International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229572

Predicting Steel Consumption for India using Multi-Variate Regression Analysis of Data from Similar Countries

Raghav Agarwalla[1], Dr. Sunil Chinta[2]
[1]Senior, UWCSEA Dover Road, Singapore
[2]IIT Madras PhD, India

Article Info
Volume 9, Issue 5
Page Number : 387-395

Publication Issue
September-October-2022

Article History
Accepted : 02 Oct 2022
Published : 14 Oct 2022

ABSTRACT
An attempt is made to create a statistical model for predicting finished steel consumption for India in the medium term by studying historical data from similar sized steel economies. Economies were selected based on their size and steel consumption profile. These were further narrowed down to those economies where reliable economic data was available for a per capita GDP range where India has been in the recent past and would be in the medium term. Finally the data for China and South Korea were found suitable. The data was split in periods where said economies were in the per capita GDP range of $500-2000 and $2000-10,000 (constant 2021 US$). India is currently at a per capita GDP of $2000. Thereafter, utilising regression analysis, starting with a single variable regression and advancing into regularised multi-variate regression, an attempt was made to narrow down the economic predictors for steel consumption. Finally, a model was derived after multiple rounds of data fitting that could predict the per capita steel consumption for India. To arrive at a steel consumption number for a particular year, the economic variables responsible for steel consumption were forecasted for the year and a range was arrived at. **The resultant range for steel consumption for India in 2030 is 157-188 mn MT.**

Keywords: Steel Economies, PCASTECONS, GDP, PCAPGDP

Economic importance of steel

An economy grows by boosting demand and supply simultaneously. Both these need capital investments. Each step of economic growth and resultant activity consumes steel because of its ubiquity. Therefore more than any other single product, steel demand is the greatest indicator of a growing economy. Steel contributed 3.8 % of the world's GDP and supported 96 million jobs in 2017. The entire supply chain of steel and its consuming sectors contributed 10.7% of the world's GDP and supported 259 million jobs in 2017.

When economies are poor (per capita GDP < $2000), the per capita steel consumption may not be as high since the individual is spending on sustenance items such as food and clothing. Individual savings are low and capital expenditure (consumer durables) at a family level is scarce. Therefore per capita steel consumption is very low.

When economies are emerging out of poverty and can be generally classified as emerging economies ($2000<per capita GDP<$10,000), there are multiple drivers for steel consumption. Families having satisfied their basic needs are steadily investing into capital goods such as appliances, automobiles and homes. Public expenditure shifts from subsidies to infrastructure creation and private expenditure is high into manufacturing and capacity creation. There is an acceleration in overall steel consumption and the per capita steel consumption exponentially increases.

As economies mature (per capita GDP>$10,000) and people become wealthy, public infrastructure investments are limited to repairs, private capital consumption is limited to replacements and drop in manufacturing competitiveness due to high labour costs/currency appreciation inhibits investment in capacity creation.

Therefore, we may see three different trajectories for steel consumption during the three different phases of an economy. However at each stage steel consumption is an important indicator.

India

As a steel consumer, India is the second largest in the world at 106 million tons of finished steel in 2021 but overshadowed by China by a factor of 10. India's per capita consumption at 76 kgs is still well below China's 667 kgs. Steel manufacturing contributes 2% to India's GDP and provides 29 million jobs. Given the current path of the economy, there are potentially going to be significant changes to this figure.

Hypothesis

Per capita steel consumption [PCASTECONS] is primarily driven by per capita GDP [PCAPGDP]. There might be additional drivers that represent consumer durables, investment and trade. PCASTECONS would have a different trajectory at different PERCAPGDP levels as indicated by a poor, emerging or developed economy.

India's PCASTECONS in the future can be predicted using a similar economy's path mapping the primary driver PERCAPGDP and additional drivers as listed above.

Approach

Using correlation and regression analysis, starting with a univariate and continuing onto multi-variate analysis, a function for India's medium term steel consumption to be derived. Use this function combined with the reader's prediction of the variables at different timelines to predict India's steel consumption at the same timelines. For the purposes of this paper we shall predict the steel consumption for the year 2030.

