MAY 25, 2020

A KIND OF STATISTICAL MODELLING TO ESTIMATE THE PRICE OF 1800 MHZ SPECTRUM

VJ CHRISTOPHER
JOINT WIRELESS ADVISER, WPC WING
Regional Licensing Office, Mumbai: 400067

Abstract

Any commodity is priced based on its value. The radio spectrum is a natural resource, but its availability is scarce in the frequency bands used for mobile communication. The pricing of spectrum is a method to ensure its effective and efficient utilization owing to its scarcity. The regulators all over the world use different mechanisms to find a reserve price that is neither too high nor too low, but the right one that would result in an economically balanced state of demand meeting the supply when it is auctioned. In India, 1800 MHz was auctioned 4 times at a reserve price determined by Telecom Regulatory Authority of India(TRAI). The auction though fetched substantial revenue, but more often than not the spectrum was sold at the reserve price indicating a lower competition in the auction. TRAI uses different metrics and techniques to arrive at the reserve price for each of the LSA.

Notwithstanding of the outcome of auction, it is intuitive that the selling price is related to the TRAI's evaluations which is based on different financial models viz. revenue surplus, production function, opportunity costs etc. That relationship is explored in this research using, OLS parametric regression modelling. Apart from the TRAI metrics, few other indicators are also selected as regressors. The regression result showed that the selling price of 1800 MHz spectrum is associated with one of TRAI's evaluation metric and two other inputs. The fitted model with the parameters explain significant association of dependent variable with the independent variable. The result showed a realistic and reliable association. The prediction of the spectrum price using this fitted model is compared with the TRAI's evaluation of 1800 MHz from the recommendations of 2018. The prediction is close to the TRAI determined price and the performance of the model is acceptable. Though TRAI itself used linear regression in some of its recommendations, it is for the first time that the OLS is used across the four auctions squarely to find the association between the TRAI's estimators and the auction price.

Contents:

1. Introduction to this paper:

In this paper, an attempt has been made to predict the price of 1800 MHz spectrum in India using statistic modelling with the help of the data collected from the four spectrum auctions conducted by the Department of Telecom (DoT) in the years 2012,2014,2015 and 2016. The data that we are interested here is the price estimation metrics TRAI has evaluated as part of its consultation process and the reserve price so recommended to DoT in these auctions. This would form the major chunk of the input data. The actual selling price (aka auction price or winning price) of the same spectrum in these auction would be the output data. In a nutshell the idea is to find the correlation between the auction price which is the response variable and the estimated prices by TRAI which are the input variables with the help of classical statistical methods and data from the 4 years' auction.

Apart from the price information which is our primary input variables, it is also explored about the presence of other extraneous variables that could affect the winning price. TRAI has in fact considered many telecom indicators and demographic information while estimating the price. Without looking into the data resources whatever TRAI has depended to produce its estimations, here we would also include external data aside from TRAI, available from the data repositories in the public domain that are important for the study.

The data from the auction history of the four spectrum auctions were compiled into 88 records database with 22 records from each year representing the 22 LSA (Licensed Service Area). After dropping the eight missing value records, the record count rounded up to 80. Additional telecom and demographic indicators were also added to make it a respectable dataset for running the regression for fitting a model. Though this moderate data set constitute the total population (from the four auctions), however from a statistical point of view, we would assume that this is an ideal sample representing the whole population including the data that is likely to be generated in future. The output data that is used in this paper is a kind of longitudinal (4 years) observational data arising out of a natural process (auction) without having any hierarchical structure and the input data is secondary data as a result of consultative process undertook by TRAI. The understanding that we would like to proceed with at this stage is that each observation (record) is considered as independent and identically distributed (IID) sample because the consultations done by TRAI in those four occasions are independently placed.

In order to keep the analysis simple, here the linear model only has been considered to find the correlation between the Winning Price and the input variables mentioned above. The Ordinary Least Square (OLS) algorithm, which is the most often used statistical method for linear regression is used to establish the association here. The multiple regression process generates an OLS equation with its coefficients representing the best estimates of the whole population of auction data. This equation so derived in the OLS process can be used to describe the relationship between the input and output variables. The regression function also can be used to predict the future price of spectrum based on the new input variables, if available. Here in this paper, after generating the regression equation the spectrum prices are also predicted with the help of available input data.

As a supplementary method to the statistical method that predicted the price of spectrum, two of the popular Machine Learning (ML) algorithms are also used to predict the price of spectrum using the same data set as that used for OLS method. The Machine Learning method is used only for prediction rather than finding the correlation equation and so here too it is used for predicting only. The premise of ML fundamentally is that the bigger the data set, the better the prediction accuracy. The records we are having is only 80 samples database which is considered a sparse dataset for applying machine learning. However, the two ML algorithms selected here are KNN regression and Lasso regression which are considered decent with regard to the sparse data set.

2. Research question: -

Let the objective of this study or the research question be framed as follows:

"To find the association, if any, between the Auction price of 1800 MHz spectrum in India in terms of the TRAI estimated price metrics and other demographic, macro economical and telecom predictors available for the years 2012, 2014, 2015 and 2016 using linear regression method and if so, predict the auction price of the same spectrum for the new (future) data"

As hinted in the research question, the association is assumed to be linear between the predictors and the response variable which is to be established in the study. Please note that the response variable is continuous (quantitative variable) and hence the linear regression is selected. So there is a working theory to start with that the Spectrum price is having a linear relation between the different price estimators used by TRAI. This theory is required to be tested.

If a linear model can be fitted to the available data using the linear regression such that it satisfies all the conditions of OLS regression, then the correlation between the independent or exploratory variables and dependent variables or response variable would confirm existence of the linear relationship which is our assumption in the research question.

That the relation between TRAI estimated prices and the auction price going to be linear is quite intuitive since both are about the spectrum prices. But at this stage it may be noted that there are a number of spectrum price variables some of which are likely to be collinear. Multicollinearity is a challenge that could adversely affect the regression, due to the interaction between the exploratory variables may need special attention.

3. Spectrum auction in India

i) The road to auction in India: -

When the telecom revolution started in India, licenses were issued administratively as auction was not an institutionalized method. It may be recalled that the first two cellular licenses issued in 1994-95 were decided on beauty contest principle in the case of Metros and through simple bidding in case of other Service areas. The Government reserved the right to bring in the third operator as part of a migration package offered to the beleaguered telecom operators under the ambit of a telecom policy. As a consequence of such an agreement between the Government and the operators, the PSU operators MTNL and BSNL were introduced in the years 1997 and 2000 respectively. The fourth cellular operator was chosen through a multi-stage bidding in the year 2001 and licenses were issued in 2001/2002.The Unified Access Services License was introduced in 2003. Licenses were issued in November 2003, January 2004, December 2006 and March 2007, and January 2008, in accordance with the Guidelines for the Unified Access Services.

ii) The 3G Auction: -

The 3G spectrum was auctioned in May 2010 based on a recommendation given by TRAI which got approved after going through a channel of the constitutional bodies consisting of Telecom Commission, Empowered Group of Ministers, and Union Cabinet. The 3G auction was considered a huge success at that time for filling the Government cashbox, but the Winner's Curse problem or the 'Irrational Exuberance' as was referred later by TRAI on account of the huge price paid to acquire spectrum in the metro cities got transpired on the telecom operators as a 'curse' and is said to continue to rattle them.

iii) TRAI recommending for no auction:

In the May 2010 recommendations, TRAI has written against conducting auction of the 2G bands (800 MHz/900 MHZ/1800 MHz) putting emphasis on the (lack of) level playing and has stated that, allocation through auction may not be possible as the service providers were allocated spectrum at different times of their license and the amount of spectrum with them varies from 2X4.4 MHz to 2X10 MHz for GSM technology and 2X2.5 MHz to 2X5 MHz in CDMA technology. Therefore, to decide the cut off after which the spectrum is auctioned will be difficult and might raise the issue of level playing field. Further since there were 12 to 14 players in each circles, TRAI felt that there was no dearth of competition, which is the impetus required to conduct auction.

Therefore, in 2010 instead of auction, TRAI has recommended a price determined through administrative mechanism from the 4th cellular operator's price fixed in 2001. Three methods were considered by TRAI. (i). Weighted average cost of Capital using NPV (Net Present Value) using 15% as pre-tax rate. (ii) PLR (SBI Prime Lending Rate) ratio of 11.09% used to work out the present value. (iii) Based on growth of AGR (Aggregated Gross Revenue). It was found that AGR increased about 5 times in 2009-2010 from the AGR in 2002-2003. As the third method yielded maximum effect, which was almost approaching the 3G auction price for 2100 MHz in the year 2010, this was selected as the price to be paid for acquiring the spectrum by new operators.

iv) Intervention by Supreme Court: -

When the 2010 recommendations were under the considerations of the Government, the Supreme Court (SC) in the year 2012 on legal considerations through an unprecedented action cancelled 122 FCFS licenses and decided that Auction is the only transparent method to sell spectrum. Accordingly, SC asked the TRAI to give fresh recommendations for auctioning the 2G spectrum considering the experience of TRAI in recommending the 3G auction price (of year 2006) that resulted in the lucrative 3G auction of 2010. That forced TRAI to seriously look at the whole valuation process. TRAI estimated the price of spectrum vide their recommendations dated 23rd April 2012. TRAI while evaluating the price of 1800 MHz spectrum discarded its earlier methodology of deducing the price from the 4th cellular price of year 2001 on the premise that 2001 price is no longer relevant in year 2012. It is also important to note that for the first time TRAI exhorted that auctioned spectrum be liberalized so that its use could be considered as technology neutral and service neutral thereby the spectrum would be capable of deploying for different applications and services. TRAI therefore recommended for auction of 1800 MHz spectrum for the first time obliging with the observation of Supreme Court.

After the 2012 auction, auction has been accepted as the de-facto method of selling spectrum by the Government. The whole process of auction also got standardized.

The auction process can be chronicled as follows: Govt ask TRAI to give recommendations to conduct auction. TRAI calls for public consultations. TRAI then draft the recommendations and send it to DoT who internally discusses it and send the same to Telecom Commission (TC). TC then moderates this price as recommended by the internal technical committee. The recommendation then goes to the Cabinet which ultimately approves the TC version either as it is or with modifications. This is the final reserve price which get notified setting the stage for auction. As can be seen the TRAI recommended reserve price is not to be considered as sacrosanct, but undergoes many changes before becoming the reserve price. As there is no fixed formula for introducing such subjective changes in the price, the final reserve price will have no relation whatsoever with the TRAI's figures assessed by its original consultative methods.

The auction of 2014 was held on the basis of TRAI recommendations of September 2013. TRAI has attributed the muted response in the 2012 auction and the null response for the 1800 MHz spectrum in 2013 auction to the very high reserve prices in the auction.

"... This is a reality that need to be factored into the current exercise. Equally, the Authority is conscious of the need to avert any possible collusive activity. That said, it also need to be accepted that reluctance to bid in auction with a reserve price does not necessarily represent collusive intent.; if the reserve price is set too high, it may dispel all bidders."

Another reason alleged for the lukewarm response in 2012 auction was due to the collusive cartel behavior of the telecom operators. But TRAI rejected it by the argument that some TSP s bought spectrum in large quantities and one TSP did not buy spectrum even though their license was not cancelled by the Supreme Court. Therefore, according to TRAI, there was no collusive behavior. TRAI also admitted in the recommendations that fixing the reserve price of 1800 MHz from 3G price was not a correct method as it did not yield an apple to apple comparison. Besides, 3G price of Mumbai and Delhi in the year 2010 was an aberration on account of 'irrational exuberance'.Another point noted by TRAI is that indexing the price from the price paid to the 4th Cellular price was too simplistic and restrictive. TRAI therefore changed its methodology to 'bottoms up' approach, focusing to each LSA and then arriving a PAN India price than the 'top down' approach it did in the year 2012 recommendations and in the 3G auction recommendations of 2006. TRAI continued the 'bottom up' approach for all its subsequent auctions

v) Expert Committee:

TRAI while giving recommendations in the year 2010, it has been referring to the 4th Cellular price of 2001 and 3G auction price of 2010. However, it felt that the price of 1800 MHz is not simple as that, but there are various issues involved. In order to further study independently, TRAI assigned this exercise to an expert body consisting of technologists and economists. The expert group studied the various issues involved and submitted a report in January, 2011. The committee explored the subject from the technical (in terms of traffic carrying capacity) and commercial angle (economic modelling) and gave the recommendations on both counts. The technical methodology involves primarily on comparing the price of 1800MHz with 2100 MH. The committee concluded that the price of 1800 MHz band should be between 1.5 to 2.67 times of that of price of 2100 MHz band (3G band). In the commercial estimation the committee suggested two methods after splitting the spectrum into two flavours- Bundled spectrum of 6.2 MHz and incremental spectrum beyond 6.2 MHz. A method known as Discounted Cash Flow (DCF) was used to estimate price of both flavours of spectrum and another method Production Function which is a kind of Substitution method was used to evaluate only incremental spectrum only.

The expert committee mentioned above submitted the report to TRAI on the price of 1800 MHz. Some of the economic methodologies such as DCF, Production function etc. in the report was so espoused by TRAI as its de facto standard of estimation that it used these methods in all of its subsequent recommendations. Therefore, from the 2013 recommendations onwards, TRAI shifted towards quantitative analysis and articulations based on the analysis rather than mere deliberations and discussions based on the technology and best practices.

Accordingly, TRAI began looking at the value the spectrum using multiple approaches as it got convinced more and more on the fact that the value of spectrum could be a combination of market information, technological factors and micro and macro-economic factors. Therefore, a single deterministic approach was dumped, but an approach that use different methods and to arrive at a probabilistic average. Different valuation such as empirical data from the past auctions for same or similar assets; using opportunity cost as an alternative cost providing input to commercial services just like in a production process; a calculation based on projection of future traffic or revenue could be applied and then to take an average of the different methods as the valuation of spectrum.

4. Literature Review

a) The methods used by TRAI to evaluate spectrum after the 2012/2013 auction:

(Refer to the TRAI recommendations reference no 1)*

TRAI has stated that the value of spectrum could be a function of market information, technological factors and micro-macro-economic factors. This understanding has led TRAI to suggest different methods towards the goal of seeking the right value for the reserve price. The following account gives a brief on the different methodologies adopted by TRAI to arrive at a probabilistic reserve price. The principle used in these methods to be:

(a) based on empirical data from the past or

(b) opportunity cost principle or

(c) based on the projections of future traffic/revenues.

i) Production Function approach (also known as Substitution or opportunity cost approach)

The basis of this method is Cobb-Douglas production function:

$X=A (y^{**}\alpha \ X \ z^{**}\beta)$ where y represents spectrum input and z represents BTS input, an optimal mix is used as trade-off.

For arriving at one possible value of 1800 MHz spectrum band, spectrum and BTS can be taken as two distinct inputs to estimate a production function to produce mobile traffic or expressed as minutes of usage. This approach is based on the assumption that the two inputs (spectrum and BTS) can be substituted for each other over a range of output. An optimal mix will be used by the TSPs to produce the required traffic and this optimal mix is determined by input prices. A higher charge for spectrum will induce TSPs to substitute spectrum for the less expensive BTS to produce the same number of minutes, and vice versa. One way of estimating the value of 1800 MHz spectrum is to take a panel data set of minutes of traffic, spectrum allocated and BTS in different LSAs over certain period in the past and estimate the coefficients of the production function which can then be used to derive the value of spectrum across LSAs.

Production function approach provides a reasonable approximation to equivalent cost savings on BTS conserved by deploying an additional unit of spectrum.

ii)Revenue Surplus Model.

The value of spectrum could be estimated from the perspective of an access service provider willing to invest in spectrum to realize the net revenue potential/revenue surplus from the GSM segment across the span of license period of 20 years for acquiring 1800 MHz spectrum.

To determine the value per MHz of 1800 MHz spectrum, the NPV of revenue surplus of each LSA is divided by the total equivalent available spectrum in that LSA.

iii) Producer Surplus Approach.

Spectrum can also be valued on the basis of this approach. As there is an inverse relationship between the quantum of spectrum allocated and the expenditure on radio access network (RAN) required for serving a particular level of demand, the allocation of additional spectrum to an existing TSP will create a Producer Surplus. The model is a bottom-up approach to determine the opportunity of cost savings to an average Telecom Service Provider (TSP) upon expenditure in the Radio Access Network (RAN) and spectrum usage charge (SUC) during the next 20 years upon getting additional spectrum (opportunity/MHz). The opportunity of the net savings in expenditure made by the TSP has been termed as 'Producer Surplus'

To determine the value under this approach, assumptions and projections are abundantly used. By making an assumption on the growth of number of subscribers and having the information on the number of voice MOU, number of SMS and amount of data usage over a certain period, TRAI through a projection methodology, the producer surplus on account of additional spectrum has been calculated. Mathematically it is expressed as:

Present Value of (expenditure on BTS in urban area and SUC (spectrum usage charge) during the next 20 years without additional spectrum - expenditure on BTS in urban area and SUC (spectrum usage charge) during the next 20 years with 'x' additional spectrum in 1800 MHz)

iv) Discounted Cash Flow (DCF): -

This method uses the projection based on the NPV of cash flows for 20 years.DCF represents the maximum amount which the investor would like to pay for acquisition of an asset (in this case access to spectrum) for a period of 20 years- license period. In this method we compute the value of a block of spectrum by determining the Net Present Value (NPV) over the spectrum license period of next 20 years of the cash flow that a mature operator presently would command by virtue of holding the corresponding block of spectrum. In the case of contracted spectrum, the cash flow accruing from the possession of 6.2 MHz is equal to the revenue earned from subscribers less the costs: the sum of the license fees, the spectrum charges, administrative, marketing and personnel costs, and the cost of the physical network1, i.e.

