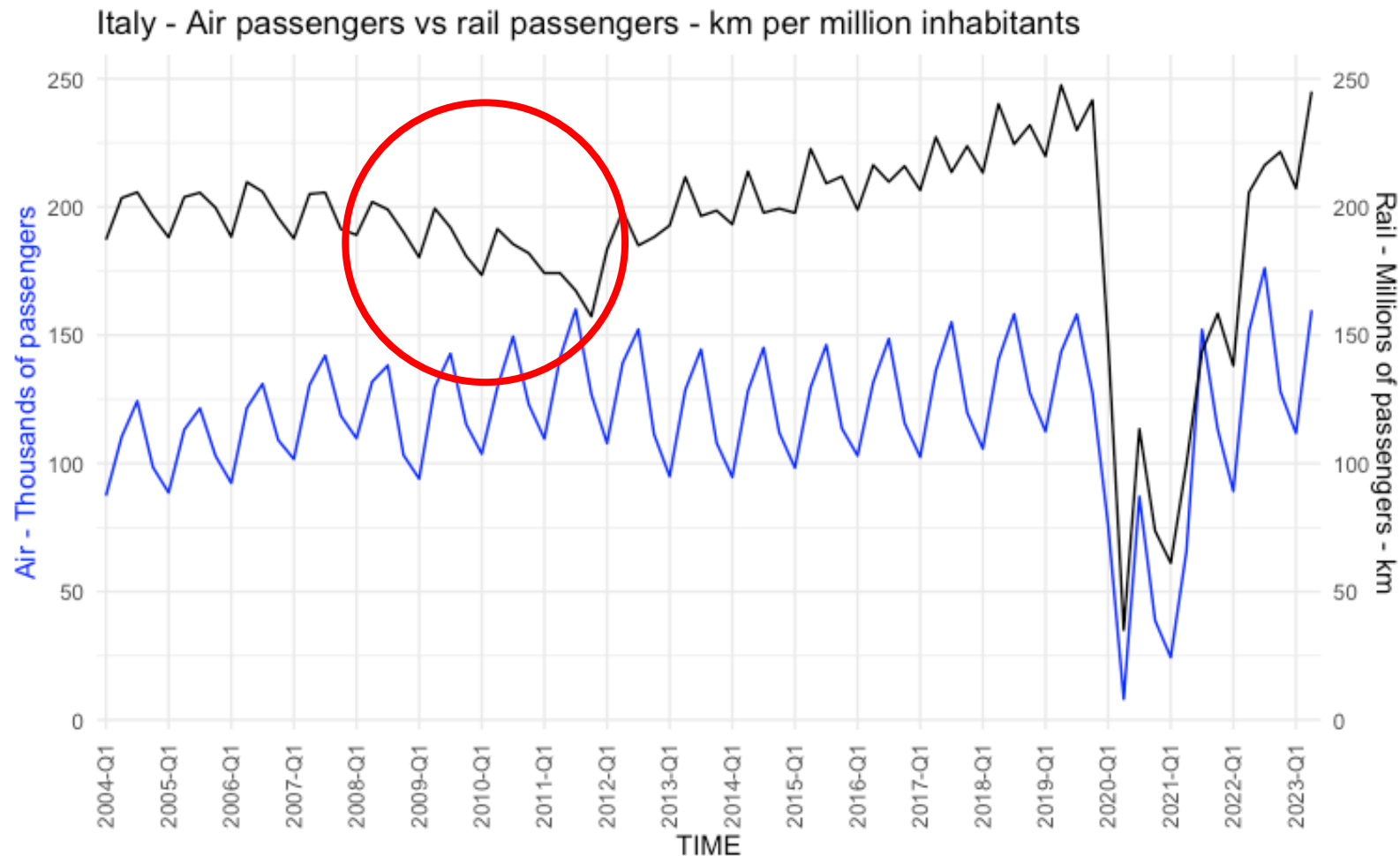
Apr 2012: Italo begins operations with the new high-speed trains.



