INDIAN INSTITUTE OF TECHNOLOGY

KANPUR

SOLAR RADIATION OVER INDIA

PANEL ANALYSIS

ECO342A: APPLIED ECONOMETRICS

SUBMITTED TO

DR. SOMESH K. MATHUR

SUBMITTED BY

ANSHUL GOEL (150110)

DATE OF SUBMISSION: April 18, 2018

INTRODUCTION & MOTIVATION

Estimation of solar radiation finds an importance in the wide domain of fields, from agriculture, construction to renewable energy to heating and cooling situations. Though the temperature, pressure and wind speed data are accurately available for the usage, solar radiation data is not available for every location and is required to be estimated. The accurate predictions of solar radiation holds a strong significance as it can help in overcoming the intermittency involved in solar power. There is a strong impetus given to renewable sources of energy especially solar energy. The main focus areas have been South Asia especially India and China. The demand is called for due to various environmental reasons and overcoming the dependence of oil and coal deficient countries like India on oil-coal imports.

Talking about the environmental concern, the major contributors to pollution are carbon dioxide emissions from automobiles and power sector. Talking about Power generation, electricity is an important ingredient in making development a reality. Generating electricity is the leading cause of pollution, emission of CO_2 in the environment is causing the problem of toxic pollutants in the air as well as chronic global warming. Nearly 80% of all Coal power plant emits carbon dioxide which is a Greenhouse gas. Considering the case of US will give a picture of impact coal has on the environment. Coal power plants produce less than half of the electricity in the USA but account for nearly 80% of power plant carbon emissions. The case though discusses the USA but is also relevant to the case of Asia. the rise of Asia was marked from the dawn of the 1990s with GDP of Asia becoming the largest in the world and incidentally, also became the largest emitter of Carbon. CO_2 emissions today from human activities today are higher than ever in human history.

Global CO_2 were 150 times more in 2011 than in 1850. This huge shift can be attributed to rapid economic development initiated by Europe and North America followed by Asia and Latin America. The concentration effect of emissions can be seen from the fact that top 10 emitters contribute 78% of global emissions, all of which are large economies with large population and energy consumption. As per reports suggested too much dependence on industries for economic development leads to exploitation of the resources. However, there is always a dichotomy between environment and economic development. Saving one side will give an impact on the other.

Thus, to maintain balance between economic development and environmental sustainabiliy, it becomes imperative to employ renewable resources of energy for power needs. India has been focus of my study as much work is needed to be done in the context of India and due to the geographical advantage India holds in terms of solar power. Since Independence, the focus of India has been on development. Although it has brought many benefits to India by boosting our economic growth, it is clear that we are far away from achieving the basic objectives of any society i.e. security of water, health and food. Over the last seventy years, India’s strong growth has been able to pull out much of our population from the poverty by increasing employment opportunities but this has resulted in degrading our environment too. Rapid economic development has led to

growing scarcity of our natural resources. Poverty and the degraded environment are closely interrelated because a major chunk of our society depends on the environment and its degradation has seriously affected our development in the long term. The impact of past industrialization and exploitation of environment shattered our dream to become an eco-friendly nation. Mrs. Indira Gandhi in the Stockholm Declaration of 1972 discussed the doctrine of "Sustainable Development". Thereafter, in 1987 a report was issued by the World Commission on Environment and Development in which they tried to link environment protection and economic development. Also, the Rio Declaration 1992 structured the principles of Sustainable Development. The main purpose of economic development is to provide a quality standard of living since industrial development creates more jobs than any other field which in turn increases the possibility of its adverse effects on the environment. Dust, smoke, and toxic gases originating from thermal power plants, mines and other factories make this environment hazardous to live. The United Nations Conference on Sustainable Development in 2012 issued 17 Sustainable Goals (SDGs) to tackle environmental economic and political challenges. Thus, India is an emerging economy have to maintain a strong balance between economic development and environment sustainability and to counter the intermittency of renewable sources of energy, accurate predictions are required not only for a particular location but an accurate estimate which can be applied PAN-India. Various measures have be taken by the government in this direction. India has launched Jawaharlal Nehru National Solar Mission with the target of generating 20,000 MW of Solar Power by 2022
This requires the data which is available for every location in India and also gives fairly accurate values to gain feasibility.

