ANALYSIS OF TIME SERIES

FEMALE LABOUR FORCE PARTICIPATION RATE IN INDIA

Team Members:

1. Ananya Gupta (544)
2. Isha Agrawal (469)
3. Ishita Agarwal (528)
4. Kriti Maheshwari (474)
5. Pragya Raj (921)

ACKNOWLEDGEMENT

We would like to express our special gratitude to Mrs. Renu Kaul, of the Department of Statistics for providing us the great opportunity to work on the analysis of the *Female Labour Force Participation Rate in India*. Your valuable teaching and guidance helped us in various phases of the completion of this project. We will always be thankful to you in this regard.

We would also like to sincerely thank our principal, Dr. Suman Sharma of Lady Shri Ram College, Delhi University for providing all the required facilities in completion of this project.

AGENDA

- Introduction to the Problem Statement
- Setting the Hypothesis
- About the Data

- Methodology: Overview of the Analysis
- Measurement of Trend by Method of Curve Fitting
- Measurement of Trend by Moving Average Method & Spencer’s Formula
- Forecasting by Simple Exponential Smoothing
- Comparison of Models

- Results and Inference
- Limitations
- References

SECTION I

INTRODUCTION TO THE LABOUR FORCE TRENDS IN INDIA:

FOCUSING ON WOMEN PARTICIPATION

Introduction to the Problem Statement: Understanding Female Labour Force Trends in India

With the ongoing placement season, we were interested to know about the change in labour participation of women in India

This is against the backdrop of:

1. **Fertility rates have declined** steadily over the past several decades, which should support increased female participation in the labour market
2. **Public programmes** promoting employment in rural areas (Mahatma Gandhi National Rural Employment Guarantee Act, **(MGNREGA)**) in 2005
3. **Increasing female population**

Female population		**Fertility rate for women**	
1990: **420 M**	2021: **669 M**	1990: **3.9**	2021: **2.1**

On the flip side, there is evidence for decline in female labor force as:

1. The official reports suggest that the overall decline in participation is due to an **increased attendance of young adults** in educational institutions
2. **Increased household income**, which reduces the need for female labour and deems them economically inactive
3. Insufficient job opportunities for women compared to rise in working population leads to:
   - i. **increased competition** with men for scarce jobs
   - ii. increasing reluctance of women to take up **informal work**

Introduction to the Problem Statement: Setting the Hypothesis

Introduction ▶ Hypothesis Setting

Keeping our current observations and our personal experiences as evidence, we form the hypothesis as follows:

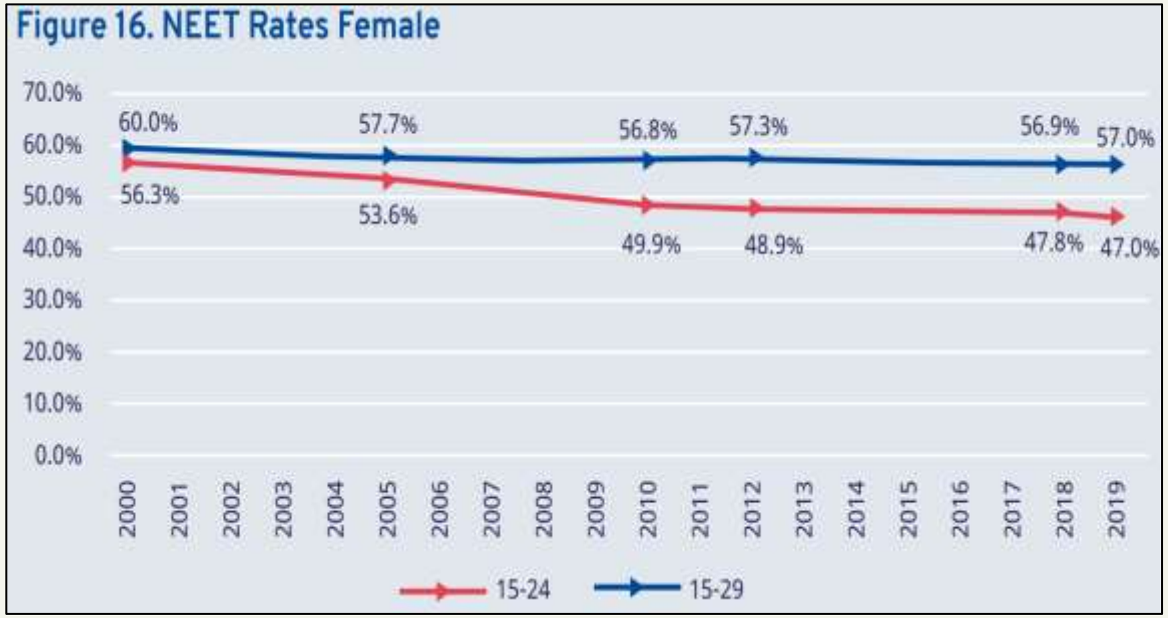
H_o: There has been a significant increase in the female participation in the labour force in India

H_1: There has been a significant decline in the female participation in the labour force in India (**Composite hypothesis**)

Given below are 2 graphs to show NEET (**Young Persons Not in Employment and Education in India**: 2000-2019) Gender Divide:

Graph 1.1

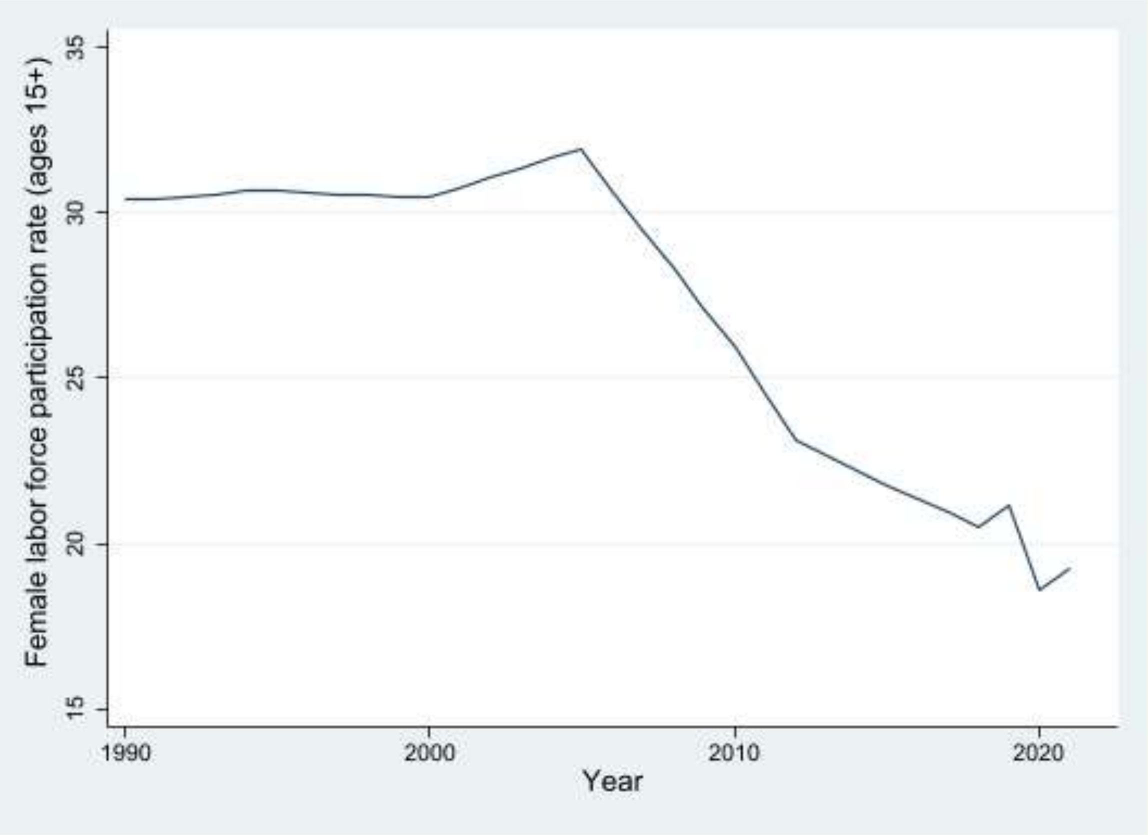
Graph 1.2



Introduction to the Problem Statement: About the Data

The data on 'Female labor force participation' used in our analysis has been collected from the Gender Data Portal of the World Bank, as published on January 10, 2022. The data has further been used in the International Labor Organization's (ILO) research brief.

Graph 1.3



1. The data for the Labor force participation rate, female (% of female population aged 15+) for the years **1990 – 2021** has been used to conduct the analysis.

2. Labor force participation rate is the proportion of the population ages 15 and older that is economically active: all people who supply labor for the production of goods and services during a specified period.

3. Graph 1.3 represents the values of the female labor force participation rate, in India against time (years).

4. The graph shows a downward trend i.e. there has been a decreasing female participation in the labour force in India, with a steep decline seen after the year 2006. A similar decline is observed in the year 2019-20 due to the effects of the pandemic.

5. We will now analyze the data by attempting to estimate the trend, fit the curve and further forecast the value for the year 2022; understanding the reasons behind the decline in the past.

SECTION II

ANALYSIS OF THE FEMALE LABOUR PARTICIPATION RATE IN INDIA:

UNDERSTANDING THE DECLINE AND FORECASTING THE FUTURE

Methodology: Overview of the Analysis

The analysis can be broadly divided into three approaches:

Curve Fitting (SPSS)

1. We attempt to fit the following curves to the data: Linear, Quadratic, Cubic, Logarithmic, Growth and Exponential.
2. We further compare the R^2 of all the curves to determine the model of best fit.
3. Using the best fit curve, we find the predicted values of women participation in labor force for the year 2022.

Moving Average & Spencer's Formula (Excel)

1. We plot an ACF graph to find the significant peaks in the data and determine the parameter of the MA model.
2. Using the method of iterated averages, taking the extent of moving average from above, we calculate the estimated values to smoothen the curve.
3. We further use Spencer's 15-point and Spencer's 21-point formula to estimate the trend values.

Exponential Smoothing (STATA)

1. We determine the optimum smoothing factor (alpha) by minimizing the sum of squared errors.
2. Using the smoothing factor, we assign relative weights to the observations and hence plot the smoothened curve.
3. This curve is then used to forecast the value of the successive year, 2022.

We compare the forecasted values from all the three methods to determine the most suitable approach

Measurement of Trend: Method of Curve Fitting

The following type of curves may be used to fit the given data:

- **1st Degree Polynomial Curve (Linear Curve):** $Y_t = a + bt$
- **2nd Degree Polynomial Curve (Quadratic Curve):** $Y_t = a + bt + ct^2$
- **3rd Degree Polynomial Curve (Cubic Curve):** $Y_t = a + bt + ct^2 + dt^3$
- **Exponential Curve:** $Y_t = a \,.\, b^t$
- **Logarithmic Curve:** $Y_t = a + b\,(\log t)$
- **Growth Curve:** $Y_t = e^{(a + bt)}$