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



Railway traffic - Italy

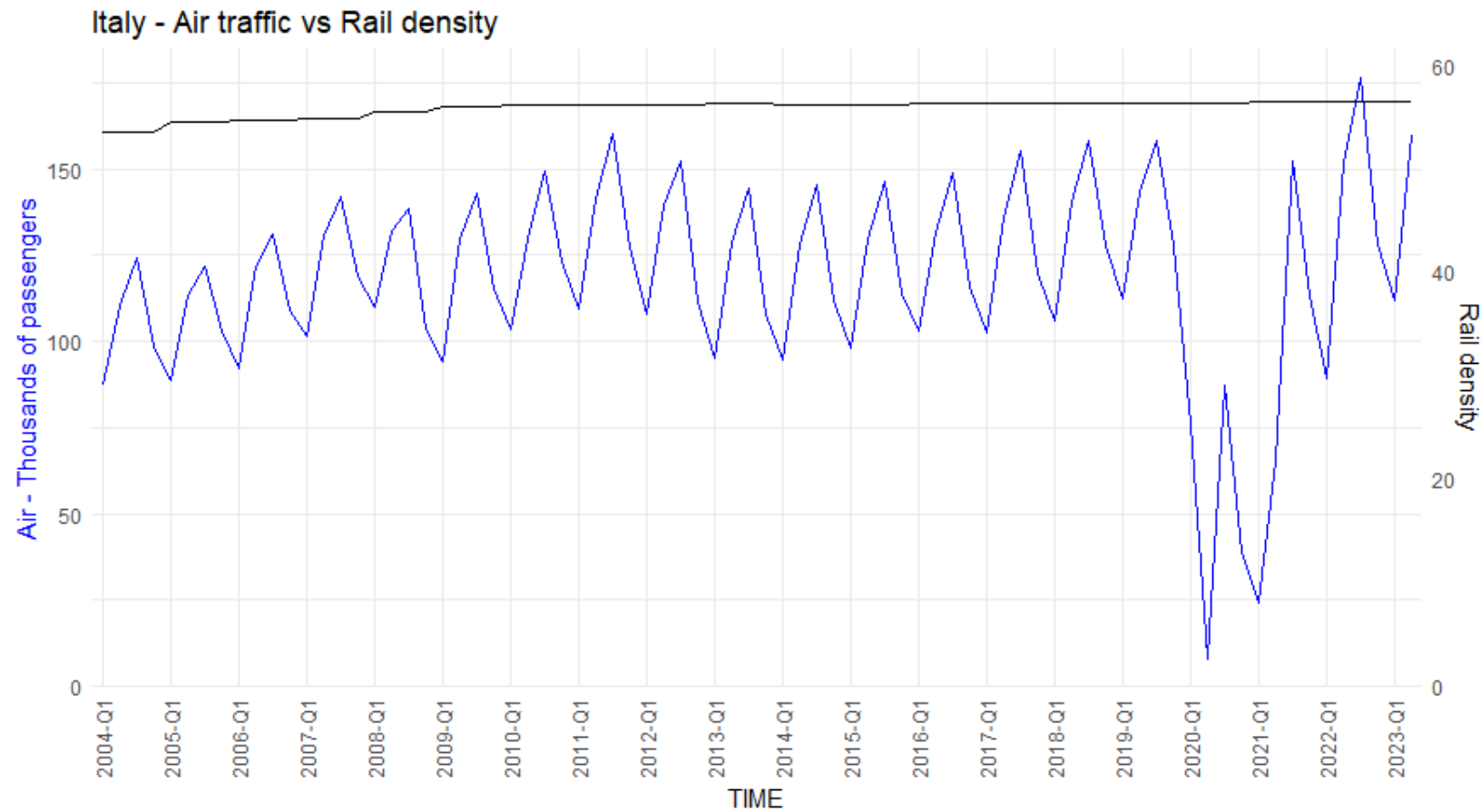


- **2008** Global Financial crisis
- **2011** Italy's debt crisis

* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



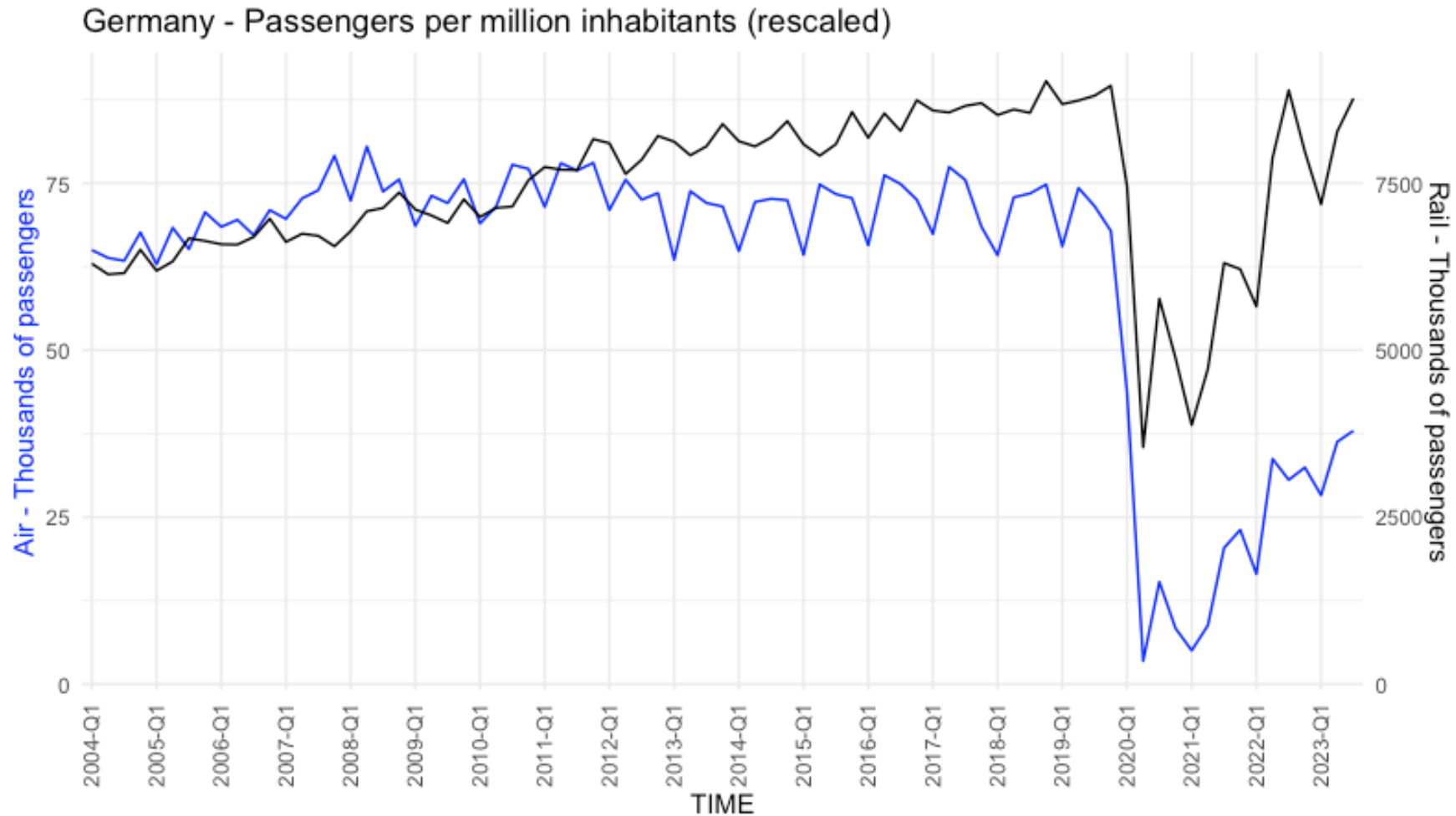
Rail density - Italy



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



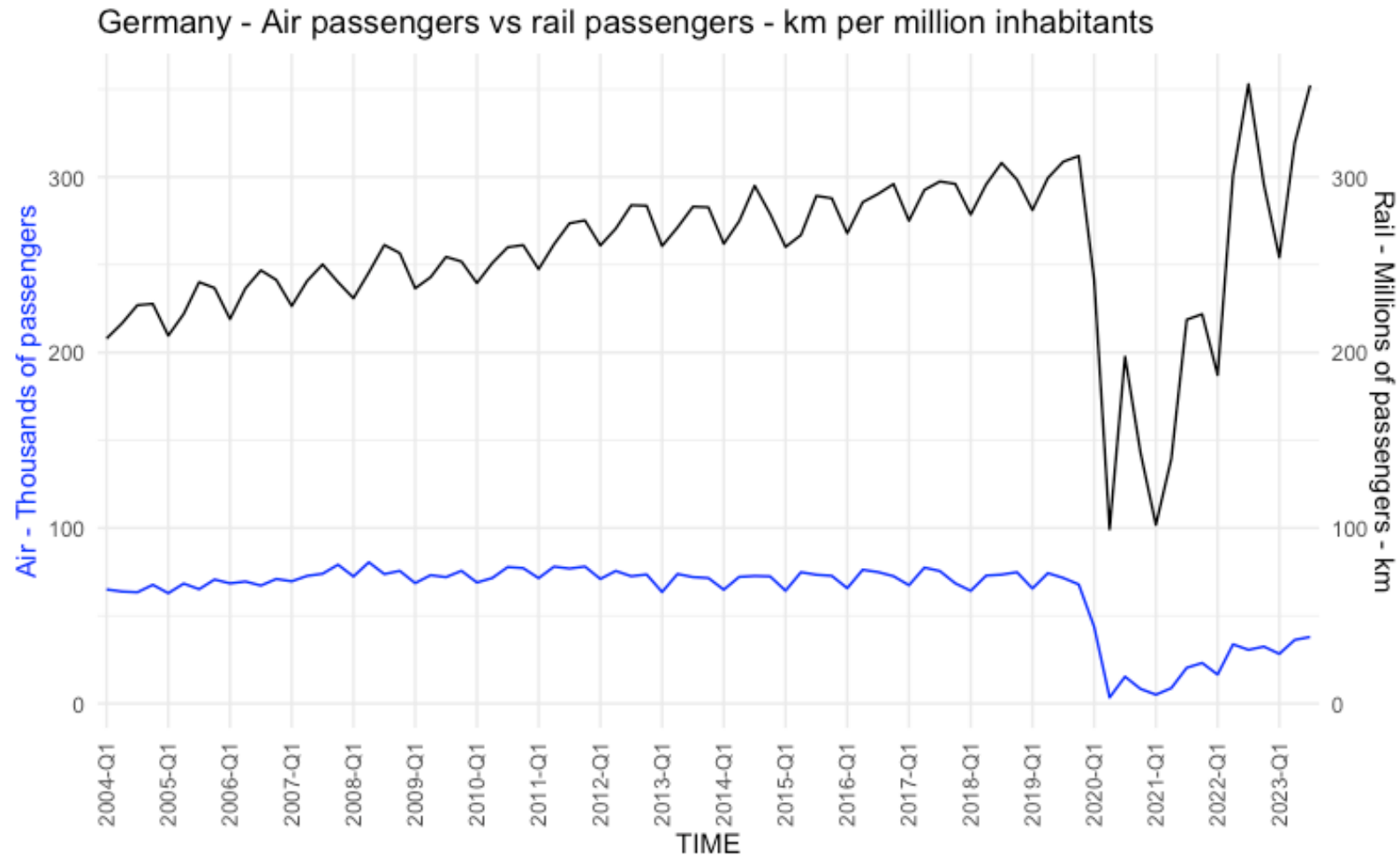
Railway traffic - Germany



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



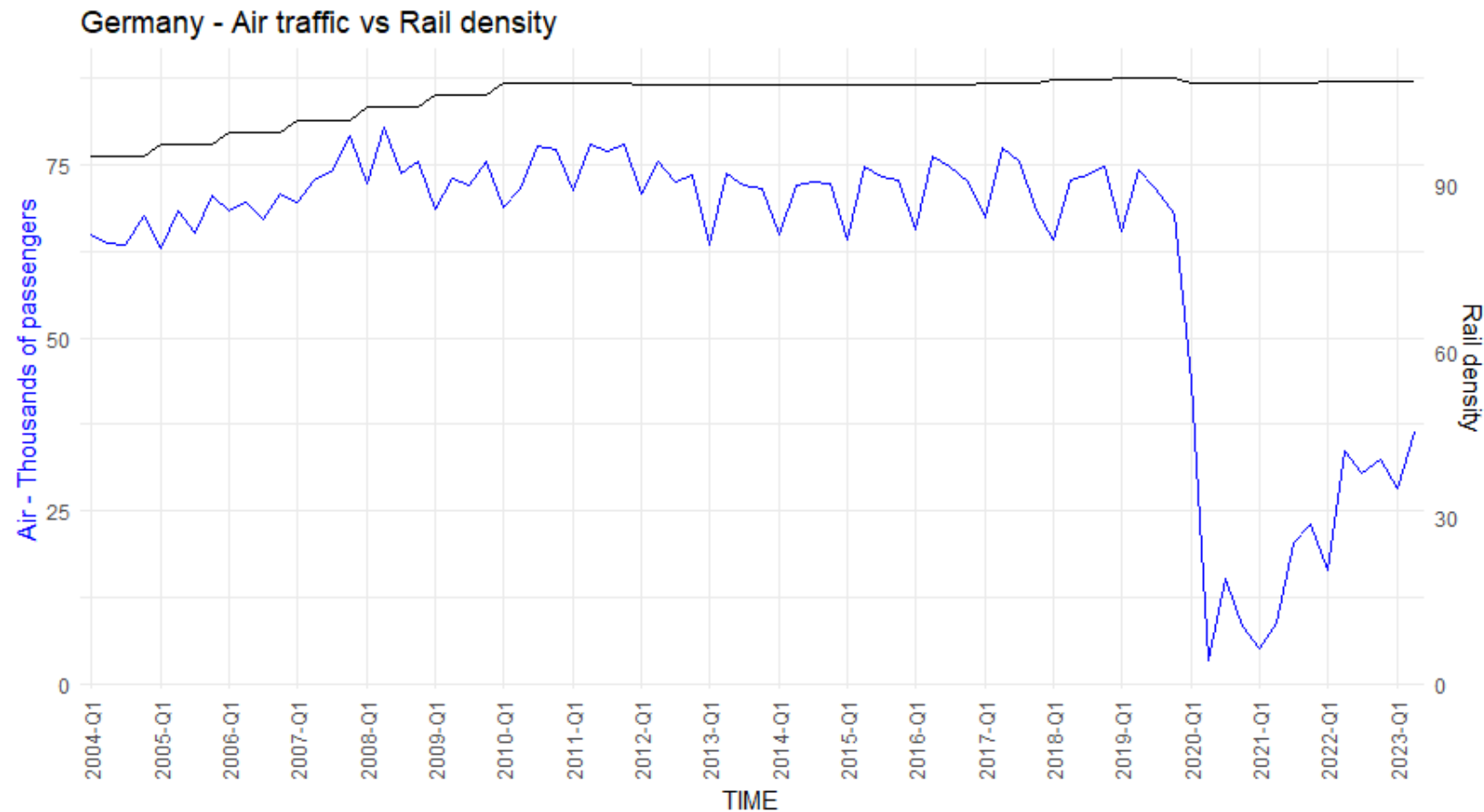
Railway traffic - Germany



* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



Rail density - Germany

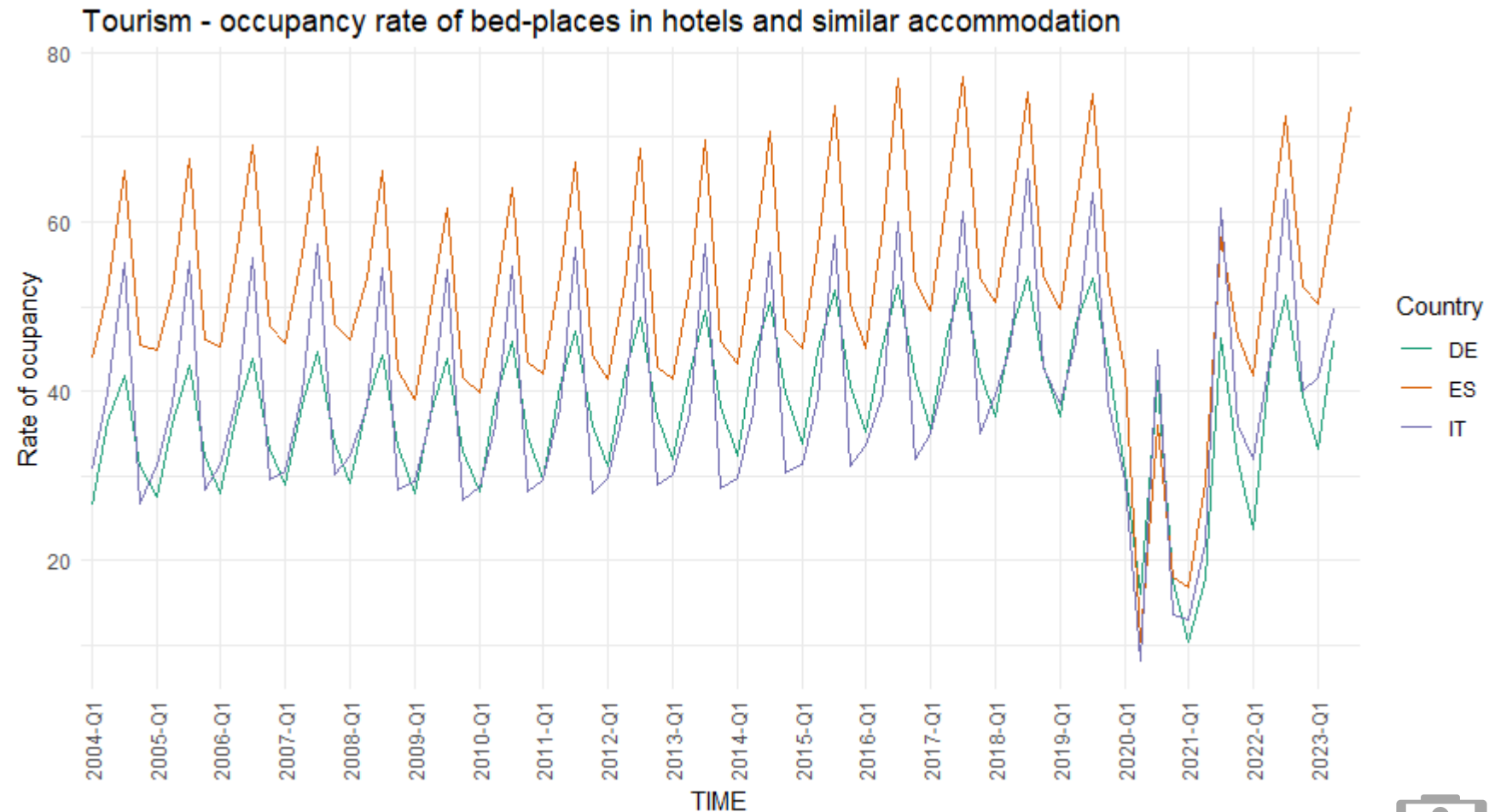


* For all the plots that compare two series, we tweaked the scale of the axes to facilitate the visualization of the series' shape.



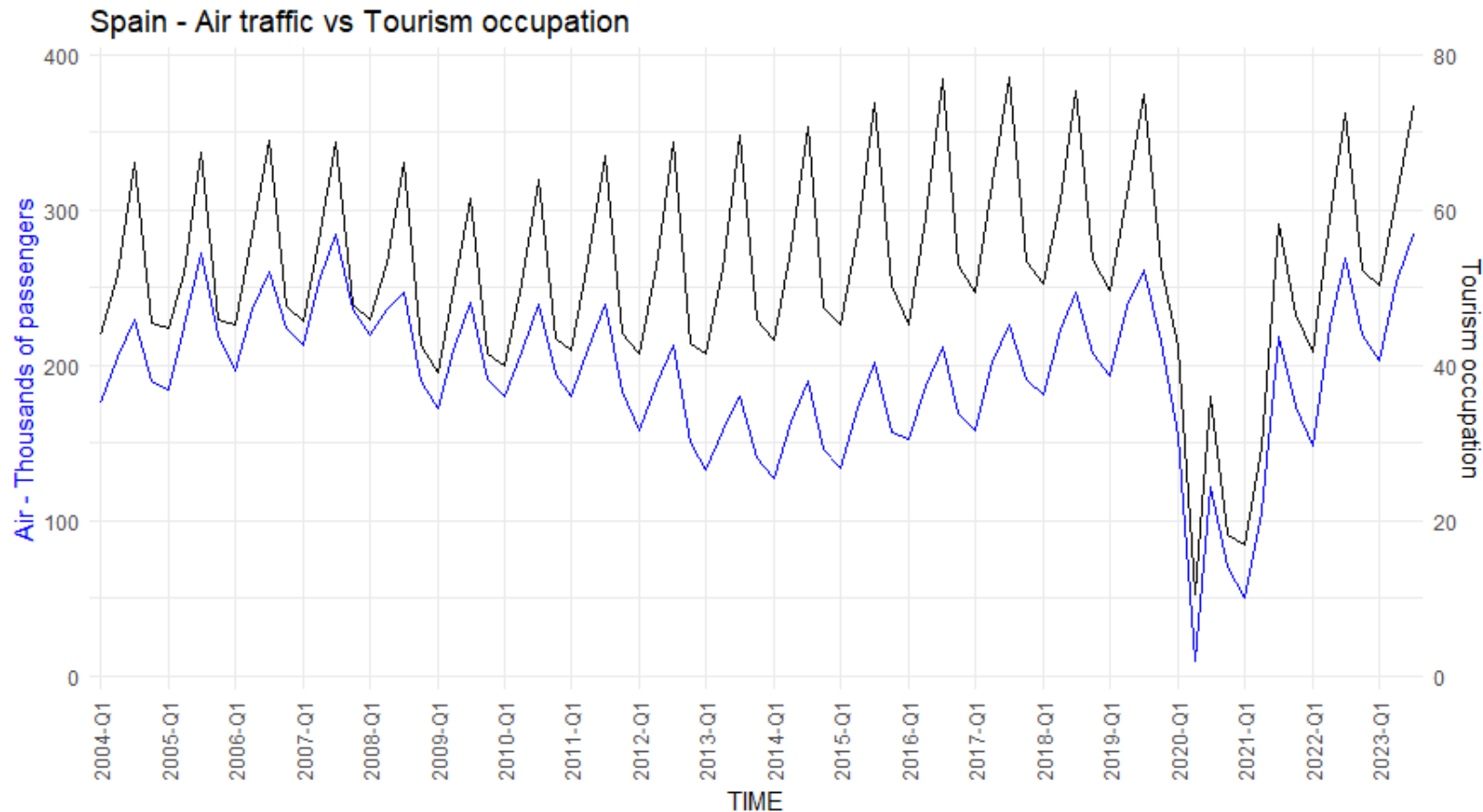
Tourism occupation

- Flights are usually related to travelling with touristic purposes. Is tourism by itself affecting air traffic?
- We used the occupancy rate of **bed-places** in hotels and similar accommodation from [Eurostat](#).
- Monthly data grouped as the average by quarter.



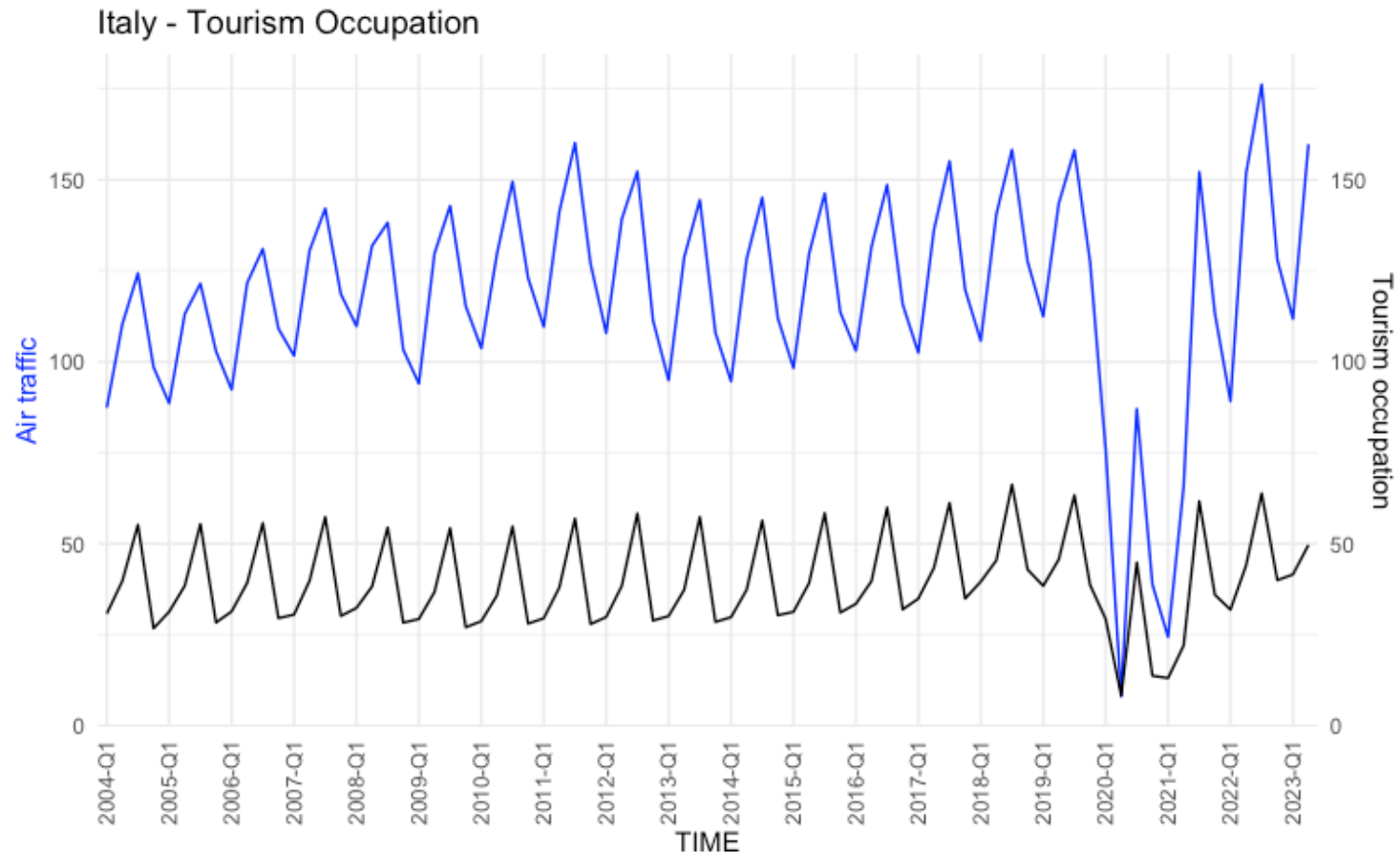
Tourism Occupation - Spain

- Kind of reminds the GDP plot (adding seasonality). Spain is a touristic country.
- Do we need both variables in this case? GDP is more robust, but we will see.



