UNIL | Université de Lausanne
HEC Lausanne

Project II

Study of weekly product sales

How can we best forecast the weekly product sales in a B2B context?

Forecasting II

Spring 2019

Group C

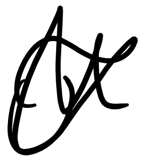
HEC Lausanne

Submitted on May 22, 2019

DECLARATION OF AUTHORSHIP

This project was written by us and in our own words, except for quotations from published and unpublished sources, which are clearly indicated and acknowledged as such. We are conscious that the incorporation of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism, subject to the custom and usage of the subject, according to the University Regulations. The source of any picture, map or other illustration is also indicated, as is the source, published or unpublished, of any material not resulting from our own research.

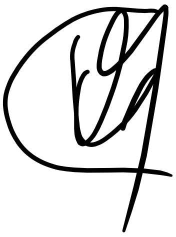
Rita Sefraoui Tahiri (10812097)



Charlotte Fleury(14608749)

Alexandre Schroeter (15400716)

Clément Perez (14406482)



SUMMARY

In the context of the forecasting class, we want to forecast the sales of 30 different anonymized products for the next 24 weeks. Our dataset contains observations from week 1 of 2016 to week 7 of 2018. We first perform an exploratory analysis to get a broader view at the total income level and the product group level. We notice a decreasing trend over the weeks with unequal variance at the overall level and no apparent seasonality.

We then focus on each specific category of products. Seasonality remains hard to spot but we can see slight trends of a few products.

After this exploratory analysis, we first divide our data into a training and a test set. The former contains the first 78 weeks of observations, while the latter has the last 33 weeks. This allows us to determine the predictive capacity of the different models we will use. We then test and use hierarchical methods, namely bottom-up, top-down (Gross-Sohl method F, Gross-Sohl method A, Forecasted Proportions) and the middle-out approach. For each hierarchical method, we test different models such as ARIMA, ETS and Random Walk. We then compute the ME, RMSE, MAE, MAPE, MPE and MASE for each method to compare the different models. As some scores may be conflicting, we decide to base our decision on the lowest RMSE.

With that in mind, we choose to select the random walk with a top down Gross-Sohl method A as it has the lowest RMSE . Using this method, we forecast a prediction for the different levels (total income, product groups and products) for 17 weeks, from week 7 to week 24 of the year 2018.

The results were surprising in different ways. First, because of the shape of forecast. It is linear and lower than the mean value at the total income level. Secondly, even if we notice a trend in the overall level, our prediction is flat which seems counter-intuitive. Thirdly, we did not expect to use a random walk model but rather a more advanced forecasting method.

When we focus on the group level, we see a similar pattern, with flat and below average predictions.

We provide a prediction interval at 80% and 95% for the overall level. It is cone shaped with the most probable values closer to our forecasted prediction.

When we check the residuals of our “best” model, we notice high autocorrelation of the residuals and thus rejection of the Ljung-Box test, which is the opposite of what we wish. The residuals could be stationary if it were not for the autocorrelation.

To conclude, even with our best model or the least bad model, our predictions remain somewhat flawed, with a poor predictive capacity. When we replace these results in a business context, little insight can be provided with whichever method used here. Unfortunately this does not help the managers to make sensible decisions.

TABLE OF CONTENT

INTRODUCTION

As part of the course Forecasting in the master of management of the University of Lausanne, we have learned the main methods to forecast chronological data using time series. In order to practice these newly learned methods, we were asked to perform an analysis on some given dataset.
We receive a dataset showing the incomes of 30 products in CHF. Those incomes are measured weekly and come from a larger database containing more products and categories. The dataset includes the incomes of each product for each week from week 1 of 2016 to week 7 of 2018. The products are identified by an anonymized identifier (SKU).
We are also given a “match table” that contains the various categories and products which were anonymized as well. This table allows us to group the different products according to their category thanks to their SKU identifier. With this dataset and “match table”, we aim at predicting the future values of product sales for the next 17 weeks using hierarchical methods.
The question we are trying to answer in this report is the following: “Can we produce meaningful forecasts of product incomes from week 8 of 2018 to week 24 of the same year using a hierarchical method?”
The first step is to match the dataset with the “match table” and to get a general overview of our dataset by exploring it in order to see if certain patterns can be observed. Then we create a hierarchical time series and fit different models on it using bottom up, top down, middle out and optimal reconciliation approaches. We forecast values for all models and evaluate their quality of prediction by computing accuracy measures and comparing them. We select the one with the lowest RMSE.
The first part of this report contains the data description as well as the methods used in order to conduct the analysis. The second part presents the results, beginning with an exploratory analysis followed by models fitting and forecasting and finally models evaluation and selection. The last part consists of a discussion of the results and conclusion.

