

Team Identification Number

16926

M3 Challenge 2023

Executive Summary

Title

Exec Summary

Keywords: Keywords, More Keywords

Contents

1	Q1: The Road Ahead	3
1.1	Defining the Problem	3
1.2	Assumptions	3
1.3	Variables	3
1.4	Models	4
1.5	Results	5
1.6	Model Revision	5
1.7	Evaluation	5
1.8	Technical Computing	5
2	Q2: Shifting Gears	6
2.1	Defining the Problem	6
2.2	Assumptions	6
2.3	Variables	6
2.4	The Model	6
2.5	Results	6
2.6	Model Revision	6
2.7	Discussion	6
2.8	Sensitivity Analysis	6
2.9	Technical Computing	6
3	Q3: Off the Chain	7
3.1	Defining the Problem	7
3.2	Assumptions	7
3.3	Variables	7
3.4	The Model	7
3.5	Results	7
3.6	Model Revision	7
3.7	Discussion	7
3.8	Sensitivity Analysis	7
3.9	Technical Computing	7

4	References	8
4.1	Bibliography	8
4.2	Program Code	8

1 Q1: The Road Ahead

1.1 Defining the Problem

In Problem 1, we were tasked with producing a short-term predictive model for e-bike sales. More specifically, we were asked to develop projections for total sales volume 2 and 5 years into the future respectively.

This is a time-series forecast problem, where we predict a single variable - e-bike sales - based on a single input - time. We will take a general approach of regressing particular equations against the existing data. Multiple types of equations will be used and the most promising model will be chosen.

1.2 Assumptions

Assumption 1: There will be no major legislative changes, governmental campaigns and/or 'black swan' (i.e., highly unpredictable and consequential) world events that significantly impact the market for e-bikes within the next five years.

Justification: in practice, it is impossible to account for rare or extreme events within the constraints of a mathematical model; the implications of such events cannot be predicted with accuracy.

Assumption 2: The market for e-bikes in the European Union behaves comparably to that of the United Kingdom; therefore, British and European sales can be considered to be in direct linear proportion.

Justifications:

- a) Of the data provided for European sales, several figures appear to include sales made in the UK (CITE EBICYCLES.COM). Therefore, UK consumer behaviour is partially accounted for even within the larger dataset.
- b) E-bicycles have only begun gaining traction as a mode of transport in relatively recent years; as a result, UK-specific consumption data is largely unavailable to the public.
- c) To a large extent, the UK and EU follow similar urban planning practices that include pedestrian walkability and bicycle access. In other terms, city layouts support the practical use of e-bikes. For this reason, population-scaled EU predictions can be considered appropriate substitutes for UK-specific predictions. By contrast, most American cities use car-centric design, frequently involving longer commute distances and poor bike access. This renders the United States hostile to the adoption of e-bikes in a way that the EU and UK are not. For this reason, we chose to exclude the US from our analysis, instead focusing on the UK and EU.

1.3 Variables

See table 1.1:

Variable	Definition
y_i	(actual) e-bike sales in year i
\hat{y}_i	predicted e-bike sales in year i
z	description

Table 1.1: Variables in the Model

1.4 Models

Linear

A linear regression is performed first due to its simplicity and ability to help pick more complex models.

Linear regression is an approach used to model a linear relationship between an independent variable x and a dependent variable y by finding the slope of the trend and initial value (y when x is 0). It is used to represent existing data and predict future values; linear models are used both for interpolation and extrapolation. In this case, the model will be fit to existing data and used to predict future sales of e-bikes. A linear growth function takes the general form of (1.1):

$$f(x) = \alpha x + \beta \quad (1.1)$$

where α is the coefficient, or growth rate; and β is the y-intercept, or initial value. The values of α and β are “optimized” using an algorithm to model a given dataset with the minimum “error”.

One of the most common and simplest methods used to calculate the coefficient and intercept of the regression line is the Ordinary Least Square (OLS) *optimization* method. In short, OLS minimizes the Square-Error for each point against a given linear function by adjusting the function's parameters, which in the end produces optimal parameters for a equation in the form of a linear line of best fit.

The Square-Error function, which is what OLS *optimizes*, is simply a summation of the squares of the difference between actual values and predicted values, over all data points (1.2):

$$E = \sum (y - \hat{y})^2 \quad (1.2)$$

Because the predicted values \hat{y} for a linear model is modelled as $\alpha x + \beta$, the Square-Error function for a linear model can be more specific (1.3):

$$E = \sum (y - (\alpha x + \beta))^2 \quad (1.3)$$

OLS calculates the values of α and β which minimize S in the summation above. Unlike the generic differential method described above, OLS is specialized for linear functions and can calculate the optimal parameters in one stop, using summation ratios. The coefficient α , or the linear trend of the dataset can be calculated with (1.4):

$$\alpha = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (1.4)$$

where n is the number of data points.