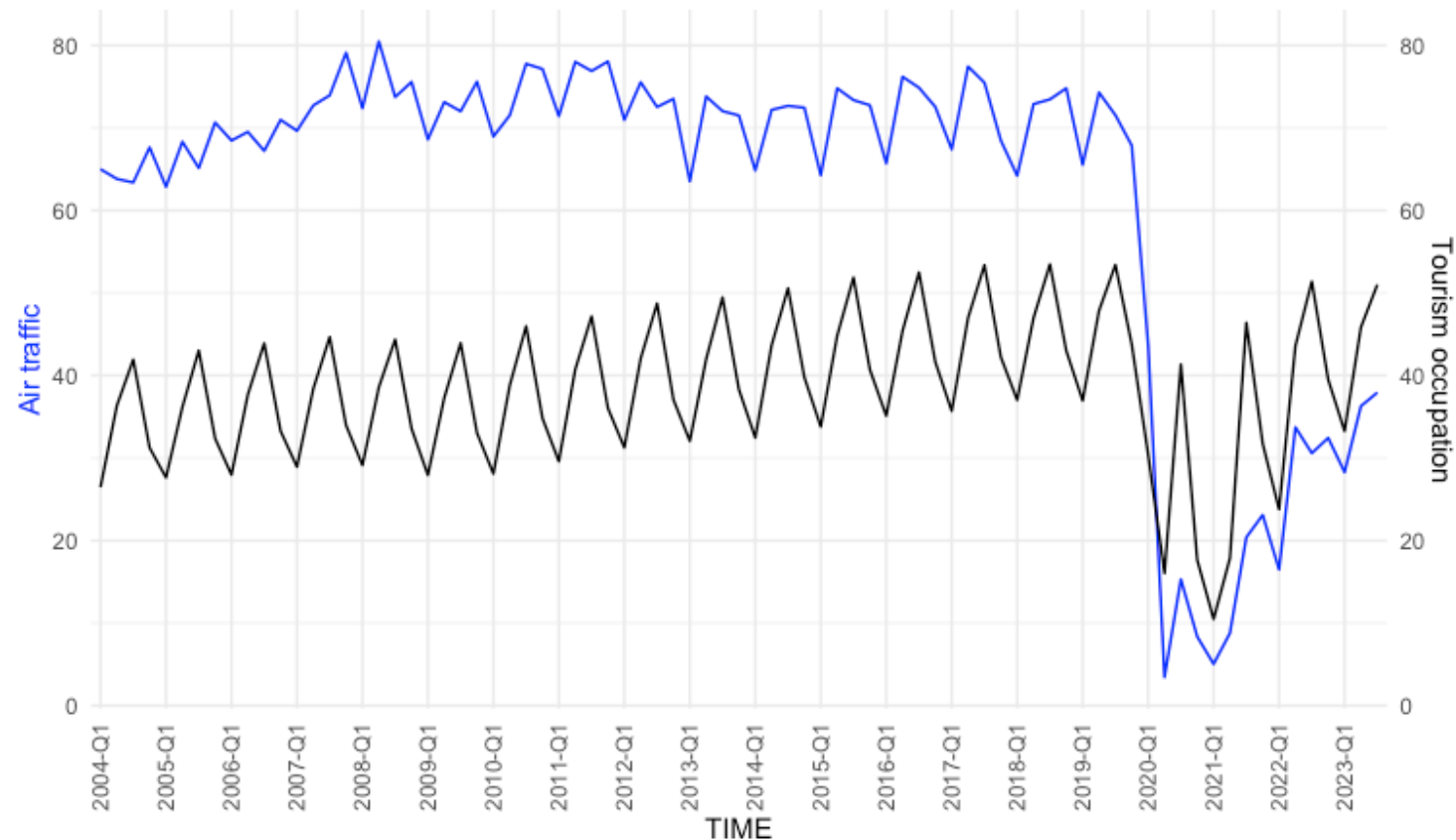
Tourism Occupation - Italy

- Seems a little flat, apart from seasonality.



Tourism Occupation - Germany

- No clear pattern.

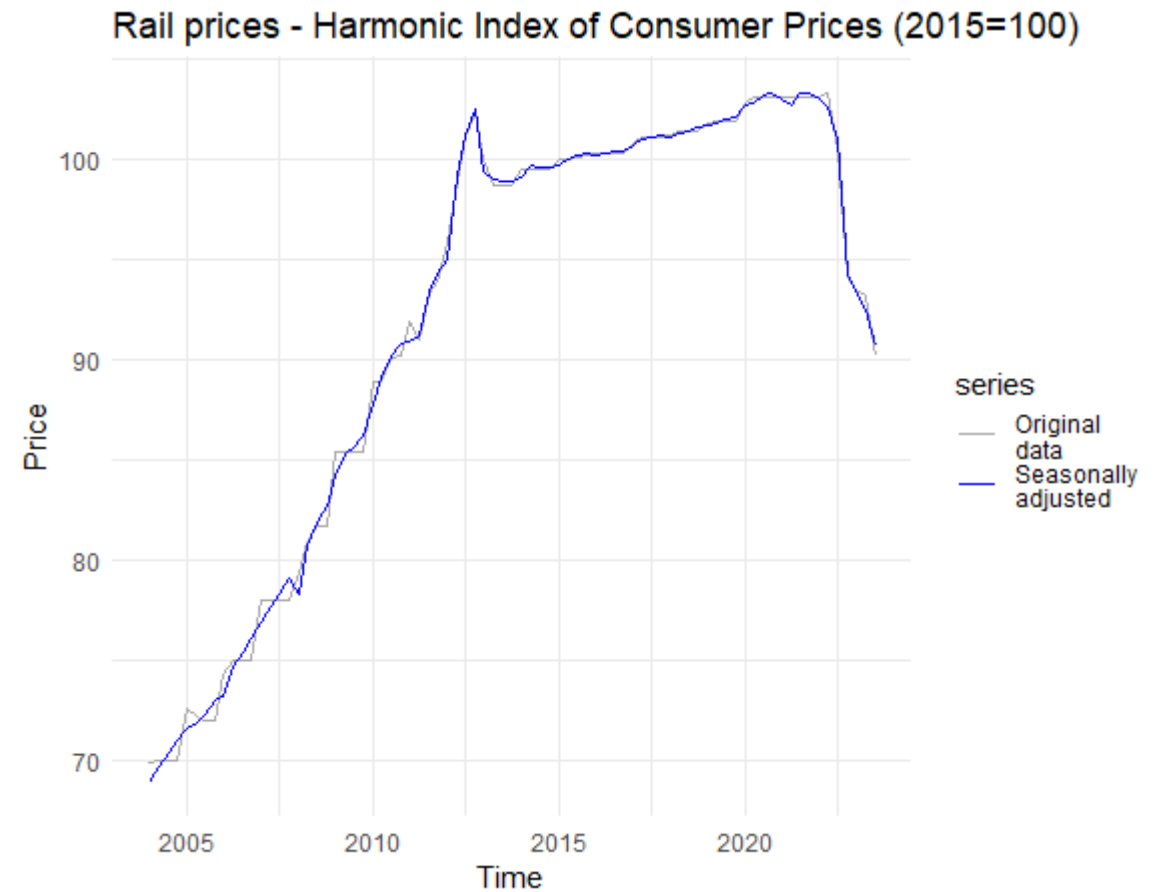
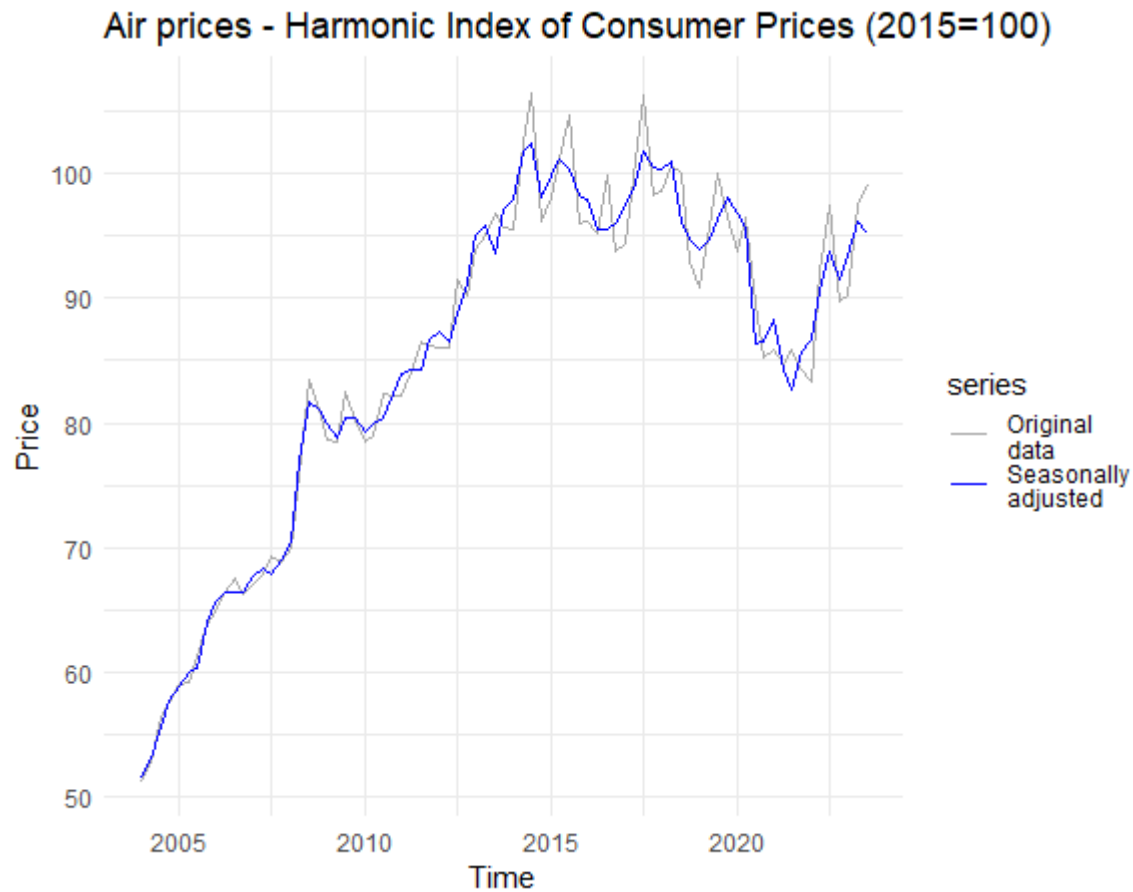


Time dependencies

- For the models that we are going to see next, we included explicit variables to account for **trend** and **seasonality**.
- Also, the type of models that we used aim to capture other kinds of time dependencies (more on this later).
- Since some of our potential predictors are also affected by seasonality, we considered a **seasonally adjusted version** of these variables. We used the **x11 decomposition** for this purpose.



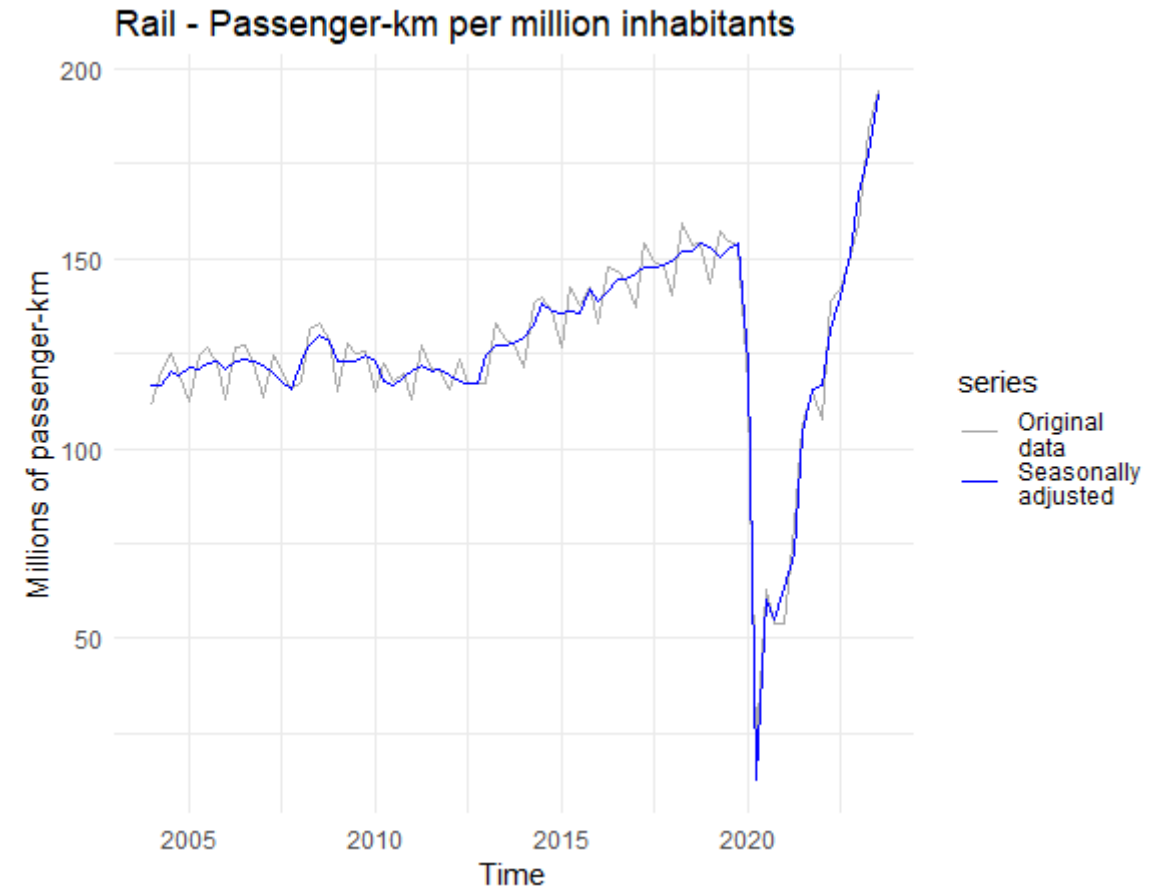
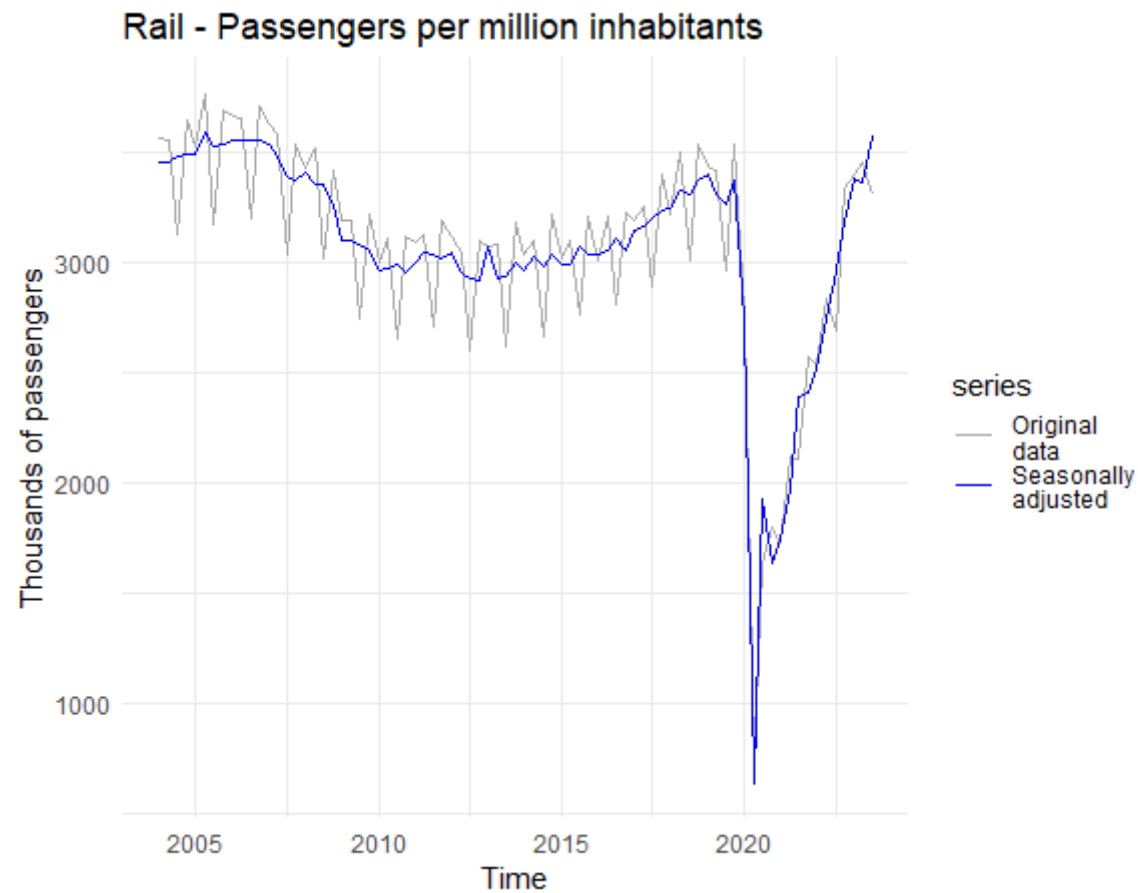
Seasonal adjustments



