

Net Zero Emissions Estimation Research

Group 09 : Cindy Aurelia Ursia | Justin Hong Jing Cheng | Lim Chang Yu | Tan Qi Yun | Yan Jie
Department Supervisor : Prof Adam Ng | Prof Ang Beng Wah
Industry Supervisor : Jillian Teo | Dr Robert Li

About Company

Standard Chartered Bank (SCB) operates in over 59 markets worldwide and aims to achieve net zero emissions through its financing activities. However, Oil & Gas (O&G) is the largest emissions contributor in its portfolio at 30.3% (13.7 MtCO₂e). SCB thus seeks to manage its industry exposure using emissions estimation to better adjust their portfolio accordingly.

Introduction

SCB Net Zero Methodology

Measuring Baseline → Projection → Target Setting → Enhancement

- This project focuses on the 1st phase: **Measuring Baseline Emissions**
- SCB currently obtain baseline emissions in the two ways shown below:

Data Coverage

39% (Regression Analysis)

Revenue-based Model
Has a 15% error rate and is prone to economic fluctuation!

61% (S&P Trucost)

Direct Reporting

Objectives

- Enhance SCB's current data coverage with publicly available data sources
- Explore on more estimation methods for corporate emissions
- Incorporate both financial and non-financial based features as proxy for companies without emission data

Methodology Overview

Linear Regression Methodology is Employed for 2 Level Approach*

Phase 1: Data Collection

Companies Data collection on both Financial & Non-Financial Predictive Features

Phase 2: Statistical Check

Ensure features to be used does not violate model assumptions

Phase 3: Model Enhancements

Use of modeling techniques to improve models' performance

Phase 4: Results & Analysis

Comparison between models used

Phase 5: Recommendations

Propose relevant features & models for use along with key takeaways

Industrial Level

Oil & Gas

Gas Manufacturing & Distribution

Natural Gas Extraction

Oil Extraction

Petroleum Refinery

Support Activities for Petroleum & Natural Gas Extraction

* This study specifically focuses on the 3 ticked sub-sectors

Methodology

Phase 1: Data Collection

- Consist of 60 O&G companies from different sub-sectors & regions with following features:

Financial Predictive Features

- Total Assets (TA)
- Market Capitalization (MCap)
- Revenue
- EBITDA
- EBIT
- Capital Expenditure (Capex)

Non-Financial Predictive Features

Operational Features

- Headcount
- Energy Consumption (en_con)
- Renewable Energy Consumption (reEn)
- Non-Renewable Energy Consumption (NonReEn)

Categorical Features

- Region
- Sub-sector
- Daily Oil Production (Oil Prod)
- Daily Gas Production (Gas Prod)
- Daily Petroleum Refinery Production

Other Features

- ESG Risk Rating (esgRating)

Phase 2: Statistical Check on Features for both Levels

Pairwise Plots (Histogram & Scatter Plots)	Correlation Coefficient (Features vs Emissions)	Correlation Coefficient (between features) & VIF	One-way ANOVA
--------------------------------------------	-------------------------------------------------	--------------------------------------------------	---------------

Phase 3: Model Enhancement Techniques

Stepwise Regression	Recursive Feature Elimination (RFE) with Cross Validation	Extreme Gradient Boosting (XGB)	Ridge/Lasso Regression (for sub-sector levels only)
---------------------	-----------------------------------------------------------	---------------------------------	-----------------------------------------------------

Phase 4: Model Performance Indicators for Results Analysis

Root Mean Squared Error (RMSE)	Adjusted Test R ²
--------------------------------	------------------------------

Industrial Level

Finalized Features after Checks:

S/N	Model Type	Revenue	Mcap	Capex	Headcount	en_con	Region	Sub-sector
1	Revenue-based	✓						
2	Financial	✓	✓	✓				
3	Non-Financial				✓	✓		
4	Financial & Non-Financial	✓	✓	✓	✓	✓		
5	Financial (with Categorization)	✓	✓	✓			✓	✓
6	Non-Financial (with Categorization)				✓	✓	✓	✓
7	Financial & Non-Financial (with Categorization)	✓	✓	✓	✓	✓	✓	✓

Models Performance:

S/N	Model Type	RMSE (%)	Adj R ² (%)
K Fold Cross Validation			
1	Revenue-based	15.13	0.53
2	Financial	15.00 (▲1%)	0.44 (▼17%)
3	Non-Financial	14.09 (▲7%)	0.58 (▲9%)
4	Financial & Non-Financial	11.61 (▲23%)	0.65 (▲23%)
5	Financial (with Categorization)	15.32 (▼1%)	0.31 (▼43%)
6	Non-Financial (with Categorization)	13.52 (▲11%)	0.56 (▲6%)
7	Financial & Non-Financial (with Categorization)	10.84 (▲28%)	0.65 (▲23%)
8	RFE	9.51 (▲37%)	0.69 (▲30%)
Shuffle Split Cross Validation			
1	Revenue-based	14.88	0.53
2	Financial	14.78 (▲1%)	0.42 (▼21%)
3	Non-Financial	13.94 (▲6%)	0.56 (▲6%)
4	Financial & Non-Financial	11.53 (▲23%)	0.63 (▲19%)
5	Financial (with Categorization)	15.13 (▼2%)	0.23 (▼57%)
6	Non-Financial (with Categorization)	13.44 (▲10%)	0.53 (0%)
7	Financial & Non-Financial (with Categorization)	10.81 (▲28%)	0.61 (▲15%)
8	RFE	9.36 (▲37%)	0.70 (▲32%)
XGB Regression Results			
1	XGB (GridSearch)	14.56 (▼2%)	0.52 (▼1%)
2	XGB (RandomSearch)	12.48 (▼16%)	0.65 (▲23%)