The literature suggests that the majority of work has been done by people of engineering background and I was unable to find any work on estimation by an econometrician, though it involves linear regression, Autoregressive processes, etc. This motivated me to choose this topic and apply my econometric tools. Another major factor for choosing the domain is that there has been work in time series and stationary aspects of the data but not in terms of panel analysis. I was not able to find any work with special focus on panel effects. This called for applying the panel data analysis which is performed by this paper.

OBJECTIVE

The objective is to estimate diffuse solar radiation over India by using panel data analysis The analysis involves data for 15 years from 2000-2014 on monthly basis. Thus, there are 180 time values available for each location. The analysis involve 23 locations all over India over which the data is available. The locations covers all the geographies of the country for which the data was accurately available. The wide range of locations include Srinagar in upmost North to Shillong in the most eastern parts to Ahmedabad in West to places like Thiruvananthapuram and Chennai in the south. Data has also been used for island states of Andaman and Nicobar (Port Blair) and Lakshadweep (Minicoy). As India is a land of such wide geographies, the work can be used to extend to multiple places in the world. Thus, this variation would be the ideal sample for the panel data analysis. The raw data was present for every hour for 15 years. Such an accurate prediction of hourly basis is better performed under the domain of Machine Learning and Time series and is very time consuming to get a cross sectional and time trend. Thus, the data was used which is month based which turns out to clearly satisfy the required goals.

DATA SET

· NSRDB (National Solar Radiation Data Base), a part of international activities of NREL (National Renewable Energy Lab, USA)
https://www.nrel.gov/international/ra_india.html
· MNRE (Ministry of New & Renewable Energy)
http://mnre.gov.in/file-manager/UserFiles/solar_radiant_energy_over_India.pdf

As the data from Indian Meteorological Departments is very expensive, I will employ the data from NREL which claims that the data have been collected at a sufficient number of locations and temporal and spatial scales to accurately represent regional solar radiation climates. The majority of work done in Indian context is through the data of MNRE (Ministry of New and renewable Energy) but it includes monthly average data for one representative year which was chosen to account for variations for atleast 15 years, thus the data is not rich enough. The NREL data consists of variation of most required parameters for 15 years on 30 min difference which is required for our estimations and predictions. As the work is based on engineering and geographical domain, I will try to briefly explain the content of relevance and will include the appendix for details, if required.

LITERATURE REVIEW

There was no literature available for the panel analysis of solar radiation over India or any other place. Though, while searching for it, I got to learn much about the whole domain and found the basis of my understanding and current work, thus suiting to the definition of literature review.

The research paper "**Predicting monthly mean daily diffuse radiation for India**" by Indira Karakoti , Prasun Kumar Das and S.K. Singh formed the very basis of my work. The paper is called the base of my work as it highlights the importance of my work as "The resource assessment is the most important exercise toward project evaluation. Due to intermittent nature of renewable energy sources, the resource assessment makes major impact on the overall viability of the project. Long term solar radiation and meteorological data (especially in Typical Meteorological Year format) is the most preferred data in large scale solar energy projects. Due to unavailability of long term measured data of solar radiation and other climatic parameters in India; the project developers are facing lots of problem towards financial closure of the projects and impact of uncertainty of re source on bankability of the projects." The monthly average daily

diffuse transmittance (ratio of diffuse-to-extraterrestrial solar radiation) which is the diffuse radiation predicted in my work. It was found to be correlated with sunshine fraction, temperature and relative humidity through the linear and quadratic equations for India. The data for sunshine fraction was not free and hence, I could not employ that factor but used other factors like solar zenith angle to counter to the deficiency.
The data of 18 stations among the 23 stations are combined and employed pooled OLS for the estimation of diffuse radiation. This should be kept in mind that they have taken the monthly average values for one particular representative year. Thus, they used only 12 values for one location while I employed 180 locations for each location, thus creating a richer data set.