***Cash Flow* = *Revenue* – (*License Fees* + *Spectrum Charge* + *Network Cost* +**

Administration, Marketing, & Personnel Cost)

v) Multiple Regression Approach

Linear regression establishes a relationship between a scalar dependent variable denoted as Y and one or more explanatory variables denoted as X. If only one explanatory variable is used, it is called simple linear regression; for more than one explanatory variable, it is called multiple linear regression.

The prices realized through last Auction in the LSAs can be correlated with other relevant variables to estimate the values of spectrum in the LSAs where spectrum was auctioned. The exercise can be done using multiple variable regressions.

vi) Last market determined price

Market revealed price (duly indexed) wherever available for respective band can serve as a benchmark price representing a lower bound while estimating the valuation of spectrum frequency bands such as 800, 900, 1800 and 2100 MHz. Market price revealed as an outcome of a competitive, transparent bidding process is best available value placed on the spectrum. Therefore, TRAI felt that the market revealed prices in previous auction(s) held in last two years can be considered for the reserve price estimation.

TRAI had taken a view that the valuation of a specific spectrum band in India cannot be compared with price realized from the auction of such band in other countries. Therefore, no indexing is done with regard to the international price of the spectrum due to completely different demographic and economic conditions prevailing in other countries.

While taking the previous auction price few points are to be noted. Using the previous auction price with indexing was supposed to bring the reserve price to a realistic price. But in Mumbai and Delhi the 3G price can be seen casting a shadow over such calculations starting with the 2012 auction onwards. Because of the high Reserve Price of Mumbai and Delhi, spectrum was not sold in 2012 and again in 2013 auction it remained unsold. Then the price was reduced by more than 50 % in the 2014 auction to be able to sell the spectrum. In 2015 there was no spectrum in 2015 in these Service areas. In the year 2016, the spectrum was sold in Delhi at the reserve price and in Mumbai, it fetched a higher price than fixed reserve price, but both Delhi and Mumbai never was able to match the 3G price of the year 2010 proving again that the 3G price of 2010 was in fact 'Irrational Exuberance' as TRAI has termed it.

It should also to be noted that TRAI has not used all the above methods uniformly in all auctions. In its wisdom, TRAI selectively applied the different methodologies described above in different auctions to bring out a probabilistic average value of spectrum.

vii) Rationale by TRAI for fixing the final recommended price not equal to the average valuation:

The higher frequency bands are used by the service providers mostly for capacity purpose and lower frequency bands are used mostly for coverage purpose. Therefore, an optimum combination of such higher and lower frequency bands is desirable to have an optimum mix of capacity and coverage. Therefore, the cost of spectrum has to be reasonable to provide the desired socio-economic benefits to all the sectors of the society.

The demand for spectrum as a natural resource is not a direct one like for most commodities. It is derived from the demand for final goods and services that are produced using spectrum as an input. There are many different users of spectrum supplying these final goods and services (e.g. telecom service providers (TSPs), broadcasters, aeronautical users, scientists, the military, etc.). In the case of TSPs, it is telecom consumers who, through their demand for telecommunication services, create a demand for spectrum. The greater the demand for telecom services, the greater will be the demand for spectrum by the TSP s. The demand for spectrum is a derived demand. Valuation of spectrum is determined to a large extent by its demand which, in turn, depends on the willingness and ability to pay of a large number of spectrum users or TSPs who use it as an input in the production of telecom services.

Since the demand for spectrum is a derived demand on the basis of the telecom demand, TRAI felt that fixing the average valuation as the reserve price is dangerous because at the recommendation stage it is not sure that this is the best starting point. If this price is fixed as the reserve price, it may end up fixing the reserve price on a higher side and auction may not kick off. Therefore, TRAI has been fixing the reserve price at 80 percent of the average valuation as the safe bet in all its recommendations.

b) The price of spectrum: -

*(**Refer: 2, 3, and 4)*

	Type of factor	Factor
Intrinsic		Propagation characteristics
		Sharing capacity
		Profusion of uses
		Global and regional harmonization
		International constraints
Extrinsic:	Physical factors	Geography Climate
	Socioeconomic factors	Demographics
		Population density
		Income distribution
		Economic growth rate
		Political stability
		Absence of corruption
		Rule of law
	Policy and Regulation	Favourable investment and customs laws
		Independent regulatory agency
		Competition policy
		Infrastructure sharing
		Rules of protection of the public against
		electromagnetic waves
		Open access rules
		Technology neutrality
		Limitation of and protection against interference
		Coverage obligations
		Spectrum caps
		Auction rules and bidding credits/set-asides
		Transparency
		Licensing framework
		Dispute-resolution mechanisms

Fig: A block diagram depicting the intrinsic and extrinsic factors affecting the spectrum price

(Reference: *****7)

The value of spectrum would rely to a large extent on the network infrastructure utilized which lies on the expenditure side and the traits of the particular market which in turn falls on the revenue side. In looking at both potential costs and revenues, these can be divided into two categories:

i) Intrinsic factors –

These are factors that pertain to the spectrum itself and cannot be changed by any particular government:

- factors stemming from laws of physics, or
- as a result of worldwide trends (e.g. harmonization), or
- international obligations (e.g. frequency allocations, bi-lateral or multi-lateral frequency coordination agreements).

Propagation characteristics, sharing capacity, Profusion of uses, Global and regional harmonization and International constraints are the examples of intrinsic factors that contribute to the cost of spectrum. For example, the value of frequency bands used for mobile access is very high compared to the price of frequency band used for point to point microwave links.

ii). Extrinsic factors –

These are factors that apply differently in each country, whether because of physical or demographic characteristics, historical, cultural or legal heritage or more pertinently, as a result of national government policies and regulations. Examples include Geography, Climate Demographics, Population density, Income distribution, Economic growth rate, Technology neutrality, protection against interference, Coverage obligations, Spectrum caps, Auction rules and bidding credits/set-asides, Transparency, Licensing framework, Dispute-resolution mechanisms. It is due to these extrinsic factors, that a universal formula to determine the price of spectrum is impossible to achieve. Each country will have a different polity, cultures, socio-economic factors that may constitute the local considerations which would influence the cost of natural resources including Spectrum. The extrinsic factors may prevent us from deducing the cost of spectrum in India from the spectrum selling history of other developed countries.

c) Econometric Approach to resolve the price of spectrum: -

*(***Reference: 4 and 5)*

Reserve Price:

A reservation (or reserve) price is a limit on the price of a good or a service. On the demand side, it is the highest price that a buyer is willing to pay; on the supply side, it is the lowest price at which a seller is willing to sell a good or service. Reservation prices

are commonly used in auctions, but the concept is extended beyond. (*Wikipedia*).
Fixing the reserve price of spectrum in an auction is an insurmountable task because as stated earlier spectrum is not directly invested as an end product, but the telecom service and its offshoots decide how much quantum and on which technology to induct the spectrum. The presence of intrinsic and extrinsic factors mentioned above make the pricing very complex and difficult task. Econometrics approach can come to help. This approach integrates mathematics, statistics, and economic theory together and applies this towards the estimation of spectrum value. Econometric modelling approach requires the real data from the past in order to conceive a mathematical model for showing relationships between dependent variables and various factors in the form of independent variables. Moreover, this is a technique that estimates both demand and supply variables, which impact the spectrum valuation. Supply variables, include number of spectrum slots to be auctioned per license, total number of spectra to be auctioned, duration of license, etc. As for demand variables, these include Gross Domestic Product (GDP), Per Capita GDP, number of mobile phone users, education level of people, and ratio of telecommunication industry revenue to total revenue. Besides, variables representing change in technology also required as they have direct impact on spectrum valuation in the long run.

All in all, Econometric approach assesses spectrum value by using various factors to create a model to estimate the spectrum value, which concentrates on spectrum grouping, spectrum license issuing process, conditions for spectrum fee payment, as well as the impact of domestic economy and market state of business on spectrum valuation. The econometric approach for spectrum value estimation is arranged into 5 following processes:

i) Model Specification – determines the relationships in a form of Linear Model or Nonlinear Model

ii) Data Collection – gathers past data for variables. Mostly in the form of panel data, this comprises of Time-Series Data and Cross-Sectional Data.

iii)Estimation – estimates coefficient value according to the specified model, which may use Ordinary Least Squares, Generalized Least Squares, Two-Stage Least Squares or other techniques.

iv) Evaluation of Estimated Model – assesses reliability of model and parameters achieved, which may use t-statistics, F-statistics, R-squared and Adjusted R-squared.
v) Forecasting – predicts the spectrum value by taking various factors (which impact on estimated value of the model) into consideration. Moreover, in considering at both potential costs and revenues, there are various factors that cannot be ignored.

d) OLS Regression

(**Reference 2,3)

Regression analysis is a technique for modelling data. In general, the goal of any statistical approach to modelling data is to take a sample of data that represents a population, then to use that sample to estimate some facet of the population. In the case of regression analysis, the focus is on the "conditional mean value" of a single dependent variable y corresponding to a given set of predictor (independent) variables $x_1, \ldots, x_p$. The conditional mean is a function that represents the mean of the dependent variable conditional on the dependence over the predictor variables.

There are many forms of regression analysis, and new approaches are being invented all the time. Here the focus is on what is perhaps the oldest but still most widely-used method of regression analysis, known as linear least squares. Linear least squares represent the conditional mean function as a linear function of the predictor variables. Suppose our dependent variable is y, and the predictors are x_1 and x_2. Then in linear least squares, we model the conditional mean function as $E[y \mid x_1; x_2] = b_0 + b_1x_1 + b_2x_2$: Here, the notation $E[y \mid x_1; x_2]$ is read "the conditional mean of y given x_1 and x_2" - "E" stands for "expectation" which is a synonym for "mean" (also equivalent to "average value"). The symbols b_0, b_1, and b_2 here are called the regression parameters. The regression parameters are numeric values that define the conditional mean function. They are not known, and must be estimated from the data.

i) Interpreting the regression parameters

The parameter $\mathbf{b_0}$ in a linear model is called the intercept, it is special because it is not multiplied by any predictor variable. The other parameters are often called slopes. The slopes tell us how much the average value of y differs when comparing individuals in the population that differ by one unit for a particular predictor variable, but are the same in terms of all other predictor variables. In many cases the intercept parameter is not very interpretable. The intercept is always equal to the conditional mean of the dependent variable when all of the predictor variables are equal to zero or at their mean.

An alternative way to express a regression model is in "generative form":

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + \epsilon$$

In the expression above, ϵ represents unexplained variation (also called "random variation", "error", or "noise"). The presence of the error term ϵ is necessary since the data we observe in the real world will never follow a linear relationship exactly. These "error terms" are used to represent the part of the data that cannot be explained using the predictor variables.

ii) **Linearity**

The role of linearity in a regression analysis is subtler than it appears on the surface. The conditional mean model used in a linear least square analysis is linear in the sense that the regression parameters are linear. The fact that the conditional mean function is linear with the regression parameters (b_0, b_1, ...) is that what allows us to use linear least squares to estimate these parameters from the data. It is only important that the conditional mean function must be a linear function of its regression parameters if we plan to use linear least squares for estimation.

iii) **Variation**

Almost of equal importance, is the conditional variance function, which is denoted as Var [y| x1...... xp]. The conditional variance function quantifies the degree of scatter in the data around the conditional mean function just like the ordinary (unconditional) variance quantifies the variance of the data around an ordinary (unconditional) mean. The conditional variance function, like the conditional mean function, depends on its arguments (x1, x2, ...). However, in some populations, the conditional variance turns out to be nearly constant in these arguments. That is, the scatter around the conditional mean has similar magnitude regardless of the values of the predictor variables. This property, known as homoscedasticity, does not need to hold in order to conduct a regression analysis. But the most common and basic methods for conducting regression analysis work best when the conditional variance function is approximately constant in this sense.

iv) **Causality**

Since the conditional mean is viewed as being a function with the predictor variables as inputs and the dependent variable as the output, it is tempting to think of this as reflecting a mechanism in which changes in the predictor variables can cause changes in the dependent variable to happen. In general, however, statistical analysis does not support such causal interpretations. When describing the regression model with conditional mean function, say, E [y| x1; x2] = 1 + 3 x1 − 2 x2, it is usually better to avoid saying something like "when x1 goes up by 1 unit and x2 is held fixed, y goes up by 3 units on average." Instead, it is usually better to say "when comparing two individuals whose x1 values differ by 1 unit, and whose x2 values are the same, the value of y will differ on average by 3 units." Similarly, instead of saying something like "x1 affects y" it is better to say that 'x1 is associated with y, after controlling for x2".

v) **Estimation**

Any statistical model must be fit to the available data, a process often referred to as parameter estimation. If we model the population conditional mean as

$$E[y|x1; x2] = b0 + b1x1 + b2x2,$$

then the estimated conditional mean function will be written as

$$E1 = \hat{b}0 + \hat{b}1\, x1 + \hat{b}2\, x2.$$

Here $\hat{b}0$ $\hat{b}1$ and $\hat{b}2$ are estimated parameters, which will rarely be exactly equal to their population counterparts (e.g. $\hat{b}1$ will differ from b1). We aim to recover these parameters as accurately as possible from the available data. As noted above, this approach has successfully been used to fit linear models for well over 100 years. It is remarkable that a modern computer can easily fit very large models with linear least squares - for example, a data set with 1 million rows (cases) and 20 predictor variables can be fit in well under three seconds. Linear least squares work best when the conditional variance is constant. However, the estimates of the model parameters produced by linear least squares generally remain accurate when the conditional variance is not constant. More advanced approaches to regression analysis address more completely the challenges of working with non-constant variance (heteroscedasticity).

The field of statistics focuses on estimation, but also places great importance on “quantifying uncertainty" - that is, characterizing the likely degree of discrepancy between the parameter estimates and their corresponding population values. First, any statistical estimator will exhibit some combination of bias and estimation variance. Linear least squares is usually unbiased, which is one of its favourable attributes. However even in linear least squares, bias can sometimes result due to mis-representativeness of the data relative to the population of interest, or to systematic measurement errors in the data. Nevertheless, in basic usage bias is normally not a major concern with linear regression analysis.

Estimation variance is inevitable in any statistical analysis. It reflects the fact that we can never recover a population exactly using a finite amount of data. The estimation variance is determined predominantly by the sample size - the more data we have, the lower the estimation variance will be. Estimation variance in a regression model is also strongly influenced by three other characteristics - the level of conditional variance, the variance of the predictor variables, and the correlations among the predictor variables.

The conditional variance is the “scatter" in the data around its conditional mean. This type of variance is “bad variance" in the sense that greater conditional variance results in greater uncertainty about the regression parameters.

The variance of the predictor variables refers to how dispersed the values of the different predictor variables are within the data set being used to fit the model. The primary purpose of a regression model is to establish how differences in the values of a predictor variable relate to differences in the expected values of the outcome variable. The predictor variables is “good variance," since greater variance in the predictor variables result in less uncertainty about the regression parameters.

Correlation among the predictor variables is referred to as collinearity, and plays an important role in regression analysis. Note that we are not saying here that predictor variables must be perfectly uncorrelated (i.e. that they must have zero correlation). We are only saying that greater correlation among the predictor variables tends to result in greater uncertainty in the parameter estimation. This can be overcome by collecting more data, but for a fixed amount of data, our estimates of the regression parameters will be less precise when substantial collinearity is present. We can view correlations between predictor variables as” bad correlations", because they adversely impact our ability to fit a regression model. Conversely, correlations between predictor variables and the dependent variable are “good correlations", because they allow us to fit models that do a better job of explaining the variation in the dependent variable.

vi) **Explained variation**

One way to view a regression analysis is as an effort to “explain the variation" in the dependent variable, using the predictors as explanatory factors. This allows us to establish a link between linear regression analysis and the more basic idea of Pearson correlation (i.e. the familiar correlation coefficient). Suppose we have fit a linear model to data using linear least squares, and have thereby obtained parameter estimates $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$.... We can use these parameter estimates to produce fitted values. To form the fitted value for a particular observation with covariate values x_1, x_2.... we form the linear combination $\hat{b}_0 + \hat{b}_1 x_1 + \ldots$. This expression follows the form of the population conditional mean function, substituting the parameter estimates for the true parameter values (which are not known), and substituting the predictor variable data for one specific case for the arguments of the conditional mean function.