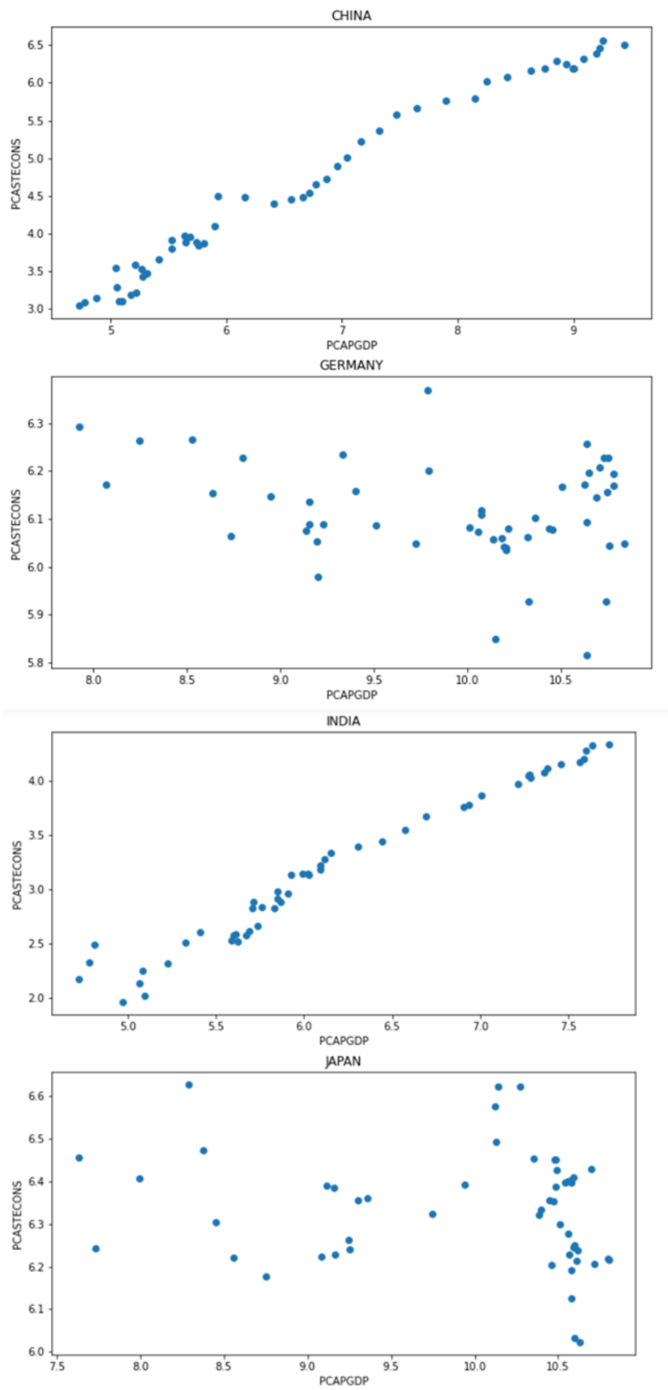
Finding the right comparable economies

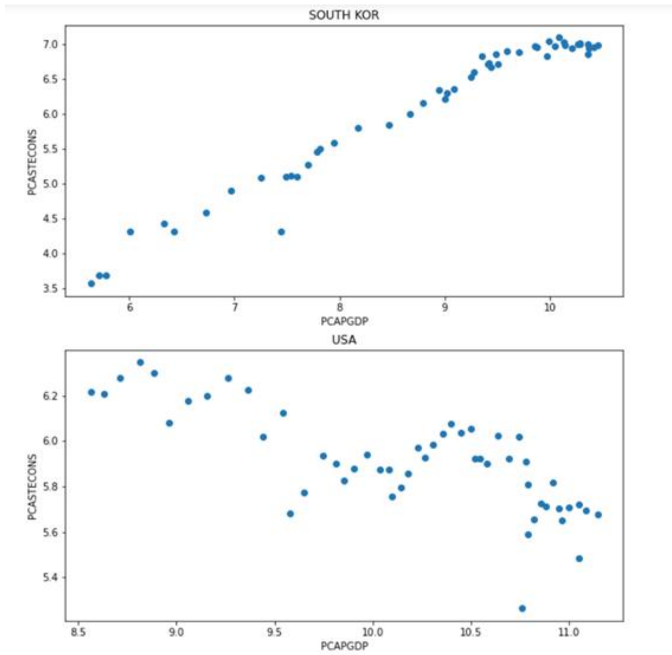
We chose the following economies:

USA/Japan/Germany/China/South Korea/Russia/India

All the above are large economies as also large producers of steel. They are at different stages of development and belong to different geographies. Russia was subsequently discarded as the data was corrupted due to the dissolution of the Soviet Union as also critical gaps in the 70s thereby rendering the data series not credible.

A simple correlation analysis was conducted between PCASTECONS and PERCAPGDP for the six economies and the results are provided on a log scale for better readability:





[All six graphs can be used or all six eliminated. Size should be small if used]

The Plots show a clear correlation for countries China, India and South Korea and poor correlation for USA, Germany and Japan. This clearly supports the hypothesis that the correlation shall not be consistent at various stages of development of an economy and it will get poor as the economy matures. Doing the correlation for the three mature economies at pre-developed stage was not possible due to inavailability of a robust data series as also the fact that the data would be so dated (pre-80s) that technological evolution would affect the outcomes.

Therefore the latter 3 data sets are also eliminated and we are left with 3 data sets only. This is sub-optimal but there are no other large economies where we have robust data from sub $2000 levels upto $10,000 levels.

Decoding the relationship between PCASTECONS and PCAPGDP

We proceeded to fit a Linear regression model using individual country's data. This approach involves fitting a log-log linear model between PCASTECONS and PERCAPGDP. Separate models are fit for each country and for the GDP ranges of 500-2000 and 2000-10,000. 30% of data within each slab is randomly taken for testing the model performance.

	Per Capita GDP $500-2000			Per Capita GDP $2000-10,000	
Coeff	India	China	South Korea	China	South Korea
PCAPGDP	0.6406	1.2241	0.5694	0.4557	0.7184
INTERCEPT	-0.6238	-3.5888	0.8459	2.1874	-0.1485
Mean % error	2.22	3.30	32.12	4.57	8.44

It is evident from the above examples that PCAPGDP is a consistent driver of the PCASTECONS. However, for each country the slopes and intercepts change and for the same country the regression output changes at different PCAPGDP slabs. This supports part of the hypothesis but does not give a verifiable predictor of PCASTECONS in the future. There is a higher than acceptable error for South Korea for PCAPGDP at 500-2000 and the reasons for this are not explained at this time.

<u>Attempt to improve the regression by adding another variable</u>

It is well know that Gross Fixed Capital Formation has a multiplier effect on economic growth representing investments into infrastructure, manufacturing, construction. These are sectors with highest impact on steel consumption from a visible standpoint. GFCF has been normalised to per capita GCFC (PCAGFCF) and a multi-variable regression between PCASTECONS and PCAPGDP+ PCAGCFC is attempted. A log-log linear model between PCA Steel Consumption and PCAPGDP+PCAGFCF is fitted. Separate models are fit for each country and for the PCAPGDP ranges of 500-2000 and 2000-10,000. 30% of data within each slab is randomly taken for testing the model quality.

Further, since it is well understood that GDP and GFCF are not independent variables, rather highly correlated variables, the output of multi-variate regression using co-dependent variables creates the problem of multi-collinearity. This makes the results statistically non-dependable. Therefore, this task is conducted using regularised regression.

Regularized regression is a type of regression where the coefficient estimates are shrunk towards zero. The magnitude (size) of coefficients are penalized. Complex models are discouraged, primarily to avoid overfitting.