Second paper of interest, "Time series analysis of hourly global horizontal solar radiation" by J.M. Gordon and T.A. Reddy, focussed on time series of analysis. I used a dynamic panel to get the approximate effect but the paper provided insight into the time trend of global horizontal solar radiation while I employed to work on diffuse horizontal radiation. The paper discussed sequential and stationary characteristics which focuses on hour-to-hour variation and individual dependent effects respectively. Thus, the paper also tried to get a cross sectional and time trend implications which I captured through panel analysis.
Accurate information on sequential and stationary properties of hourly solar radiation for systems with memory (small power backups for solar energy generating systems with buffer, certain solar absorption AC systems, certain control problems in solar building).
The aim of the paper to determine an analytical model, otherwise these things are done with large scale calculations. The paper was interesting in the sense that very few considered sequential of hourly data like this paper but majority of papers focus on longer time frames of daily, monthly etc. The paper includes many interesting implications, for example, sequential characteristics of wind speed, air temperature, relative humidity have strong correlation at the one-hour time lag. The database covered climatic condition from tropical to temperate which is also true in my work too. The objective behind discussing sequential properties was to generate and examine hour-to-hour variation in solar radiation, persistent times and persistence strengths. One may want to have different auto-correlational relations for different hours of the day, but it was found not to vary much. So, can ignore hour-to-hour variation in rho(d) and use the mean of all hours of the day. Though it seems that the whole day hours have impact, but they may be due to 1 hour lag effect using partial autocorrelation function. Though in majority, lag 2 can be ignored but not in 0.05 significance.

Finally, the third paper that I particularly want to mention is "Improved statistical procedures for the evaluation of solar radiation estimation models" by R.J. Stone. The majority of the statistical work done in the field of solar radiation estimation depends on Root Mean Square Error (RMSE) and Mean Bias Error (MBE) as defined below:

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} {d_i}^2 \right)^{\frac{1}{2}}$$

$$MBE = \left(\frac{1}{n} \sum_{i=1}^{n} d_i \right)$$

Where d_i is the difference between true and predicted value

The paper discusses the use of RMSE and MBE and also introduce the t statistics, its significance and how it can be related to MBE and RMSE to get better gauge for models. RMSE is used to check for short time performance performance of error but suffer from some drawbacks like (i) few large errors in the RMSE expression can produce a significant increase in RMSE, (ii) does not differentiate between under and over-estimation as it is the whole square of errors. MBE is used to gauge the long term performance of the model but suffers from the fact that overestimation of one is cancelled by underestimation of other is it doesn't involve absolute terms and is sought to work like a moving average. There are also drawbacks in their combined use as they are conflicting values to each other which calls for requirement of other instruments which is a subjective and time consuming process. Second issue is that RMSE and MBE have dimensional values. The solution came out by applying T statistic which can be defined in terms of RMSE and MBE as follows:

$$t = \left(\frac{(n-1)MBE^2}{RMSE^2 - MBE^2} \right)^{\frac{1}{2}}$$

Smaller the value of t-statistic better is the performance.

Though I have mentioned only 3 particular papers in literature review, these are not the only papers I considered but just the most important papers. The comprehensive list is given in references.

WORKING MODEL

$$(H_D/H_0)_{it} = \alpha + \beta_1 T_{it} + \beta_2 P_{it} + \beta_3 (R_h)_{it} + \beta_4 (W_s)_{it} + \beta_5 (W_D)_{it} + \beta_6 (\theta_z)_{it} + \beta_7 D_{it} + \beta_8 (T^2)_{it} + \beta_9 (P^2)_{it} + \beta_{10} (R^2{}_h)_{it} + \beta_{11} (W^2{}_s)_{it} + \beta_{12} (W^2{}_D)_{it} + \beta_{13} (\theta_z{}^2)_{it} + \beta_{14} D^2{}_{it} + \beta_{15} Lat_i + \beta_{16} Long_i + \varepsilon_{it}$$

where

H_D is diffuse solar radiation

H_0 is the global extraterrestrial radiation

T_{it} is the temperature

P_{it} is the pressure

$(R_h)_{it}$ is the relative humidity

$(W_s)_{it}$ is the wind speed

$(W_D)_{it}$ is the wind direction

$(\theta_z)_{it}$ is the solar azimuth angle which is the angle made by the horizontal projection of Sun from the North-South line

D_{it} is the dew point

Lat_i is the longitude of the location

$Long_i$ is the longitude of the location

I sequentially worked on the models of estimation, starting with pooled OLS and then to Fixed Effect and Random effect which gave Fixed Effect model using Hausman Test. Then, I checked for time-fixed effects, i.e. whether the time dummies for all years are required by using the null hypothesis that all time dummy terms are zero which came was rejected. Then, I checked for dynamic panel which came out to be the case but upto lag 1. So, I used dynamic panel model and used Arellano-Bover estimation. To check for Adjusted R Square, I stored the predicted rel_dhi and regressed it on dependent variable. Thus, I got an adjusted R Square of **0.93**. The procedure have been explained in detail in Methodology.