Once we have the fitted values in-hand, we can take the Pearson correlation coefficient between these fitted values and the observed values of the dependent variable (y). The fitted values are intended to track with the dependent variable. The closer they do so, the better the apparent explanatory performance of the model. The squared Pearson correlation coefficient between the fitted values and the observed value of the dependent variable is called the R-squared, the “proportion of explained variance," or the “coefficient of determination." The R-squared falls between 0 and 1. In general a higher R-squared is seen as reflecting a better fit of the model, but this interpretation should be qualified in two ways: first, “goodness of fit" refers to more than just the mean function - to have a model that fits well, we would like to capture the variance structure as well as the mean structure; second, higher R-squared can reflect “overfitting," in which the model fits the data in-hand better than it will fit equivalent data that we observe in the future.

vii) **Statistical inference**

The tests of standard errors, confidence intervals, and hypothesis are three ways to quantify the accuracy of estimated population parameters. All three of these concepts can also be used in a linear regression analysis. For each estimated regression parameter, say $\hat{b}_1$, we have a standard error s_1. Roughly speaking, the value of s_1 is the average discrepancy between $\hat{b}_1$ and its population value, which is b_1. Each regression parameter estimate will have its own standard error, reflecting the unique level of information about each parameter in the data. As noted above, several factors influence the uncertainty in a parameter estimate, including primarily:

(1) sample size,

(2) conditional variance in the dependent variable,

(3) variance of the predictor variable, and

(4) collinearity.

Note that factors 1-2 impact all parameters equally, while factors 3-4 impact different parameters in a model to different extents. Once the parameter estimates are in-hand, we can construct 95% confidence intervals for each regression parameter. Assume the regression parameter was estimated to be a, and the 95% confidence interval estimated to be have a range of 'd', then any value between 'a-d' and 'a+d' would be consistent with the observed data.

5. Auction outcome of the four auctions held during the period 2012-2016:

We will go through each of the auction of 1800 MHz spectrum in the various auctions held in years 2012, 2014, 2015 and 2016 to understand how the reserve price of TRAI was fixed and how much it differed from the reserve price of Govt and the Winning Price. The 2013 auction draw blank without any bid for 1800 MHz. It is generally considered as the extension of 2012 auction and so this will not be considered in this study.

(i) 23rd April 2012 recommendation (auction held in November 2012 for 1800 MHz):

None of the methods discussed in section 4 above were used to calculate the reserve price of spectrum in 2012. Instead the 3G auction determined price was used to derive the price of 1800 MHz reserve price on the criteria of the technical efficiency factor. TRAI has taken the 3G auction (2100 MHz) price of May 2010 as the base price to fix the reserve price for 1800 MHz. TRAI stated "the Authority is convinced that for the determination of the reserve price for spectrum in the 1800 MHz band, the reference price should be the price discovered in the year 2010 for 3G spectrum and not the price discovered in the year 2001 for the 4th Cellular License." TRAI also decided that the 1800

MHz band is 1.2 times more efficient than 2100 MHz band for the purpose of calculating the reserve price in this band. Since the 3G auction was held two years prior to the 2012 auction, the price has to be indexed using SBI average PLR rate @ 12.63%.

TRAI realised while fixing the reserve price that it shall be risky and detrimental if the reserve price would exceed the market price, which remains as an unknown entity until the auction is actually held. But TRAI deliberated that it is a good case for setting the minimum price as high as uncertainty about market value allows. That means picking a minimum price that is as high as possible, but the risk of exceeding market value being acceptably low. TRAI accordingly decided to use a factor of 0.8 times of the fixed prices to reach the reserve price. The reserve price of 1800 MHz spectrum is arrived from the 3G price as:

[apply SBI PLR of 12.63% on 3G price of 2010 for one-year gap] X
1.2 (efficiency factor for 1800 MHz over 2100 MHz) X
0.8 as the factor for keeping the reserve price below the expected market price.

In order to justify that the reserve price so calculated will not force the lowering of competition and the operators does not raise the tariffs, TRAI did some analyses projecting the reserve price in a 20 years' window from 2012-13 to 2032-33 using all the available spectrum in 1800 MHz as on that date. In this analysis, while considering the annualized EMI/ Minutes of Use (MoU) at 15% interest showed that the factor EMI/MoU gradually coming down in the calculation thus proving its affordability. Similar EMI analysis was done for a single operator purchasing the 5MHz spectrum in 1800 MHz at the fixed price. In this case also, TRAI concluded that the results are similar and according to TRAI, affordability was ensured at the price fixed by the above formula.

The auction was held in November 2012 and the table below shows the valuation conducted by TRAI while arriving at the reserve price. The last two columns in the table shows the Govt approved Reserve Price and the winning price of the auction for each of the service area. The auction lasted only 2 days with 4 unsold LSA s and spectrum selling at higher than reserve price only in one LSA s.

LSA	TRAI fixed Reserve Price.	Govt Fixed Reserve price	Winning price
Andhra Pradesh	296.93	229.53	229.53
Assam	8.97	6.94	6.94
Bihar	43.99	34.01	37.14
Delhi	717.26	554.45	**Not Sold.**
Gujarat	232.69	179.87	179.87
Haryana	48.14	37.22	37.22
Himachal Pradesh	8.05	6.22	6.22
Jammu & Kashmir	6.55	5.06	5.06
Karnataka	341.64	264.1	**Not Sold.**
Kerala	67.58	52.24	52.24
Kolkata	117.69	90.98	90.98
Madhya Pradesh	55.87	43.19	43.19
Maharashtra	271.99	210.25	210.25
Mumbai	702.14	542.76	**Not Sold.**
North East	9.15	7.07	7.07
Orissa	20.98	16.22	16.22
Punjab	69.63	53.82	53.82
Rajasthan	69.42	53.66	**Not Sold.**
Tamil Nadu	316.78	244.87	244.87
Uttar Pradesh (E)	78.83	60.94	60.94
Uttar Pradesh (W)	111.16	85.93	85.93
West Bengal	26.74	20.67	20.67

(ii) Recommendations dated 9th September 2013 for 1800 MHz band (Auction held in Feb.2014 for 1800 MHz and 900 MHz): -

TRAI has done an analysis in 2013 whether the price of 1800 MHz could be derived from the price of 3G auction in 2100 MHz of May 2010. TRAI has noted the following:

Overall market conditions (both economic and financial) have materially altered during the period in question viz. from May 2010 till 2013. These changes can be seen in (a) the deteriorating financial performance and overall financial position of the sector, (b) the general slowdown in the economy and other macro-economic developments and (c) expectations of the future which have altered radically.

The falling trend of profitability- Earnings Before Interest Tax Depreciation and

Amortization (EBITDA), Profit Before Interest and Tax (PBIT) and the rising trend of debt (long term) of telecom access service companies point towards the overall weakening financial health of the TSPs. The share of stressed assets in respect of loans to telecom companies on the books of the banks has increased from 1.3% in March 2011 to 15.64% in March 2013.Moreover, the fall in the number of subscribers, minutes of usage per subscriber per month and the declining average revenue per user (ARPU) has adversely impacted the growth of revenue of TSPs.

In addition to the sector-based measures of growth, the overall economic slowdown has also impacted market conditions in 2013. The GDP growth rate has declined from 8.5% in 2010-11 to 4.8% at the end of 2012-2013, declining in each successive quarter since March 2011.

To assess the value of spectrum, various approaches have been adopted rather than selecting one particular methodology of valuation, as it is simply not possible to say deterministically that any one valuation is the 'right' valuation. Each model has certain strengths as well as limitations. Where some models better capture intrinsic technical features, others are more strongly grounded in economic and market realities. No one model completely captures every variable- technical, economic, sectoral, geographic and regulatory- that influences the valuation of spectrum. These recommendations, therefore, present a reasonable valuation obtained from an appraisal of the results of different models, which, to the best of the Authority's belief, has a high probability of realization in the actual world.

Taking into consideration of the economy and overall deteriorating condition of the telecom industry, TRAI desisted drawing the price of spectrum from the 3G auction price. Further concerned with the various possibilities that the Reserve price may vary wide off from the real value of the spectrum TRAI has decided that a fresh valuation to be done using multiple methods and that the reserve price for the auction should be fixed at 80% of the average valuation (simple mean) of all such methods.

In the 1800 MHz band, spectrum in LSAs of Delhi, Karnataka, Mumbai and Rajasthan failed to find any bidder in the auction held in November 2012 and March 2013. TRAI has done multiple regression and simple linear regression by correlating the auction determined price of 18 LSAs to compute the reserve price of 1800 MHz spectrum in the in these 4 LSA s. In addition to the regression method, two other methods namely Producer surplus and productions function method using the opportunity cost principle were also used by TRAI to assess the reserve price.

A fourth method which is a revenue surplus method known as Discounted Cash Flow (DCF) is also used to value the spectrum. In this method the net present value (NPV) was arrived at using DCF methodology. This NPV represents the maximum amount which the investor would like to pay for acquisition of an asset (in this case access to spectrum for a period of 20 years).

The auction was held in February 2014 and the table below shows the valuation conducted by TRAI to arrive at the reserve price. The table shows the Govt approved Reserve Price and the winning price of the auction for each service area.

Service Area	TRAI fixed Reserve Price	Govt Fixed Reserve price	Winning price (In Rs. Crores)
Andhra Pradesh	130	163	163
Assam	7	7	36.1
Bihar	37	37	43.1
Delhi	175	219	364
Gujarat	115	143	237.8
Haryana	27	27	27
H.P.	6	6	6
J & K	5	5	6.1
Karnataka	124	155	155
Kerala	52	52	52
Kolkata	59	73	73
M.P.	43	43	50.4
Maharashtra	138	173	290.35
Mumbai	165	207	272
North East	7	7	7
Orissa	16	16	16
Punjab	54	54	54
Rajasthan	26	26	26
T.N.	166	208	208
U.P.(E)	61	61	64
U.P.(W)	62	62	94.95
W.B.	21	21	24.6

The auction lasted for 11 days and the spectrum was sold in all the service areas.

(iii) Recommendation of Oct.2014 (Auction in March 2015 for 1800 MHz): -

For the estimation of reserve price, TRAI dropped the single /multiple variable regression method citing the limited number of data points (12) whereas it may be recalled that in the 2013 recommendations, TRAI conducted the multiple variable regression with 14 data points. The DCF method of the expert's panel of 2011 also not used for the calculation as the license conditions have changed since 2011 but also the same principle is being used in the Revenue Surplus model presented here. The auction determined price of 2014 auction also be considered as one of the possible candidates for estimation of the Reserve price. Three other methods used to calculate the reserve price are the following: -

i)Producer Surplus,

ii)Production Function method

iii)Revenue Surplus method employing DCF

TRAI considers the 2014 auction as the most successful auction as major part of the spectrum (about 82%) was sold and the highest till that date. Therefore, TRAI retained the same methodologies and process to calculate the reserve price for the 2015 auction. The mean value of the auction price and the valuation of each of the three methods (i.e. Producer Surplus, Production function and Revenue surplus) completed the valuation of the 1800 MHz spectrum. TRAI considered the price obtained in the 2014 auction as an authentic representation of the price of spectrum. Therefore, the higher of the two figures- Feb.2014 auction price and 80 percent of the mean value obtained in the estimation- is the reserve price that was fixed for the 1800 MHz spectrum. Unlike the 2013 recommendations, where the lower value of the two figures was used, here the higher of the two figures was taken as the reserve price of Spectrum.

The auction was held in March 2015. The table shows the 4 estimates done by TRAI along with the Govt approved Reserve Price and the winning price of the auction for each service area.

It may be noted that in the auction reserve price for North East has been reduced by 50% and for Rajasthan it is discounted by 30%. Spectrum availability in West Bengal and Maharashtra is limited that auction is not recommended in these LSAs.

Service Area	TRAI fixed Reserve Price.	Govt Fixed Reserve price	Winning Price
Andhra Pradesh	163	169	242.8
Assam	36.1	**Not fixed**	**No auction**
Bihar	61.51	61	**No Bid**
Delhi	364	**Not fixed**	**No auction**
Gujarat	237.8	238	238
Haryana	31.87	32	46.6
H.P.	9.48	9	15.9
J & K	24.63	**Not fixed**	**No auction**
Karnataka	155	185	185
Kerala	75.09	75	83.45
Kolkata	73	73	149.1
M.P.	68.83	**Not fixed**	**No auction**
Maharashtra	290.4	**Not fixed**	**No auction**
Mumbai	272	**Not fixed**	**No auction**
North East	21.06	11	11
Orissa	23.37	23	33.1
Punjab	70.7	71	71
Rajasthan	85.89	60	72.75
T.N.	208	225	225
U.P.(E)	97.32	97	106.95
U.P.(W)	94.95	95	95.95
W.B.	35.12	**Not fixed**	**No auction**

The spectrum was either limited in quantity or not available for auction in seven LSA s. Therefore, there was no auction for LSA s Maharashtra, West Bengal, Assam, J&K, Madhya Pradesh, Mumbai and Delhi.

The auction lasted for 22 days.

(iv) Recommendations of January 2016 (Auction in 2016 for 1800 MHz and other bands):

If the result of 2015 auction is examined, it can be seen that only in 11 LSA s the Auction price (winning price) exceeded the Reserve price. This number is insufficient to do the regression analysis. Therefore, the regression method was not used in the valuation exercise. The following methods were used to do the valuation:

1. Producer surplus;
2. Market determined price of 2015 or prior value indexed properly;
3. Production function approach;
4. Revenue surplus model;

A simple mean of the above 4 methods was taken to arrive at an average valuation.

Then the reserve price is chosen with the help of the following criteria;

(i) should be higher of the two figures – 80% of the average valuation of spectrum band in the LSA or the price realised in the March 2015 auction /February 2014 (duly indexed with SBI base rate) auction;

(ii) in LSAs where no spectrum was offered in March 2015 and February 2014 auctions, reserve price should be 80% of average valuation; and

(iii) in LSAs where spectrum was offered in March 2015 auction but remained entirely unsold, the reserve price should be lower of the figures – 80% of average valuation or the reserve price as fixed in March 2015 auction.

	Recommended Reserve Price (Rounded off)	Govt Fixed Price.	Winning price
Andhra Pradesh	243	243	243
Assam	40	40	40
Bihar	62	62	62
Delhi	399	399	399
Gujarat	238	238	238
Haryana	47	47	49.3
H.P.	16	16	16
J & K	13	13	13
Karnataka	185	185	No Bid
Kerala	83	83	83
Kolkata	149	149	151.15
M.P.	83	83	83
Maharashtra	318	318	318
Mumbai	298	298	489.2
North East	11	11	11
Orissa	38	38	No Bid
Punjab	77	77	77
Rajasthan	91	91	91.9
T.N.	225	**Not fixed**	**No auction**
U.P.(E)	115	115	133.15
U.P.(W)	96	96	100.3
W.B.	46	46	46

The auction lasted for only 6 days.

(v) Compilation of the estimation methods used in the last four auctions: -

Different methods were employed by TRAI in its recommendations to determine the reserve price of spectrum in the auctions held in 2012,2014,2015 and 2016. Now these methods, which is covered earlier, could be consolidated on the basis of year of auction as:

1. Production Function method: These methods were used in the recommendations of 2014 Auction, 2015 Auction and 2016 Auction.
2. Revenue Surplus method: - Used only in 2015 Auction and 2016 Auction.
3. Producer Surplus method: - Used only in 2014 Auction and 2015 Auction.
4. Multiple Regression: - This method was used in the recommendation for 2014

Auction for finding the price of 4 service areas where spectrum could not be sold in the previous auction. (i.e. 2012 Auction). The regressor used are GSDP per Capita, Residual Teledensity, Population and AGR per population.

5. Discounted Cash Flow: - It has similarities to the method of Revenue Surplus where the NPV is used to calculate the cash flow from the next 20 years of license period. This method per se was not used by TRAI in any of the four auctions. But the 2011 expert committee report calculate the value of spectrum using this method.

6. Model and the variables in this study: -

In this study we identify the variables that would possibly can be used to predict the price of spectrum. The estimated value of spectrum was determined by TRAI through a consultative cum participatory process. In these methods, TRAI has used different mechanisms like projection of Cash flow, NPV based on the traffic and other telecom parameters, Marginal price with regard to reduction in the RAN (Radio Access Network) when one unit of spectrum added, Substitution of spectrum with BTS as used in the Cobb-Douglas equation etc. While developing the formula for the price, TRAI also taken into consideration the various telecom and non-telecom pointers that would possible influence the auction price. Hence we may assume the final price estimate as derived by TRAI using these mechanisms reflect the assimilation of socio, economic and telecom factors.

Two methods are selected here for using econometric modelling - the linear regression method OLS using the parametric modelling and as a supplementary method using Machine Learning. Both these methods are on the premise of predicting the future price with past history of auction data.

i)e-Auction

As was mentioned earlier, the objective of capturing the Auction price of 1800 MHz with the help of data in 4 auctions is to be realized through the statistical method of linear regression. In this paper the Ordinary Least Square method is used to identify and establish a linear association of the Dependent variable namely Auction Price with the right input variables. The market price of spectrum is the final price or winning price obtained in the e-auction, an online process called SMRA (Simultaneous Multiple Round Auction) conducted for each of the 22 Service areas as per the following process.