DATA DESCRIPTION

The dataset (.RData) consists of weekly product incomes in CHF from week 1 of 2016 to week 7 of 2018. It is structured as follows:

- The first column refers to the year in which the income was made.
- The second column refers to the week in which the product was sold. The weeks are numbered from 1 to 52 for each year.
- The remaining 30 columns contain the actual incomes in CHF of the different products. Each column corresponds to a specific product whose name corresponds to an anonymized SKU identifier.

All of them are numerical values. This dataset is an actual subset of another larger one. The original larger dataset contained the weekly sale volumes of products from week 1 of 2016 to week 7 of 2018. To make the products comparable, an average price in CHF was computed for each product. This average price was then multiplied by the weekly sale volumes in order to get an income of each product for each week of the studied period. A unique subset of 30 products from this original transformed dataset was then given to us.
A “match table” containing all different SKU identifiers and their respective group was provided. This “match table” was modified in order to have the same number of characters for all the products’ group and name so that we could use the appropriate R method. This was then used to change the columns’

names of the dataset containing the product incomes. These modifications allowed us to create a hierarchical time series with the function hts().

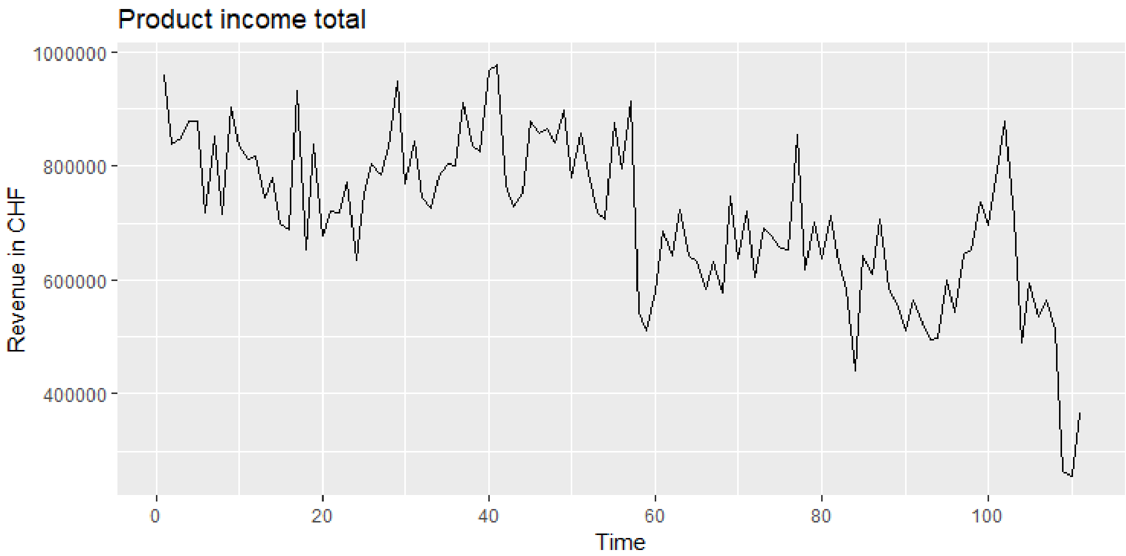
METHODS

We first construct a time series for each product with the function ts(). Then we create a hierarchical time series with the product time series using the function hts() from the package "hts".
We start by performing a short exploratory analysis of the time series using the function autoplot() for the total and category levels.
We create a training set and a testing set. We choose here to keep the first 78 weeks in the training set and that the testing set would be the data from the 79th week onwards. Indeed, as we are dealing with a time series, we cannot take samples at random. We choose to have a training set with two thirds of our data and to test the models on one third. It is not as efficient as a cross-validation method, but still will give us a base on which to compare the efficiency of our different models. We then fit and test different models on the training set in order to get the best one to forecast the wanted values.
As the aim of this project is to make forecasts using a hierarchical method, we decide to fit a bottom up approach using Arima, ETS and Random Walk methods. We produce forecasts for each methods at each level using the function *forecast(object, method = "bu", h = 33, fmethod = "")*. The bottom up approach consists of base forecasts that are created for the bottom level (here the products) and that are summed up to construct forecasts for the higher levels (here the groups and the total).
We also fit top down approaches "tdfp" (top-down forecast proportions), "tdgsa" (top-down Gross-Sohl method A), "tdgsf" (top-down Gross-Sohl method F) using the function *forecast(object, method = "approach", h = 33, fmethod = "")* and follow the same procedure as for the bottom up approach. The top down approaches construct base forecasts for the top levels (here the total). Those base forecasts are used to create forecasts for the lower levels by distribution.
We finally fit an optimal reconciliation and a middle-out approach: *forecast(object, h = 33, method = "comb" , weights = "wls", fmethod = "")* and *forecast(object, h = 33, method = "mo" , level = "", fmethod = "")*. The middle-out approach is a combination of bottom up and top down approaches. It consists of choosing a middle level for the base forecast and use a bottom up approach for higher levels and a top down approach for lower levels forecasts.
To evaluate the quality of each fitted model and their respective forecasts, we compare them to the test set and compute accuracy measures using the function *t(accuracy.gts(object_forecast, test_set, levels =""))*. The RMSE and MAE were computed but we decide to base our choice on the RMSE in order to minimize extreme errors.
Finally to compute and represent the forecasts, the forecast() method is used in conjunction with autoplot(). We first create the method with the forecast as follows:
prod.tdgsa.fct.rw <- forecast(object = prod.hts, h = 17, fmethod = "rw", method = "tdgsa")
And then use this method in argument of the function autoplot() and autolayer(), specifying each time the level forecasted. The level 0 is the total of the products, the level 1 is the group of the product and finally the level 2 is the individual product.