Model Summary of different curves and their Parameter Estimates

Dependent Variable: wParticipation

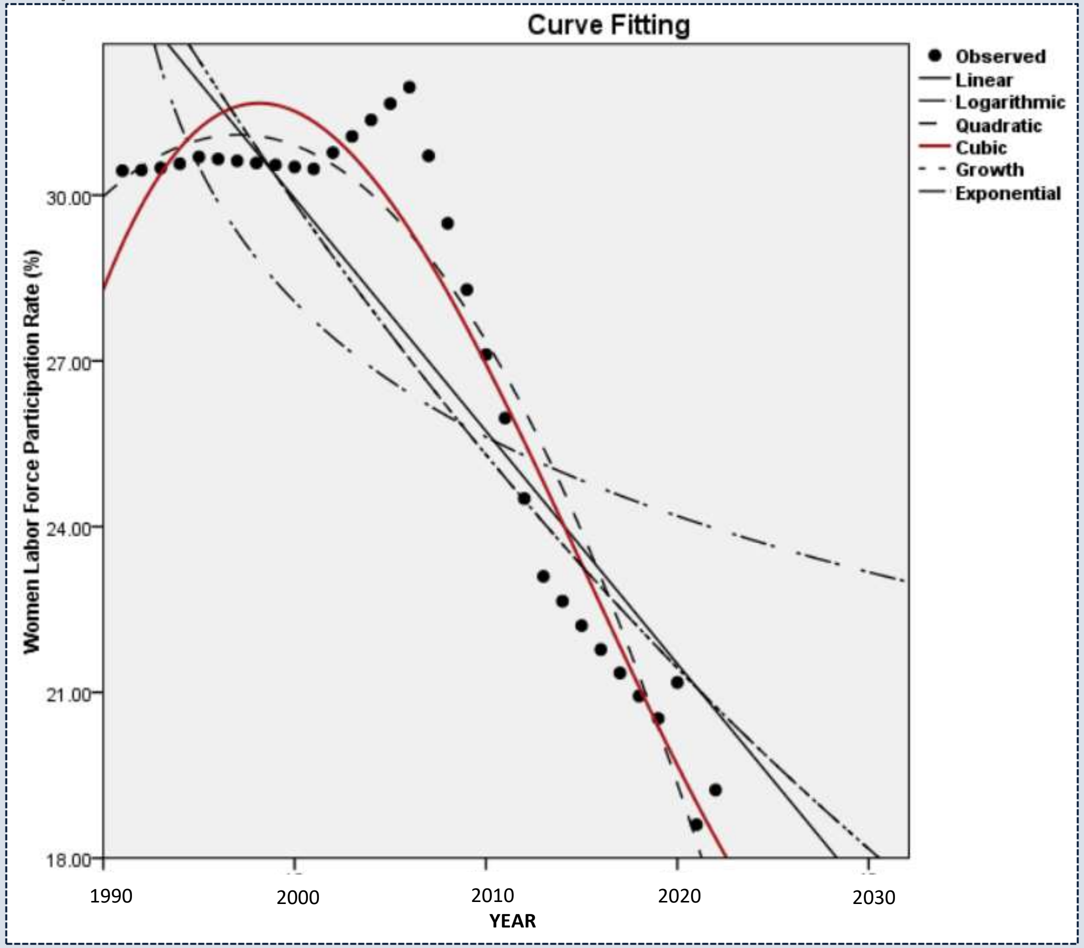
Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.778	105.338	1	30	.000	34.154	-.421		
Logarithmic	.455	25.075	1	30	.000	36.204	-3.532		
Quadratic	.927	185.240	2	29	.000	29.973	.317	-.022	
Cubic	.942	150.635	3	28	.000	28.293	.884	-.065	.001
Growth	.777	104.299	1	30	.000	3.562	-.017		
Exponential	.777	104.299	1	30	.000	35.244	-.017		

Measurement of Trend: Method of Curve Fitting

Analysis ▶ Curve Fitting ▶ Graphical Representation

The following figure represents the fitted curves of different types for the given time series data:

Graph 2.1



Inferences from the graph:

- The **quadratic curve** and the **cubic curve** visually highly resemble the actual values.
- From the curve representing actual observed values, we can observe that there has been a **decline in the Women's Labour Force Participation Rate after 2006**.
- Further, to statistically compare the best fit of the competing models, we compare the values of R^2, which has been tabulated in the following slide.

The fitted equation for the quadratic curve is given by:

$Y_t = 29.973 + 0.317 * x + -0.022 * x^2$

The fitted equation for the cubic curve is given by:

$Y_t = 28.29302465 + 0.88433028 * x + -0.064718061 * x^2 + 0.00085575014 * x^3$

Measurement of Trend: Method of Curve Fitting

Analysis ► Curve Fitting ► Comparing the Models

Fitting of Quadratic Curve

Quadratic

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.963	.927	.922	1.248

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	576.909	2	288.454	185.240	.000
Residual	45.158	29	1.557		
Total	622.067	31			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Case Sequence	.317	.099	.663	3.212	.003
Case Sequence ** 2	-.022	.003	-1.593	-7.717	.000
(Constant)	29.973	.705		42.490	.000

R2 = 0.927, Adjusted R2 = 0.922

Fitting of Cubic Curve

Cubic

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.970	.942	.935	1.139

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	585.773	3	195.258	150.635	.000
Residual	36.295	28	1.296		
Total	622.067	31			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Case Sequence	.884	.235	1.852	3.763	.001
Case Sequence ** 2	-.065	.016	-4.610	-3.943	.000
Case Sequence ** 3	.001	.000	1.890	2.615	.014
(Constant)	28.293	.909		31.110	.000

$R^2 = 0.942$, Adjusted $R^2 = 0.935$

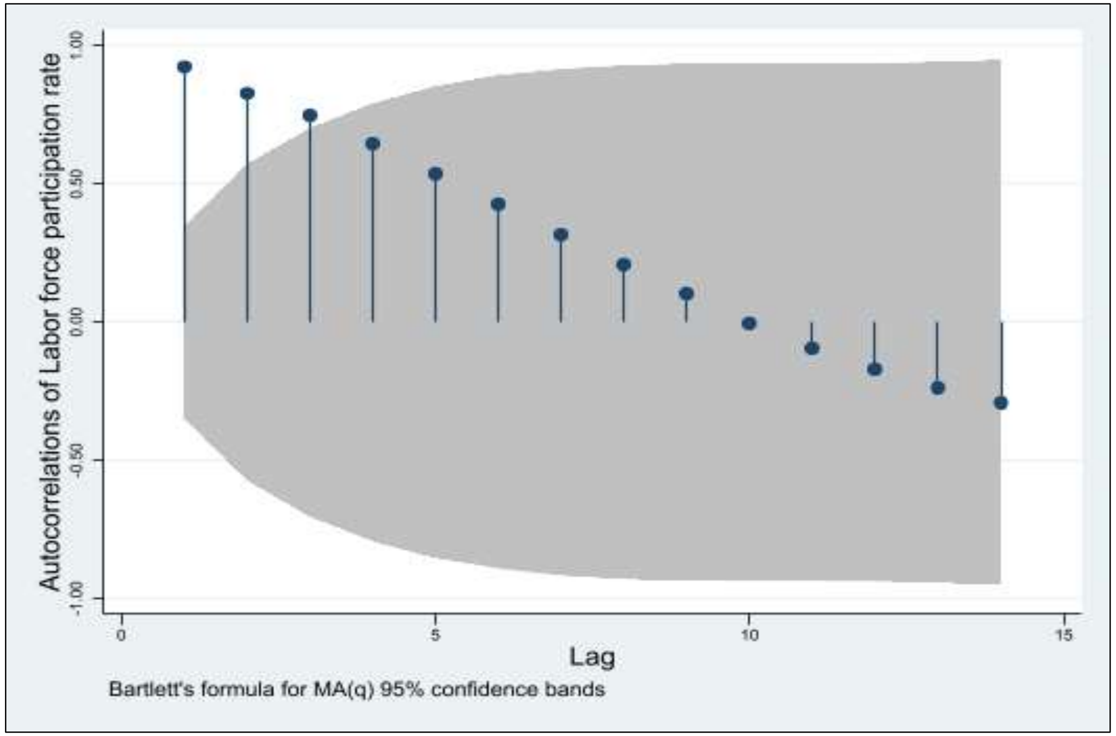
Here, $R^2_{cubic} > R^2_{quadratic}$. Therefore, the cubic curve is the best fit for the given data.