	Per Capita GDP $500-2000			Per Capita GDP $2000-10,000	
Coeff	India	China	South Korea	China	South Korea
PCAPGDP	0.9052	0.1815	0.0589	0.2388	0.6288
PCAGFCF	0.0761	0.8210	0.7739	0.7455	0.3887
Mean % error	2.05	3.17	37.87	3.38	7.89

As is evident from the above table, adding GFCF to the analysis does not deteriorate the quality of the output if we are tackling multi-collinearity effectively. There has been a slight, albeit statistically insignificant, improvement in the analysis by adding the variable. Adding a variable is not important singularly from the point of view of reducing error margins, the exercise helps us understand the impact of multiple drivers on the demand of commodity especially one as ubiquitous as steel that has myriad applications and in turn drivers. The ultimate quality of prediction can improve manifold if we can use multiple drivers that may remain symmetric or become divergent in the future. E.g. an economy may show consistent headline growth but may turn its focus from exports to domestic consumption. Something like automobile production may see downward shifts if the policy starts taxing automobile sales harshly to fund public mobility.

Steel Intensity

Our objective at this stage is to find the right variables for predicting steel consumption in general and those that may be significant for India. It is clearly understood that steel consumption would be driven by Headline growth (GDP). However it is boosted significantly by Investment (GFCF). Therefore these are two fundamental variables that cannot be ignored. Further, indicators such as Manufacturing as a % of GDP, Construction as a % of GDF, Auto production, merchandise exports are all drivers of steel demand. There may be less significant indicators but reliable, long term, multi-lateral data is not available on them.

It is important to take a break here and talk about steel intensity. Steel intensity is the concept that explains the difference in steel demand for similarly sized, similarly populated economies. This is an observed phenomena that debunks GDP being a unilateral indicator of steel consumption. An emerging nation saving hard, investing aggressively, focussed on manufacturing and exporting ruthlessly can have significantly higher per capita steel consumption than an economy largely agrarian, leading in services, ignoring public capital expenditure and not focussed on exports. These are not hypothetical examples. At around $2000 per capita GDP the per capita steel consumption of the selected economies were as follows:

Country	Year	Per Capita GDP	Per Capita Steel Cons	Steel intensity Kgs/$1000 GDP
South Korea	1983	2199	195	89
China	2006	2099	288	137
India	2019	2072	75	36

It is well known that India was saving less, investing less, manufacturing less and exporting less on a per capita basis than the Asian Tiger Korea and the Chinese Dragon. Therefore, at the same income level, its steel consumption was 40-75% less than these players.

Therefore deriving any formula by ignoring indicators representing investment, consumer durable spend and trade can be debunked straight away without checking their statistical integrity. The future prediction of steel consumption shall require an estimation of steel intensity.

Finding the right suite of variables

We ran multiple regressions using base variables PCAPGDP+PCAGFCF and additional variables from the following list:

Manufacturing as a % of GDP/Industrial as a % of GDP/Auto production/Merchant Exports/Urban population %

Additional variables could have been thought of but they were either included in the above (e.g. Construction which is included in Industrial) or may be insignificant in the overall picture (no. of air conditioners purchased) or may be highly correlated (auto components to auto production) or simply unavailable.

Basis, the overall exercise the best regression fits were obtained by using the following variables:

Per Capita GDP (PCAPGDP)

Per Capita GFCF (PCAGFCF)

Per Capita Automobile production per capita (PCAAUTOPROD)

Per Capita Merchandise Exports per capita (PCAMEREX)

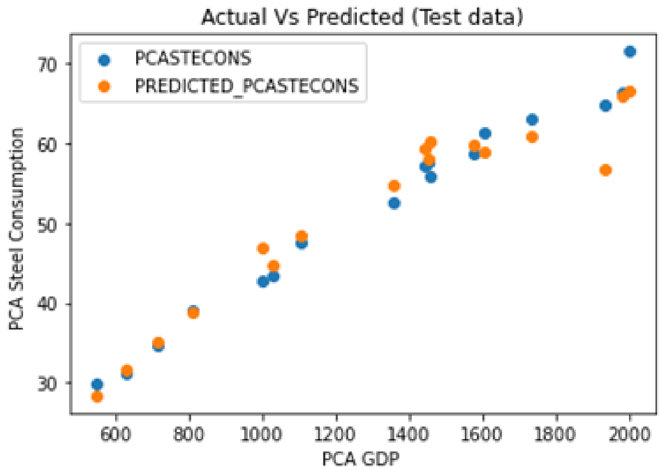
We continued to use regularized regression in order to deal with the multi-collinearity between the variables. Separate models were fit for each country and for the GDP ranges of $500-2000 and $2000-10,000. 30% of data within each slab is randomly taken for testing the model quality.