RESULTS

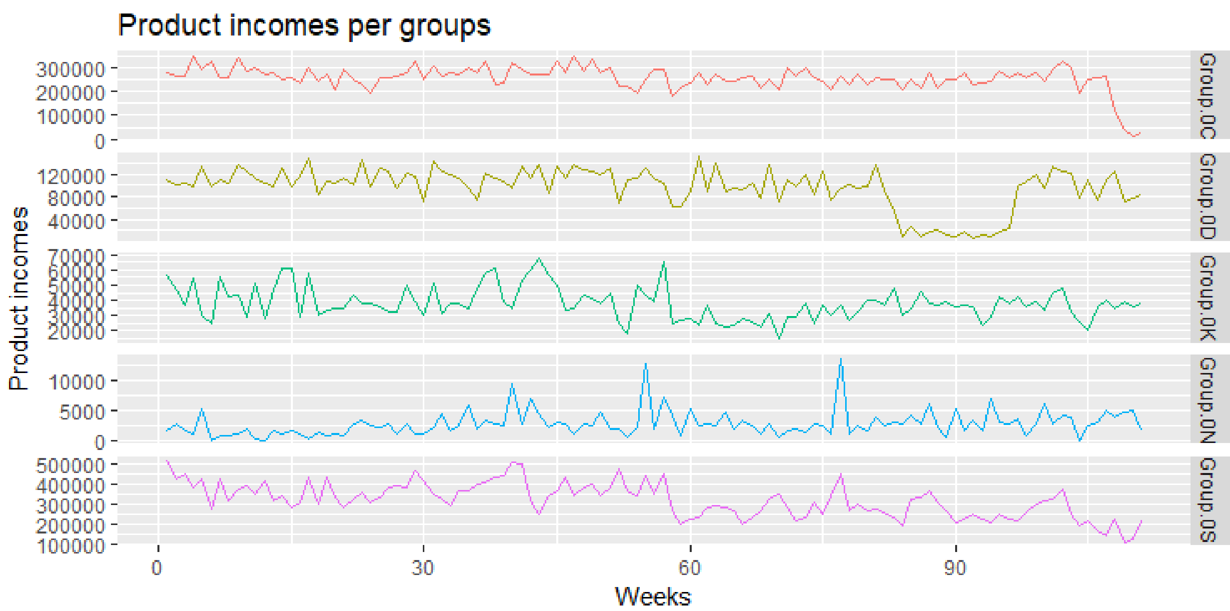
i) Exploratory analysis

Total level



To get a general overview of the time series, we use the function autoplot(). We can notice that the time series consisting of all products shows a clear decreasing trend with unstable variance and no clear visible seasonality. The variance also seems to increase, the peaks look smaller at the beginning of the series and bigger towards the end.

Category/group level



When we look at the different groups of products, we can observe different types of behaviours:

Group0C:

The series seems to have no particular trend until week 105 where we can observe a drop. The variance looks relatively constant. There is no clear seasonal pattern.

Group0D:

We can notice a relatively constant variance, no particular trend and no apparent seasonal pattern. There is however a drastic decrease between weeks 80 and 100.

Group0K:

There seems to be no trend, no seasonality nor constant variance. We can however notice that there is a change in variance from week 60 onwards and that there might be a drop in the income values.

Group0N:

There might be a very slight increase with a variance that is not constant. There seems to be no clear seasonal pattern.

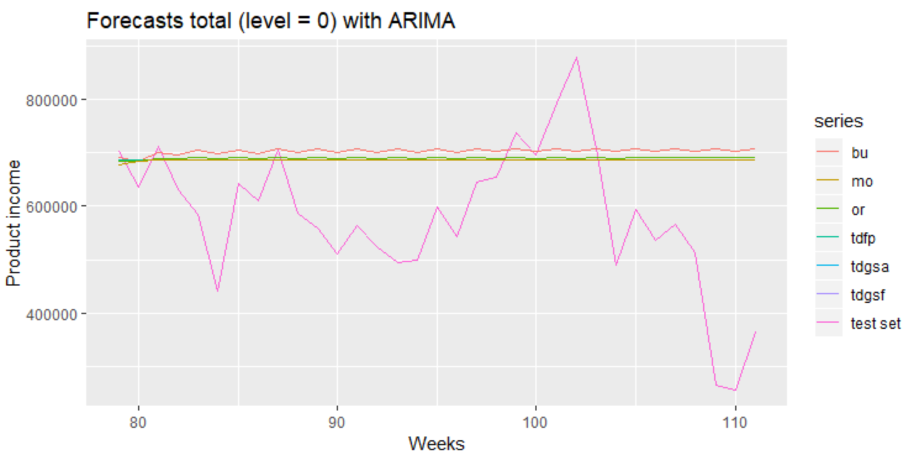
Group0S:

We can see a slight decreasing trend with unstable variability. No seasonal pattern seems to be prominent.

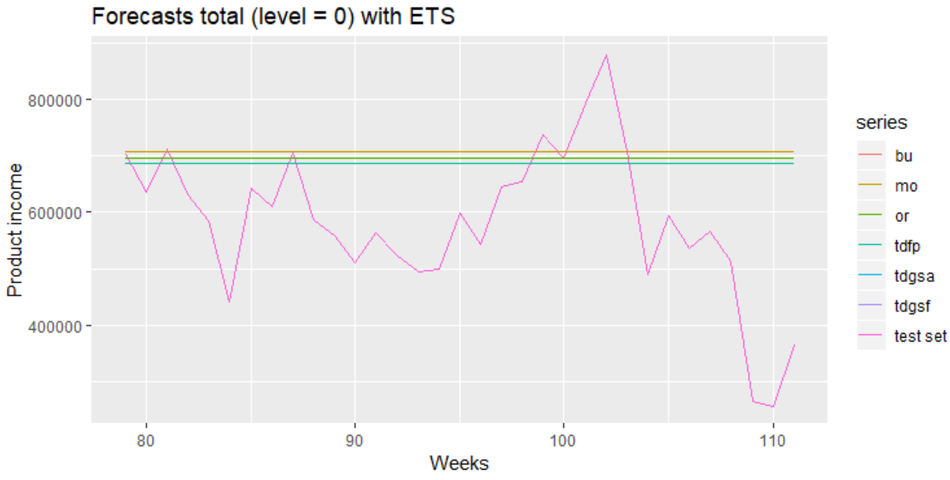
Overall some products have more impact than others because of the difference of magnitude in the sales the groups of products generate. Indeed, the group 0C shows revenues around 200 000 .- to 300 000.- until the week 105, whereas the group 0N represents revenues between 0.- and 5000.- . It would therefore make sense that some groups of products will influence more our forecasts.

ii) Forecasting methods

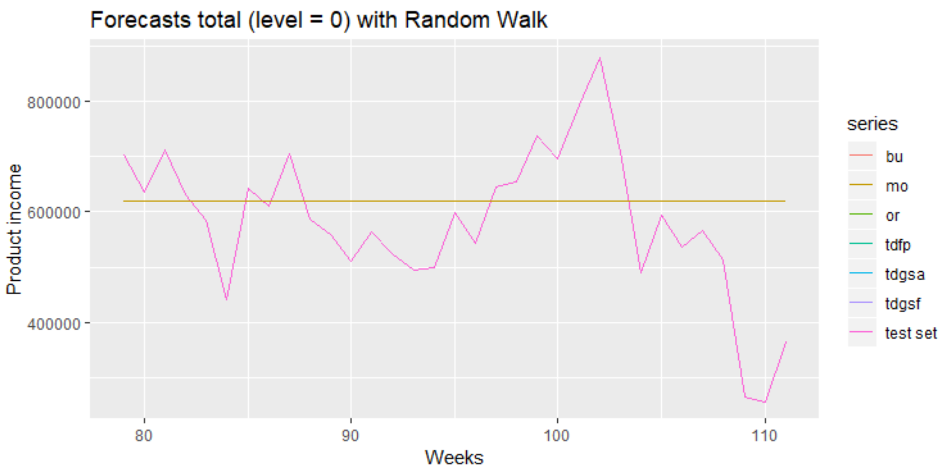
In order to choose the best model to forecast future values, we divide our time series into a training and a test set. We then fit models on the training set and forecast values for the test set. We plot the results of each method along with the real values of the test set into the same graph.



Looking at the graph of forecasts with an ARIMA method here we can notice that all predictions are very close to each other. They seem quite linear, although we can observe a slight pattern in the bottom up approach.



The graph of forecast with an ETS method shows again close linear predictions for all approaches. The results look relatively similar to the ARIMA method as the predicted product incomes are, in both cases, around 700’000 CHF.



The last graph shows that all approaches seem to forecast the same values with a random walk. This method differs from the two others in terms of predictions’ value. Indeed, the product incomes are closer to 600’000 CHF.