METHODOLOGY & EMPIRICAL RESULTS

The data comprises for 23 locations in India and has been assigned a number as follows:

Ahmedabad	1	Jodhpur	9	Pune	17
Bangalore	2	Kolkata	10	Ranchi	18
Bhavnagar	3	Minicoy	11	Shillong	19
Bhopal	4	Mumbai	12	Srinagar	20
Chennai	5	Nagpur	13	Thiruvananthapuram	21
Goa	6	New Delhi	14	Varanasi	22
Hyderabad	7	Patna	15	Visakhapatnam	23
Jaipur	8	Port Blair	16		

The data is monthly wise for 15 years from year 2000 to 2014, thus 180 values for each location for each variable.

I first tried to check between the Fixed Effect, Random Effect and Pooled OLS. So, I followed the technique taught in the course ECO342A as shown in figure 1.

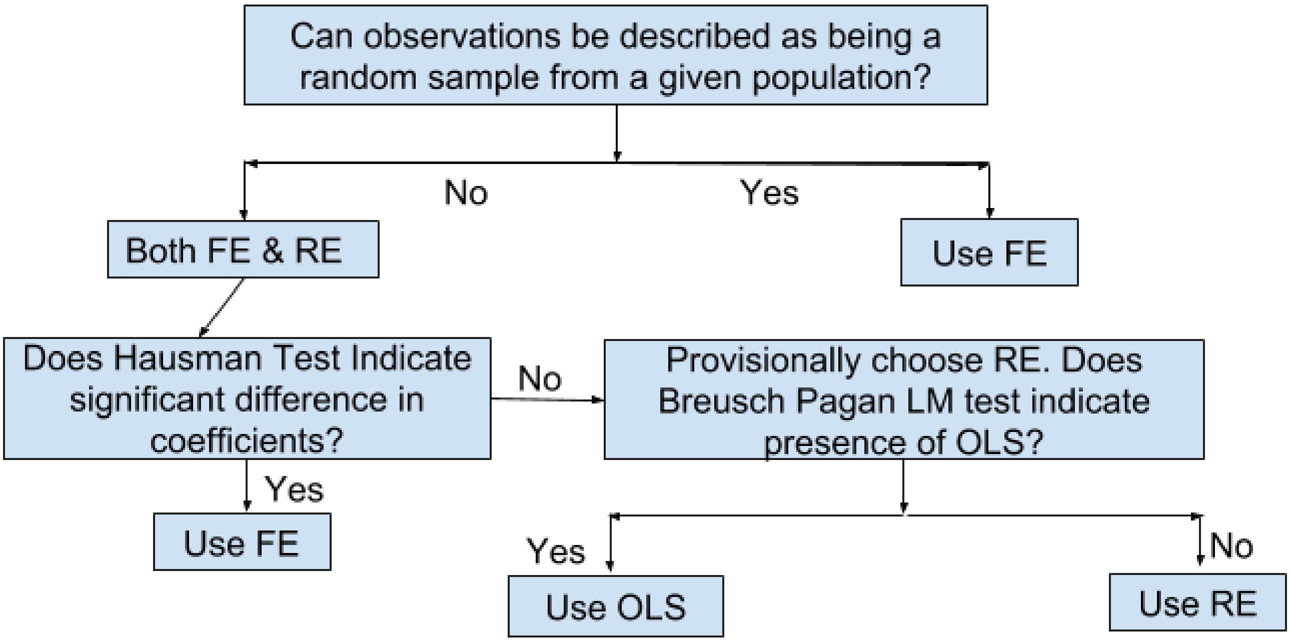

Figure 1

Hausman Test gave fixed effect with P> chi2 = 0.0000. Thus, we checked later between Fixed effect and pooled OLS using F test in which Prob > F = 0.0000. Thus, Fixed effect was used.
Next step was to check for time fixed effects.
To test for time fixed effects, I run the fixed effect estimate but this time include time dummy variables as covariates. Then, I employed the f test to determine if all the dummies for time are simultaneously zero using the command **testparm**. Thus, I ran the command **testparm i.time** after the fixed effect regression and found that Prob > F = 0.0000. Thus, we need to include time effects. Also, the thing remained now was to think about the dynamic nature of panel. Thus, to get an idea of some possible correlation of diffuse radiation with its past values, I checked the correlation with lagged 1 and lagged 2 values and found out to be 0.8293 and 0.6192 respectively.This hinted towards dynamic panel as a result, I ran Arellano Bover estimates. I stored the estimate as rel_yhat and checked the fit with our dependent variable, rel_dhi using regression command:
Regress rel_dhi rel_yhat
This gave a high adjusted R square of 0.9297. I checked Arellano-Bover estimates for lag 2 also but the lag 2 term came out to be insignificant.