After calculating the slope of the trend, the intercept β , is calculated by (1.5):

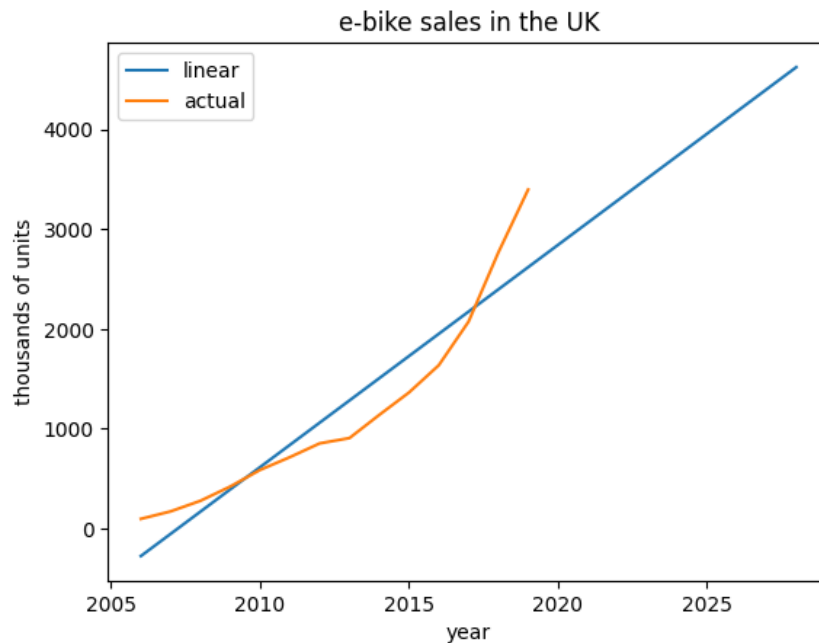
$$\beta = \bar{y} - \alpha \bar{x} \quad (1.5)$$

OLS was applied to the given data set to obtain the coefficient α and the y-intercept β - 222.6 and 446810, respectively - which corresponds to the following linear equation:

$$\hat{y}_i = 222.6i - 446810 \quad (1.6)$$

where i is the year.

Linear regression is useful in relation to the problem as it is simple to interpret and portray, allowing the prediction of data to be accurate during interpolation. However, if the data to be predicted is outside the range, i.e. predicting future e-bike sales, extrapolation may be inaccurate due to a false assumption of the trend. Furthermore, if the variables plotted provide a non-linear relationship, a linear regression line may inaccurately represent and predict values, which is the case in the data provided. Statistical error of the linear regression model against existing data shows a good but not perfect accuracy (Table 1.2).



Metric	Value
MSE	108671
RMSE	329.65
MAE	268.82
PCC	0.9386
R^2	0.8810
CVAR	867138

Table 1.2: Linear Model Statistics

Figure 1.1: Regressed Linear Model

1.5 Results

Results

1.6 Model Revision

Model Revision

1.7 Evaluation

Strength 1: asdf

Strength 2: asdf

Weakness 1: asdf

1.8 Technical Computing

Technical computing

2 Q2: Shifting Gears

2.1 Defining the Problem

Defining the Problem

2.2 Assumptions

Assumption 1: Statement **Justification:** blah blah

Assumption 2: Statement **Justification:** blah blah

Assumption 3: Statement **Justification:** blah blah

2.3 Variables

See table 2.1:

Variable	Definition
x	description
y	description
z	description

Table 2.1: Variables in the Model

2.4 The Model

Model

2.5 Results

Results

2.6 Model Revision

Model Revision

2.7 Discussion

Strength 1: asdf

Strength 2: asdf

Weakness 1: asdf

2.8 Sensitivity Analysis

Sensitivity Analysis

2.9 Technical Computing

Technical computing

3 Q3: Off the Chain

3.1 Defining the Problem

Defining the Problem

3.2 Assumptions

Assumption 1: Statement **Justification:** blah blah

Assumption 2: Statement **Justification:** blah blah

Assumption 3: Statement **Justification:** blah blah

3.3 Variables

See table 2.1:

Variable	Definition
x	description
y	description
z	description

Table 3.1: Variables in the Model

3.4 The Model

Model

3.5 Results

Results

3.6 Model Revision

Model Revision

3.7 Discussion

Strength 1: asdf

Strength 2: asdf

Weakness 1: asdf

3.8 Sensitivity Analysis

Sensitivity Analysis

3.9 Technical Computing

Technical computing

4 References

4.1 Bibliography

4.2 Program Code

Result data generated:

```
text data stuff
```

Python program code:

```
# pass
```