We create tables displaying ME, RMSE, MAE, MAPE, MPE and MASE of each method (ARIMA, ETS, Random Walk) and of each approach (bottom up, top down, etc) (Figure 1: Tables of the ME, RMSE, MAE, MAPE, MPE and MASE of each approach with the ARIMA method, ETS method and Random walk method on the total level). For ARIMA and ETS, we find the RMSE to be lower for the top-down approaches and higher for the optimal reconciliation, middle-out approach (for ETS) and bottom-up approach.
We see that the lowest RMSE of all is the one for the random walk method, and that it is the same for all approaches. Taking a look at the MAE, the results are here consistent with the RMSE, with the lowest MAE being for the random walk method. Moreover, the MAE of the top-down approaches of the ARIMA methods and ETS methods are lower than that of the other approaches.
This first result indicates that random walk should be the best, and then it would be ETS and the worst would be ARIMA. We need to see the results for the group levels and product levels before selecting a definitive way to forecast our products sales.

Category/group level

	RMSE arima 1		RMSE ets 1		RMSE rw 1
bu	46893.69	bu	46526.37	bu	41546.40
tdgsa	45425.85	tdgsa	45374.38	tdgsa	41066.81
tdgp	45603.83	tdgp	45552.51	tdgp	41237.04
tdgsf	45554.69	tdgsf	46334.24	tdgsf	41546.40
or	45714.78	or	46830.53	or	41546.40
mo	46893.69	mo	48195.65	mo	41546.40

For the group level, we look at the sum of the RMSE throughout the groups for each approach as presented in the table above. The random walk method gives the lowest total RMSE. The lowest of all is the one from the top down Gross-Sohl method A (tdgsa) which is coherent with our results of the total level where the random walk method was selected. The MAE also shows that the random walk method gives the lowest scores (Figure 2: Tables of the MAE of each approach with the ARIMA method, ETS method and Random walk method on the category level). The last step of our forecast selection is to see if our previous results are coherent with the product level RMSE on the test set.

Product level

	RMSE arima 2		RMSE ets 2		RMSE rw 2
bu	9855.960	bu	10257.038	bu	9494.735
tdgsa	9892.605	tdgsa	9886.080	tdgsa	9359.779
tdgp	9911.728	tdgp	9905.206	tdgp	9377.904
tdgsf	9704.701	tdgsf	10231.153	tdgsf	9494.735
or	9733.073	or	10284.080	or	9494.735
mo	9701.328	mo	10471.085	mo	9494.735

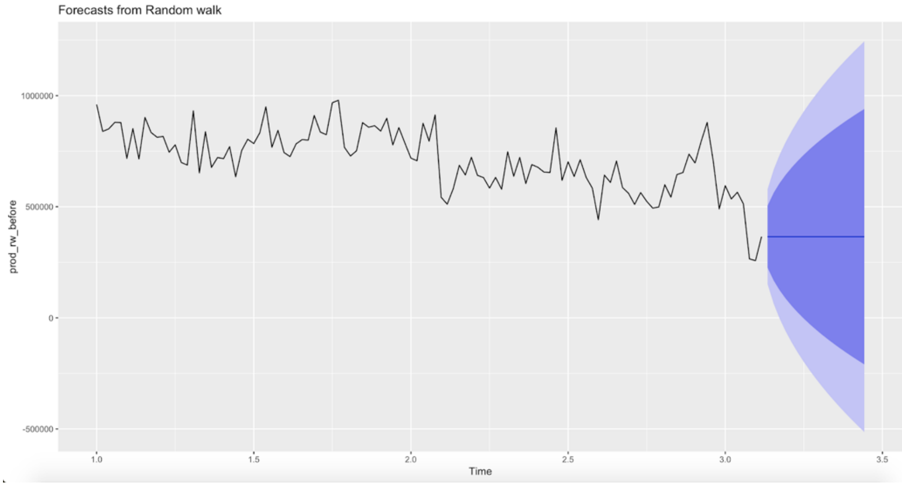
Here at the product level, we see the best method is still random-walk, with the top down Gross-Sohl method a being the best with the lower RMSE. The MAE confirms that random walk is still the best method (Figure 3: Tables of the MAE of each approach with the ARIMA method, ETS method and Random walk method on the products level).

Given all the concurrent results, we choose the random walk method with a top down Gross-Sohl method A to forecast our data.