Measurement of Trend: Method of Moving Average

Analysis ▶ **Moving Average Method**

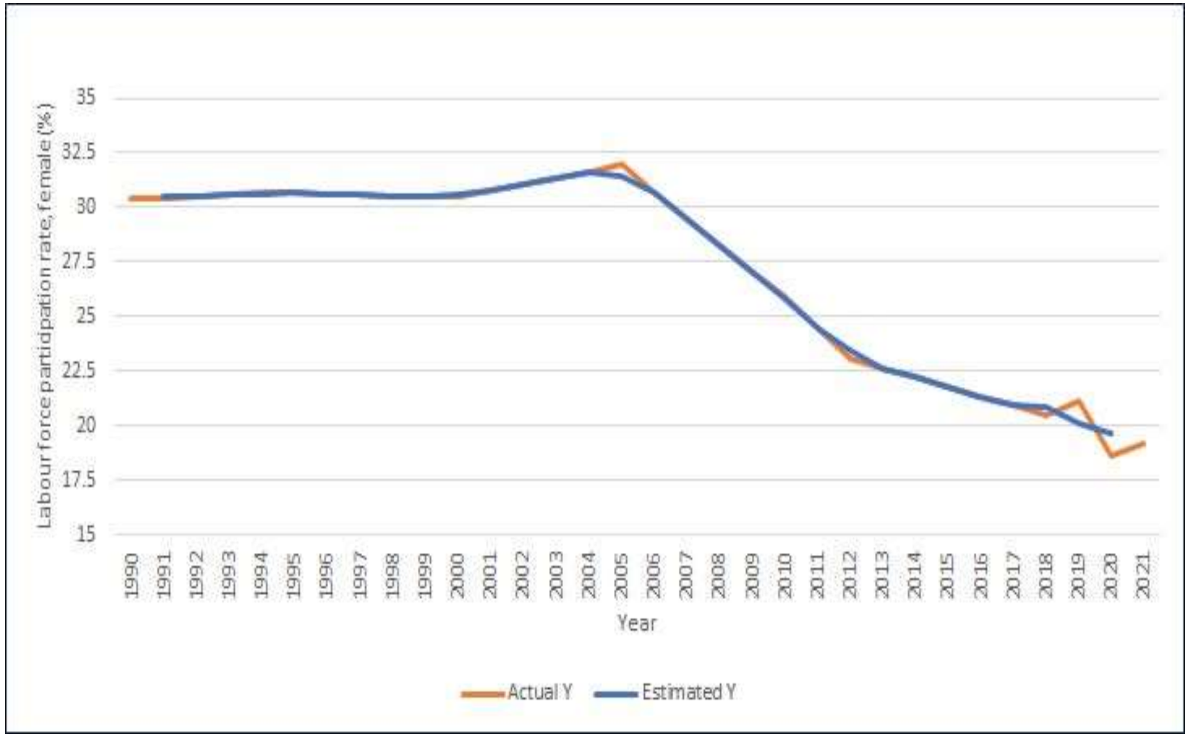
The ACF Plot is shown in the figure below:

Graph 2.2



The graph for the MA(3) model is shown in the figure below:

Graph 2.3



From the figure, we conclude that 3 lags lie outside the 95% confidence interval, hence we take the extent of the Moving Average model to be 3.

After removing the local shocks, we see that there is a decreasing trend for the labor force participation of women in India. Visually, MA(3) seems a good fit to the original data but poses certain limitations.