Sub-sector Level

Finalized Features after Checks:

S/N	Sub-sector	Capex	EBITDA	Headcount	en_con	Region	esgRating
1	Oil Extraction	✓			✓	✓	
2	Gas Extraction				✓		✓
3	Petroleum Refinery		✓	✓			✓

S/N	Sub-sector	Gas Prod	NLG Prod	Daily Oil Prod	Fuel Oil Prod	Gasoline Prod	LPG Prod	Diesel Prod
1	Oil Extraction			✓				
2	Gas Extraction	✓	✓					
3	Petroleum Refinery				✓	✓	✓	✓

Models Performance:

S/N	Model Type	RMSE (% improvement from Revenue-based Model)		
		Gas Extraction	Oil Extraction	Petroleum Refinery
Linear Regression				
1	Revenue-based	41.13	4.325	12.28
2	Stepwise	23.50 (▲43%)	5.35 (▼24%)	18.41 (▼50%)
3	RFE	22.71 (▲45%)	5.66 (▼31%)	16.83 (▼37%)
Ridge Regression				
1	Revenue-based	31.18	4.22	12.46
2	Stepwise	18.72 (▲40%)	5.14 (▼22%)	9.43 (▲24%)
3	RFE	20.07 (▲36%)	5.75 (▼36%)	9.35 (▲25%)
Lasso Regression				
1	Revenue-based	35.37	4.24	12.16
2	Stepwise	12.04 (▲66%)	5.19 (▼22%)	11.22 (▲8%)
3	RFE	11.71 (▲67%)	5.61 (▼32%)	12.65 (▼4%)

Summary of Best Performing Models:

Level	Model Type	RMSE (%)
Industrial	RFE with Multi-Linear Regression	9.36 (▲37%)
Sub-sector: Gas Extraction	RFE with Lasso Regression	11.76 (▲67%)
Sub-sector: Oil Extraction	Revenue-based with Ridge Regression	4.22
Sub-sector: Petroleum Refinery	RFE with Ridge Regression	9.35 (▲25%)

Recommendations

Level	Model to Use	Key Features to Use
Industrial	RFE with Multi-Linear Regression	<ul style="list-style-type: none"> Energy Consumption Revenue Region: China & Europe Headcount Sub-sector
Gas Extraction	RFE with Lasso Regression	<ul style="list-style-type: none"> Energy Consumption
Oil Extraction	Revenue-based with Ridge Regression	<ul style="list-style-type: none"> Daily Oil Production Revenue
Petroleum Refinery	RFE with Ridge Regression	<ul style="list-style-type: none"> Gasoline Production EBITDA Headcount

Overall

RFE should be used alongside regression models since it yielded the most optimal results. (Based on RMSE and % improvement)

Sub-sector

Ridge/Lasso regression is best used at handling the smaller datasets which tends to overfit. (As compared to industrial level)

Future Direction of Studies

- Further expand dataset to include more companies**
More data points can display stronger patterns, hence improving model performance. (i.e. purchase databases)
- More robust algorithms can be explored**
May be better in handling more complex relationships and outliers in a larger dataset, thus improving estimation accuracy. (i.e. Random Forest, Deep Learning, AdaBoost)
- Extend study to related sectors (Power, Metals & Mining)**
Proposed method can be extended to other sectors for greater accuracy of SCB's financed emissions reporting. Non-financial features in these sectors are to be explored.

Achievements

Our results showed a **significant reduction in error rates** as compared to SCB's model. These findings were **well received** by our industry supervisors and also **brought to the upper management's attention**.

As the next step, SCB will be **further extending its collaboration period** with NUS ISEM. The team seeks to further build upon our current approach based on our proposed future studies.

Key Takeaways

- There is a no one-size fits all approach to emission estimation.** Accurate industry and sub-sectors estimations require different features combinations due to their varying nature of activities.
- Global problems requires economy-wide solutions.** Joint efforts to expand datasets is instrumental to amplify SCB's results applicability & help our world transit to a net-zero future.

Key Skillsets

Python

Data Processing and Cleaning, Feature Engineering, Estimation Models Development

ESG

Accounting of Greenhouse Gases (GHG) emissions and ESG reporting standards

Statistics

Statistical Analysis Techniques (p-value, ANOVA, VIF) and Regression Techniques

Research

Industry Research, Competitor Analysis, Project & Stakeholder Management and Powerpoint