Thus, we employed Arellano Bover estimate to the model which gave a good fit with adjusted R^2 of 0.93.

The coefficients of explanatory variables are given in the table:

Variable	Coefficient	Standard Error	P > |z|
L1.rel_dhi (lag of dependent variable)	.1822735	.0092349	0.000
Lat (latitude)	.2031326	.0537415	0.000
D (longitude)	.1822497	.0634645	0.004
Monthly_dew (monthly average of dew point)	-2.382772	.1565837	0.000
Monthly_pre (monthly average of pressure)	.3228634	.1408564	0.022
Monthly_tem (monthly average of temperature)	-2.636447	.1814493	0.000
Monthly_wsd (monthly average of wind speed)	-6.09804	.5843516	0.000
Monthly_rhd (monthly average of relative humidity)	.5239126	.1057516	0.000
Monthly_sza (monthly average of solar azimuth angle)	8.860246	.681355	0.000
Monthly_wdn (monthly average of wind direction[angle])	-.0580452	.0114129	0.000
Monthly_dew_sq (square of monthly average of dew point)	.0304323	.0016646	0.000
Monthly_pre_sq (square of monthly average of pressure)	-.0001455	.0000799	0.069
Monthly_tem_sq (square of monthly average of temperature)	.0290083	.0018772	0.000
Monthly_wsd_sq (square of monthly average of wind speed)	.7849555	.0740541	0.000
Monthly_rhd_sq (square of monthly average of relative humidity)	-.0048812	.0005422	0.000
Monthly_sza_sq (square of monthly average of solar azimuth angle)	-.0565826	.0038498	0.000
Monthly_wdn_sq (square of monthly average of wind direction[angle])	.0002445	.0000302	0.000
_cons (constant term)	-433.266	68.56696	0.000

Figure 2 helps us to understand the fit of values.