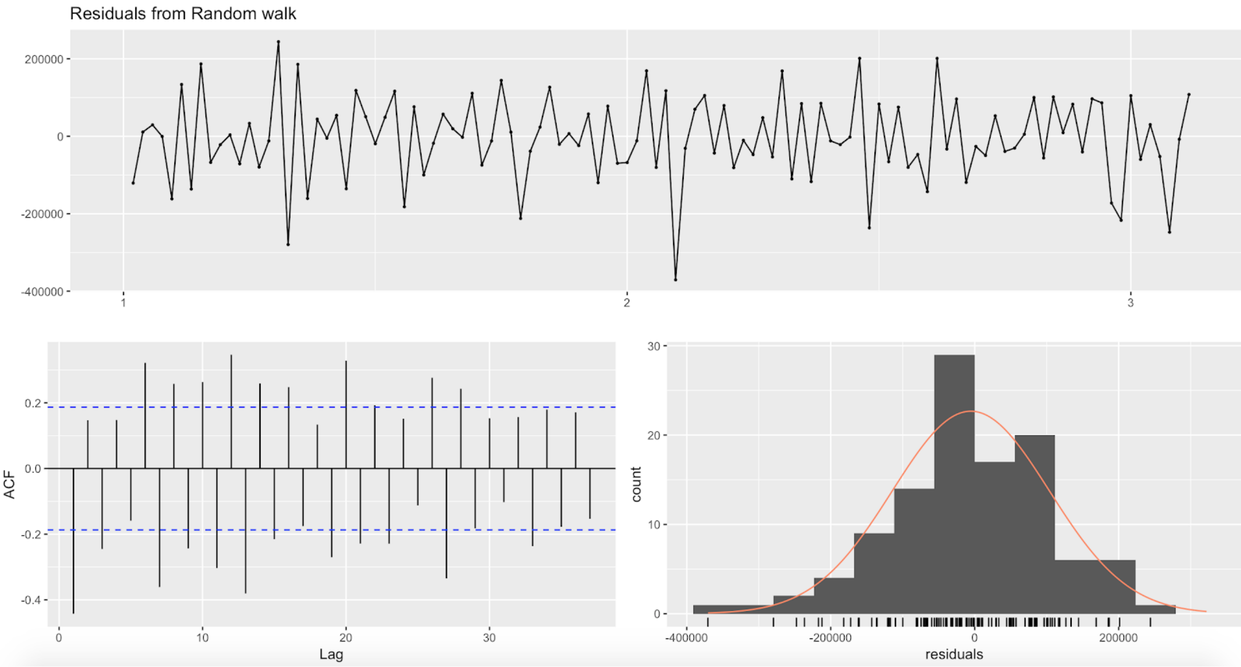
iii) Forecasted values

We represent below the forecasts for 17 weeks, up to week 24 of the year 2018. We see all the forecasts are linear, horizontal, no matter what the level is. It seems that it is an improbable forecast, even though it minimizes the RMSE and MAE. Particularly, all the forecasts seem lower than the average level of the data forecasted. This is explained by the decreasing trend of the group products 0S and 0C which decrease overall the total.

Total level

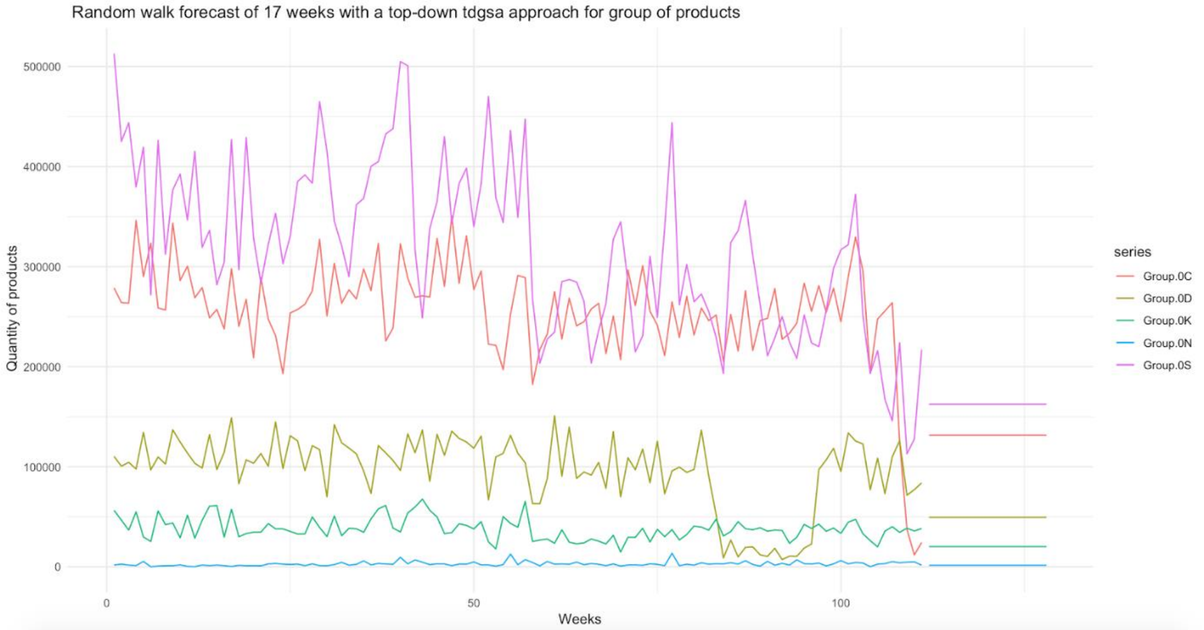


We see the 95% and 80% intervals of prediction for the total forecast. As expected, it is cone shaped, with the most probable values closer to our last data and the farther we move away from our last measurement the larger the intervals.