	Per Capita GDP $500-2000			**Per Capita GDP $2000-10,000**	
Coeff	**India**	**China**	**South Korea**	**China**	**South Korea**
PCAPGDP	0.7263	0.1967	0.2565	0.2761	0.3680
PCAGFCF	0.1095	0.2519	0.3973	0.6852	0.4153
PCAAUTOPROD	0.1489	0.2781	-0.0685	0.1508	-0.0345
PCAMEREX	0.0078	0.2647	0.1739	-0.1212	0.2583
Mean % error	1.65	2.23	35.58	5.22	7.54

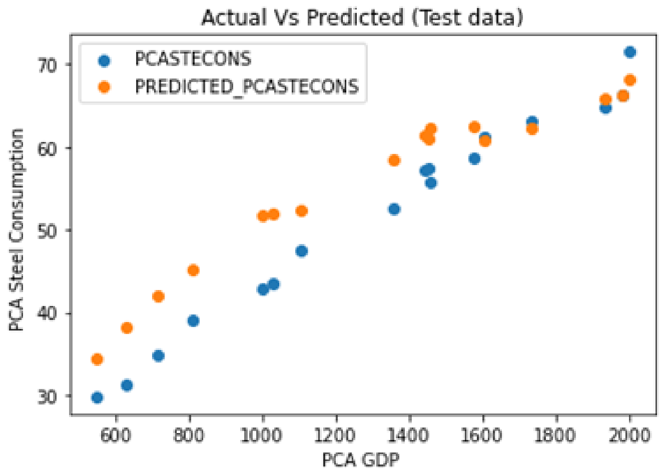
Cross Prediction

Our previous tests have been to test a data series against itself, i.e. short/medium term predictions for the same economy where expectedly there would not be a sudden change in economic profile, e.g. a manufacturing intensive economy would not overnight become agrarian or a net exporter is not suddenly going to have a major trade deficit. Therefore quality of fits would be and have been good. The litmus test is to take the variables for one country and fit it to another country at a similar stage in its economic lifecycle. We took the output of China and Korea for the PCAPGDP slab $500-2000 individually and applied them to India's data from the same economic period and fit the predicted steel consumption to the actual figure. The results were as follows:

Prediction for India using China's data



Prediction for India using South Korea's data



Mean % error= 8.79

The results are encouraging as the error levels are quite low. The results allow a correction for difference in steel intensities of similarly income level economies by incorporating the variables responsible for differences in steel intensity. The error for Korea-India fit is higher than China-India fit because Korea data in the 500-2000 slab had a comparable higher error across analyses. This may be related to reliability of data that pertains from the period 1974-1983 when South Korea grew at an unprecedented pace.

Pooling of Data-diversification of data risk

We then attempted pooling of data between China and South Korea using a regularized Ridge Regression to fit the model. As the outputs from individual regressions in the above step were varying from one another despite providing fairly good predictions, an attempt was made to come up with a unifying equation. This exercise, if successful, would provide the reader a logical and concise output.

The model takes the following form:

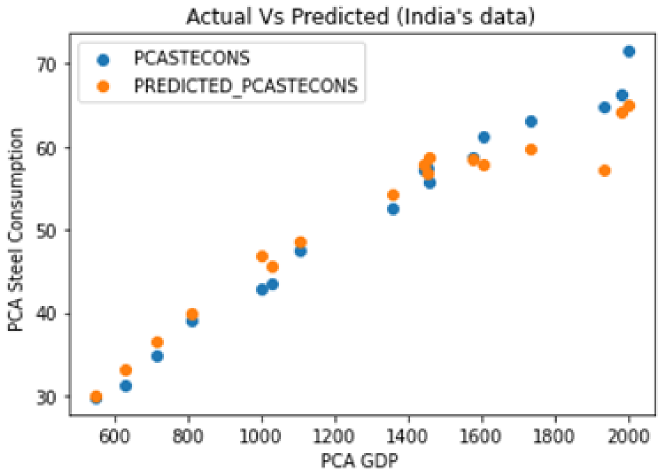
$$\ln\left(\frac{y}{\tilde{y}}\right) = \beta_{gdp}\ln\left(\frac{X_{pgdp}}{\tilde{X}_{pgdp}}\right) + \beta_{pauto}\ln\left(\frac{X_{pauto}}{\tilde{X}_{pauto}}\right) + \beta_{pgfcf}\ln\left(\frac{X_{gfcf}}{\tilde{X}_{gfcf}}\right) + \beta_{pmerex}\ln\left(\frac{X_{pmerex}}{\tilde{X}_{pmerex}}\right)$$

y: Steel Consumption

x̃: Geometric mean of the respective feature and country

We used China's and South Korea's data for PCAPGDP between 500 and 2000 to fit a regression line, and test the performance against India's data for the same slab.

Coeff	Value
PCAPGDP	0.2324
PCAAUTOPROD	0.1954
PCAGFCF	0.2702
PCAMEREX	0.1904



Mean Error % = 4.65

As is evident the error of 4.65% is no worse than 4.20% and 8.79% from the previous step, it does provide a unifying equation for prediction of steel demand for India.

Final Step

Using the regularized Ridge Regression from the previous step but this time for a PCAPGDP slab of $2000-10,000, we arrive at the following equation that hopefully should provide a reliable predictor for steel demand/consumption for India in the next 20 years or till it reaches a per capita GDP of $10,000.

Coeff	Value
PCAPGDP	0.2960
PCAAUTOPROD	0.0826
PCAGFCF	0.4346
PCAMEREX	0.1525

Forecasting India's Steel Consumption in 2030

Population forecast for 2030

According to the latest edition of the United Nations' World Population Prospects, India is expected to overtake China as the most populous country in this decade and grow from 1.4 billion currently to 1.515 billion by 2030.

GDP Forecast for 2030

GDP is expected to grow between 6 and 8% in real terms per annum for the next decade as per Goldman Sachs provided it does or does not do certain things to boost it's economy. This is the range that has been reiterated by multiple economists and agencies around the world.

There is a lot of pent up energy in the economy on the back of two lost years of the pandemic as well as major structural reforms implemented in the years preceding Covid that are expected to play out going forward. However Quantitative Tightening and associated phenomena around the world seems to be pushing the global economy into a short term recession and this may play a spoilsport for the Indian economy as well. Therefore, GDP is projected to grow from 3.17 tn in 2021 to between 5.4 tn and 6.3 tn by 2030 in constant 2021 US$.

Gross Fixed Capital Formation forecast for 2030

GFCF as a % of GDP in India between 2000-2010 grew from 25% to 40%. This was on the back of a credit fuelled commodity-manufacturing boom. However, starting 2010 private companies with highly leveraged balance sheets and the public exchequer with its burgeoning deficit were forced to cut down on investments. From a high of 40% in 2010, GFCF fell to 30% by 2016. This has been a painful period of consolidation for the Indian economy. China on the other hand has maintained a GFCF of 40%+ since 2000 in an unprecedented investment boom, the likes of which the world has never seen in its history.

Going forward, the tailwinds for GFCF are a pickup in private sector investments as capacity utilizations are running high and private corporate debt is immensely manageable. The Government has indicated a $1.3 tn infrastructure plan (National Investment Pipeline) for the years 2020-25 and seems serious in implementing the same. Notwithstanding the expected reduction in global capital flows, the level of investment in the coming decade shall be higher than the preceding decade and GFCF should be 35% of the GDP for this decade. For 2030, a bad case scenario is 32% of GDP and a good case scenario is 38% of GDP has been estimated.