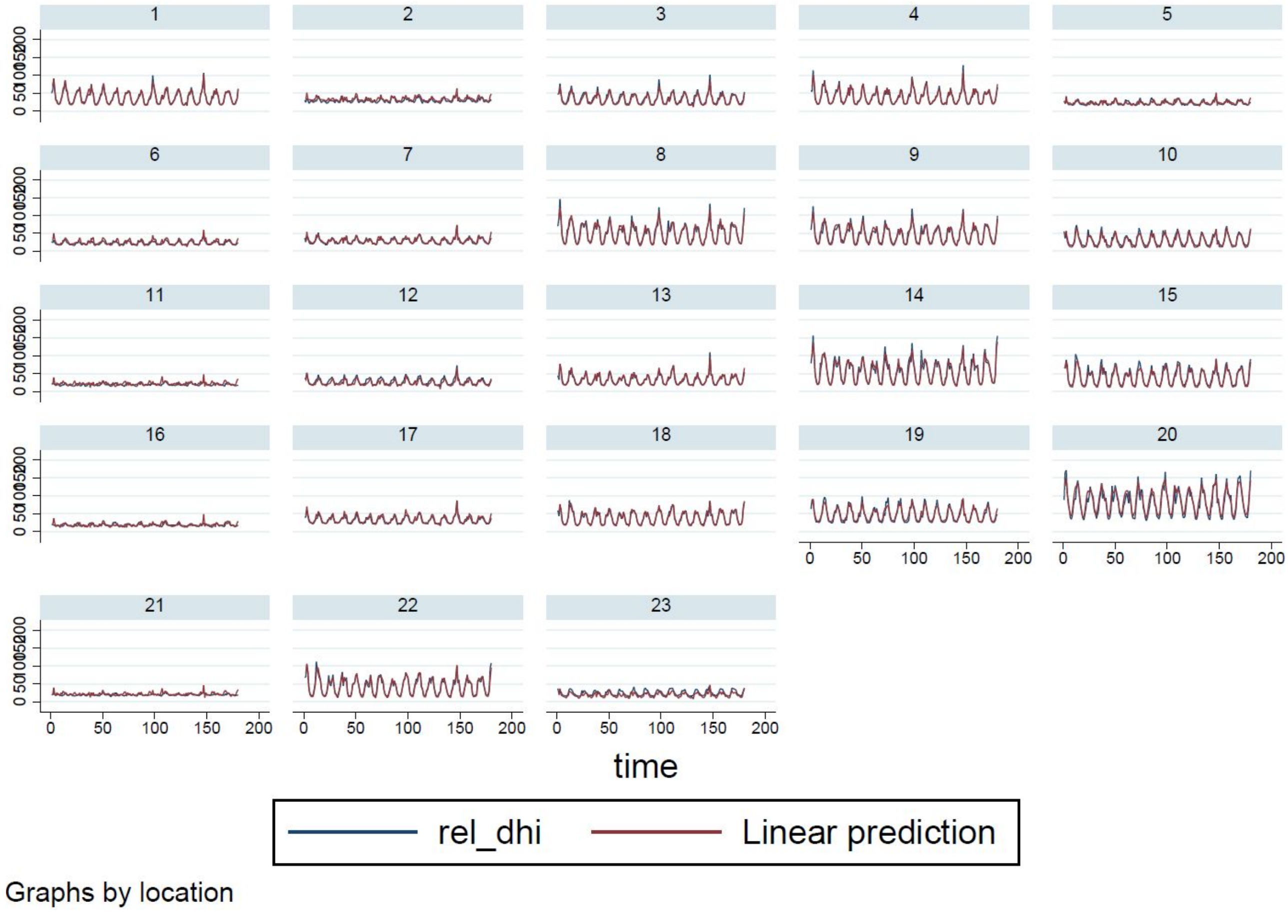

Figure 2

Scatter plot is shown in figure 3. To see the fit for a single year, we checked the fir for year 2000 (in-sample) to get an idea as shown in figure 4.