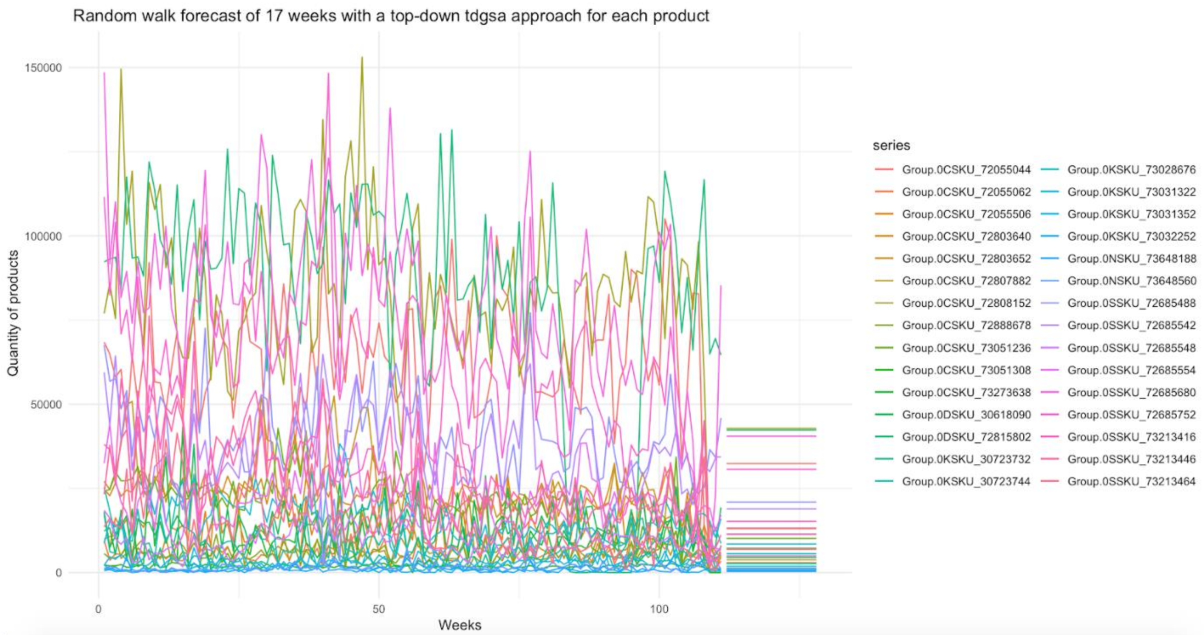
To evaluate the overall method, we check the residuals with the random walk forecast chosen on the total product level. It seems like there is still a very high autocorrelation and residuals that are not normally distributed. Moreover, the Ljung-Box test is rejected, confirming our high autocorrelation. It seems that even though it is the best method compared to the others on the basis of the RMSE and MAE on the test set of our data, it is not a satisfactory method to compute the forecast. This means that these values are to be used with a lot of precaution, and that we would use this method only to forecast very short-term value, or only for the products that already have a more horizontal trend in their sales.

Category/group level



We see that at the group level too, the forecast is a horizontal line for each group.
The random walk method tends to converge towards a lower bound value for each of these groups in this case. Indeed, all forecasts look lower than the original values. Moreover, no trend is shown, and for example the group 0S which has a decreasing trend presents a forecast that is not decreasing, which shows this method does not give the most realistic values to expect at the group level either.

Product level



Finally, at the product level, the same behaviour is observed. All the forecasts are represented as horizontal lines, and the random walk method gives a quite unrealistic forecast. If we observe particularly the group.0KSKU_30723744 product (the teal line in the upper part of the figure), we see that the forecasted value looks much smaller than any value it already had. It is coherent with what we found before that all the forecasted lines seem lower than the level of their original values.

Because we used a top-down method, and the top-level (total products) is decreasing, this decrease is first computed at the top-level and then distributed to the other sublevels (group and product levels). This is manifested through all the forecasts getting slightly lower than their historical values to be coherent with the top level that must be lower than its historical values too to get the best results according to our selection criteria (the RMSE).

DISCUSSION/CONCLUSION

For this study, we use a training set and a test set on which we apply some forecasting methods we know for hierarchical methods. Our aim is to select the best model with the data at our disposal. We look at the lowest RMSE to minimize the extreme errors at each level of the forecast.

The random walk method with a "top-down Gross-Sohl method A" approach is the best for the total level, the group level and the product level according to our selection criteria. Thus, it is the one we select.

The model used predicts a weekly product sale from week 7 to 24 of the year 2018 lower than the average sales from the past year.