Merchandise Export Forecast for 2030

This is the most difficult forecast as it just does not map India's manufacturing competitiveness as also the potential of the world economy to absorb more goods. This is further affected by policy variables such as trade barriers and trade agreements. From a qualitative standpoint, India is figuring out the key levers to being more competitive in merchandise exports. There are a slew of trade agreements, either inked or in the pipeline, a domestic push towards manufacturing in terms of the PLI scheme and structural reforms such as GST that are expected to boost exports significantly in the coming decade. On the other hand, an impending global slowdown and the increasing nationalistic tendency to onshore production by developed countries may play spoilsport to India's export ambitions. Last, the challenges faced in the recent past by importers from the world's factory China due to geopolitical issues and China's zero covid policy has birthed a paradigm popularly known as China+1. Many companies are moving production from China to countries such as India and Vietnam. This phenomena can make these countries the new Asian tigers in terms of exports.

India is currently exporting just short $40 bn per month in 2022 demonstrating a growth of nearly 15% over last year. An expectation of real growth for India in merchandise exports for the next decade would range from a pessimistic 7% due to a slowdown in global trade to a wildly optimistic 12% that could only be realised if India manages to attract record FDI in manufacturing and gets its act right. Basis the above, the range of merchandise export in 2030 (constant 2021 US$) could be $725 bn to $1100 bn.

Automobile Production Forecast for 2030

India produced 4.4 million vehicles (excluding 2/3 wheelers) in 2021. India is growing both as a vehicle market as well as an export hub, especially for small cars. Further the trend of moving upward from 2/3 wheelers to 4 wheelers as per capita income increases, is a major tailwind for this sector. It is expected that Indian auto production shall grow between 10-12%

for the next decade. Therefore, the range of auto production in 2030 could be 10.4 to 12.2 million vehicles.

Final Projections for 2030

Parameter	Unit	Low Case	High Case
Population	million	1515	1515
GDP	$ Billion	5361	6342
GFCF	$ Billion	1715	2410
Merchandise Exports	$ Billion	725	1100
Auto Production	million	10.40	12.20
Per Capita GDP	$	3540	4185
Per Capita GFCF	$	1130	1590
Per Capita Merchandise Export	$	480	725
Per Capita Auto Production	Nos.	0.007	0.008
Per Capita Steel consumption	Kgs	103	124
Gross Steel consumption	Mn MT	157	188

The steel consumption for India by 2030 is expected to range between 157 and 188 million MT per annum.

Notes:

$ denotes 2021 constant US$

Mn is million

MT is Metric Tonnes

REFERENCES

[1]. https://worldsteel.org/steel-topics/statistics/steel-statistical-yearbook/
[2]. https://data.worldbank.org/
[3]. https://statisticsanddata.org/data/top-countries-by-motor-vehicle-production-1950-2020/
[4]. https://worldsteel.org/wp-content/uploads/The-role-of-steel-manufacturing-in-the-global-economy.pdf
[5]. https://worldsteel.org/steel-by-topic/statistics/annual-production-steel-data/C_asu_fsp_pub/CHN/IND
[6]. https://www.weforum.org/agenda/2022/08/world-population-countries-india-china-2030/
[7]. https://www.financialexpress.com/economy/india-may-grow-at-8-2-over-next-decade-goldman-sachs-chalks-out-4-growth-levers/2622997/
[8]. https://economictimes.indiatimes.com/news/economy/indicators/india-may-be-on-the-cusp-of-private-investment-revival/articleshow/93795139.cms
[9]. https://en.wikipedia.org/wiki/National_Infrastructure_Pipeline
[10]. https://www.investindia.gov.in/production-linked-incentives-schemes-india
[11]. https://commerce.gov.in/international-trade/trade-agreements/
[12]. https://en.wikipedia.org/wiki/China_Plus_One
[13]. https://www.siam.in/statistics.aspx?mpgid=8&pgidtrail=9

Cite this article as :

Raghav Agarwalla, Dr. Sunil Chinta, "Predicting Steel Consumption for India using Multi-Variate Regression Analysis of Data from Similar Countries", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 387-395, September-October 2022. Available at doi : https://doi.org/10.32628/IJSRST229572

Journal URL : https://ijsrst.com/IJSRST229572