Each auction has two stages- clock round stage and frequency identification stage. In the clock stage, the price is determined for the block size in the particular band in which the bid take place for each service areas. This is a generic stage in which every slot is auctioned treating them as equivalents without frequency location information. Bidders will bid for a block (i.e. right to a single spectrum block not linked to any specific frequency). The Clock Stage will consist of a number of rounds (the “Clock Rounds”). These rounds will stop once (i) for every service area where spectrum is being auctioned, the number of Bids at the prices set in the last completed Clock Round is less than or equal to the number of blocks available; and (ii) there are no opportunities for Bidders to increase their demand in accordance with the Activity Rules. The Clock Stage will establish a common Winning Price for all blocks within a service area and who are the Winning Bidders in each service area. In India SMRA method was used successfully. (Please refer to another paper by this author "A short analysis of Spectrum Auction in India" *[10] to see the mapping of clock stage in the auction).

The starting point of any auction is the Reserve Price on which the bidders put their bids on. If the Reserve Price is wide off the mark on the higher side, the auction will fail to fire. If it is on the lower side, there is a risk of collusion among the bidders. Finding the right reserve price is critical so that it will prompt the bidders to have enough cushion to enable them to put the bid on and on. In the auction in certain service areas, it can be seen that due to lack of demand or the reserve price being on the higher side, the auction prematurely ends in the first clock cycle itself and the reserve price becomes the auction price.

ii)Auction Price is a function of reserve price and the evaluated price:

From the foregoing argument it may be logical to assume that the Auction Price would be a function of the Reserve Price or the Reserve Price itself, if there is no great demand in an LSA. This Reserve Price was deduced from the valuation process conducted by TRAI using different analytical methods after taking a mean of all these methods. The mean, which TRAI calls the probabilistic average, so arrived at is either depreciated or discounted to fix the Reserve Price. By extending this logic, it can be argued that these methods used by TRAI decided the Reserve Price and that finally determined the Auction Price. In other words, the various prices determined in those analytical methods are going to decide the Auction Price. That is, we can assume that the Auction Price is mathematically related to the independently calculated prices from the different methods of TRAI. Therefore, the Price obtained thru Production function, Producer Surplus, Revenue Surplus/DCF, Previous winning price etc. are all determinants or Regressors of the final Auction Price, which will be Dependent variable.

iii)Ordinary Least Square Equation

Having examined the role of these prices on the Auction Price, the question is whether all these variables have influence on the Auction Price. It is not possible to predict this at this stage. The question of their significance can be found only after the statistical regression is run and the model is fitted.

The Auction Price, AP can be mathematically expressed as

$AP(i)=a.X1(i)+b.X2(i)+c.X3(i)+ d.X4(i)+e.X5(i)$

Where (i) represents the 22 Service areas

X_1=Production Function based price.

X_2=Producer Surplus Price

X_3=DCF Price/Expert Price/Revenue Surplus Price (all these methods are similar)

X_4=Previous Auction based Price

X_5=Telecom Commission Fixed Price.

As shown above, X_3 represents the three methods as they have common principle linking all of them. Therefore, for simplification we will refer X_3 as DCF method only. The 'a' to 'e' are the parameters that will be determined thru the OLS regression.

It is our assumption that the relationship between the input and output variables is linear so that the OLS can be applied. The linear assumption mandates the parameters of our OLS model for the Auction Price to be linear. Apart from the TRAI estimators, the Auction Price can also be related to other inputs like demographic variables, GDP of the State, telecom indicators etc. in a linear. Some of the potential regressors that could influence the Auction Price we can think of:

1. GSDP
2. Population
3. AGR
4. GINI Index
5. HHI Index.
6. ARPU (Rs. /Subscriber/month)
7. MOU (Minute of Use)
8. No of Mobile subscribers. (In Millions)
9. Tele density (per 100 people)
10. No of participants in the auction
11. No of activity rounds in auction.
12. Some sort of spectrum scarcity index in a LSA.
13. Freq. Band weightage factor in a LSA.

The influence of predictors like AGR, GINI Index and HHI index and technical figures like Spectrum Scarcity index, Band weightage factor etc. are left out from the study for future consideration. The number of activity rounds (clock rounds) shows how much vibrant the auction had been. It is directly related to the demand for spectrum. However, it has been found that there is no Service areas wise data, it may not be useful as it is. The number of participants are few in numbers in all the auctions varying from 5 to 8 same as the no of days' auction was held. We have four auctions and there are only four number of figures which is unlikely to influence the LSA wise auction price and therefore dropped.

iv)The final list of the variables: -

The following is the final list of explanatory variables and the response variable used in the study:

a) Response Variable:

Auction Price (AP) in Rs. Crores per MHz

b) Explanatory Variables:

1. The price of spectrum calculated by Production function method in Rs. Crores per MHz
2. The price of spectrum calculated by Producer Surplus method in Rs. Crores per MHz
3. The price of spectrum calculated by DCF method in Rs. Crores per MHz
4. The price of spectrum calculated based on the previous auction price method in Rs. Crores per MHz
5. The price of spectrum fixed as the final Reserve Price in Rs. Crores per MHz
6. GSDP per capita in Lakhs of Rupees per population.
7. Population of each LSA.
8. ARPU expressed in Rs. /Subscriber/month
9. Minutes of Use (MOU) expressing the mobile telephone traffic in Minutes.
10. No of Mobile subscribers expressed in Millions.
11. Tele density expressed as the number of mobile connections per 100 people.

Please note that the above variables are collected for each of the 22 Licensed Telecom Service Areas. The independent variables are quantitative (continuous) and the response variable is also quantitative.

7. Where the data came from: -

Having finalised the input variables and response variables, the next step is finding the data for the collection. The auction estimates are collected from the TRAI recommendations in their website. The telecom indicators from the DoT Telecom annual reports are collected for the four years data. For the demographic data and macro-economic data, websites of Niti Ayog, RBI and Other State Government websites were explored and collected. The details of the data collected from the various sources are elaborated below:

a) TRAI data from its different valuation exercises: -

The data that is extracted from TRAI website was subjected to some pre-processing steps as described below:

i)TRAI Recommendations for 2012 auction:

As was noted, the TRAI fixed the reserve price on the basis of 3G price by applying some transformations on it. There were no production Function, Producer surplus and Revenue Surplus as none of these methods were used in the estimation of Reserve price. Therefore, these are missing values in the 2012 auction. For this paper these columns are to be furnished with the values to proceed further. To work around this missing value problem, data by comparison is used with reference to the Expert Committee report and to substitute the missing values with those values. The following scheme has been used to capture data from the Expert Committee in order to fit into the 2012 auction table:

1. The Expert Committee has used DCF method to calculate both the contracted spectrum (up to 6.2 MHz) price as well as incremental spectrum (beyond 6.2 MHz) price whereas a second method, Production function method, is used to calculate the price of only incremental spectrum.

2. Therefore, for Metro and Category A and B LSA s, the price of incremental spectrum assessed by the Expert group using the second method can go to the Production Function column of 2012 data. The category C LSA values in the Production function is adapted from 'Final price' table for incremental spectrum in the Expert group report.

3. For Metro and Category, A and B LSA s, the price of contracted spectrum assessed by the Expert group using the first method can go to the DCF column of 2012 data table. The category C LSA values in the DCF column is taken from 'Final price' table for contracted spectrum in the Expert group report.

4. Now we are left with the Column of Producer Surplus. If it is not filled up, there will be missing data problem which can affect the regression seriously. Data that was not used yet, in the above steps, from the Expert group report is the calculation done for the incremental spectrum by the first methodology. Therefore, these incremental spectrum charges from Expert report can fill up the Metro Class A and B LSA s of the Producer Surplus column. As for the Class C LSA s, there are no more estimated values left in the report. Therefore, those values to fill the DCF column in the data is repeated in the Producer Surplus column. Though debatable, but for the wholeness of data for the year 2012, it may be an adjustment (for the six LSA) rather than dropping these columns as missing values.

5. In addition to the three methods elaborated above, the previous auction price value is also inserted which TRAI found out by moderating the 3G value with relevant indexing.

6. Here is the fully transformed records of the different columns for the year 2012 that would be used in the regression:

LSA	Producer surplus (Rs. Crore)/MHz	Production Function.(Rs. Crore)/MHz	Expert Price-11/Revenue Surplus/(DCF)(Rs. Crore)/MHz	Previous AP.(Rs. Crore)/MHz	80% valuation. (Rs. Crore)/MHz	TRAI (Rs. Crore)/ MHz
Andhra Pradesh 2012	391.98	471.93	153.77	309.31	296.93	296.93
Assam 2012	10.4	31.33	10.4	11.208	8.97	8.97
Bihar 2012	51.04	153.69	51.04	54.996	43.99	43.99
Delhi 2012	229.63	269.83	149.78	896.604	717.26	717.26
Gujarat 2012	287.84	422.91	149.87	290.868	232.69	232.69
Haryana 2012	59.12	156.68	14.5	60.168	48.14	48.14
HP 2012	9.34	28.12	9.34	10.068	8.05	8.05
J&K 2012	7.6	22.89	7.6	8.196	6.55	6.55
Karnataka 2012	252.2	439.65	136.16	427.068	341.64	341.64
Kerala 2012	199.89	264.43	73.98	84.468	67.58	67.58
Kolkata 2012	43.48	51.71	49.48	147.12	117.69	117.69
MP 2012	203.72	305.18	87.71	69.84	55.87	55.87
Maharash. 2012	302.08	446.85	117.14	340.008	271.99	271.99
Mumbai 2012	139.83	174.85	101.11	877.716	702.14	702.14
North East 2012	10.61	31.95	10.61	11.436	9.15	9.15
Orissa 2012	24.33	73.26	24.33	26.22	20.98	20.98
Punjab 2012	154.24	206.88	72.86	87.048	69.63	69.63
Rajasthan 2012	231.22	326.46	106.03	86.784	69.42	69.42
TN 2012	361.61	490.49	187.38	395.988	316.78	316.78
UP(E) 2012	299.58	337.95	151.76	98.544	78.83	78.83
UP(W) 2012	168.27	336.84	60.11	138.948	111.16	111.16
WB 2012	161.26	272.65	44.79	33.42	26.74	26.74

ii) TRAI Recommendations for 2014 auction: -

In this auction, we have the following categories of estimations for Spectrum Price:

1. The Production Function and Production Surplus methods are used by TRAI to independently evaluate the Spectrum Price.
2. The 2011 Experts Price by the DCF method is also used with the price of 1800 MHz as of year 2011 indexing for the year 2013.

3. There is one more method used by TRAI which is the multiple regression for calculating only the price of 4 Service Areas based on the price of 2012 auction for other areas. The values so derived are used to fill the missing 4 values in the last auction value column.

4. There is no pre-processing required here. There are no missing values also. The data is directly tabulated

Service Area	Producer Surplus	Production function	DCF	Realized price per MHz of 1800 MHz spectrum (2012)	Mean	80% of average valuation per MHz	TRAI fixed Reserve Price=lower of (80% of the Mean, 2012 Auction price)
Andhra Pradesh	101.59	129.75	227.1	229.53	162.62	130.1	130
Assam	12.25	10.73	15.36	6.94	11.32	9.06	7
Bihar	88.59	52.59	75.38	37.14	63.42	50.74	37
Delhi	170.92	251.85	221.2	388.11	218.9	175.12	175
Gujarat	91.07	101.07	221.34	179.87	143.39	114.71	115
Haryana	42.95	32.27	21.41	37.22	33.46	26.77	27
H.P.	7.33	9.63	13.79	6.22	9.24	7.39	6
J & K	41.62	7.83	11.22	5.06	16.43	13.14	5
Karnataka	157.97	118.9	201.09	184.86	155.21	124.17	124
Kerala	71.59	55.8	109.26	52.24	72.22	57.78	52
Kolkata	47.88	39.67	73.08	90.98	73.13	58.5	59
M.P.	74.66	78.69	129.54	43.19	81.52	65.22	43
Maharashtra	170.74	137.64	173	210.25	172.91	138.33	138
Mumbai	109.51	238.42	149.33	379.93	206.74	165.39	165
North East	27.14	10.93	15.67	7.07	15.2	12.16	7
Orissa	19.57	25.07	35.93	16.22	24.2	19.36	16
Punjab	87.12	45.77	107.6	53.82	73.58	58.86	54
Rajasthan	118.9	66.29	156.59	58.17	76.63	61.3	26
T.N.	276.74	111.95	276.73	244.87	207.89	166.31	166
U.P. (E)	125.92	83.7	224.13	60.94	123.67	98.94	61
U.P. (W)	71.75	64.4	88.77	85.93	77.71	62.17	62
W.B.	21.34	53.84	66.15	20.67	40.5	32.4	21

iii) Recommendations for 2015 auction: -

In this auction, we have the following categories of estimations for Spectrum Price:

1. The Production Function method

2. The Producer Surplus method.
3. The DCF method.
4. The spectrum price in the last auction of the year 2014.
5. There are no missing data as the estimates from TRAI recommendation is directly usable. Here is the tabulation for the year 2015.

	Producer Surplus	Production Function	DCF	Average (mean) Value.	Achieved Price-2014 auction	Higher of (mean or the price realized in 2014 auction).	TRAI Reserve Price.
Andhra P.	123.38	164.12	226.78	169.32	163	163	163
Assam	17.42	12.85	64.13	32.62	36.1	36.1	36.1
Bihar	103.76	62.97	97.72	76.89	43.1	61.51	61.51
Delhi	247.61	349.34	213.03	293.49	364	364	364
Gujarat	154.46	133.75	132.25	164.57	237.8	237.8	237.8
Haryana	56.53	33.13	42.68	39.84	27	31.87	31.87
H.P.	11.51	11.53	18.39	11.86	6	9.48	9.48
J & K	53.81	9.38	53.85	30.78	6.1	24.63	24.63
Karnataka	213.1	161.15	209.76	184.75	155	155	155
Kerala	92.15	60.4	170.88	93.86	52	75.09	75.09
Kolkata	74.22	62.13	51.51	65.21	73	73	73
M.P.	117.07	81.62	95.05	86.03	50.4	68.83	68.83
Maharashtra	305.07	182.33	221.79	249.89	290.35	290.35	290.35
Mumbai	217.08	338.02	139.86	241.74	272	272	272
North East	43.8	13.09	41.43	26.33	7	21.06	21.06
Orissa	26.29	30.02	44.55	29.22	16	23.37	23.37
Punjab	145.66	51.34	102.52	88.38	54	70.7	70.7
Rajasthan	203.59	74.51	125.33	107.36	26	85.89	85.89
T.N.	247.3	185.59	260.74	225.41	208	208	208
U.P. (E)	194.56	84.9	143.16	121.66	64	97.32	97.32
U.P. (W)	122.09	62.7	93.42	93.29	94.95	94.95	94.95
W.B.	39.45	53.81	57.75	43.9	24.6	35.12	35.12

iv) TRAI Recommendations for 2016 auction: -

In this auction, we have the following categories of estimations for Spectrum Price:

1. The Production Function method

2. The Producer Surplus method.
3. The DCF method.
4. The spectrum price in the last auction of the year 2015. If the 2015 value is not available, then then previous achieved price (February 2014 duly indexed) of 1800 MHz auction has been used.
5. Here is the data extracted directly from the TRAI recommendation. No case of missing data here:

LSA	Achieved Price - March 2015 (Or 2014 duly indexed)	Producer Surplus Model	Production function Model	Revenue Surplus Model	Average (mean) Value	Average Value per MHz	Reserve Price (as calc)
Andhra P.	242.8	133.85	134.07	234.14	186.21	242.8	242.8
Assam	39.54	60.27	20.23	71.49	47.88	47.88	39.54
Bihar	47.21	182.2	48.59	126.09	101.02	101.02	62
Delhi	398.71	185.02	204.21	208.4	249.09	398.71	398.71
Gujarat	238	246.98	130.58	150.18	191.43	238	238
Haryana	46.6	76.24	25.54	49.03	49.35	49.35	46.6
H.P.	15.9	18.63	6.28	21.89	15.68	15.9	15.9
J & K	6.68	48.27	15.43	59.79	32.54	32.54	13.02
Karnataka	185	191.56	129.21	212.6	179.59	185	185
Kerala	83.45	92.07	35.79	165.9	94.31	94.31	83.45
Kolkata	149.1	98.02	48.51	57.44	88.27	149.1	149.1
M.P.	55.21	174.21	62.12	122.16	103.42	103.42	82.74
Maharash.	318.04	248.04	171.84	249.34	246.81	318.04	318.04
Mumbai	297.94	233.62	142.15	142.93	204.16	297.94	297.94
North East	11	45.78	5.68	38.5	25.24	25.24	11
Orissa	33.1	83.86	15.2	58.33	47.62	47.62	38.1
Punjab	71	180.68	36.23	97.1	96.25	96.25	77
Rajasthan	72.75	182.98	60.36	138.88	113.74	113.74	90.99
T.N.	225	128.61	191.23	290.19	208.76	225	225
U.P. (E)	106.95	238.18	73.17	156.04	143.59	143.59	114.87
U.P. (W)	95.95	160.95	54.73	105.86	104.37	104.37	95.95
W.B.	26.95	87.34	48.12	65.99	57.1	57.1	45.68

b) The Govt fixed Reserve Price and the Winning Price: -

The recommendations fixed by TRAI may undergo few changes when it is discussed by the Telecom Commission (TC) within the ambit of its Telecom Policy and the constraints imposed by such policy in a public auction. Naturally this had a bearing on the reserve price also. TRAI has criticized the adjustment of the reserve price by DoT as 'cherry picking'.