As we can see, the forecast is horizontal and quite low, reflecting the overall decrease in the top-level of the forecast (total). This decrease is shown after in every sublevel as their forecasted lines are always lower than their historical values.

When analysing the best method here, we perform a residuals' analysis and we see that their ACF graph shows a high autocorrelation. This indicates that our model does not explain very well the behaviour of our variables. Indeed, there seems to still be seasonality, cycles, or another form of autocorrelation that we could not comprehend, and that the method does not cater for. This confirms that our forecast is not optimal, even if it is the best we found with our methods. The absence of trend in the top-level forecast is also indicative of a defective method.

Pragmatically, it is hardly usable as it is for business purposes, but it is still an informative parameter as we see this overall decrease and expectation for lower revenues in the future.

We might have found better forecasts with other methods or selection criterion. Indeed, the RMSE might not have been the best choice, as it minimizes the extreme values, it might have forced our forecasts into a more linear or smooth line to avoid bigger square errors. Choosing a less penalizing criterion, like the MAE, might have let us keep some patterns and seasonality in our forecasts.

The fact that we have chosen a training set of the first two thirds of time series, and a testing set of the last third might also be a reason why we have a non-optimal behaviour. Indeed, a cross-validation method adapted to the time series format would have also taken into account the later measures of our time series, and therefore the part where the trend is steeper. It could have given us a forecast with a better trend for example.

APPENDIX

Figure 1: Tables of the ME, RMSE, MAE, MAPE, MPE and MASE of each approach with the ARIMA method, ETS method and Random walk method on the total level

ARIMA:

	ME	RMSE	MAE	MAPE	MPE	MASE
bu	-119298.0	177418.6	138414.9	31.4073	-29.0925	1.6293
tdgsa	-104373.3	167390.5	130568.5	29.5868	-26.3186	1.5369
tdgp	-104373.3	167390.5	130568.5	29.5868	-26.3186	1.5369
tdgsf	-104373.3	167390.5	130568.5	29.5868	-26.3186	1.5369
or	-106618.0	168966.9	131868.3	29.8862	-26.7437	1.5522
mo	-103573.8	167193.3	130643.1	29.5808	-26.1893	1.5378

ETS:

	ME	RMSE	MAE	MAPE	MPE	MASE
bu	-113419.4	173175.5	135234.5	30.6548	-27.9808	1.5918
tdgsa	-103791.2	167028.1	130268.6	29.5181	-26.2116	1.5334
tdgp	-103791.2	167028.1	130268.6	29.5181	-26.2116	1.5334
tdgsf	-103791.2	167028.1	130268.6	29.5181	-26.2116	1.5334
or	-111905.0	172187.4	134448.4	30.4753	-27.7025	1.5826
mo	-123890.9	180207.5	141649.3	32.0362	-29.9049	1.6673

Random walk:

	ME	RMSE	MAE	MAPE	MPE	MASE
bu	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021
tdgsa	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021
tdgp	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021
tdgsf	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021
or	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021
mo	-35563.15	135611.7	102124.3	22.556	-13.6748	1.2021

Figure 2: Tables of the MAE of each approach with the ARIMA method, ETS method and Random walk method on the category level

```
> knitr::kable(MAE.arima1, digits = 4)
```

	MAE arima 1
bu	34205.42
tdgsa	33940.61
tdgp	34252.37
tdgsf	33598.09
or	33599.43
mo	33608.00

```
> knitr::kable(MAE.ets1, digits = 4)
```

	MAE ets 1
bu	34333.70
tdgsa	33918.03
tdgp	34227.48
tdgsf	34791.51
or	34736.09
mo	35731.54

```
> knitr::kable(MAE.rw1, digits = 4)
```

	MAE rw 1
bu	31625.62
tdgsa	32903.56
tdgp	33210.41
tdgsf	31625.62
or	31625.62
mo	31625.62

Figure 3: Tables of the MAE of each approach with the ARIMA method, ETS method and Random walk method on the products level

```
> knitr::kable(MAE.arima2, digits = 4)


|       | MAE arima 2|
|:------|-----------:|
|bu     |    7537.509|
|tdgsa  |    7722.982|
|tdgp   |    7753.258|
|tdgsf  |    7413.613|
|or     |    7426.846|
|mo     |    7407.004|
> knitr::kable(MAE.ets2, digits = 4)


|       | MAE ets 2|
|:------|---------:|
|bu     |  7987.104|
|tdgsa  |  7717.892|
|tdgp   |  7748.232|
|tdgsf  |  7992.524|
|or     |  8021.575|
|mo     |  8193.096|
> knitr::kable(MAE.rw2, digits = 4)


|       | MAE rw 2|
|:------|--------:|
|bu     | 7355.592|
|tdgsa  | 7370.217|
|tdgp   | 7404.507|
|tdgsf  | 7355.592|
|or     | 7355.592|
|mo     | 7355.592|
```