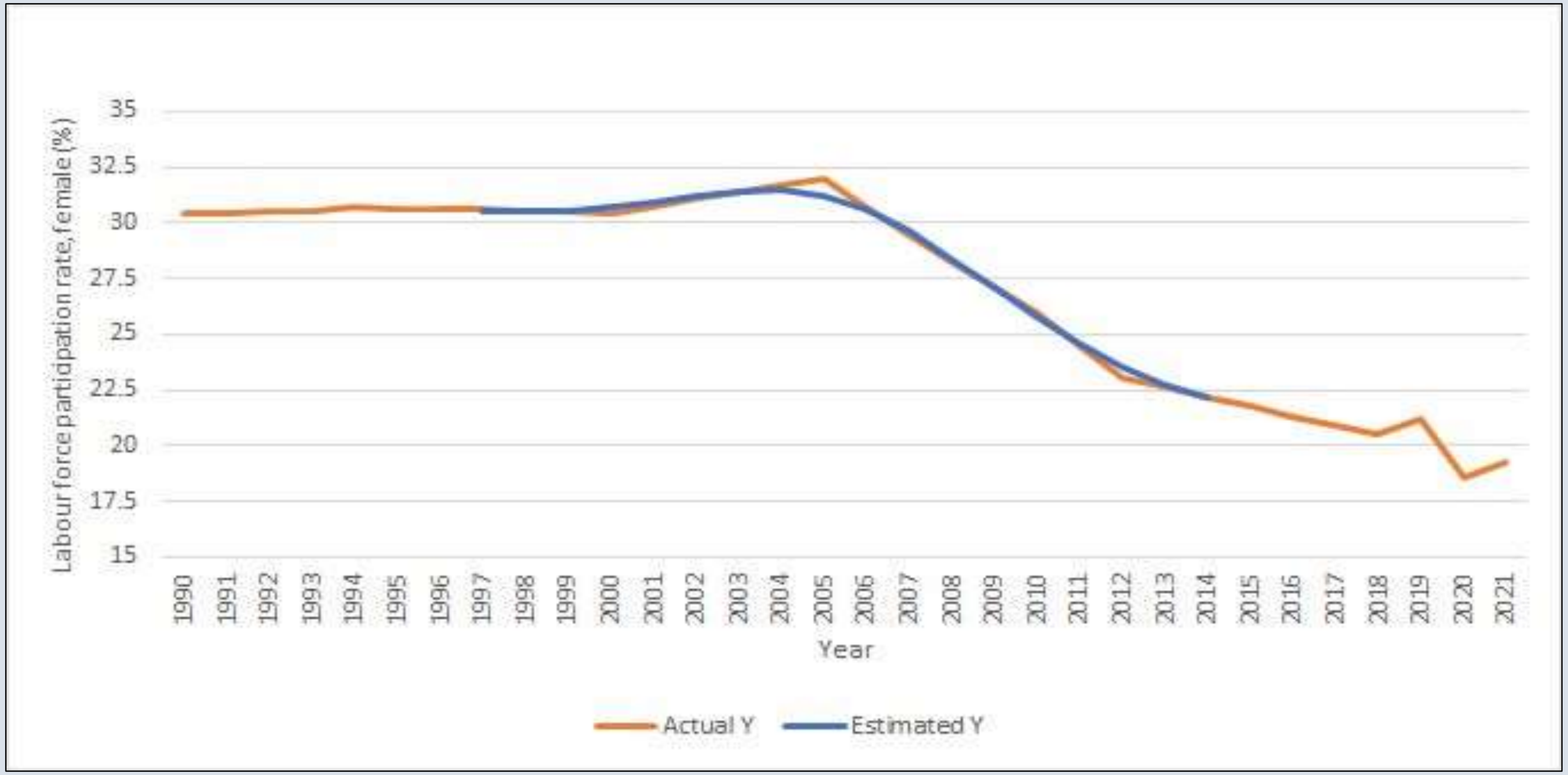
Measurement of Trend: Spencer's Formula

Analysis ▶ Spencer's Formula ▶ 15-Point

As method of simple Moving Averages assigns equal weightage to all the lagged values, we apply Spencer's 15-point and Spencer's 21-point formula to give relatively more weights to the recent observations.

The graph for Spencer's 15-point formula is shown below:

Graph 2.4



The formula for Spencer's 15-point formula is shown below:

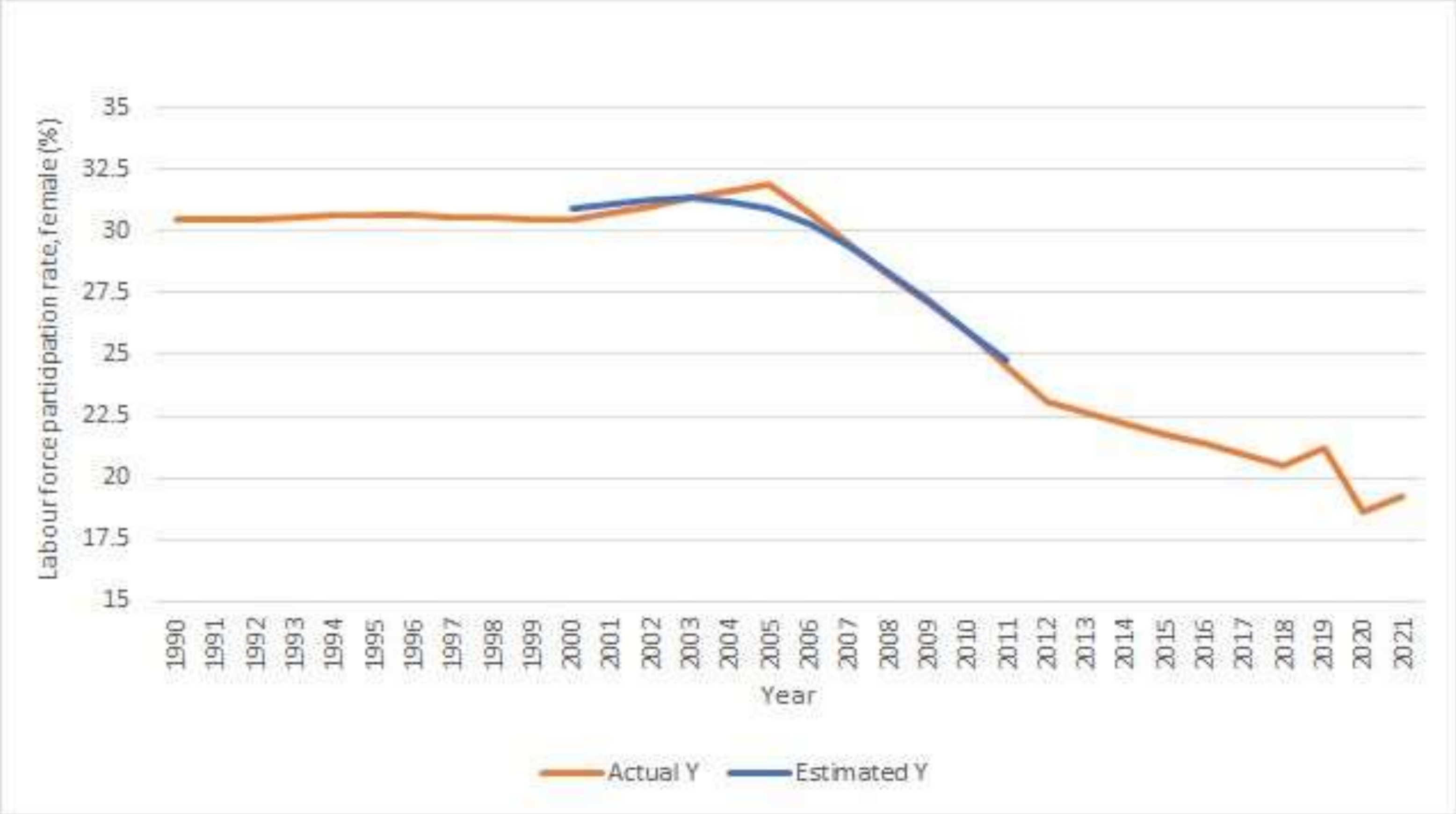
$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74, \ldots]$$

Measurement of Trend: Spencer's Formula

Analysis ▶ Spencer's Formula ▶ 21-Point

The graph for Spencer's 21-point formula is shown below:

Graph 2.5



The formula for Spencer's 21-point formula is:

$$\frac{[-0.5, -1.5, -2.5, -2.5, -1, 3, 9, 16.5, 23.5, 28.5, 30, \dots]}{350}$$

Limitations

1. We notice that the data points considerably reduce after applying both the Spencer's formulae

2. These methods can be used for smoothing the curve but not predicting the future values

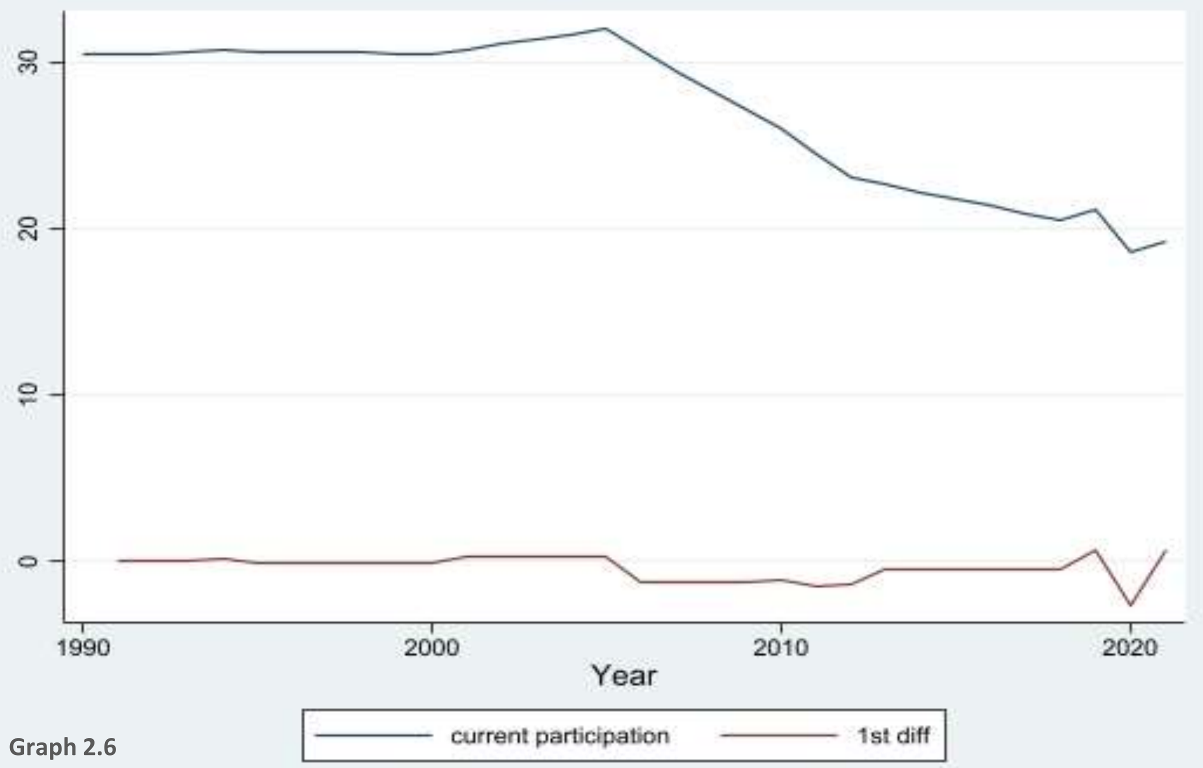
3. The weights of the data points are pre-determined, unlike exponential smoothing which assigns relatively greater weights to the more recent values

Forecasting: Simple Exponential Smoothing

Analysis ▶ Simple Exponential Smoothing ▶ ADF Test

The simple exponential smoothing technique overcomes the limitations posed by the moving average method. However, it requires the time series data to be free of trend and seasonality. In order to make our time series data stationary, we differentiate the data and further use the Augment Dickey-Fuller (ADF) Test to test its stationarity.

Graph representing the actual and differenced time series data



Graph 2.6

Augmented Dickey-Fuller (ADF) Test

H_o: $\alpha = 1$ i.e. the given time series data is not stationary
H_1: $\alpha \neq 1$ i.e. the given time series data is stationary
where α is the coefficient of the first lag on Y

```
Dickey-Fuller test for unit root                   Number of obs  = 30
Variable: partdiff                                 Number of lags =  0

H0: Random walk without drift, d = 0

                                       Dickey-Fuller
                   Test          ------ critical value ------
                statistic          1%           5%          10%
--------------------------------------------------------------
 Z(t)              -4.208       -3.716       -2.986       -2.624
--------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0006.
```

Since the p-value is less than 0.05, it is statistically significant and hence, we reject the H_o.
We may conclude that the differenced time series data is **stationary**.

Forecasting: Simple Exponential Smoothing

Analysis ▶ Simple Exponential Smoothing ▶ Determination of α

Since the time series data has now been made stationary, we can use the simple exponential smoothing technique to forecast the value of the female labour force participation rate (%).

Determination of the Smoothing Factor (α)

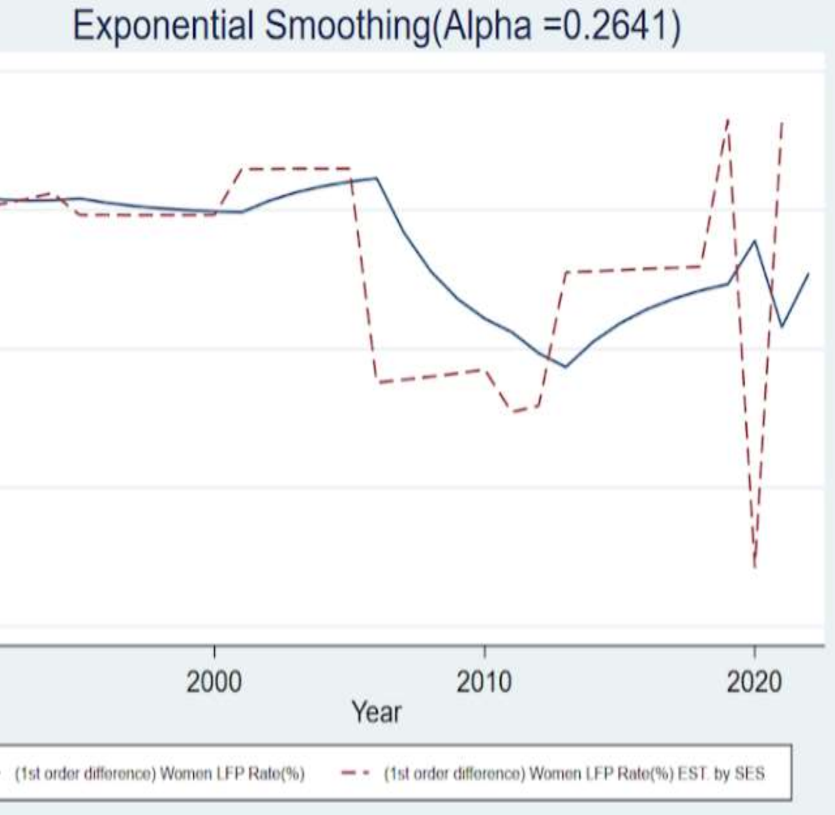
- Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.
- We use STATA to determine the optimum value of the smoothing constant (α) by minimizing the sum of squared errors.
- This parameter controls the rate at which the influence of the observations at prior time steps decay exponentially. Alpha is often set to a value between 0 and 1
- A value close to 1 indicates fast learning (that is, only the most recent values influence the forecasts), whereas a value close to 0 indicates slow learning (past observations have a large influence on forecasts).

```
. tssmooth exponential sesf=partdiff

computing optimal exponential  coefficient (0,1)

optimal exponential coefficient =        0.2641
sum-of-squared residuals        =      15.17355
root mean squared error         =      .69962087
```

Hence, the optimal exponential coefficient is obtained as 0.2641.