"Cherry-picking from the Authority's recommendations is neither right nor logically defensible. Yet, this is what DoT seeks to do. Either one accepts the valuation methodologies adopted by the Authority in their entirety or one devises a different valuation methodology altogether. It is simply wrong to select those recommendations of the Authority that are convenient, by-passing others because of a mind-set to realize a pre-determined value".

The price fixed by TC will be finally ratified by the Union Cabinet after which the price becomes the 'Govt. fixed reserve price'. The auction is then announced through the Notice Inviting Auction (NIA) where the reserve price is officially published. After the auction is completed, the winning price is declared. The values collected from the auction results of the four years from the DoT website is tabulated here. The RP stands for the Govt approved Reserve Price and AP stands for the auction price at which the spectrum was sold, both in Rs. Crores per MHz.

Overall there are 47 instances out of 88 (more than 50%) where the spectrum was sold at the reserve price (which is indicated in yellow color in the table). This means that the variable RP can mask the other input variables during regression due to its perfect association. Therefore, this variable is suppressed in the sequence we follow in the regression process.

It may also be seen from the table that there are some missing data either due to not offering any spectrum (indicated as 'No Auction') or due to lack of response in the auction (indicated as 'No Bid')

1. Assam, Delhi, J&K, MP, Maharashtra, Mumbai, West Bengal in the year 2015 and Tamil Nadu in the year 2016 – Total 8 cases of 'No Auction'.
2. Delhi, Karnataka, Mumbai and Rajasthan in the year 2012; Bihar in the year 2015; Karnataka and Orissa in the year 2016 – Total 7cases of 'No Bid'. (shown in red colour)

Auct.Year ->	2012		2014		2015		2016	
LSA ∨	RP	AP	RP	AP	RP	AP	RP	AP
Andhra Pradesh	229.53	229.53	163	163	169	242.8	243	243
Assam	6.94	6.94	7	36.1	No Auction.		40	40
Bihar	34.01	37.14	37	43.1	62	No Bid.	62	62
Delhi	554.45	No Bid.	219	364	No Auction.		399	399
Gujarat	179.87	179.87	143	237.8	238	238	238	238
Haryana	37.22	37.22	27	27	32	46.6	47	49.3
H.P.	6.22	6.22	6	6	9	15.9	16	16
J & K	5.06	5.06	5	6.1	No Auction.		13	13
Karnataka	264.1	No Bid.	155	155	185	185	185	No Bid.
Kerala	52.24	52.24	52	52	75	83.45	83	83
Kolkata	90.98	90.98	73	73	73	149.1	149	151.15
M.P.	43.19	43.19	43	50.4	No Auction.		83	83
Maharashtra	210.25	210.25	173	290.35	No Auction.		318	318
Mumbai	542.76	No Bid.	207	272	No Auction.		298	489.2
North East	7.07	7.07	7	7	11	11	11	11
Orissa	16.22	16.22	16	16	23	33.1	38	No Bid.
Punjab	53.82	53.82	54	54	71	71	77	77
Rajasthan	53.66	No Bid.	26	26	60	72.75	91	91.9
T.N.	244.87	244.87	208	208	225	225	No Auction.	
U.P. (E)	60.94	60.94	61	64	97	106.95	115	133.15
U.P. (W)	85.93	85.93	62	94.95	95	95.95	96	100.3
W.B.	20.67	20.67	21	24.6	No Auction.		46	46

The dataset overall consists of 88 samples (22 Service areas X 4 auctions). If we drop the above missing values of 15 data points, the sample will shrink to 73 data points. The reduction in the data points can influence the accuracy of the model. In order to circumvent this issue, 7 cases of 'No Bid' values will be filled up with the last year's data, if available. Therefore, the logic is to retain the 'No Bid' cases by replacing the missing bid value with the bid value obtained in the subsequent auction in the cases of 2012 auction (as there is no prior data), replacing with the previous data in the case of Bihar LSA in the year 2015 and the last price sold in the case of 2016 Auction since we have no further auction data. By this logic, the following values will be substituted against the missing 'No Bid' values:

Delhi 2012: 364 (the price obtained in 2014)

Karnataka 2012: 155 (the price obtained in 2014)

Mumbai 2012: 272 (the price obtained in 2014)

Rajasthan 2012: 26 (the price obtained in 2014)

Bihar 2015: 43.1 (the price obtained in 2014)

Karnataka 2016: 185 (the price obtained in 2015)

Orissa 2016: 33.1 (the price obtained in 2015)

With the above tinkering a sample size of 80 out of the whole population can be achieved. The term population, should be referred here as to representing all conceivable data related to the Spectrum Auction in India, including the data that we may likely to aggregate in all the spectrum auction that are to be held in the future. Here the assumption is that the sample size of 80 shall be assimilating the characteristics of a simple random sample so as to apply the standard statistical methods to explore for a realistic model.

At this stage it is to be mentioned that a working theory is already on the platter that the Auction price could be a function of the TRAI estimators and some other variables. This hypothesis has to be tested using a regression model with the trends in the data. It is a confirmatory study design to check the data arising out of the recommendations prior to the auction and the post auction data and fitting them to a model. The data is also assumed to be coming from an independent and identical distribution to assume the normal distribution. That might help us to explain that it would be plausible to build a model that can predict the future mean price of 1800 MHz spectrum for all those 22 Service areas with a certain margin of error.

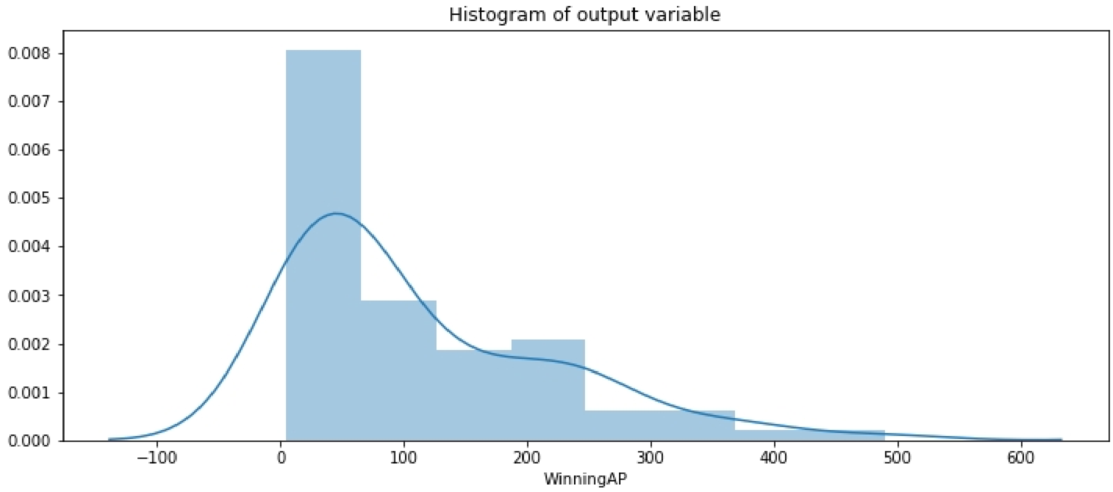
c) Other explanatory variables that could affect the Auction price: -

The information related to GSDP per Capita and population is collected from some of the web resources mentioned earlier. Remaining telecom indicators were collected from the TRAI data as well as from the Statistics data maintained by DoT. Here is a sample of the data so collected for few Service Areas.

LSA	GSDP (Rs. Lakhs)	Population	ARPU (Rs./ Subscriber / month)	MOU(Minute of Use)	No of Mobile subscribers. (In Millions)	Tele density (per 100 people)
AP 2012.	73883569	85671869	113	362	66.83	78.11
AP 2014	79707152	86725633	121	383	64.72	74.69
AP 2015	86089634	87748995	129	387	70.56	80.6
Ap 2016	95623896	88758108	136	380	74.59	84.52
Assam 2012	14317491	31608128	118	461	14.21	45.87
Assam 2014	15452540	31996908	122	438	14.66	46.49
Assam 2015	16521230	32374472	130	440	16.43	51.34
Assam 2016	19110899	32746778	126	406	17.69	54.63
Bihar 2012	39806155	138856016	79	369	63.48	48.43
Bihar 2014	43546610	140563945	90	396	58.32	43.66
Bihar 2015	46601683	142222600	93	397	65.12	48
Bihar 2016	47050319	143858160	93	418	71.76	52.27
Delhi 2012	34379750	17004505	144	411	42.49	223.29
Delhi 2014	39290838	17213660	160	436	41.05	206.73
Delhi 2015	42889924	17416781	165	414	45.35	220.4
Delhi 2016	47405772	17617074	177	409	47.87	225.97
Gujarat 2012	61560607	61813892	95	321	53.32	88.11
Gujarat 2014	73428387	62574203	98	339	52.14	84.5
Gujarat 2015	81142764	63312579	108	347	56.8	90.61
Gujarat 2016.	89446534	64040674	115	352	60.48	95.32
Haryana 2012	29753852	25678496	75	298	22.56	87.14
Haryana 2014	34679932	25994342	96	366	19.9	75.09
Haryana 2015	36663587	26301075	103	377	21.75	80.51
Haryana 2016	39964594	26603537	108	399	22.76	83.05
HP 2012	7271984	6953155	74	368	7.97	116.18
HP2014	8284669	7038679	101	413	7.03	101.13
HP 2015	8906019	7121735	106	392	7.63	108.51
HP 2016	9627406	7203635	112	363	8.63	121.85
J&K 2012	7825555	12703085	160	632	6.31	53.1
J&K 2014	8511550	12859333	173	616	7.41	61.26
J&K 2015	8237211	13011073	161	522	8.88	72.4
J&K 2016	9697785	13160700	164	515	9.64	77.77
Karnataka 2012	60600981	61883426	109	361	55.71	92.74
Karnataka 2014	70484961	62644592	122	387	53.43	87.63
Karnataka 2015	74842911	63383798	141	384	57.18	92.67
Karnataka 2016	83144922	64112712	146	371	61.45	98.67

Now that the data is complete with all the possible explanatory (regressor) variables and the dependent (response) variables, the regression process could be performed. But before doing that some plots need to be observed to have a first impression of the characteristics of data.

8. Exploration of the Data: -



Examining the above histogram plot of the auction price, we could see that the data has unimodal appearance, not the bell shaped but rather right skewed with some outliers. Most of the Auction price data concentrated between the values about 5 and 250.The Mumbai and Delhi LSA prices produce some outliers because the telecom operators had a strong preference and incentive to bid competitively compared to the other LSA s. The outliers are caused by this action. It can be seen from the histograms of the explanatory variables like Production function, TRAI price, Govt fixed price and to some extend on the population of the LSA, that each has close resemblance to the response variable. This might explain the linear association between input and output. Here is the description of the auction price data across the four years:

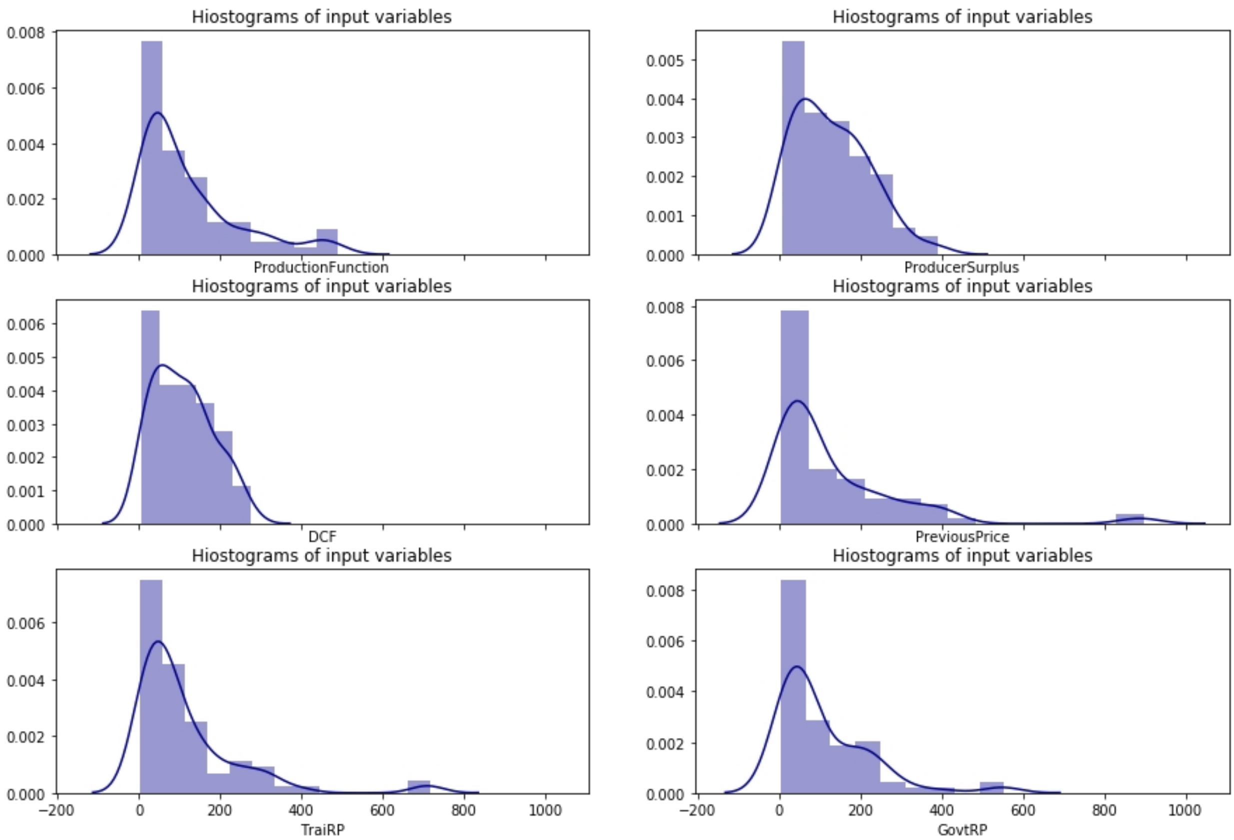
Count	80
Mean	111.97
Standard Deviation:	108.34
Minimum value:	5.06
25% of data has auction price up to:	33.10
50% of data has auction price up to(median)	71.88
75% of data has auction price up to:	181.15
Maximum value	489.2

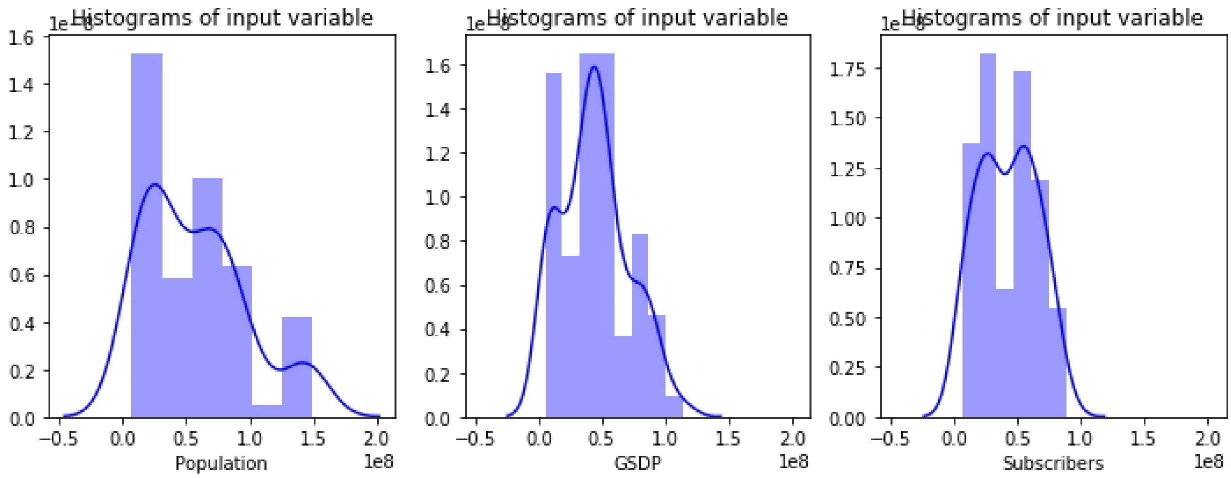
This also shows the 75% of data is within the auction price of Rs.181.15 Crores. The high value points only constitute 25% of the data. That is the same skewness we found in the histograms. It indicates a crowding of circles with Spectrum Prices up to Rs 200 Crores. LSA which are Metros and Class A stations are quite few compared to Class B and C LSA.