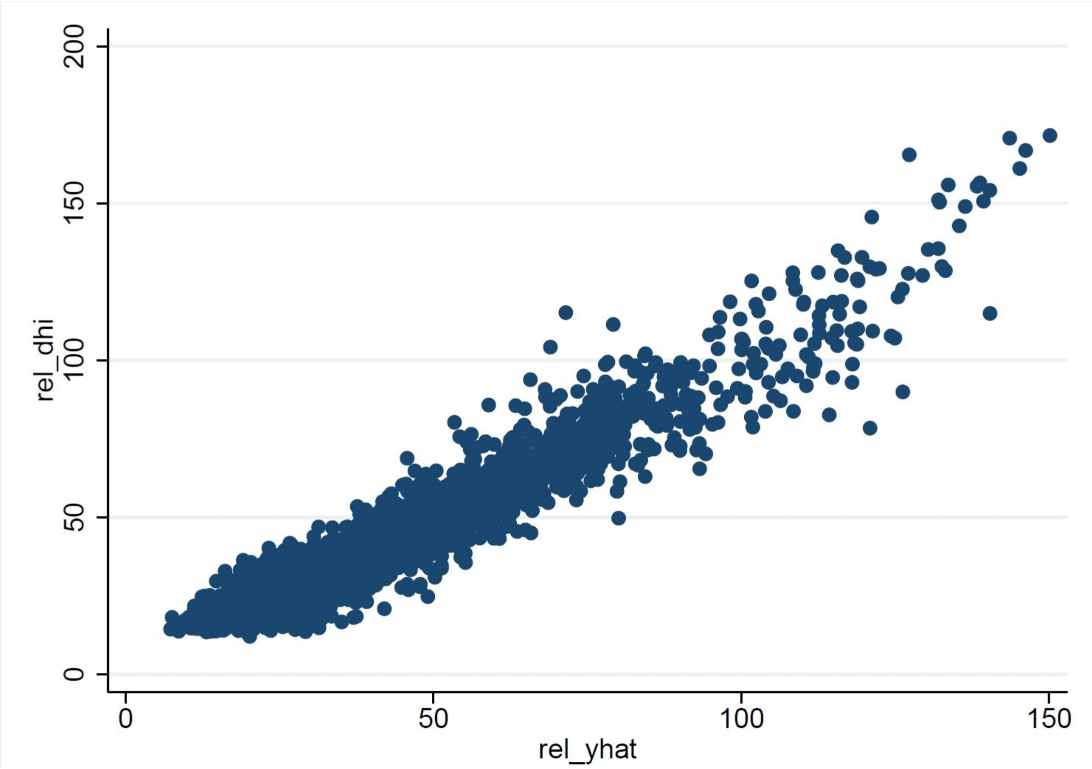

Figure 3

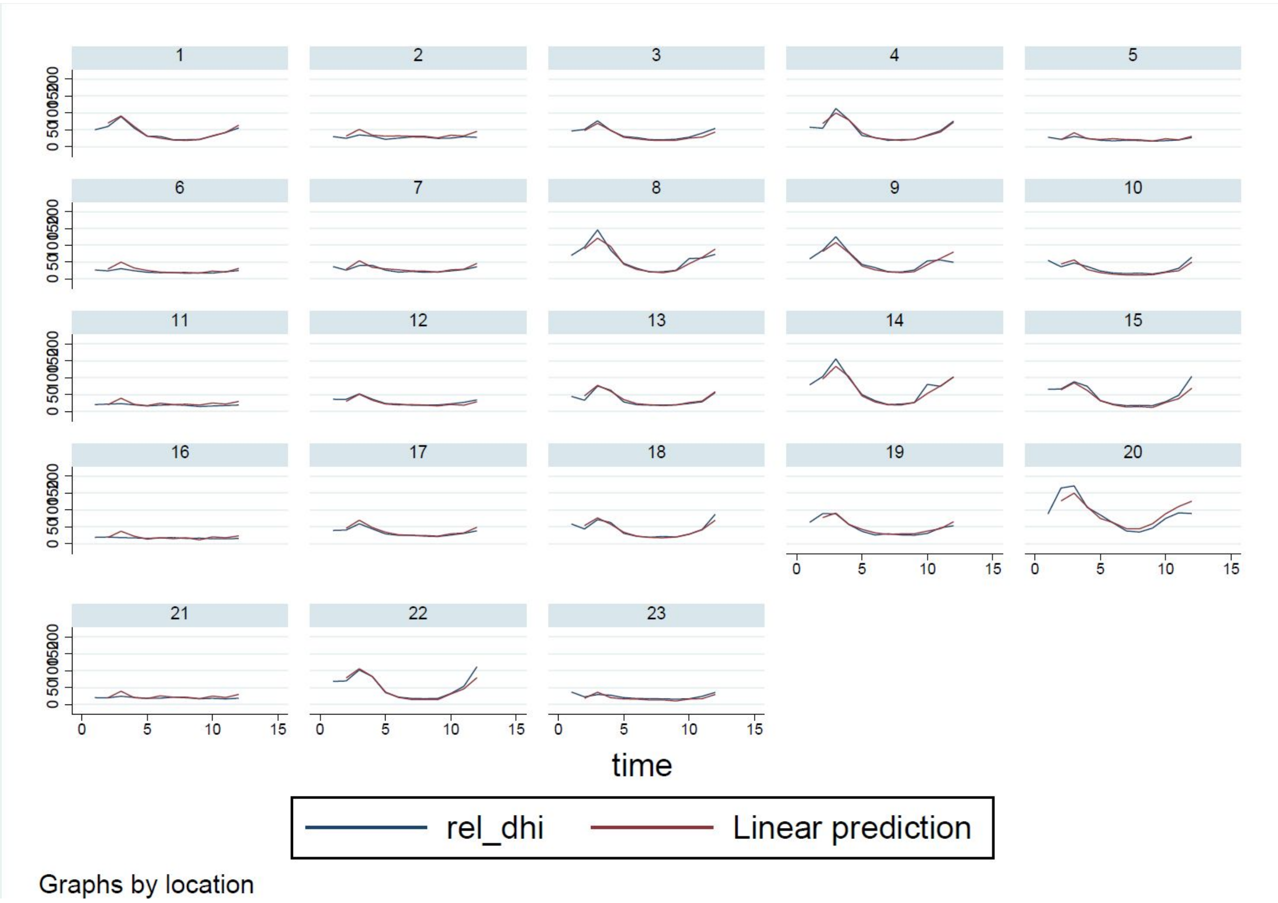

Figure 4

CONCLUSION

In this paper, I tried to work on estimation of solar radiation over India through panel data approach. I worked from the most basic model and gradually turned to more advanced models, inculcating various features and assumptions like checking for time-fixed effects, stationary or dynamic panel analysis, etc. At the end, the model is able to give a good fit with adjusted R^2 of 0.93 which is quite how. As per my knowledge goes, this is the first ever work done on solar radiation estimation using panel data and gives a great direction to further explore the field owing to its feasibility. The data used in the model is readily available for every location and thus is universal. The model covered various climate zones and other variations in geography. Also, the period of testing is also quite significant with rich monthly based data for every year. Though the results seem encouraging, the work can be extended to work further, including more assumptions and features, checking for various other issues and employing the use of Neural networks or Bayesian methods for estimation which is the inherent motive behind this work.