Graph representing the Exponentially Smoothened Curve for the differenced time series data

Graph 2.7

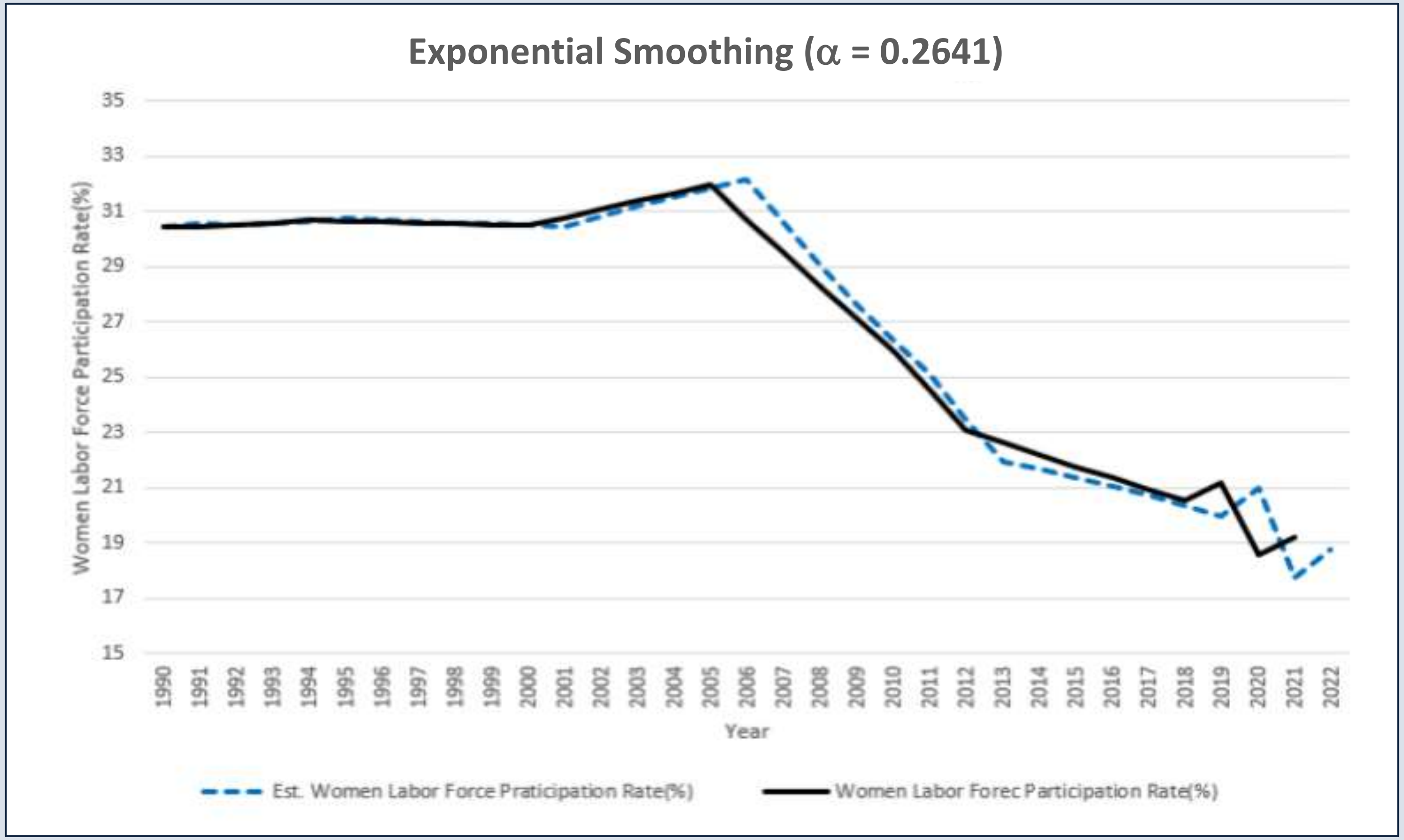


Forecasting: Simple Exponential Smoothing

Analysis ▶ Simple Exponential Smoothing ▶ Forecast

Graph representing the exponentially smoothened curve for the actual time series data