* Examples from Spain, but the same logic applies to the other countries



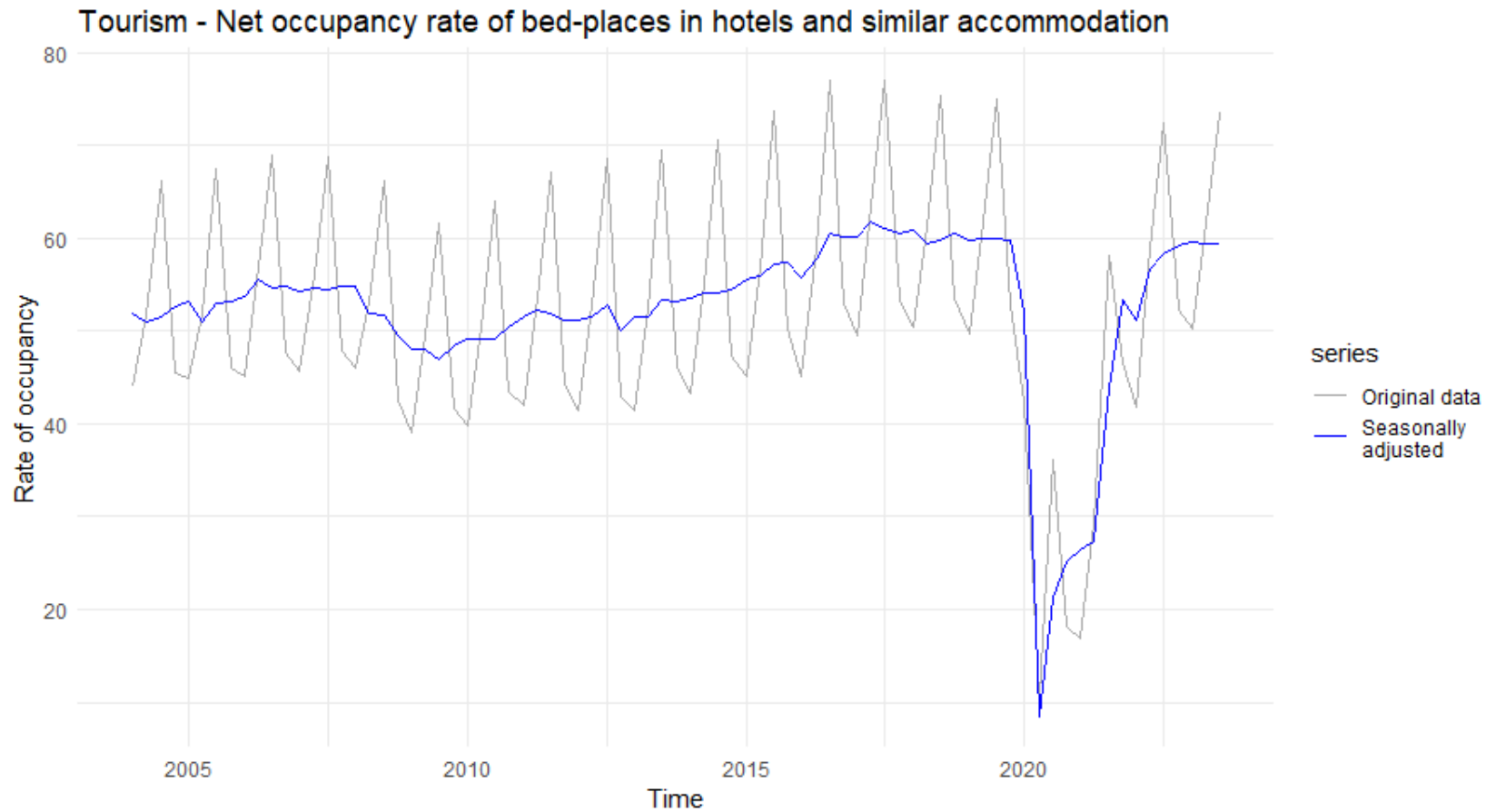
Seasonal adjustments



* Examples from Spain, but the same logic applies to the other countries



Seasonal adjustments



* Examples from Spain, but the same logic applies to the other countries



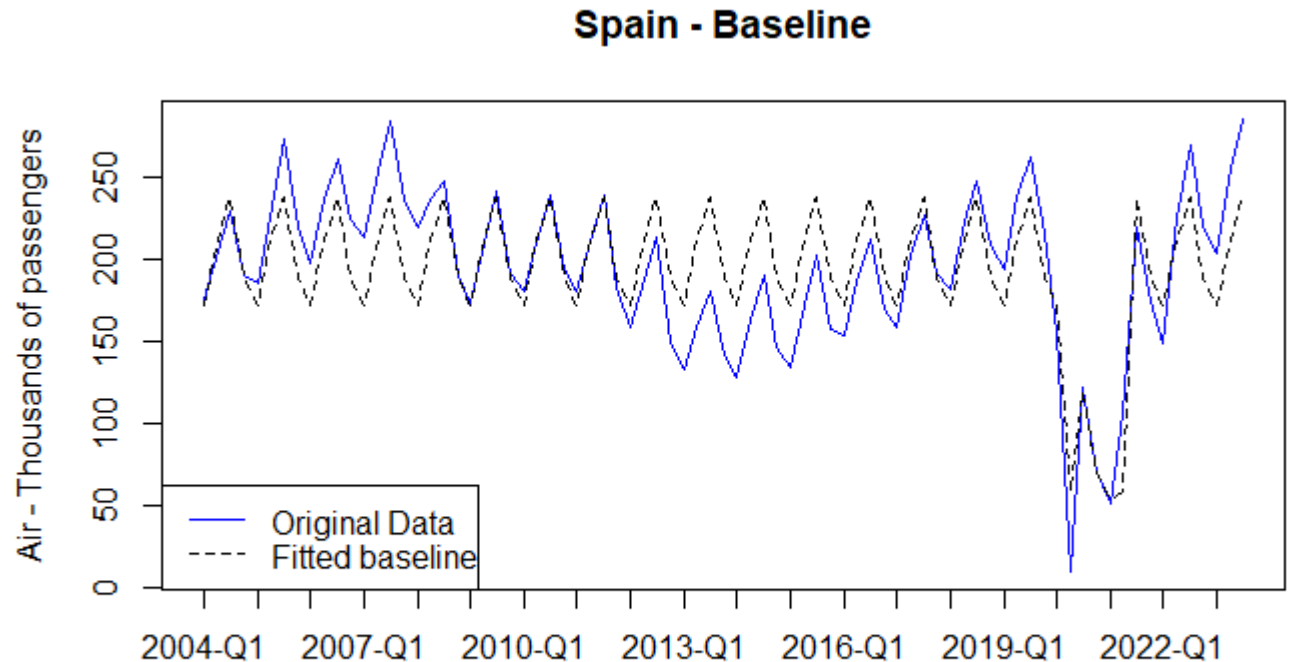
Statistical modeling

- We aim for a well-fitting model, but our primary goal is to **explain the data**.
- We avoided families of models that would be harder to interpret.
- We used an approach with two parts:
 - A **time-series linear model** to grasp the meaningful variables and understand our series.
 - A **dynamic regression with ARIMA errors** to capture the remaining time-dependencies (with the autoregression and the moving average components) and make an improvement in terms of fit
- To handle the bias-variance trade-off and avoid overfitting, we used the **LASSO and the best subset selection** approaches as feature selectors, along with information criteria.



Statistical modeling - Spain

- We started with a **very simple** linear model as a reference point. It only included the non-time-series predictors: **trend***, **quarter** (the seasonality), and **COVID restrictions**.
- These variables are **not explicit external factors**, except from COVID which we *know* that had an impact on air traffic.
- Therefore, this model serves as a **baseline** to compare and check how well our other (time-series) predictors explain the data.



Coefficients:	TIME				
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	171.754	6.559	26.186	< 2e-16	***
QUARTERQ2	39.082	9.487	4.120	9.89e-05	***
QUARTERQ3	66.198	9.195	7.199	4.45e-10	***
QUARTERQ4	16.919	9.316	1.816	0.0735	.
COVID_AVG_RESTRICTIONS_FACTOR1	-153.236	21.674	-7.070	7.72e-10	***
COVID_AVG_RESTRICTIONS_FACTOR2	-118.707	17.233	-6.889	1.68e-09	***

AICc = 765.9

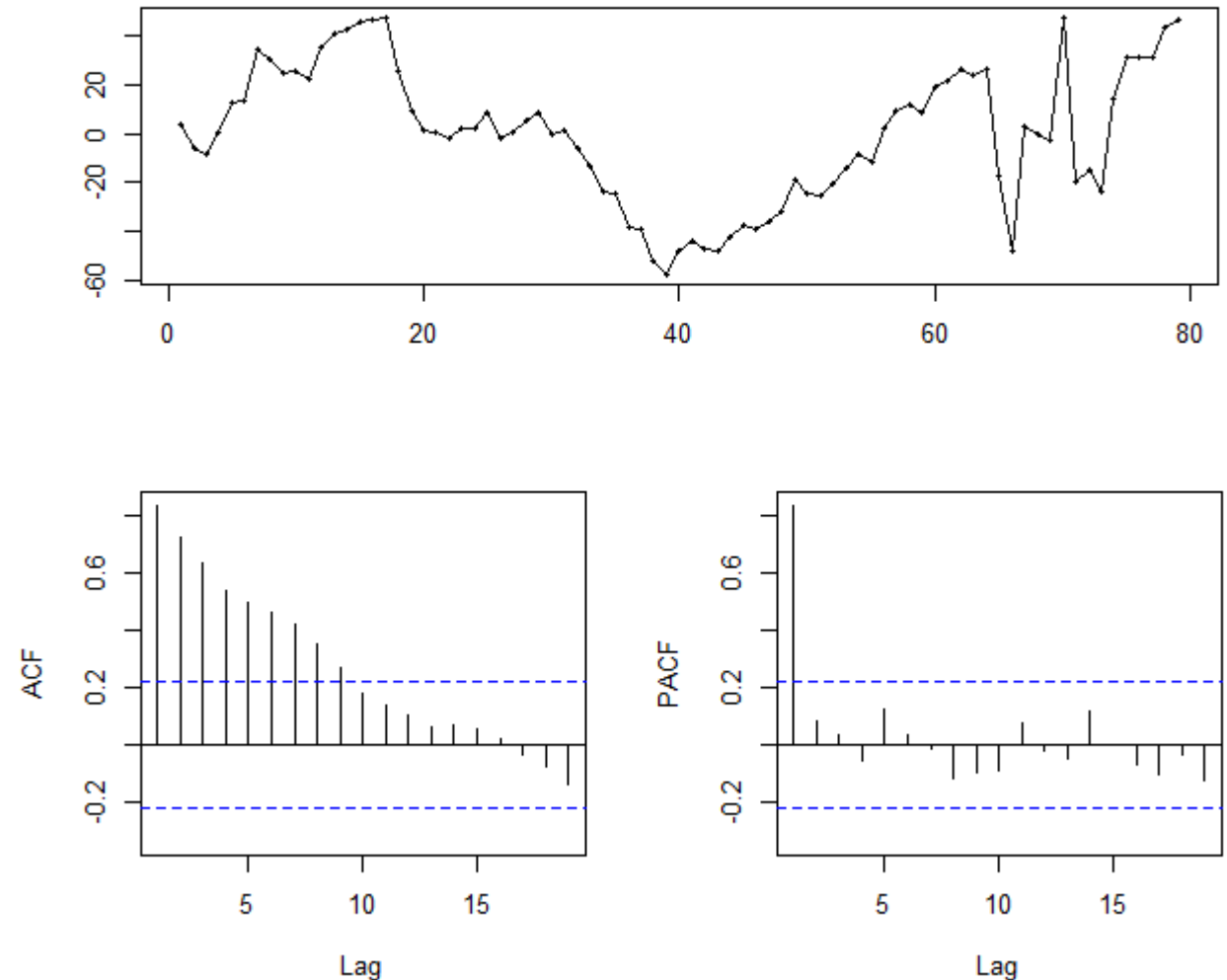
Adjusted R² = 0.65

* Trend was not significant for Spain, so we removed it

Statistical modeling - Spain

- It does not capture the time dependencies well.
- Hints of a pattern in residuals. **Not white noise.**
- Several of the first lags have a considerable **autocorrelation.**

Spain - Baseline residuals and autocorrelation



Statistical modeling - Spain

- We tried both the standard version and the **seasonally adjusted version** of our predictors. The latter worked best.
- The Lasso regression and the best subset selection with adjusted R^2 and C_p as criteria resulted in the **same model**.
- The best subset selection with BIC as criterion, omitted the passenger-km variable.
- However, we may have **collinearity** issues.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-250.5824	25.3484	-9.886	6.44e-15	***
QUARTERQ2	34.9441	3.1479	11.101	< 2e-16	***
QUARTERQ3	60.8337	3.0516	19.935	< 2e-16	***
QUARTERQ4	10.7901	3.0912	3.491	0.000839	***
COVID_AVG_RESTRICTIONS_FACTOR1	-59.0041	12.0977	-4.877	6.50e-06	***
COVID_AVG_RESTRICTIONS_FACTOR2	-61.2979	10.0742	-6.085	5.55e-08	***
R_MIO_PKM_S_ADJ	0.1979	0.1046	1.891	0.062707	.
GDP_PC	96.6077	4.8945	19.738	< 2e-16	***
RAIL_DENSITY	-6.2683	0.5126	-12.228	< 2e-16	***