Two methods are used here to reduce the skewness. The first is to introduce a Category variable, named as 'category' which will have 4 levels for the four types of Classes-Metro, Class A, Class B and Class C respectively to denote the different classes of LSA s. 3 Dummy variables will be used to accomplish this feat. This will leverage the price difference between the Metro and Class B and C LSA s. The second technique is to transform the spectrum prices into logarithmic scale which will pull down the gap between the very high prices of Mumbai, and Delhi with lower priced circles.

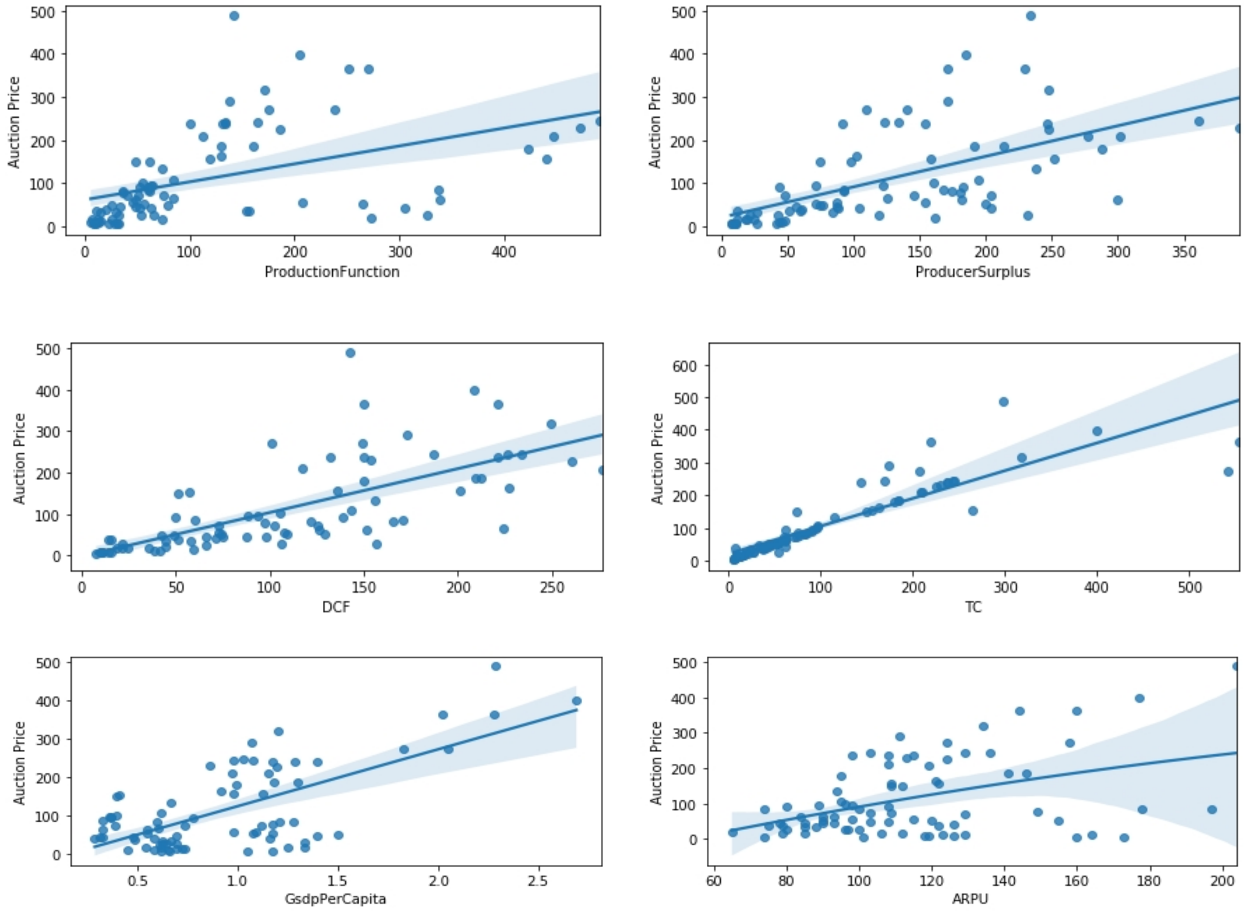
The second method is using logarithmic scale to normalize the skewness. Therefore, all the seven price input variables and one population variable and the response variable all are translated to logarithmic scale to absorb the skewness.

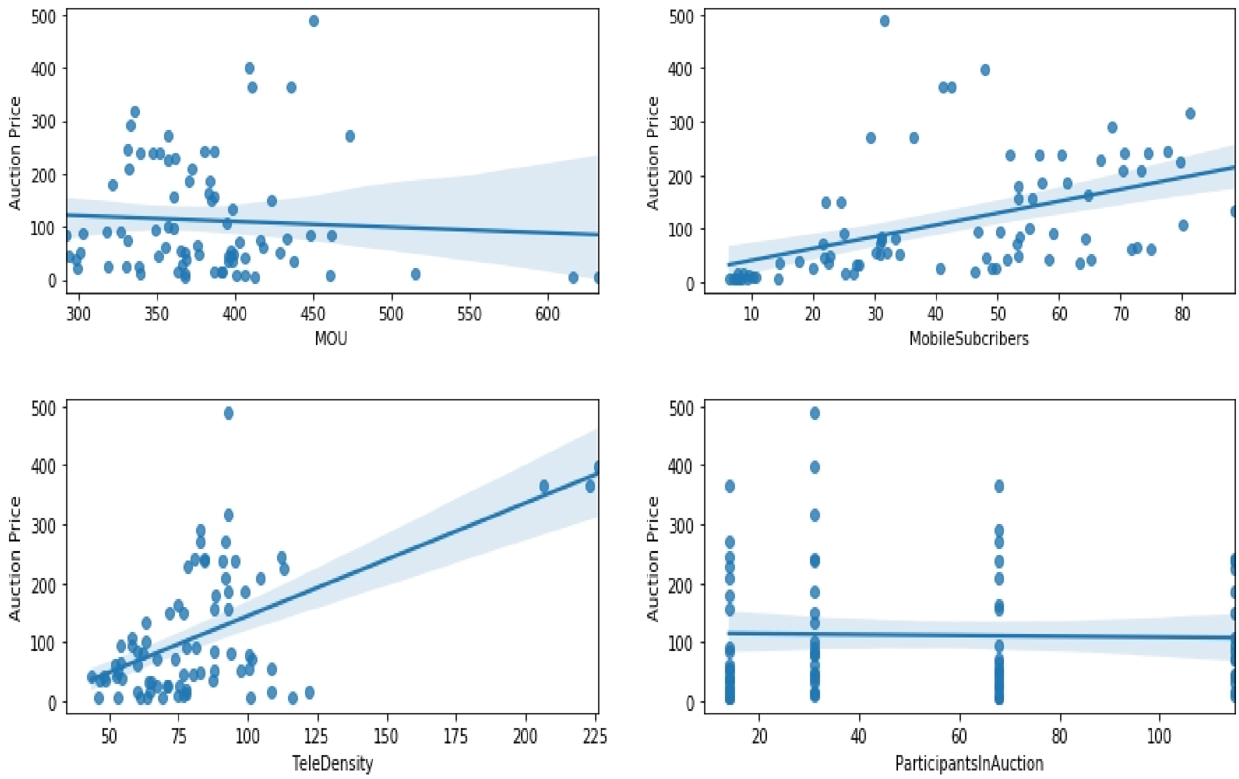
Histograms of the explanatory variables: -





Here are the scatter plots of the different variables plotted against the Auction Price.

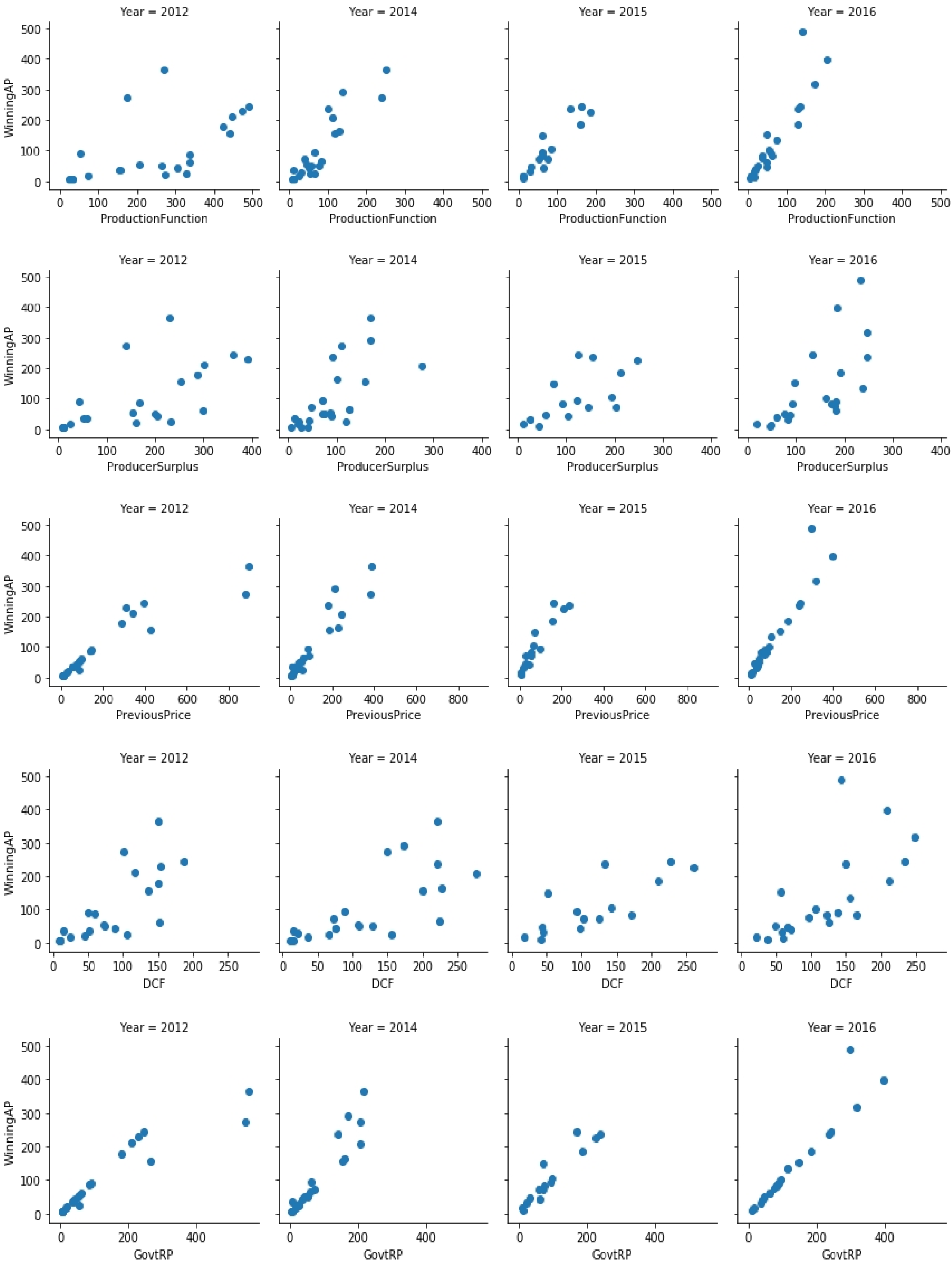




(#The TC variable is named as 'GovtRP' in some other plots)

From the scatter plots above, the relation between each of the independent variables with the dependent variable can be seen if it is linear or not. As can be seen the spectrum variables like Producer Surplus, DCF, TC and other variables like GSDP Per Capita, ARPU and Mobile Subscribers has fair amount of correlation with the Winning Reserve Price variable. The relation between the TC and Winning price is obviously linear as the final price of spectrum sold would be at the reserve price itself. Similarly, the PP is the previous auction price that also connect well with the next years' reserve price. Therefore, these variables TC and PP have predictable relationship with the response variable. These variables should be carefully observed in the regression as they could dominate the other explanatory variables easily causing strong multi collinearity.

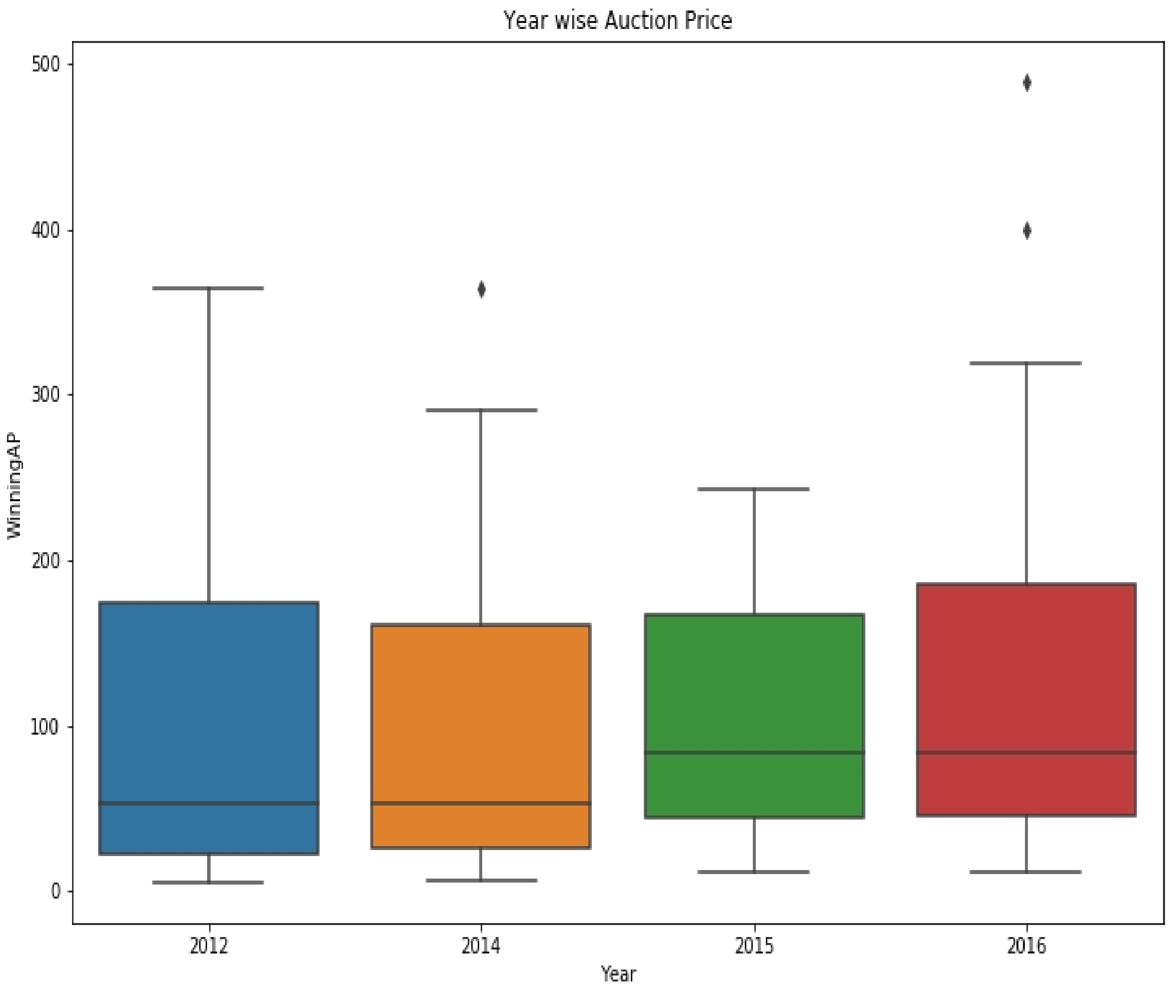
The year wise scatter plots of some of the variables gives a clearer picture of these variables, with the yearly variation though a linear pattern in most of the variables. In the plots below shows all the price variables plotted against the Winning Price, with yearly demarcation. The TC (Govt RP) and the PP (Previous auctioned price) variables follows a perfect linear pattern with the auction price response variable calling for caution. Other variables have fairly linear pattern, but its effect could not be predicted, but need to be tested.