Graph 2.8



Interpretation from the graph

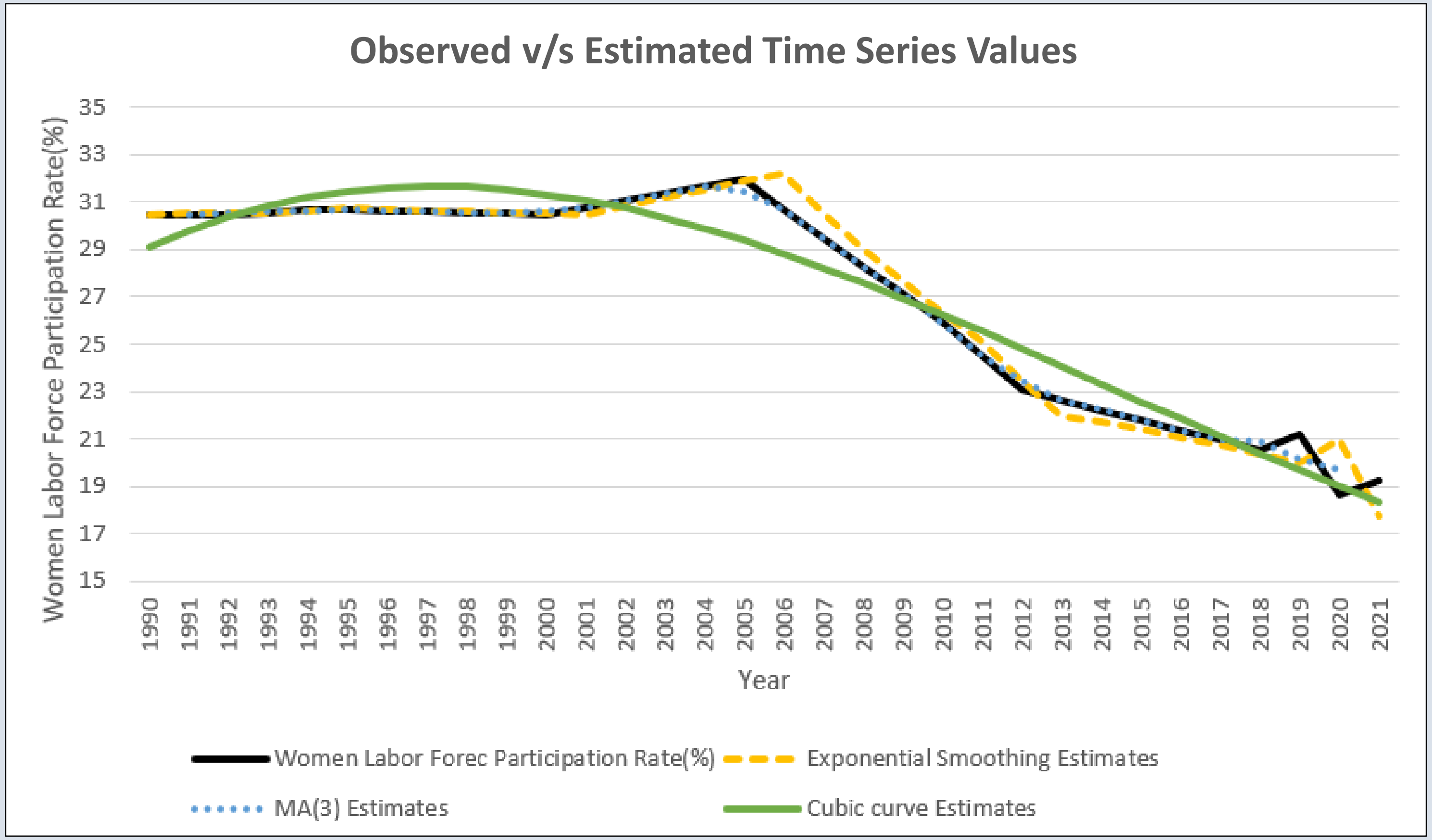
- Simple Exponential Smoothing provides a good fit for the data as the estimated values lie very close to the actual values.
- The smoothing coefficient is chosen as 0.2641, which lies between 0.1 and 0.3.
- The forecasted value for 2022 is obtained as **18.7791276%.**
- This is greater than the female labor force participation rate in 2021, which shows a positive outlook for the future.
- This upward trend can be attributed to the increase in the proportion of educated women entering the working age population and the recovery of the economy post-covid.

Comparing the Models: Estimated and Actual Values of Y

Analysis ▶ Comparison of Models

Graph representing the actual v/s estimated values for the time series data

Graph 2.9



A Comparative Analysis

1. The method of **curve fitting** is based on visualizing the underlying trend in data and making a prediction **without accounting for the variations** in the data. **SES** on the other hand, **works only on stationary data**. It smoothens the Y_t values on the basis of an exponential coefficient which gives us a rough estimate of the number of past values considered while making predictions.

2. As we can see from the graph, the **MA(3) model fits better** to the data, however it cannot be used for forecasting. The **exponential smoothing curve can be used from forecasting** unlike cubic and MA(3) curves, hence it is a better choice for the given data.

3. The predicted values for 2022 labour force participation of women via both curves are:
   SES: 18.7791276 **Cubic: 17.75104776**

SECTION III

CONCLUSION:

STATISTICAL OVERVIEW OF THE PARTICIPATION OF FEMALES IN THE LABOUR FORCE

Results and Inferences from the Analysis

We reject the H_o to conclude that there has been a significant decline in the female participation in the labour force in India

Possible reasons for this decline are:

1. During 2005-2010 there is a sharp decline in the labor force of women, this can be because the **rate of secondary education** was the highest during this time which is the period when women are most economically inactive.

2. The educated women would be ready to enter the workforce, however the **jobs available are not increasing proportionately.** Moreover as observed, **the number of competing men** leads to insufficient number of jobs for women

3. As women will be educated, they **might not want to settle** for informal/ unsatisfactory work or wages

4. As the **household income increases**, many women especially in rural areas might opt out to work, becoming inactive in the formal economy

5. An **entrepreneurial mindset** among women might also encourage them to opt out of the formal labor force

6. Given that handicraft and household small businesses are often top choices for women and lie in the **informal sector of the economy**, if this proportion increases, the women in labor force decreases

Limitations of the Analysis

Our analysis of the *Female Labour Force Participation Rate* poses the following limitations:

Data is compiled based on surveys and censuses:

- Labour force data from population censuses are often based on a limited number of questions.
- Data of unemployed workers, workers in small establishments, or workers in the informal sector is not being involved.

Reference period of a census or survey

- In some countries, survey is done during a specific period before the inquiry date, while in others data are recorded without reference to any period.
- The estimated labour force may be much smaller than the numbers actually working.

Differing definitions of employment age

- For most countries the working age is 15 and older, but in some countries children younger than 15 work full- or part-time and are included in the estimates
- Calculations may systematically over- or underestimate actual rates.

Statistical concept and methodology

- Unpaid workers, family workers, and students are often omitted.
- Labor force size tends to vary during the year as seasonal workers enter and leave.

References

References:

Books:

1. The Analysis of Time Series by Chris Chatfield
2. Basic Econometrics by Damodar N. Gujarati
3. Fundamentals of Applied Statistics by S.C. Gupta and V.K. Kapoor

Links:

1. https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-new_delhi/documents/publication/wcms_789243.pdf
2. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---inst/documents/publication/wcms_250977.pdf
3. https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?contextual=default&locations=AF-IN&name_desc=false
4. https://core.ac.uk/download/pdf/213942549.pdf
5. https://knoema.com/atlas/India/topics/Demographics/Population/Female-population
6. https://ourworldindata.org/female-labor-supply
7. https://towardsdatascience.com/time-series-in-python-exponential-smoothing-and-arima-processes-2c67f2a52788#:~:text=Exponential%20smoothings%20methods%20are%20appropriate,look%20at%20the%20differenced%20series
8. https://www.youtube.com/watch?v=rlkOjNjAPEM&ab_channel=BloombergMarketsandFinance

Before we end the presentation, we would like to show a short video to you:

THANK YOU