AICc = 595,5

Adjusted R^2 = 0.96

```
> vif(esPasBestSeasAdj)
```

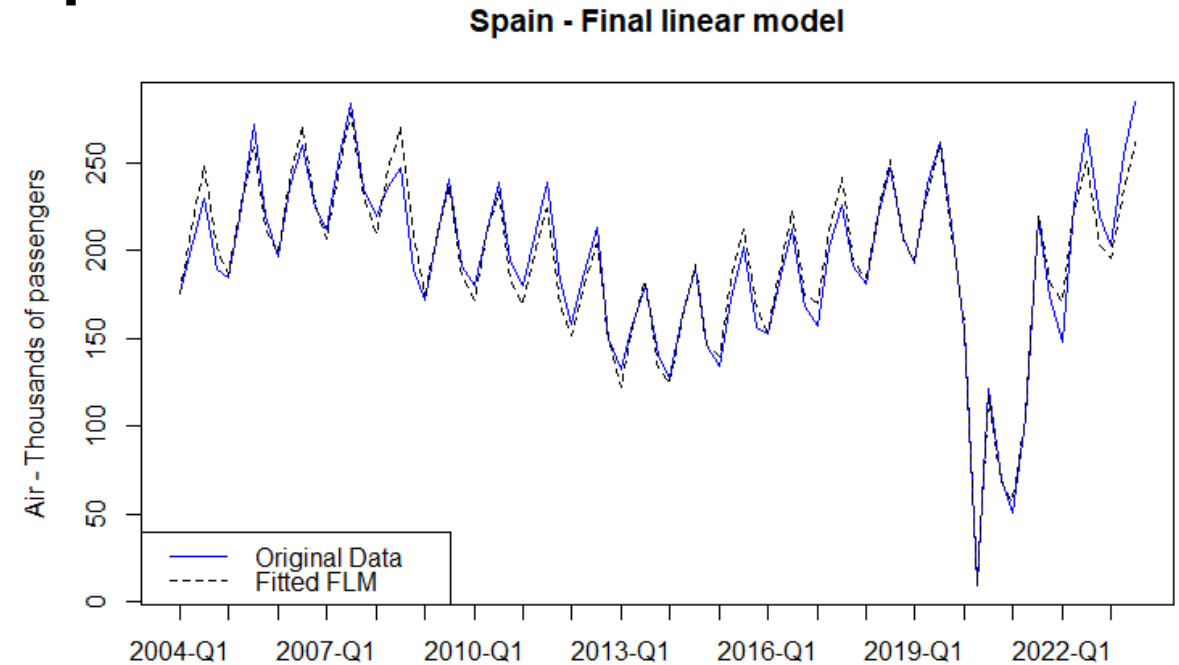
	GVIF	Df	GVIF^(1/(2*Df))
QUARTER	1.112085	3	1.017864
COVID_AVG_RESTRICTIONS_FACTOR	5.521797	2	1.532922
R_MIO_PKM_S_ADJ	5.902836	1	2.429575
GDP_PC	1.740271	1	1.319193
RAIL_DENSITY	1.973513	1	1.404818

Statistical modeling - Spain

- By **removing passenger-km**, the least significant variable from before (was even discarded by the BIC criterion), we **eliminate our collinearity** issues.
- Slightly worse AICc, but **more stable** model.

```
> vif(esPasBestSeasAdjVIF)
```

	GVIF	Df	GVIF^(1/(2*Df))
QUARTER	1.106017	3	1.016936
COVID_AVG_RESTRICTIONS_FACTOR	1.456347	2	1.098541
GDP_PC	1.313449	1	1.146058
RAIL_DENSITY	1.274264	1	1.128833



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-268.2362	23.9914	-11.181	< 2e-16	***
QUARTERQ2	35.3268	3.1979	11.047	< 2e-16	***
QUARTERQ3	61.0889	3.1035	19.684	< 2e-16	***
QUARTERQ4	10.8359	3.1467	3.444	0.000967	***
COVID_AVG_RESTRICTIONS_FACTOR1	-76.3896	8.0069	-9.540	2.37e-14	***
COVID_AVG_RESTRICTIONS_FACTOR2	-76.5526	6.1453	-12.457	< 2e-16	***
GDP_PC	101.1924	4.3286	23.377	< 2e-16	***
RAIL_DENSITY	-5.6912	0.4193	-13.573	< 2e-16	***

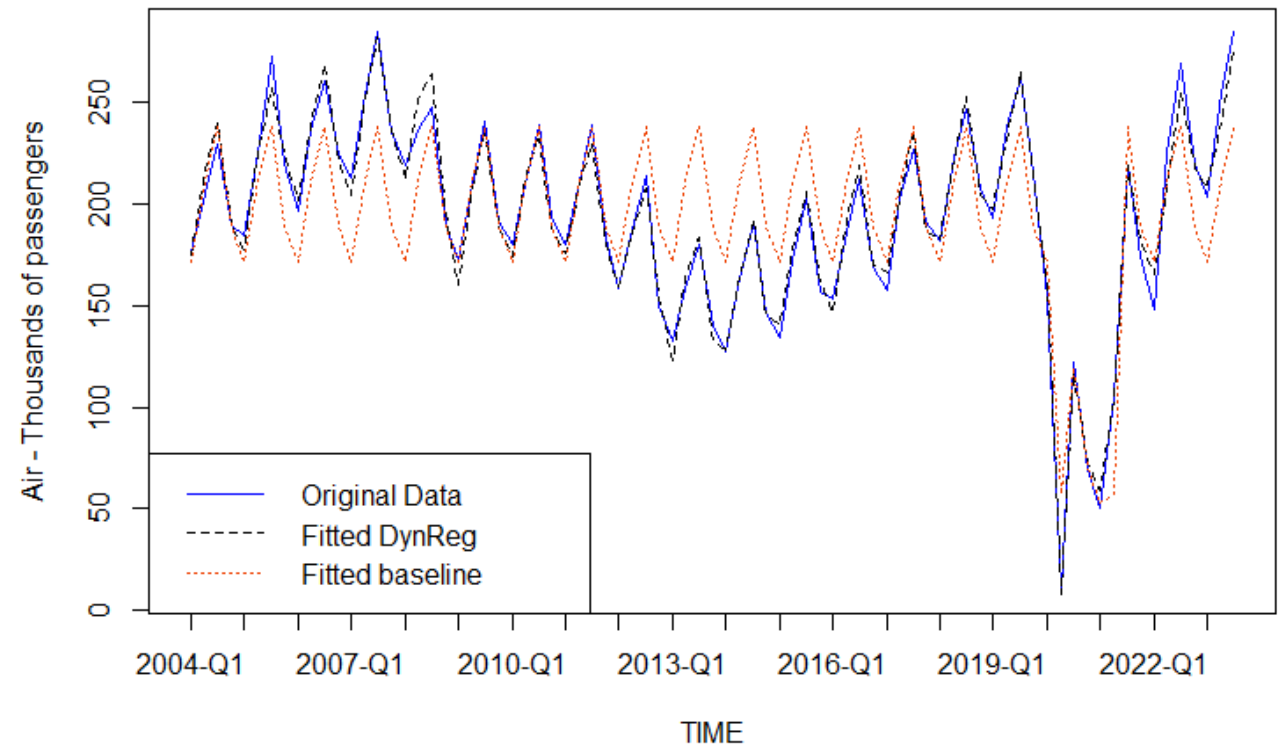
AICc = 596.8

Adjusted R² = 0.96

Statistical modeling - Spain

- The fit is already very good, but we could further improve it.
- We fitted a **dynamic regression with ARIMA errors** to account for the remaining time dependencies.
- If the goal were to make **forecasts**, this would be the model to pick.
- We used the auto.arima, which follows an algorithm to pick the best model parameters (p, d and q).
- Only resulted in one AR term.

Spain - Final model with ARIMA errors (dynamic regression)



Regression with ARIMA(1,0,0) errors

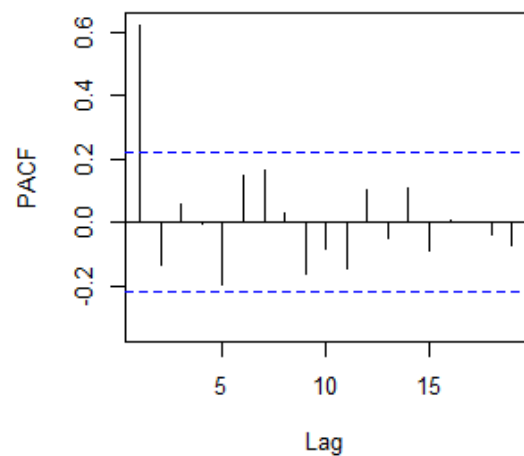
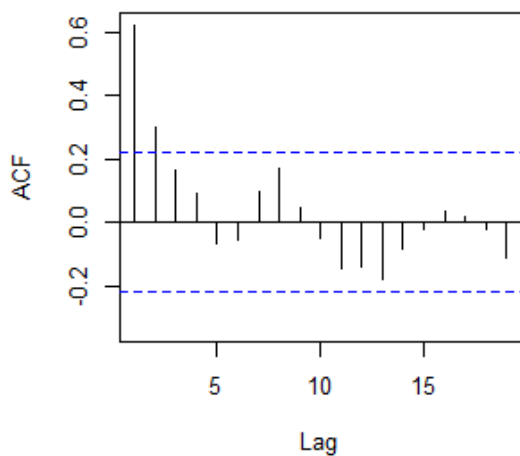
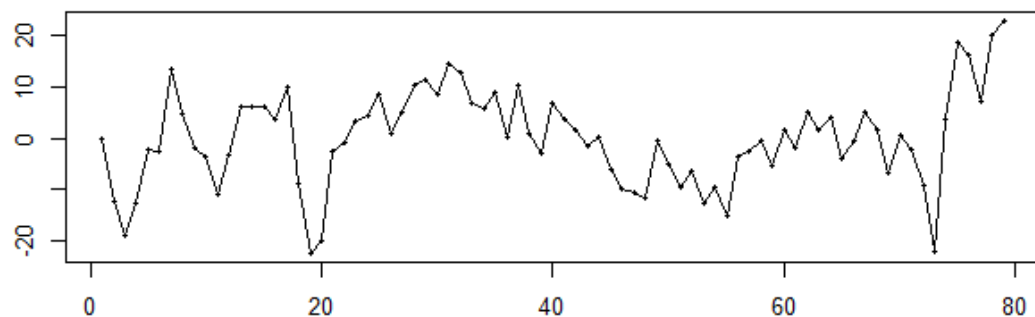
Coefficients:

	ar1	intercept	QUARTERQ2	QUARTERQ3	QUARTERQ4	COVID_AVG_RESTRICTIONS_FACTOR1	COVID_AVG_RESTRICTIONS_FACTOR2
	0.6795	-298.9944	34.8709	60.6412	11.1269	-72.3212	-75.2416
s.e.	0.0905	38.3523	1.6764	1.8806	1.6910	6.6878	7.3050
	GDP_PC	RAIL_DENSITY					
	104.4569	-5.2695					
s.e.	6.1166	0.8444					

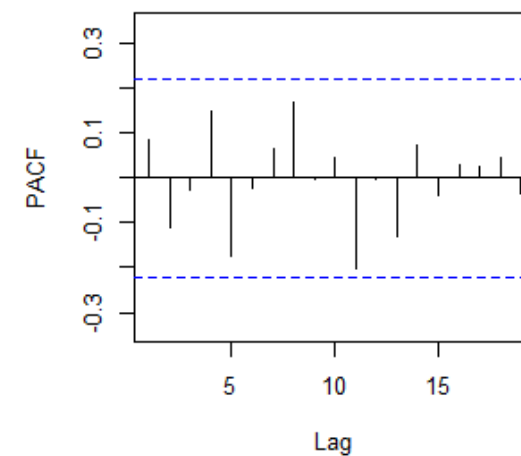
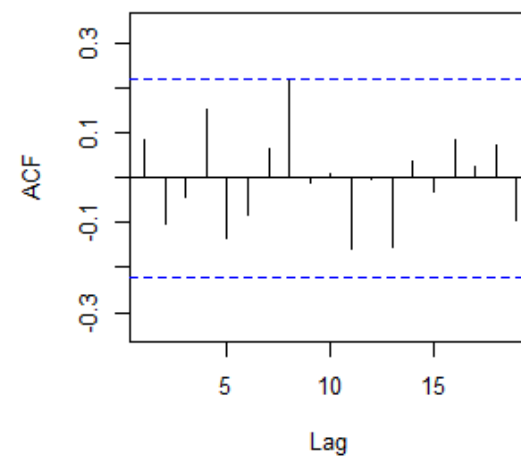
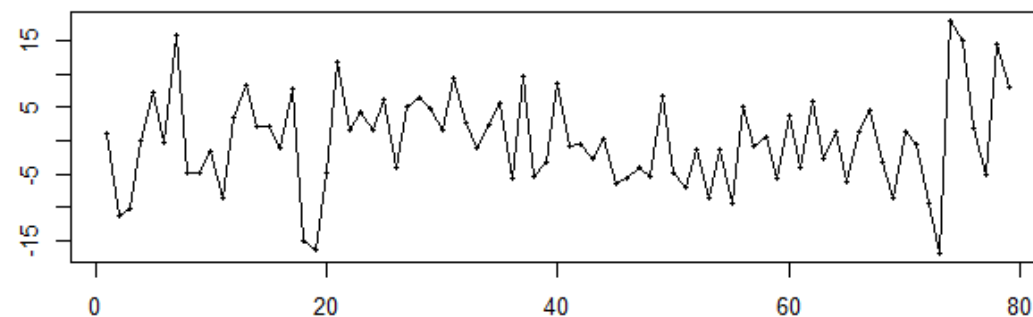
sigma^2 = 56.01: log likelihood = -266.64
AIC=553.28 AICC=556.51 BIC=576.97

Statistical modeling - Spain

Spain - Final linear model - Residuals and autocorrelation



Spain - Final model with ARIMA errors - Residuals and autocorrelation



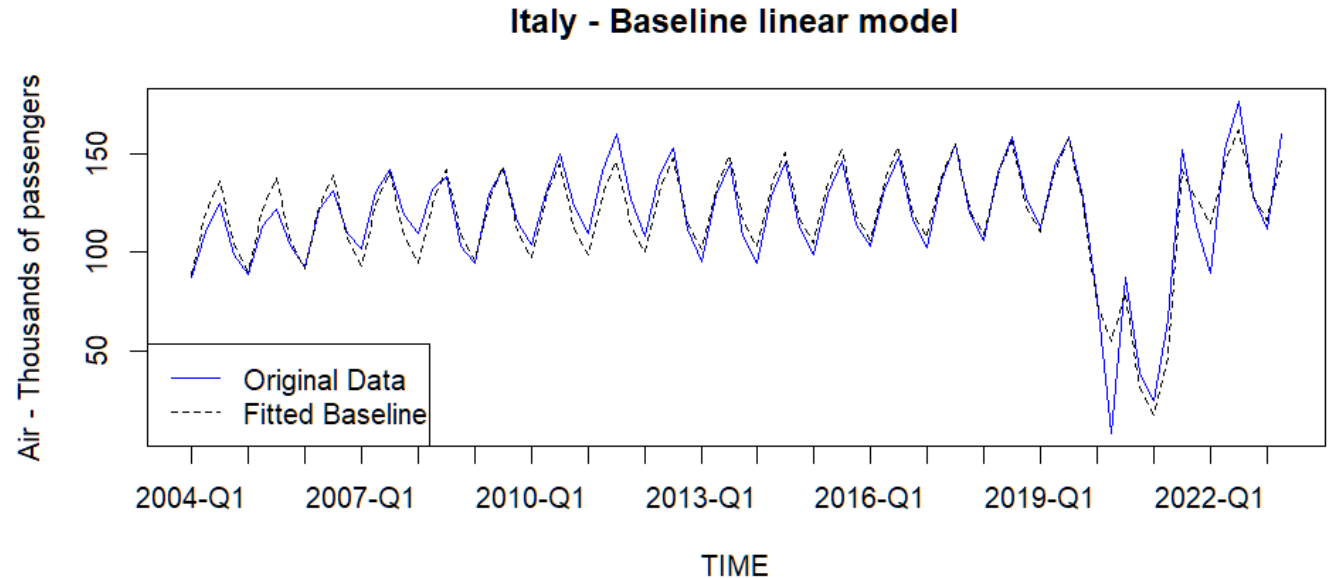
Statistical modeling - Spain

Insights:

- There is indeed a strong seasonal component where Q3 (summer) is related to a greater air traffic.
- **GDP per capita dominates** what happens with air traffic in Spain.
- Then, to better understand this phenomenon, anybody interested would need to **unwrap** what happens with the **economic activity** in the country, which is not an easy task.
- **Are railways really competing with airplanes in Spain?**
- Interestingly, it seems that air traffic in Spain is **somewhat unresponsive to the variations in price** (of both air and rail trips), although it might be too soon to evaluate recent policies.

Statistical modeling - Italy

- Baseline linear model including the non-time-series predictors: **trend**, **quarter** (the seasonality), and **COVID restrictions**.
- All parameters revealed to be significant.
- Good initial fit, because the timeseries exhibits a linear trend and consistent seasonality.