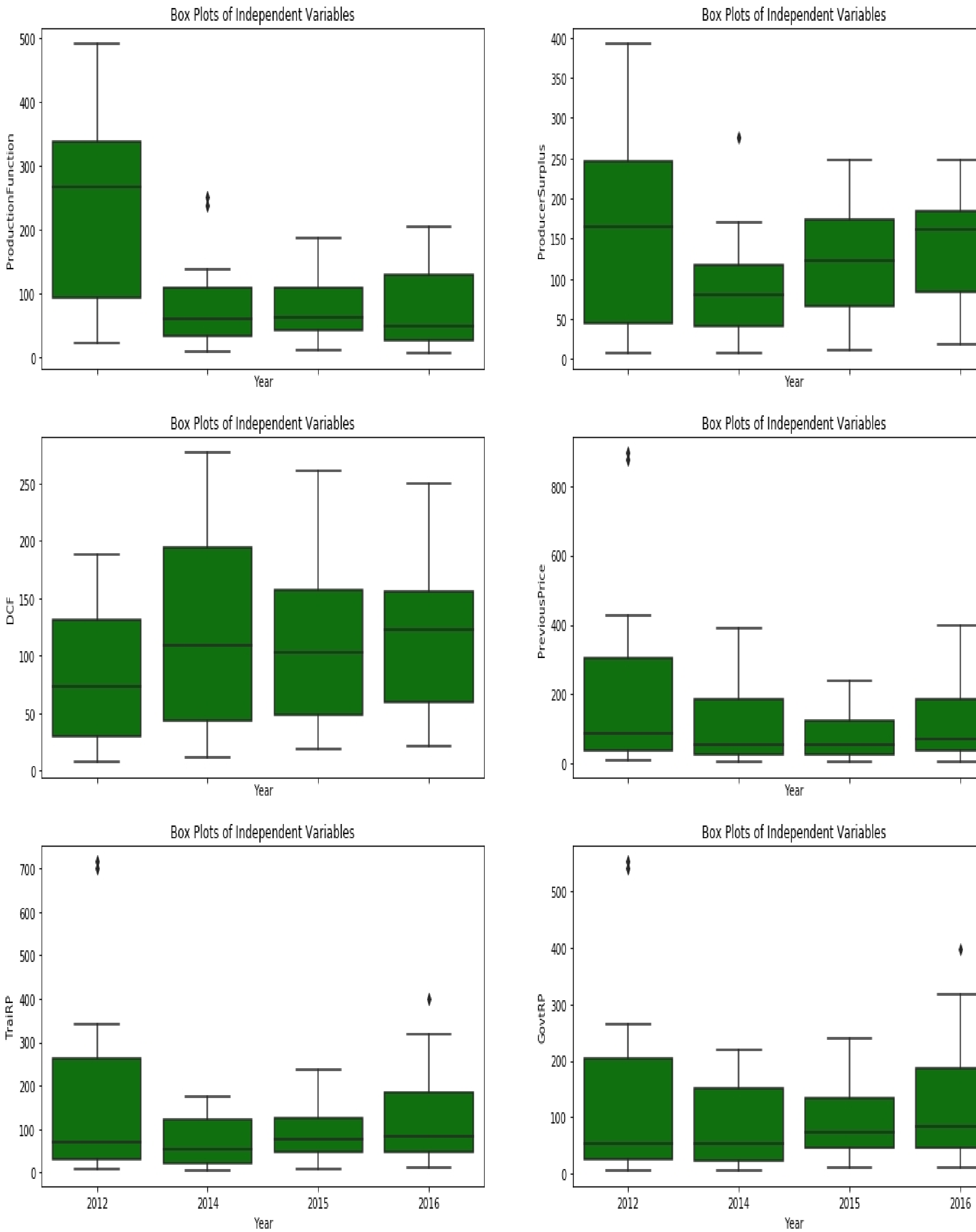


Year wise Box plots: -

Another type of plot that indicates the distribution like the histogram and scatterplots, but gives a better visual appeal. The year wise plots also show the outliers in the data. This can be seen from the plots below:

Three outliers can be spotted from the picture one in year 2014 and two in the year 2016. These are the auction prices in those years accountable to Mumbai and Delhi LSA s. Though outliers, these points have to be retained in the data set since the data points for Mumbai and Delhi for the year 2015 is already part of missing data on account of no auction. The outliers can pull the regression line from the normal path and distort the output. Year wise box plots for the explanatory variables also shows some outliers in their plot, shown as black spots in the plots below. The DCF variable has apparently no outliers.




Box Plots of Independent Variables
ProductionFunction
Year
Box Plots of Independent Variables
ProducerSurplus
Year
Box Plots of Independent Variables
DCF
Year
Box Plots of Independent Variables
PreviousPrice
Year
Box Plots of Independent Variables
TraiRP
2012
2014
2015
2016
Year
Box Plots of Independent Variables
GovtRP
2012
2014
2015
2016
Year

9. The Linear Regression: -

The OLS regression for each independent variable were individually performed (simple linear regression) and the results have been compiled as below:

```
Dep. Variable: LogAP

Model: OLS

Method: Least Squares
```

	coef	std err	t	P>|t|	CI
LogPF	0.7439	0.087	8.581	0.000	0.571-0.916
LogPS	0.9300	0.083	11.240	0.000	0.765-1.095
LogDCF	1.0727	0.078	13.812	0.000	0.918-1.227
LogTRP	0.9508	0.036	26.462	0.000	0.879-1.022
LogGRP	0.9507	0.027	35.511	0.000	0.897-1.004
LogPOP	0.5925	0.145	4.083	0.000	0.304-0.881
GsdpPerCa	1.0946	0.241	4.549	0.000	0.616-1.574
ARPU	0.0090	0.004	2.034	0.045	0.000-0.018
MOU.	-0.0054	0.002	-2.484	0.015	-0.010--0.001
MobSubsc	0.0346	0.004	7.893	0.000	0.026-0.043
TeleDens	0.0140	0.004	3.737	0.000	0.007-0.022
ActiviRou	0.0018	0.004	0.494	0.623.	-0.005-0.009

All variables with the exception of the 'ARPU' and 'Activity rounds' in auction are found to be statistically significant. The effect of price estimator variables on the response variable was expected and not surprisingly their coefficients are quite robust. It is also observed that MOU (Minutes of Use-Traffic) is negatively related to Auction Price. Individual scatter plotting also shows such an association. If more MOU is consumed in a LSA, that shows the capacity of the network. So in a big LSA more traffic will be normally generated and hence the Spectrum Price should be higher. This result therefore seems counter intuitive, but truly reflect the data. But it is not possible to draw a conclusion only on this basis of individual effect. Further it is observed that variables ARPU, MOU and Activity Rounds have a very small coefficient in the regression indicating its weak association in general.

The correlation matrix is shown here below. LogPP (last auction price in log form), LogGRP (TC or Govt fixed Reserve Price in log form) and LogTRP (Reserve Price fixed by TRAI in log form) are having a Pierson correlation index of more than 0.9. A very strong index may be a symptom of non-linear association between the continuous variables due to interaction between exploratory variables. By a thumb rule a value higher than 0.9 could be indicative of interaction of these variables and thereby of a strong possibility of multi collinearity among them. This fact has to be carefully factored while multivariate OLS is performed.

	LogPF	LogPS	LogDCF	LogTRP	LogGRP	LogPP	LogAP
LogPF	1.000000	0.753082	0.636073	0.783151	0.742615	0.820767	0.696867
LogPS	0.753082	1.000000	0.848707	0.843728	0.833561	0.792948	0.786307
LogDCF	0.636073	0.848707	1.000000	0.828934	0.858078	0.783800	0.842489
LogTRP	0.783151	0.843728	0.828934	1.000000	0.990517	0.966205	0.948565
LogGRP	0.742615	0.833561	0.858078	0.990517	1.000000	0.965522	0.970437
LogPP	0.820767	0.792948	0.783800	0.966205	0.965522	1.000000	0.926636
LogAP	0.696867	0.786307	0.842489	0.948565	0.970437	0.926636	1.000000

(PF: Production Function, PS: Producer Surplus, DCF: Discounted Cash Flow, TRP: Trai Reserve Price, GRP: Govt Reserve Price, PP: Previous Auction Price, AP: Winning Auction Price)

It is also noteworthy to reflect on the significance of these variables from a statistical angle how the non-linearity sneak in into the overall picture. As was noted earlier, TRAI uses a Probabilistic Average of its estimation methods. In simple terms, this is the mean of all the methods taken together. Since each part of this mean is separately contributing to the regression as independent variables, then average of these methods such as the TRAI reserve price, variable named as 'LogTRP' may not be as independent as an independent variable should be. Similarly, the RP fixed by Government (variable 'LogGRP') is the same as TRAI fixed price with few selective changes in the Reserve Price of some of the LSA s. While fixing reserve price such changes are made arbitrarily without following a pattern or formula, the variable LogGRP also should be fitting the pattern of a non-independent variable. 'LogPP' is another variable which relates to the auction price determined by the market. Even though the LogPP data does not fall strictly under the time series category, however there would be some undesirable effect as applied to longitudinal data due to the repetition of estimates in the data used for regression. However, it is retained and included as one of the regressors until the regression produce evidence to reject it.

10. The Regression output: -

The OLS regression has been performed using the *statsmodule* in the python. There are 80 observations with LogAP as the dependent variable which represents the auction price in natural log units. Started with all the explanatory variables, the regression finally converged on to one category variable and three quantitative variables after eliminating the statistically insignificant variables one by one. The variable 'Category' is representing the class of the LSA. The reference class is the Metro LSA (Mumbai, Delhi and Kolkata), Category T1, T2 and T3 are the dummy variables representing the Class A, Class B and Class C LSA s respectively. The other independent variables that were found to be significant are LogDCF, LogPop and GsdpPerCapita which are the DCF estimator, Population of the LSA both translated to natural logarithm and GSDP Per Capita of that LSA respectively.

The regression result explains the variation in the dependent variable, using the predictors as explanatory factors. The R-squared value is 91. The squared Pearson correlation coefficient between the fitted values and the observed value of the dependent variable is called the R-squared, the "proportion of explained variance," or the "coefficient of determination." This means that the fitted model with the parameters explain the 91 percent of the association of dependent variable with the independent variable. In other words, the proportion of explained variance is 91 percent. The data set is a sample of the population. The regression coefficients that we get from the analyse on the sample is always estimate of the true population parameters.

The VIF (Variance Inflation Factor) values also were calculated as follows: -

[3.03, 3.80, 4.23, 3.75, 3.39, 2.02]

Since these values are less than 10, as a thumb rule it confirms that the parameters determined in the regressions are quite stable.

The model can be described in the standardized notation of [$\hat{Y}=\beta_0+ \beta_1 X_1+ \beta_2 X_2$......] as shown here: -

For Metro LSA: -

log (AP)=5.4287 + 0.3795 X log(DCF) + 0.4476 X log (POP) + 0.4588 X GSDP Per Capita

For Class A LSA: -

log (AP)= -0.7249 + 5.4287 + 0.3795 X log(DCF) + 0.4476 X log (POP) + 0.4588 X GSDP Per

Capita

For Class B LSA: -

log (AP)= -1.5225 + 5.4287 + 0.3795 X log(DCF) + 0.4476 X log (POP) + 0.4588 X GSDP Per Capita

For Class B LSA: -

log (AP)= -1.9317 + 5.4287 + 0.3795 X log(DCF) + 0.4476 X log (POP) + 0.4588 X GSDP Per Capita

(Please Note that the addition of all value yields a value in natural logarithm. The sum value of log(AP) has to be converted back from natural logarithm to normal value in Rs. Crore per MHz using exponentiation)

The expression can be interpreted in the following manner:

"Keeping the population and GSDP per Capita of a Metro LSA is kept at their mean value and if the auction price of 1800 MHz spectrum is valued using DCF method, then one unit (logarithmic) increase in the DCF value so estimated would likely generate 0.375 unit (logarithmic) increase in the Auction Price. Similarly one unit of population increase (log) would result in 0.4476 units of increase in LogAP and for one unit of increase in Gsdp per capita in Rs. Lakhs would increase 0.4588 units of LogAP , both when the other two variables are kept at their mean". Thus the price is linearly related to the DCF, Population and GSDP Per Capita.

```
OLS Regression Results
==============================================================================
Dep. Variable:            LogAP        R-squared:                       0.916
Model:                      OLS        Adj. R-squared:                  0.909
Method:           Least Squares        F-statistic:                     132.2
Date:          un, 10 May 2020        Prob (F-statistic):           3.72e-37
Time:                  15:50:22        Log-Likelihood:                -27.263
No. Observations:            80        AIC:                             68.53
Df Residuals:                73        BIC:                             85.20
Df Model:                     6
Covariance Type:      nonrobust
==============================================================================
                    coef    std err          t      P>|t|      [0.025 0.975]
------------------------------------------------------------------------------
Intercept         5.4287      0.137     39.600     0.000      5.155 5.702
C(Cat)[T.1]      -0.7249      0.166     -4.363     0.000     -1.056    -0.394
C(Cat)[T.2]      -1.5225      0.160     -9.494     0.000     -1.842    -1.203
C(Cat)[T.3]      -1.9317      0.184    -10.522     0.000     -2.298    -1.566
LogDCF_c          0.3795      0.084      4.527     0.000      0.212     0.547
LogPOP_c          0.4476      0.088      5.062     0.000      0.271     0.624
GsdPerCap_c       0.4588      0.116      3.966     0.000      0.228     0.689
==============================================================================
Omnibus:                  4.890       Durbin-Watson:                    1.040
Prob(Omnibus):            0.087       Jarque-Bera (JB):                 5.001
Skew:                    -0.303       Prob(JB):                         0.0820
Kurtosis:                 4.064       Cond. No.                         8.6700
==============================================================================
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
```

```
specified.
```

What is the significance of the intercept value, β0 of 5.4287? The meaning of β0 is the estimate of the mean outcome when x= 0, and always stated in terms of the actual variables of the study. It may be worth mentioning that the quantitative input variables are normalized using centering (the c suffix in the variables), meaning that mean value of each variables is subtracted from the absolute value of that variable. The OLS regression done after centering does not change the parameter value but only shifts the intercept to a meaningful value. This intercept value can be interpreted as the Auction Price of a Metro LSA where the DCF, Population and GSDP per capita are at their mean value of the overall statistical population data. The normalization process has assumed importance because of the fact that the no independent variables can have a value of zero in its data set.