References

- Karakoti I, Das P.K. and S.K. Singh. 2011. “Predicting monthly mean daily diffuse radiation over India”. *Applied Energy* 91 (2012): 412-425

- Stone R.J. 1993. “Improved Statistical Procedure for the Evaluation of Solar Radiation Estimation Models”. *Solar Energy* Vol. 5 I: No. 4: 289-291

- Gordon J.M and T.A. Reddy. 1988. “Time Series Analysis of Hourly Global Horizontal Solar Radiation”. *Solar Energy* Vol. 41: 423-429

- Perdomo R, Banguero E. and G. Gordillo. 2010. “Statistical Modelling for Global Solar Radiation in Bogota”. *IEEE* 978-1-4244-5892-9/10

- Reyna Oscar. 2007. “Panel Data Analysis: Fixed and Random Effects using STATA”. Presentation. Princeton University

- Cameron A. and Trivedi P.K. 2009. “Microeconometrics using STATA”. Stata Press

- Course Curriculum: ECO342A, Indian Institute of Technology, Kanpur

Appendix

The STATA codes for whatever simulation is done:

```
. import excel "E:\Monthly_new\monthly_only_loc_new_sq.xlsx", sheet("com") firstrow

. destring time, replace

. xtset location time

. xtreg rel_dhi lat D monthly_dew monthly_pre monthly_tem monthly_wsd monthly_rhd
monthly_sza monthly_wdn monthly_dew_sq monthly_pre_sq monthly_tem_sq monthly_

> wsd_sq monthly_rhd_sq monthly_sza_sq monthly_wdn_sq, fe

. estimates store FE

. xtreg rel_dhi lat D monthly_dew monthly_pre monthly_tem monthly_wsd monthly_rhd
monthly_sza monthly_wdn monthly_dew_sq monthly_pre_sq monthly_tem_sq monthly_

> wsd_sq monthly_rhd_sq monthly_sza_sq monthly_wdn_sq, re

. estimates store RE

. hausman FE RE

* use FE

* FE vs OLS

* generating dummies

. tabulate location, generate(l)

. regress rel_dhi lat D monthly_dew monthly_pre monthly_tem monthly_wsd monthly_rhd
monthly_sza monthly_wdn monthly_dew_sq monthly_pre_sq monthly_tem_sq monthl

> y_wsd_sq monthly_rhd_sq monthly_sza_sq monthly_wdn_sq l1 l2 l3 l4 l5 l6 l7 l8 l9 l10 l11 l12
l13 l14 l15 l16 l17 l18 l19 l20 l21 l22 l23

. test  l1 l2 l3 l4 l5 l6 l7 l8 l9 l10 l11 l12 l13 l14 l15 l16 l17 l18 l19 l20 l21 l22 l23

* use FE

** testing for time fixed effects

. ** fe with time dummies

. xtreg rel_dhi i.time lat D monthly_dew monthly_pre monthly_tem monthly_wsd monthly_rhd
monthly_sza monthly_wdn monthly_dew_sq monthly_pre_sq monthly_tem_sq m

> onthly_wsd_sq monthly_rhd_sq monthly_sza_sq monthly_wdn_sq, fe

. testparm i.time
```

```
* time effects are needed

. xttest3

* presence of heteroscedasticity

** checking for the dynamic panel by running arellano bover estimates with time dummies

** taking the idea of dynamic by using correlate

. correlate rel_dhi L.rel_dhi

. correlate rel_dhi L2.rel_dhi

** switch to next data for getting arellano bover estimates and other procedures

. clear

. import excel "C:\Users\Anshul Goel\Downloads\data_prashast.xlsx", sheet("Sheet1") firstrow

. xtset location time

. ** arellano bover estimate is rel_yhat

. regress rel_dhi rel_yhat

. xtline rel_dhi rel_yhat

. xtline rel_dhi rel_yhat if time < 13

. ** for 1 year
```

Please find the attached STATA outputs for whole process.