Coefficients:

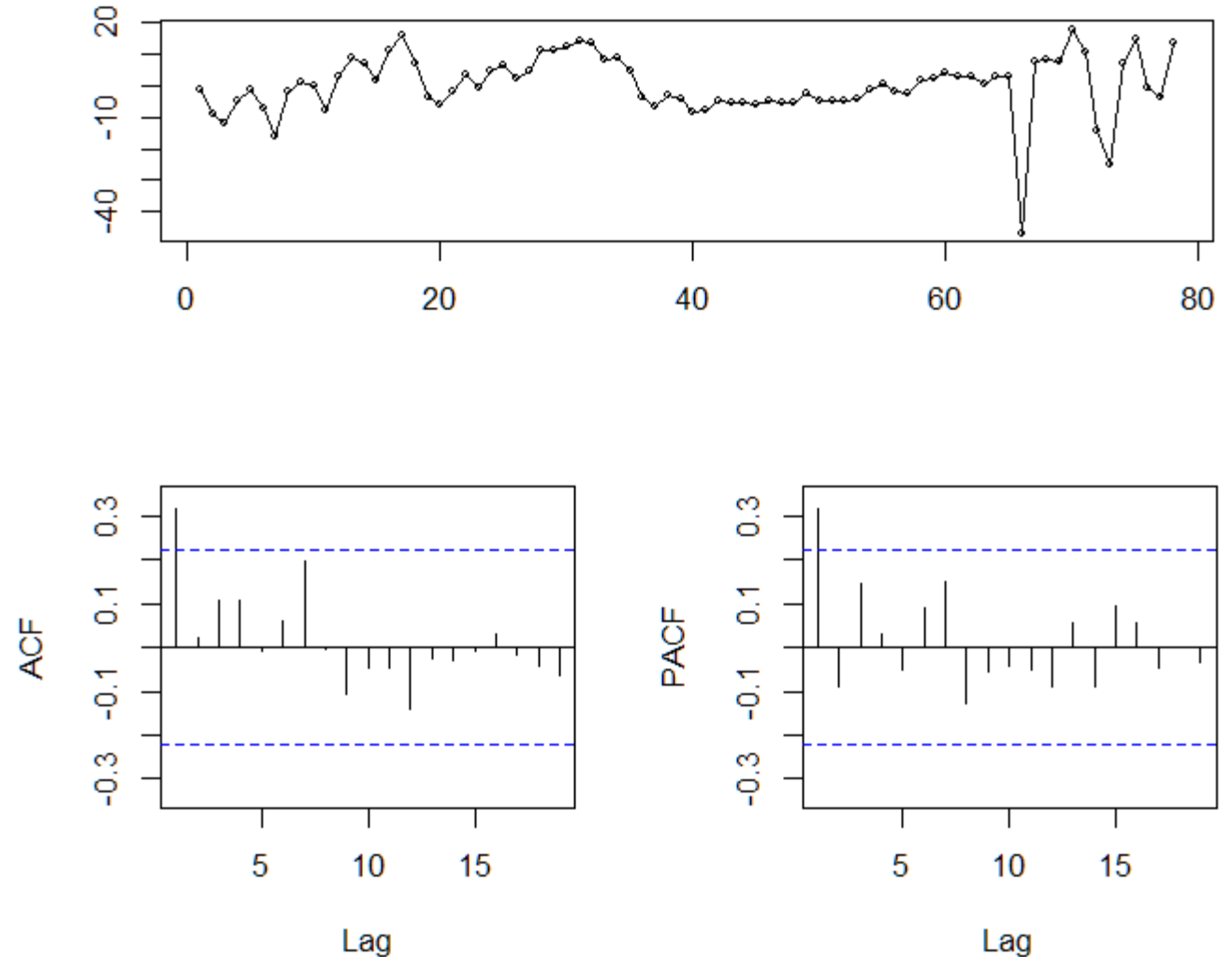
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	88.23299	2.94349	29.976	< 2e-16	***
TREND	0.35478	0.05299	6.695	4.01e-09	***
QUARTERQ2	30.26003	3.10193	9.755	8.33e-15	***
QUARTERQ3	46.95004	3.14107	14.947	< 2e-16	***
QUARTERQ4	13.83893	3.14224	4.404	3.62e-05	***
COVID_AVG_RESTRICTIONS	-47.76766	2.53122	-18.871	< 2e-16	***

AICc = 587
Adjusted R² = 0.89

Statistical modeling - Italy

- Hints for autocorrelation in the first lag.
- Durbin-Watson test reveals significant positive autocorrelation (0.317179)
- **Not white noise.**

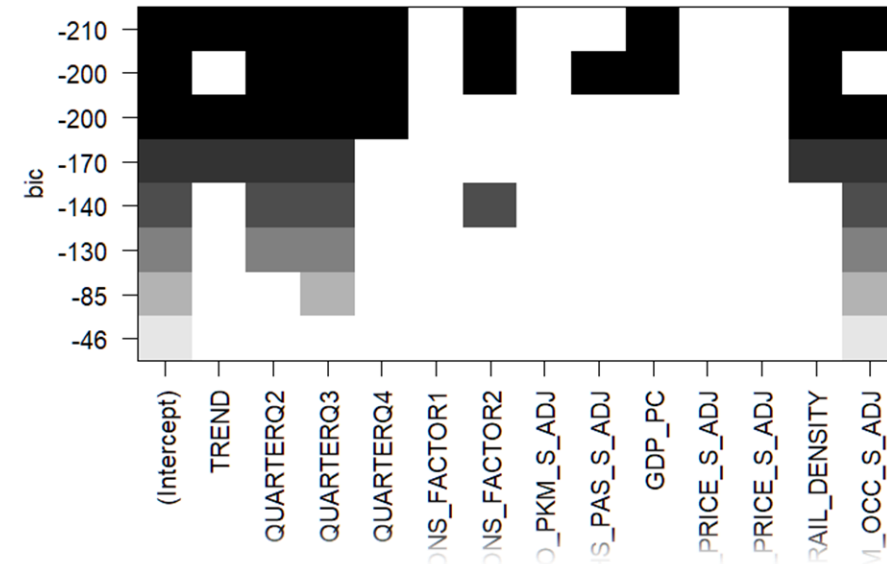
Italy - Baseline residuals and autocorrelation



Statistical modeling - Italy

- Again, the model fitted on **seasonally adjusted** parameters outperformed the standard model.
- In this case, the **best subset selection** with Adjusted R-squared/Cp/BIC criteria gave the **same output variables**.
- Best subset selection with seasonally adjusted parameters **doesn't give a model with strong collinearity issues**, opposite to Lasso and non-seasonally adjusted models.

BIC criterion for variable selection

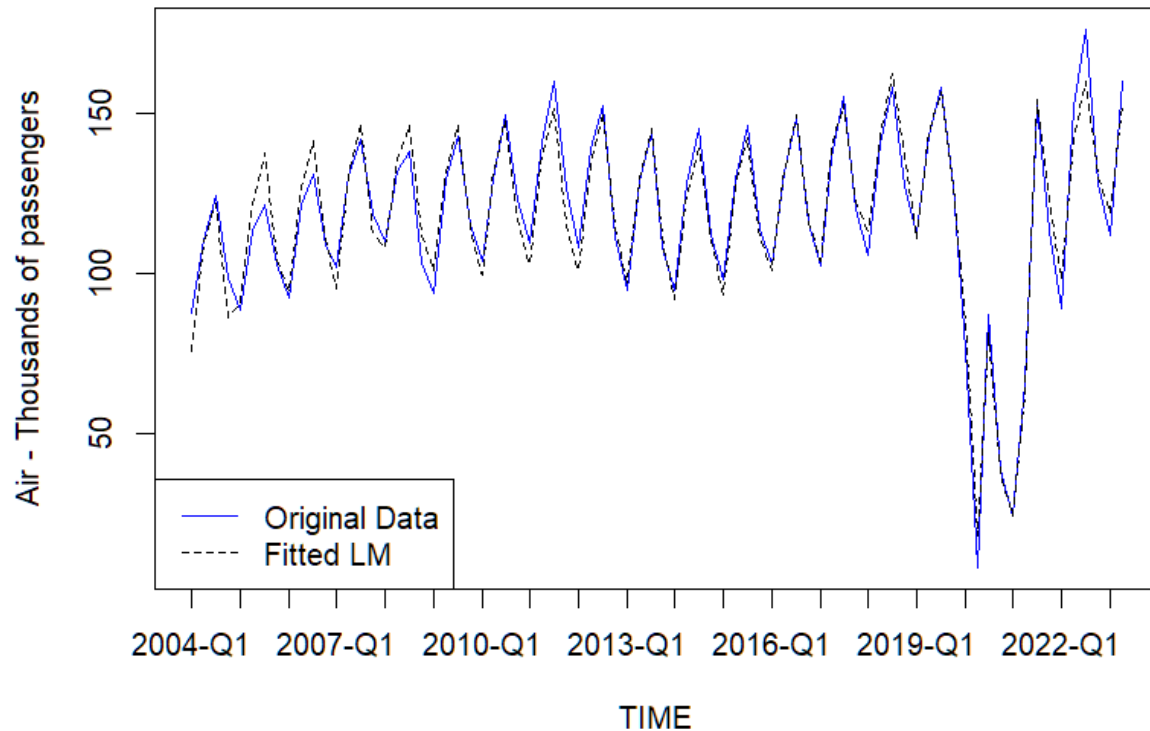


VIF for variables selected by Best Subset Selection

TREND	QUARTERQ2	QUARTERQ3
7.471362	1.495720	1.490548
QUARTERQ4	COVID_AVG_RESTRICTIONS_FACTOR2	GDP_PC
1.491282	7.460806	2.251836
RAIL_DENSITY	TOURISM_OCC_S_ADJ	
4.237445	7.988296	

Statistical modeling - Italy

Italy - Final linear model



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-914.9477	108.6292	-8.423	3.37e-12	***
TREND	-0.3657	0.0862	-4.243	6.75e-05	***
QUARTERQ2	32.2511	1.9888	16.217	< 2e-16	***
QUARTERQ3	47.2478	2.0196	23.395	< 2e-16	***
QUARTERQ4	14.8015	2.0201	7.327	3.38e-10	***
COVID_AVG_RESTRICTIONS_FACTOR2	-30.3433	7.9184	-3.832	0.000277	***
GDP_PC	13.0589	3.2675	3.997	0.000159	***
RAIL_DENSITY	15.2253	1.7853	8.528	2.16e-12	***
TOURISM_OCC_S_ADJ	2.0743	0.3153	6.579	7.65e-09	***

AICc = 521

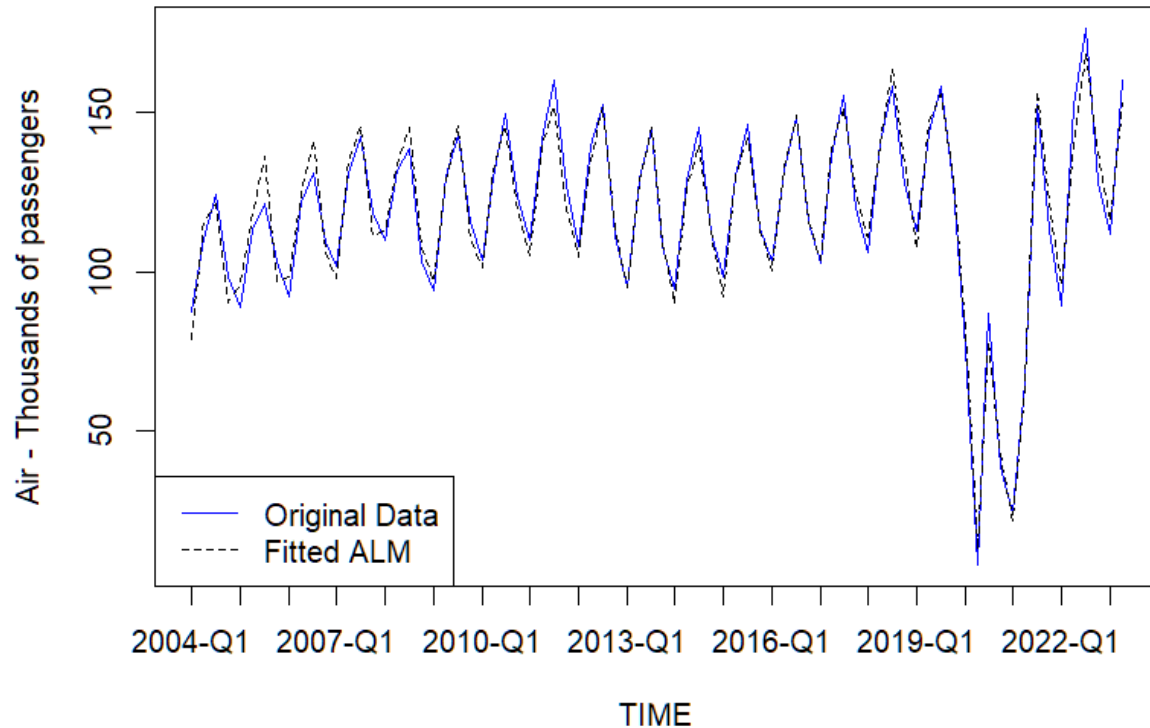
Adjusted R² = 0.95

Improvement in both Adjusted R² and AICc compared to the Baseline linear model.

All variables are **significant**.

Statistical modeling - Italy

Italy - Final model with ARIMA(0,0,1) errors



Coefficients:

	ma1	intercept	TREND	QUARTERQ2	QUARTERQ3	QUARTERQ4	COVID_AVG_RESTRICTIONS_FACTOR1
	0.5830	-842.0428	-0.3016	32.1497	47.2816	14.8166	-2.2340
s.e.	0.0898	127.3523	0.0931	1.4249	1.9104	1.4663	4.8105
	COVID_AVG_RESTRICTIONS_FACTOR2	GDP_PC	RAIL_DENSITY	TOURISM_OCC_S_ADJ			
	-31.5783	14.8646	13.6623	2.0518			
s.e.	7.4756	3.9893	2.1237	0.2898			

sigma^2 = 30.22: log likelihood = -237.89
AIC=499.78 AICc=504.58 BIC=528.06

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.04303326	5.09513	4.225892	-0.7438481	4.465491	0.1600268	0.01668633

AICc = 504

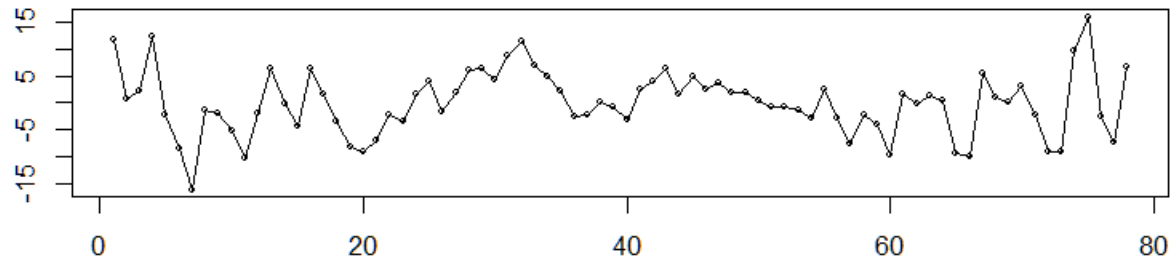
Improvement in AICc compared to the Baseline and previous linear models.

Presence of moving average component with one lag (**ma1**).