Regression without logarithmic conversion -

The exercise was repeated with the variables using as it is without translating into logarithmic value to see if the relation can sustain. The outcome is reproduced here: -

```
   OLS Regression Results
==============================================================================
Dep. Variable:          WinningAP       R-squared:                       0.920
Model:                        OLS       Adj. R-squared:                  0.912
Method:             Least Squares       F-statistic:                     118.1
Date:            Sun, 10 May 2020       Prob (F-statistic):           7.38e-37
Time:                    16:21:03       Log-Likelihood:                -386.87
No. Observations:              80       AIC:                             789.7
Df Residuals:                  72       BIC:                             808.8
Df Model:                       7
Covariance Type:        nonrobust
==============================================================================
                  coef         std err     t       P>|t|    [0.025     0.975]
-------------------------------------------------------------------------------
Intercept        234.2206       11.412    20.523   0.000   211.470    256.971
C(Cat)[T.1]      -80.2333       15.106    -5.311   0.000  -110.347    -50.119
C(Cat)[T.2]     -169.6162       13.383   -12.674   0.000  -196.295   -142.938
C(Cat)[T.3]     -147.6183       15.580    -9.475   0.000  -178.676   -116.561
Population_c       0.7478        0.131     5.695   0.000     0.486      1.010
DCF_c              0.1917        0.091     2.112   0.038     0.011      0.373
ProducerSur_c      0.1904        0.059     3.216   0.002     0.072      0.308
GsdpPerCap_c      98.9114       10.685     9.257   0.000    77.611    120.212
==============================================================================
Omnibus:                   13.634       Durbin-Watson:                   1.297
Prob(Omnibus):              0.001       Jarque-Bera (JB):               16.682
Skew:                       0.804       Prob(JB):  0.000239
Kurtosis:                   4.555       Cond. No.                         718.
==============================================================================
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

It is interesting to note that in addition to DCF, Producer Surplus variable also becoming significant in this regression. But their coefficients adding together becoming equal to the coefficients that was found with the DCF variable alone in the logarithmic version.

Therefore, assuming the DCF and Producer Surplus got estimated as almost equal, then both the regression function might produce similar values in the prediction. The Confidence-interval associated with the DCF variable shows a bigger interval indicating the higher uncertainty factor. The VIF of the variables all were found to be stable. However, in this model, it is felt that, the distortion due to non-symmetric distribution of the data (skewedness) and the outliers might be influencing adversely here. The log version model would be preferable to the normal version.

The regression function can be expressed as:

AP=234.22+ 0.75 X Population(LSA) +0.19 X DCF +0.19 X Producer Surplus+ 98.91 X GSDP Per Capita

With the Dummy variable for Class A, Class B and Class C LSA s further causing a penalty of -80.23, -169.62 and -147.63 respectively with reference to a Metro LSA.

11. **Prediction of future values using the above equations:**

In the year 2018, TRAI has given another recommendation for auction. By using their values (after indexing for one year assuming MCLR rate of 7% appreciation) and the population and GDSP values of the year 2018-19, the Auction prices for the year 2018-2019 could be predicted as below:

LSA	Class of LSA	TRAI recommended Price in Rs. Crores per MHz (2018)	Predicted Price of 1800 MHz spectrum In Rs. Crores per MHz	
			Regressed using Linear Scale. (2018-19)	Regressed using Log scale. (2018-19)
Andhra Pradesh	A	279	232.32	248.53
Assam	C	46	13.55	20.46
Bihar	C	88	116.41	43.48
Delhi	M	457	465.29	465.69
Gujarat	A	273	246.14	198.98
Haryana	B	57	92.46	40.15
Himachal Pradesh	C	18	65.84	9.90
Jammu & Kashmir	C	15	8.23	11.78
Karnataka	A	109	234.79	230.87
Kerala	B	95	98.83	71.36
Kolkata	M	173	130.45	78.32
Madhya Pradesh	B	95	78.40	66.23
Maharashtra	A	365	260.40	257.19
Mumbai	M	561	401.59	424.03
North East	C	13	9.36	12.73
Orissa	C	27	30.52	20.57
Punjab	B	88	82.73	50.16
Rajasthan	B	105	68.36	58.85
Tamilnadu	A	100	227.00	235.04
U. P. (East)	B	153	71.66	60.66
U.P. (West)	B	115	74.44	61.46
West Bengal	B	53	45.51	48.41

The difference in the predicted value of the two methods first in the log scale and the second in the log scale can be seen above in the table. Aside is also placed the TRAI recommended price. It may be seen that the log based prediction has given a lower price compared to the linear prediction. However, as explained elsewhere the log conversion helps that model able to absorb the large dynamic variation in the data when data is especially skewed and would give a more stable result reflecting the changes on account of the changes in the DCF values, population and GSDP per capita ‘rationally’. Further there is an additional contribution of Producer Surplus in the linear model though it is noticeable that the sum of the coefficients is equal in both models. Referring to the figure estimated by TRAI in the year 2018, the Producer Surplus estimation by TRAI is substantially higher than the valuation using DCF method. This is the reason for elevated prices in the linear model. For example, for Bihar the TRAI has found a price of Rs.66.5 Crore by the DCF method whereas the same LSA is valued at Rs.244.28 Crores using the Producer Surplus method with wide variation in the estimates. That may be the reason the linear model has evaluated Bihar price at Rs.116.41 Crores whereas the log model without using the Producer Surplus has evaluated it at Rs.43.48 Crores. Such variations can be seen across the LSA s. The linear model is only for comparison purpose, but for predicting the price the log model regression is recommended.

12. Machine Learning based prediction: -

*(****Reference: 6)*

Machine learning models are used for every kind of prediction in almost all fields these days. Compared to the classical statistical models, the ML models do the prediction by learning the trends in the data by memorizing it. Then the ‘learned’ model will be able to predict the outcome for future values of the input. There are various algorithms available which can be deployed based on the category, quality and quantity of the dataset available. In machine learning algorithms, the volume of data is most important factor which offers the opportunity to the model to capture the relation between the input variables (known as features) and output variable (known as target) and predict a model with low *bias (the difference between the true population parameter and the expected estimator. It measures the accuracy of the estimates.)* and low variance *(Variance, on the other hand, measures the spread, or uncertainty, in these estimates.)* Low bias represents the model to accurately predict and the low variance means the universality of prediction when it is offered input from diverse test data.

The machine learning methods does not care for any hypothesis, confidence interval, IID (independent and identically distributed) sample etc. which are crucial in a pure statistical model. Because a statistical model is most often used to understand and explain the relation, correlation or causation between the dependent variable and independent variables whereas the ML model is used for pure prediction.

In this paper two supervisory regression models are used to predict the price of Spectrum. These models are selected because they are appropriate for sparse data sets.

1. KNN (K Nearest Neighbors) Regression.
2. Lasso Regression.

Both models are used to predict value of a quantitative target variable type. A brief on both methods are given here: -

1. KNN Algorithm

A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbours. Another approach uses an inverse distance weighted average of the K nearest neighbours. KNN regression uses the same distance functions as KNN classification.

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

(Ref: https://www.saedsayad.com/k_nearest_neighbors_reg.htm)

The K in the KNN is the number of neighbors used in the algorithm. In this paper, the algorithm is run using 5 and 10 neighbors. The 5 neighbors have given a score of 90% and by using 10 neighbors the accuracy is down to 70%.

2. The LASSO Algorithm (from Wikipedia)

Lasso was introduced in order to improve the prediction accuracy and interpretability of regression models by altering the model fitting process to select only a subset of the provided covariates for use in the final model rather than using all of them.

Lasso is able to achieve interpretability at the same time avoiding overfitting problem associated with linear regression by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, which forces certain coefficients to be set to zero, effectively choosing a simpler model that does not include those coefficients. This idea is implemented in such a way that the sum of the squares of the coefficients is forced to be shrunk to zero. This is also called L1 penalty method which is controlled by the alpha parameter.

3. ML Software:

Both the ML algorithms are used in this paper using the SKLEARN package in the Python environment from https://scikit-learn.org/ .

Predicted value of Spectrum (2018-19) using Machine Learning methods	TRAI Recommended Price	Spectrum Price of 1800 MHz spectrum with KNN Algorithm		Spectrum Price of 1800 MHz spectrum with Lasso Algorithm	
		KNN-R (K=5) Score=90.2%	KNN-R (K=10) Score=78.2%	LASSO-R (Alpha=5) Score=85.6%	LASSO-R Alpha=10) Score=60.5%
Andhra Pradesh	279	210.83	213.48	163.50	165.15
Assam	46	14.40	18.64	30.87	52.62
Bihar	88	34.53	28.27	35.29	51.99
Delhi	457	352.24	221.16	293.22	208.54
Gujarat	273	178.13	213.00	160.68	161.39
Haryana	57	43.92	49.08	75.17	94.73
Himachal Pradesh	18	9.46	15.66	58.72	63.18
Jammu &Kashmir	15	7.63	14.83	32.62	53.14
Karnataka	109	183.76	213.00	170.32	167.35
Kerala	95	64.36	52.90	91.67	102.52
Kolkata	173	147.25	155.55	141.03	151.82
Madhya Pradesh	95	67.71	56.88	47.30	83.26
Maharashtra	365	194.27	213.69	166.82	164.69
Mumbai	561	352.24	230.84	260.03	196.12
North East	13	16.21	17.97	32.79	53.18
Orissa	27	22.48	17.97	31.90	52.68
Punjab	88	53.07	52.90	73.10	93.19
Rajasthan	105	52.21	51.59	52.55	85.32
Tamilnadu	100	200.23	205.48	163.68	165.15
U. P. (East)	153	77.42	56.40	54.56	85.53
U.P. (West)	115	59.79	60.07	28.91	76.59
West Bengal	53	36.46	41.80	38.10	80.51

Both the models have been used to predict the price of spectrum using the test file of the TRAI prices recommended in the year 2018, which was also used for the OLS regression model. The ML program was done with different settings of the control variable.

The result has been tabulated above:

- (All units are in Rs, Crores per MHz)
- (The Score represents the score in predicting the test score while using an 80:20 train: test ratio in the data set for both models)

The KNN is found to be a better model for sparse data sets, as seen from the output. With neighbors of 5 numbers, the KNN algorithm gives a better prediction probability of 90.2%. Therefore, KNN Machine Learning method is preferred to find the price of 1800 MHz spectrum.

4. **The Price: -**

Having completed the prediction using both the statistical method and machine learning methods, it is time to conclude the findings. The OLS method was performed using the linear values and logarithmic values and based on the performance it was decided to use the logarithmic version. As for the machine learning method, the predicted value using the KNN regression will be retained.

The values from both methods finally selected is tabulated above and to be considered as the final values for the Spectrum Price of 1800 MHz in the year 2018-2019.

Predicted value of Spectrum (2018-19) for the LSA s	By OLS Regression. (Rs. Crores per MHz)	By Machine Learning. (Rs. Crores per MHz)	TRAI Price of 2018. (Rs. Crores Per MHz)
Andhra Pradesh	249	211	279
Assam	21	14	46
Bihar	43	35	88
Delhi	466	352	457
Gujarat	199	178	273
Haryana	40	44	57
Himachal Pradesh	10	9	18
Jammu & Kashmir	12	8	15
Karnataka	231	184	109
Kerala	71	64	95
Kolkata	78	147	173
Madhya Pradesh	66	68	95
Maharashtra	257	194	365
Mumbai	424	352	561
North East	13	16	13
Orissa	21	22	27
Punjab	50	53	88
Rajasthan	59	52	105
Tamilnadu	235	200	100
U. P. (East)	61	77	153
U.P. (West)	61	60	115
West Bengal	48	36	53

13. Conclusion: -

The study was conducted to explore those factors contributing to the price of 1800 MHz spectrum in India and to explain the contribution of each factor. We have considered a number of factors based on a theory that we had at the beginning about how the discovered Auction price is related to the reserve price recommended by TRAI before the actual auction is conducted. Hooking on to this theory, data from the auction history was collected along with other inputs that are likely to influence the price. This data was then regressed with the help of statistical methods to look for evidence if any to establish the relation of the Auction price with the input variables. The null hypothesis- that there is no relation- was rejected after it was found that the coefficients are statistically significant. The plots were also checked and the assumptions that we made for conducting the linear regression were found to be stable from these plots. There is a linear association of the response variable with the explanatory variables was established through the regression equation. Three significant variables (based on p-value) are selected that had a greater influence on the response variable out of which one is the estimator of TRAI. The R Squared value is 91% which is the proportion of the variance in the response variable that can be explained by the explanatory variables.

The prediction of future values was also carried out using the TRAI 2018 estimators. The predicted values were compared to the TRAI's reserve price. Some of the LSA were having glaring difference between the predicted value and TRAI value. A reflection on that. The emphasis in this paper was on the three major price forecasting methodologies namely, DCF, Producer Surplus and Production Function adopted by TRAI during its consultation stage with the stakeholder's participation. The TRAI followed similar approaches in almost all its recommendations submitted to DoT after 2012 auction. Each of these methods are using a different model to develop the price and basically are independent of each other. As they are formulated using different approaches, these three methods are producing three distinct values for each of the LSA. TRAI then uses average of all these methods to converge to a single value as an estimated price for each of LSA s. This average value so obtained is further depreciated by 80% to arrive at the reserve price. The Government fixes this reserve price with or without certain moderation and then offered as the final reserve price for auction. That would explain why the predicted values were not able to match the TRAI values.

The value of other variables like GSDP per Capita and population figures used for prediction also can influence the predicted price. It should be observed that any deviation from their true values during the data collection can infuse unsolicited errors in the equation. Such errors can also change the predicted figures. Especially GSDP per capita figure of an LSA had higher influence on the price as per the regression formula. Such variations also can cause the predicted value change.

In order to complement the regression process, two popular Machine Learning methods were also employed to predict the price for each LSA. Their accuracy of prediction is good. The same dataset was used to predict as that was used for Regression.

The results from both of these studies shows a positive and linear relation with the predictors. Can these values be substituted as the reserve price for the next auction? The data from the regression can certainly be used for future auction because, these values are derived through statistical method which were found to be robust. The only difficulty that would arise in such exercises as was done here is ensuring the accuracy and totality of the data. Outliers and missing data can give erroneous regression line resulting in poor prediction. It is interesting to observe that in the auction data in 47 out of 88 instances the reserve price without any escalation became the final price. That point to the fact that the prices obtained in the prediction can serve as the next auction reserve price.

The induction of a category variable representing the Class of LSA in order to delineate the high spectrum cost LSA from low spectrum cost LSA s and log transformations before the regression should have minimised misspecification errors. But in a regression, the essence is data and the trends in the data brought out by the process what help us in predicting the future. If such trends are carefully captured in the final equation, these values could be used to predict the future prices after feeding the correct input.

14.Future Study: -

The study that was undertaken in this paper is at a basic level using the traditional regression route, not any 'radical' method. It may be simplistic, but the principle can be extended to a more complex data oriented process and evolve into a robust method of prediction. The 2012 data that we have used is extracted from the 'expert's report' since TRAI has not done any valuation in that year. While filling the columns for the 'C' class LSA, some extrapolation was also used. Limited data set is one of the limitation of any study, especially in studies involving prediction where (lack of) data is a critical factor. The prediction value that we have fund can be verified with the true values in an actual auction whenever it happens. In general, more number of auctions can provide us with more and more data points thereby increasing the prediction accuracy and robustness. The association that was found in this study might change to a new association when more data are added and the test is repeated. The point is that when the sample data and sample size changes, the regression output also can change. A statistical test when done over and over again, with different samples can give astonishing result. The apprehension of the goodness of fit crossing into a territory of overfitting model also would be reduced with more data and more and more testing.

1800 MHz spectrum is the key band used in almost all mobile services. Once we know the price of this important band, price of other less frequently used bands could be adjudged on the basis of the 1800 MHz band price in a similar fashion that TRAI has perfected in their recommendations.

There is further scope, to improvise the study by adding more variables that might further enhance the outcome of the study. Some of the variables were already mentioned earlier which could not be included in this study. If the regression can be repeated by including them, it might explain the proportion of variance more accurately. Another possibility is including the interaction terms of the TRAI variables in the equation to study their effects. The non-linear effects can be studied by including the quadratic terms of the regressors, but conducting the linear regression. Like that, there are various possibilities to experiment and improve the results.

TRAI can also come out with a variable or indicator like '***spectrum scarcity factor***' based on many other telecom indicators or other ITU norms as prescribed for spectrum demand. Such a 'ready reckoner' figure would be able to explain the demands in each

LSA and relate to easily with regard to the demand comparing to other countries. This factor, which can be a complex figure need to be regularly updated to maintain its dynamicity. May be this factor can be related to other telecom indicators like AGR, SUC, MOU, GDP, GINI index etc. or in other words it could become a leitmotif in the spectrum universe. In an ideal world if such an index can be created, test like regression is not required.

Another factor could be 'band value' or 'band weightage' assigning a band value to distinguish among the different bands in terms of its utility. By knowing a band value, for example it may become easy to estimate the valuation of a TSP or deduce the band value of a band if you know the band value of another band.

Tail: The code used in this paper is annexed here so that it can be used by others to critically examine the finding and also to improve the test with addition of other variables which may lead to a more robust model to predict the price of spectrum accurately. It may also help spin off the study in other directions with innovative ideas for creative and pioneering work.

15. Reference.

1. Telecom Regulatory Authority of India (https://www.trai.gov.in/)

i) Recommendations on Spectrum Management and Licensing Framework, dated 11th May, 2010.

ii) Report on the 2010 Value of Spectrum in the 1800 MHz Band, dated January 30, 2011.

iii) Recommendations on Auction of Spectrum, dated 23rd April, 2012.

iv) Recommendations on Valuation and Reserve Price of Spectrum, dated 9th September, 2013.

v) Recommendations on Valuation and Reserve Price of Spectrum: Licenses Expiring in 2015-16, dated 15th October, 2014.

vi) Recommendations on Valuation and Reserve Price of Spectrum in 700 MHz, 800 MHz, 900 MHz, 1800 MHz, 2100 MHz, 2300 MHz and 2500 MHz Bands, dated 27th January, 2016.

viii) Recommendations on Auction of Spectrum in 700 MHz, 800 MHz, 900 MHz, 1800 MHz, 2100 MHz, 2300 MHz, 2500 MHz, 3300-3400 MHz, 3400-3600 MHz Bands dated 1st August 2018.

2. Basic Econometrics – Damodar N Gujarati /Sangeetha.
3. www.coursera.com.- OLS regression fundamentals.
4. Conference Paper "Application of multiple regression analysis to the prices of the spectrum in the IMT band" by Insua, Manuel; Frias, Zoraida; Pérez Martínez, Jorge
5. Estimation of Commercial Value of Spectrum: The Approach Adopted in Thailand: Settapong Malisuwan, Jesada Sivaraks, Thitipong Nandhabiwat, Navneet Madan, and Pannakorn Laokulrat.
6. www.coursera.com.- Machine learning fundamentals.
7. Exploring the Value and Economic Valuation of Spectrum *April 2012 prepared for ITU by John Alden, Vice President, Freedom Technologies*
8. Liberalisation of spectrum in the 900MHz and 1800MHz bands ; ComReg Document

Number: 09/99c 21 December 2009
9. Spectrum value of 800MHz, 1800MHz and 2.6GHz A DotEcon and Aetha Report for Ofcom July 2012.
10.A short analysis of Spectrum Auction in India –V.J. Christopher, 2017.

16. Abbreviations: -

AGR-Aggregated Gross Revenue.

AP- Auction Price.

DCF- Discounted Cash Flow.

DoT- Department of Telecom.

GOVT- Government.

GSDP- Gross State Domestic Product.

LSA- Licensed Service Area.

MCLR-Marginal Cost of Lending Rate.

ML- Machine Learning.

NPV- Net Present Value.

OLS- Ordinary Least Square

PF- Production Function.

PLR- prime Lending Rate.

PP- Previous auction Price.

PS- Producer Surplus.

RP- Reserve Price.

TC- Telecom Commission

TRAI- Telecom Regulatory Authority of India.

WPC- Wireless Planning and Coordination.