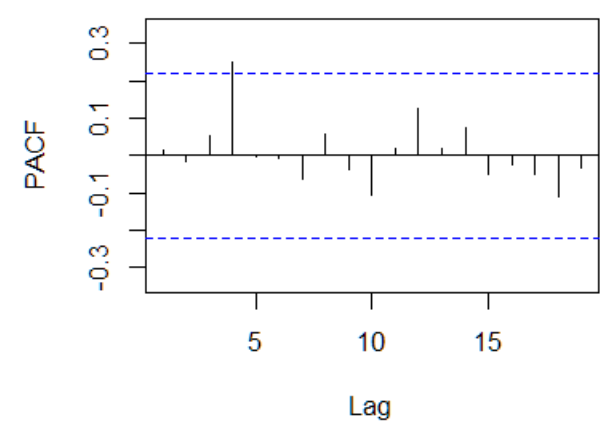
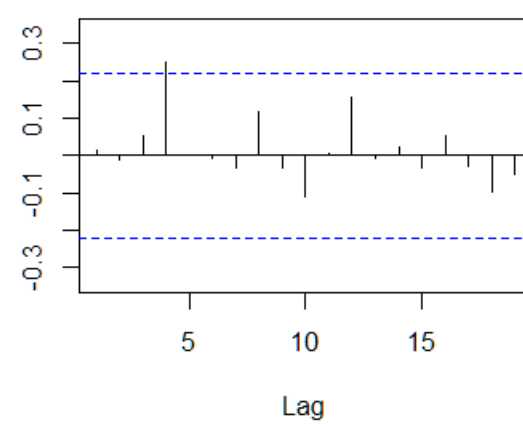
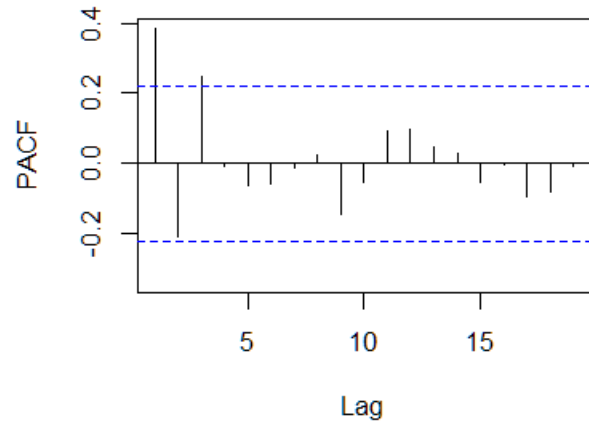
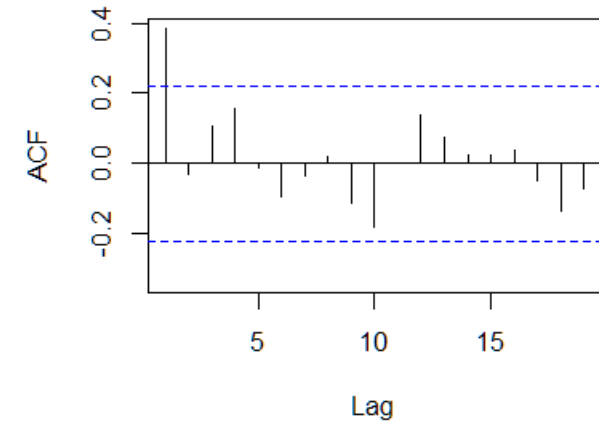
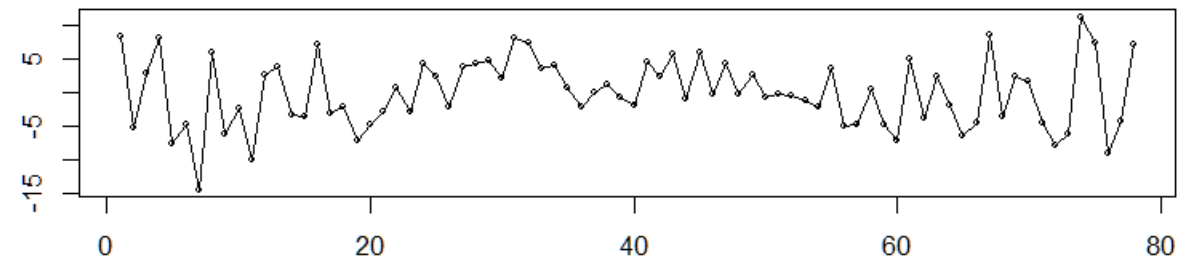
Good fit to the data.

Statistical modeling - Italy

Italy - Final linear model - residuals and autocorrelation



Italy - Final model with ARIMA(0,0,1) errors - residuals and autocorrelation



Durbin-Watson Test statistic: 1.91

Statistical modeling - Italy

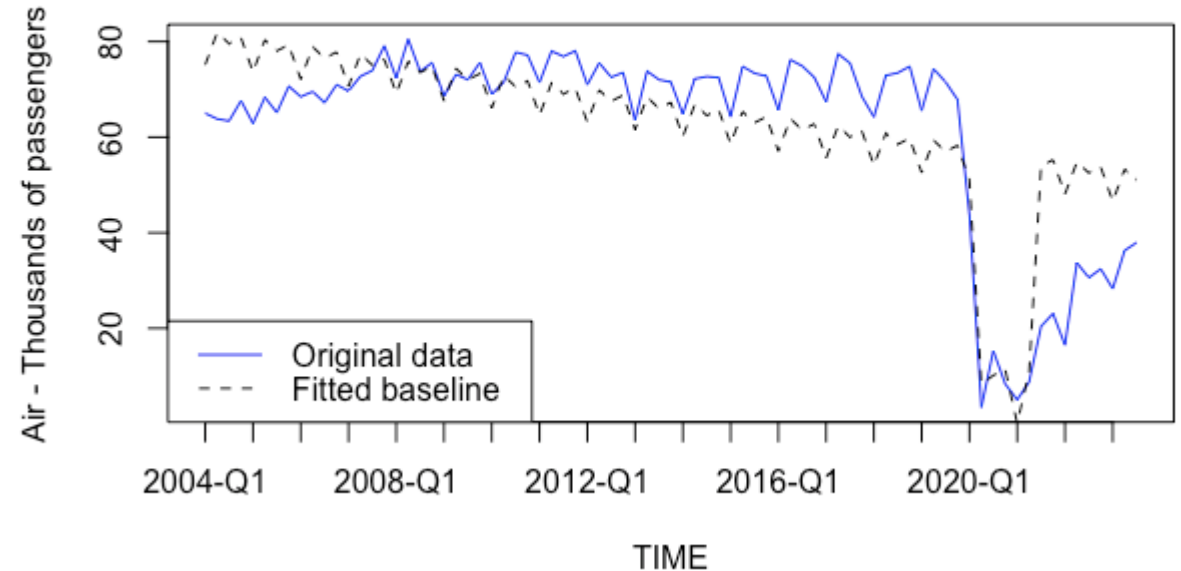
Insights:

- Variables **Trend** and **Rail Density** should be further inspected to understand their effects on the number of air passengers.
- **COVID restrictions** played a very clear role in the variations of air traffic, specially with mandatory measures.
- Naturally, there's also a strong impact of **Quarter 3** (Summer) in air traffic.
- Also the **GDP** variable has a very relevant impact, supporting the idea that the **economic activity** in the country plays an important role in domestic air traffic.
- **Air price** and **Rail price** are not considered into the best models, as well as other rail-related parameters, showing a certain independence between air and rail traffic. This aligns with the **lack of train discounts campaigns** in Italy.

Statistical modelling – Germany

- Simple linear model
predictors: trend, seasonality (quarter), COVID restrictions
- Shows different significance among the parameters
- Doesn't do a good job in fitting original data

Germany - Baseline



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	75.50162	3.64659	20.705	< 2e-16	***
TREND	-0.37593	0.06353	-5.917	1.02e-07	***
QUARTERQ2	7.12405	3.87203	1.840	0.0699	.
QUARTERQ3	5.19452	3.89830	1.333	0.1869	
QUARTERQ4	6.79517	3.94826	1.721	0.0895	.
COVID_AVG_RESTRICTIONS_FACTOR1	-45.36527	7.47373	-6.070	5.43e-08	***
COVID_AVG_RESTRICTIONS_FACTOR2	-49.43946	9.05257	-5.461	6.40e-07	***

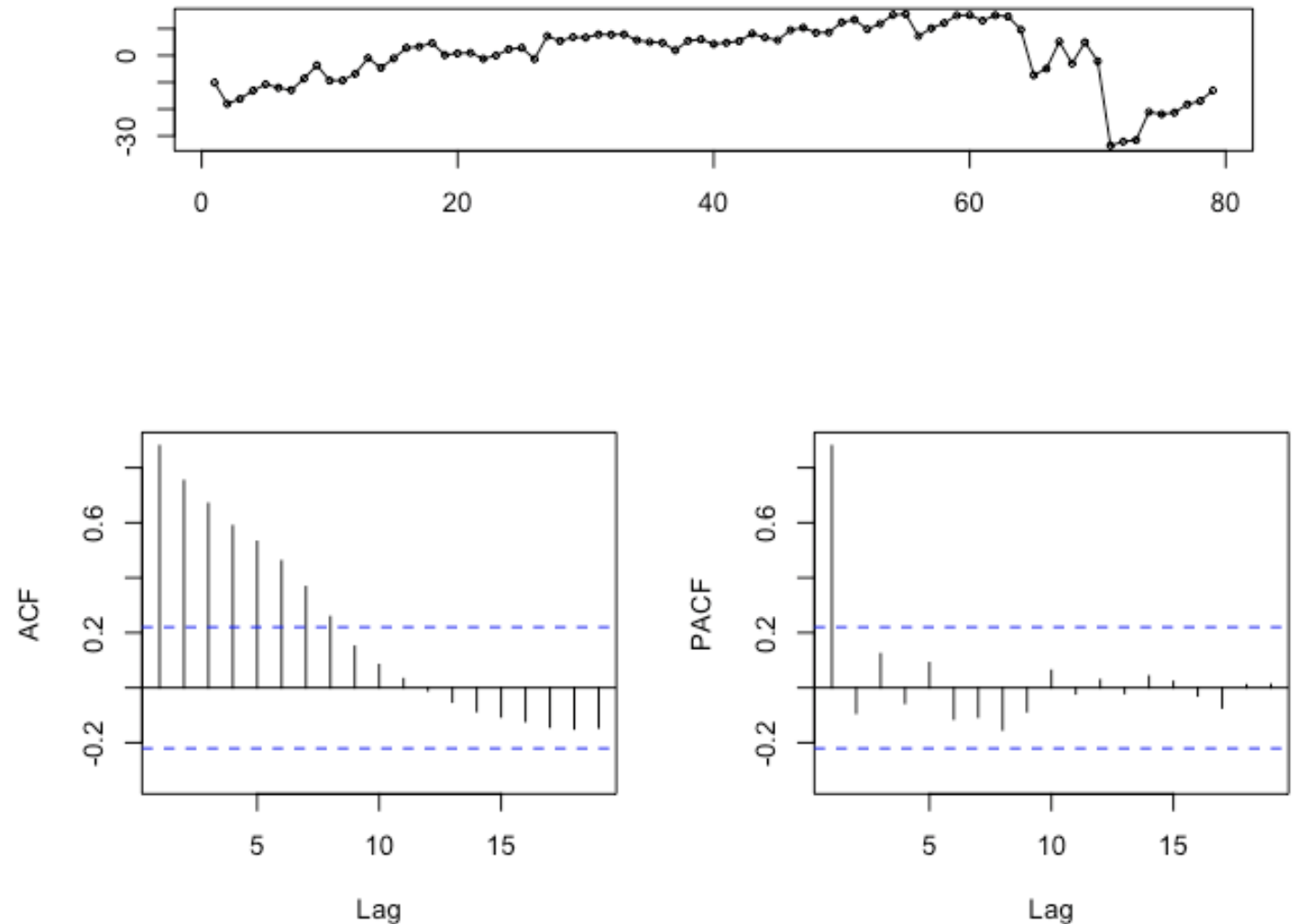
AICc = 630.0249

Adjusted R² = 0.6398

Statistical modelling – Germany

- Not white noise.
- Durbin-Watson test indicated a strong evidence of positive autocorrelation in the residuals.
- The model doesn't capture some trend or pattern.

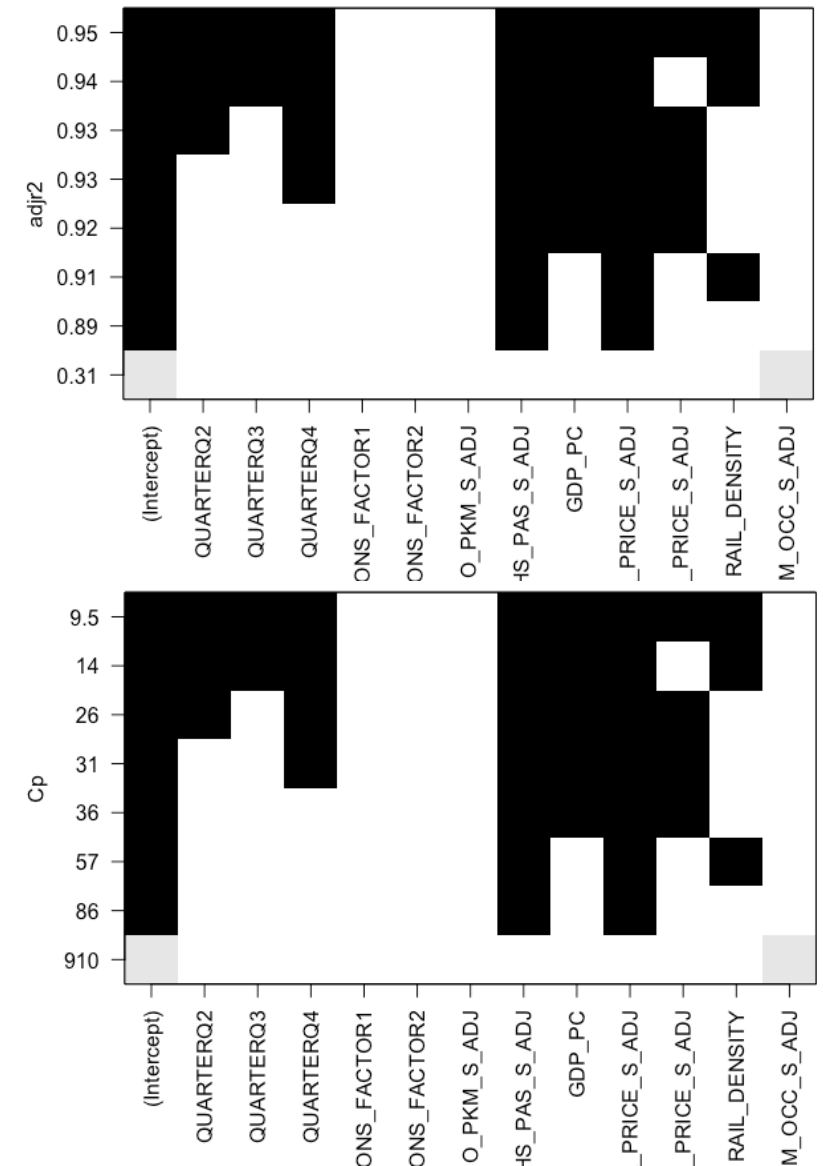
Germany - Baseline residuals and autocorrelation



Statistical modelling – Germany

- Season adjustment of the predictor improved the results of our models
- In order to choose predictors to our models we used **Lasso Regression** and **best subset selection**
- We got different results:
 - Lasso Regression: *COVID restrictions* is a significant variable
 - Best subset selection: *COVID restrictions* are **not** significant

(Intercept)	QUARTERQ2	QUARTERQ3
18.44629067	5.35120298	4.39043329
QUARTERQ4	COVID_AVG_RESTRICTIONS_FACTOR1	COVID_AVG_RESTRICTIONS_FACTOR2
6.26882216	-4.42910200	-1.07183948
R_MIO_PKM_S_ADJ	R_THS_PAS_S_ADJ	GDP_PC
0.00000000	0.01269306	-8.93349731
AIR_PRICE_S_ADJ	RAIL_PRICE_S_ADJ	RAIL_DENSITY
-0.82920904	0.40183777	0.62945452
TOURISM_OCC_S_ADJ		
0.00000000		



Statistical modelling – Germany

- After comparing the results we found not much difference (AICc: 487.3263, 489.6472; R^2 : 0.9419, 0.9425)
- We decided to keep the COVID restriction variable and look closely later
- However, collinearity problem was found

```
> vif(dePasBestSeasAdjVIF)
```

	GVIF	Df	GVIF ^{1/(2*Df)}
QUARTER	1.053874	3	1.008784
COVID_AVG_RESTRICTIONS_FACTOR	1.242625	2	1.055808
AIR_PRICE_S_ADJ	5.646416	1	2.376219
RAIL_PRICE_S_ADJ	2.821449	1	1.679717
GDP_PC	8.489331	1	2.913646

Statistical modelling – Germany

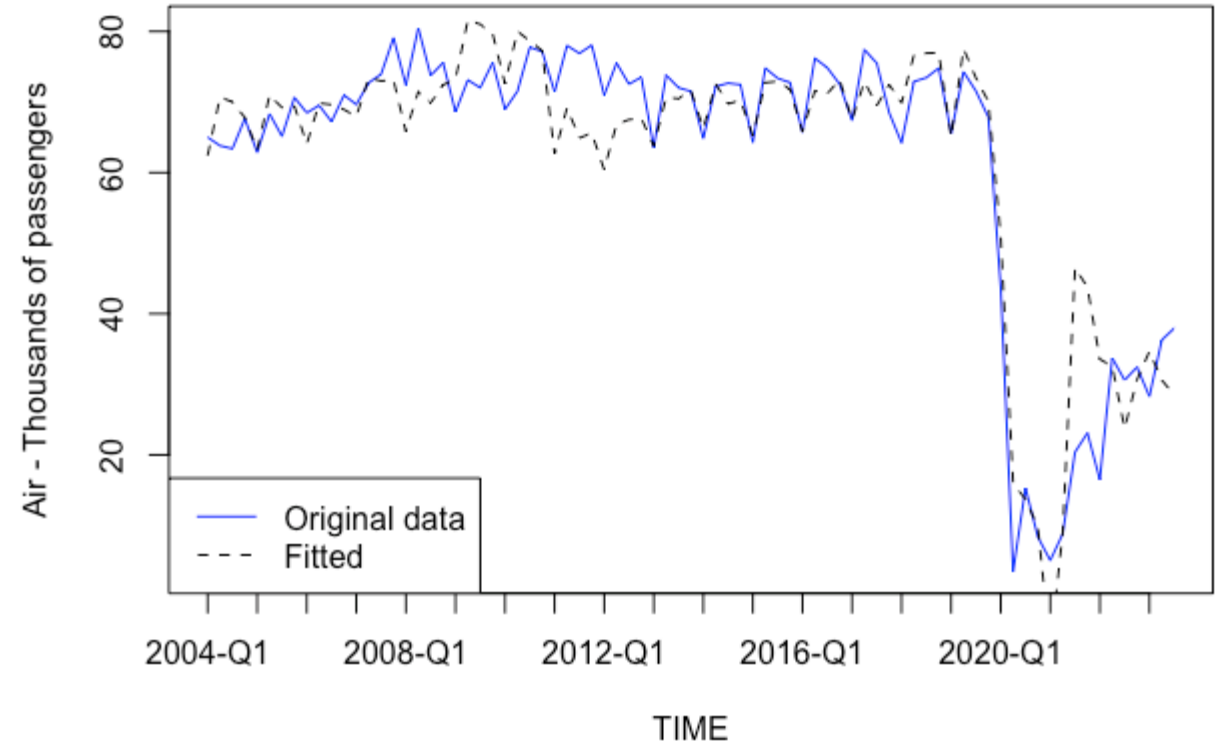
Best linear model

- Better fit than the baseline
- Higher Adjusted R^2
- All the coefficients are significant

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	82.2528	18.7003	4.398	3.81e-05	***
QUARTERQ2	6.7174	2.2662	2.964	0.00415	***
QUARTERQ3	5.4693	2.2777	2.401	0.01900	*
QUARTERQ4	5.7419	2.3071	2.489	0.01520	*
COVID_AVG_RESTRICTIONS_FACTOR1	-36.8018	4.4453	-8.279	5.60e-12	***
COVID_AVG_RESTRICTIONS_FACTOR2	-50.2180	5.3416	-9.401	4.88e-14	***
AIR_PRICE_S_ADJ	-0.6874	0.1083	-6.349	1.88e-08	***
RAIL_PRICE_S_ADJ	1.5185	0.1398	10.863	< 2e-16	***
GDP_PC	-10.6349	3.6496	-2.914	0.00479	**

Germany - Final Linear model



AICc = 547.685 Adjusted
 $R^2 = 0.8776$

Statistical modelling – Germany

- Trying to improve using the **dynamic regression with ARIMA errors**
- Significantly better results
- ARIMA model indicates 2 moving average terms

Regression with ARIMA(0,0,2) errors

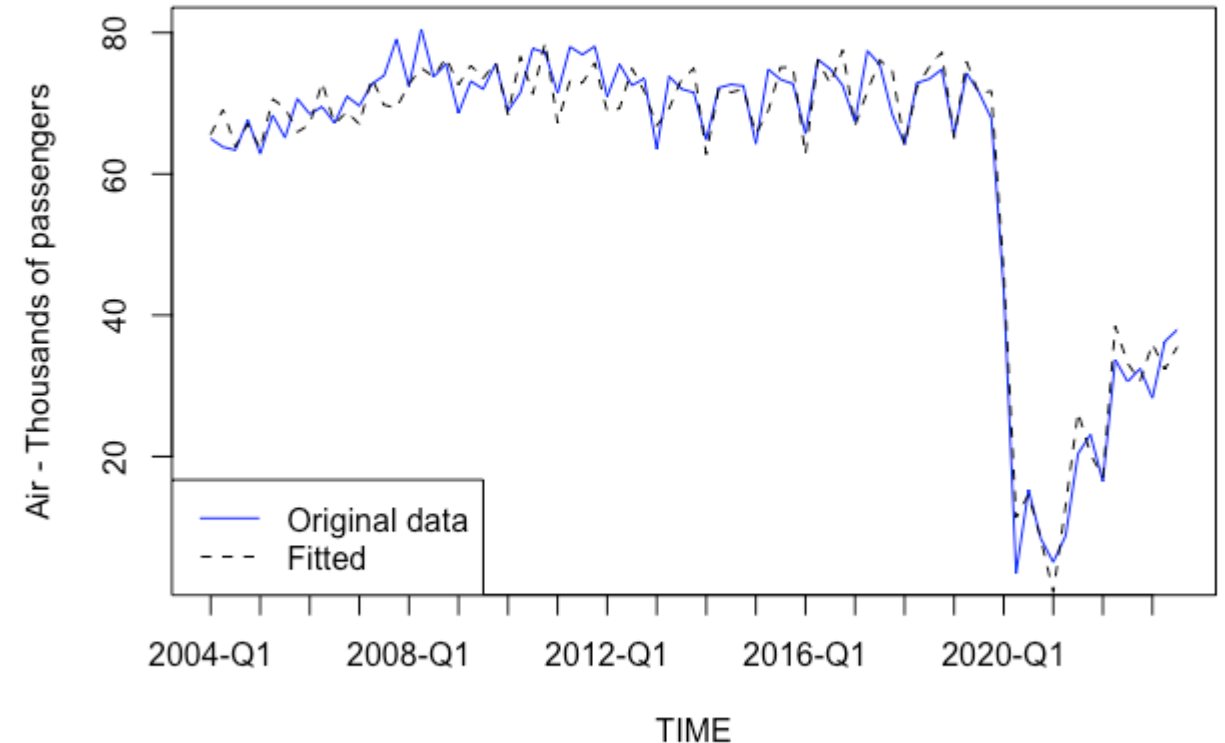
Coefficients:

	ma1	ma2	intercept	QUARTERQ2	QUARTERQ3	QUARTERQ4	
	0.8419	0.5667	91.2717	6.0826	5.2767	7.0453	
s.e.	0.1329	0.1779	43.5443	0.9789	1.0905	1.0363	
	COVID_AVG_RESTRICTIONS_FACTOR1		COVID_AVG_RESTRICTIONS_FACTOR2		R_THS		
			-5.6173		-7.9761		
s.e.			4.1083		4.3874		
	GDP_PC	AIR_PRICE_S_ADJ	RAIL_PRICE_S_ADJ	RAIL_DENSITY			
	-14.5304	-0.5848	0.8329	-0.0168			
s.e.	3.1695	0.1127	0.2020	0.5227			

sigma^2 = 14.34: log likelihood = -210.73

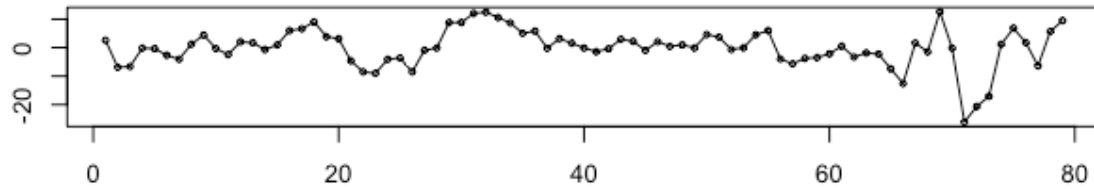
AIC=449.46 AICc=456.02 BIC=482.63

Germany - Final model with ARIMA errors

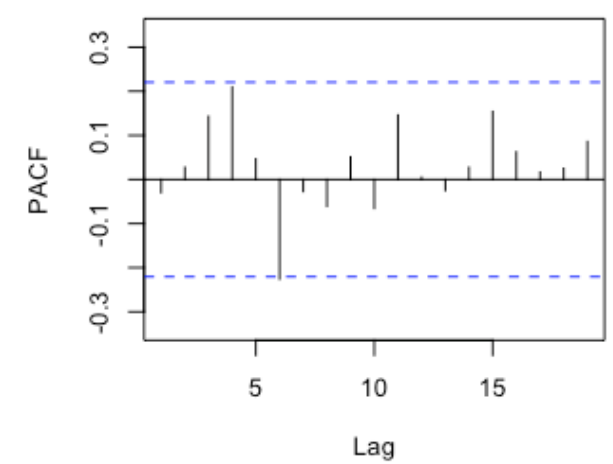
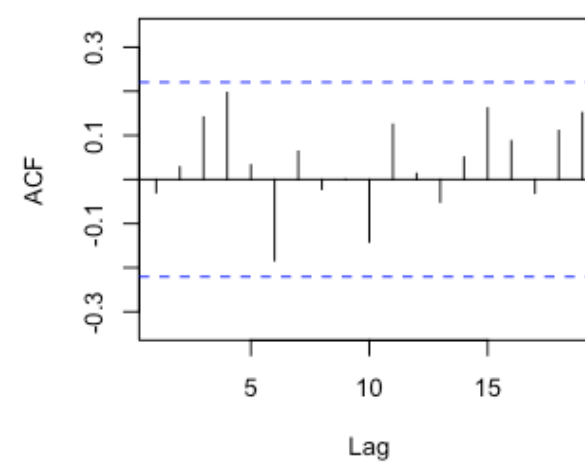
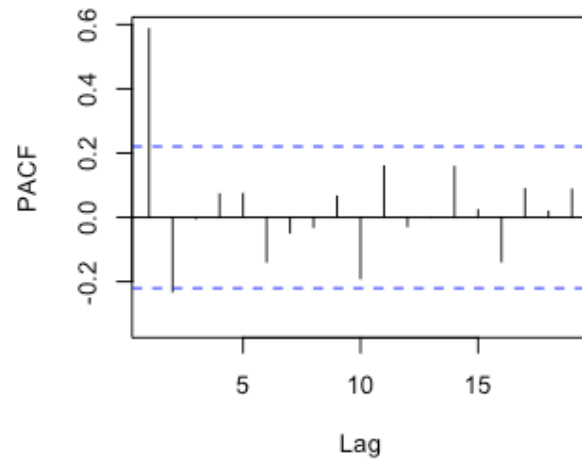
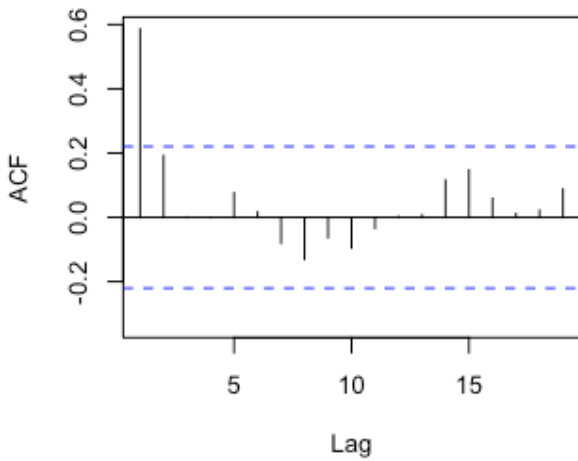
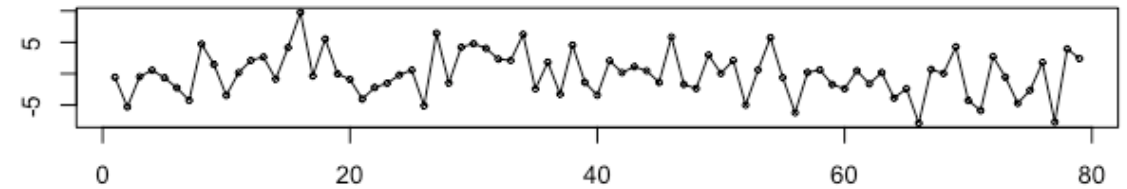


Statistical modelling – Germany

Germany - Final linear model - Residuals and autocorrelation



Germany - Final model with ARIMA errors - Residuals and autocorrelation



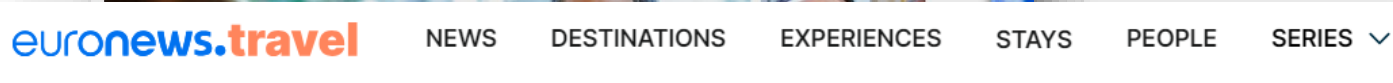
Statistical modelling – Germany

Insights:

- **GDP per capita** has the biggest impact on the air traffic in Germany, which aligns to our previous assumption: economic activity has an undeniable role in the changes of airplane activity of the country. However, GDP brought a negative effect to the air traffic in Germany.
- **COVID restrictions** brought a negative effect.
- Unlike in Spain and Italy, air traffic in Germany depend on the **prices, both, railway and airway**. Interestingly enough, rail price has a stronger impact than air price.
- There is a **seasonality** present in the data, however it is smaller comparing to Spain and Italy (that can be seen from the plots as well).

Discussion

- Policymakers are trying to react (not only in these countries), **but is it enough?**



What's behind Europe's rail revolution?

When **flight-shaming** took off in 2018, it aimed to make people feel change their travel habits by highlighting the huge emissions we get

Air travel cur expected to

But such a r realistically r away.

The end of budget flights? France pushes for minimum ticket prices in Europe

But many are wondering if these proposals are not just symbolic gestures. In particular, there is debate as to whether the suggested minimum prices are high enough to have a serious effect. Beaune has not fixed a figure for the minimum price and said that he "doesn't want to see ticket prices increase tenfold either."



Discussion

- For railway policies to make an impact on air traffic, its usage should be widely spread and deeply rooted. It **may not be easy** in all cases.
- When people are economically active, they tend to travel. It is essential to create **true alternatives** to airplanes if policymakers want to reduce their impact on the environment. Perhaps, **structural measures are needed**.
- Although there are elements in common, air traffic does not respond equally to the same factors in different countries. **There is no silver bullet**.



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
MATEMATICA



Thank you