



ISE 457 Forecasting Project Report

Duru Naz Şahinakman - [REDACTED]

Seray Cengiz - [REDACTED]

Selin Koçyiğit - [REDACTED]

INSTRUCTOR

Dr. Öğr. Üyesi Canan Pehlivan Yıldırım

TABLE OF CONTENT

1. Plot the Time Series

- 1.1 Dataset Description and Data Preparation
- 1.2 Time Series Plot

2. Create Training and Test Sets

- 2.1 Training and Test Set Division

3. Modeling the Training Data

- a. ETS Models
 - i. Identify Candidate ETS Models
 - ii. Fitting ETS Models
- b. ARIMA Models
 - i. Transformation
 - ii. Stationarity Check
 - iii. Identify Candidate ARIMA Models
 - iv. Fitting ARIMA Models

4. Model Selection on Training Data

- 4.1 Selecting the Best ETS Model
- 4.2 Residual Diagnostics of ETS(M,A,N)
- 4.3 Ljung-Box Test for ETS(M,A,N)
- 4.4 Selecting the Best ARIMA Model
- 4.5 Residual Diagnostics of ARIMA(2,2,0)
- 4.6 Ljung-Box Test for ARIMA(2,2,0)
- 4.7 Selected Candidate Models

5. Forecasting the Test Set

- 5.1 Generating Forecasts for ETS(M,A,N) and ARIMA(2,2,0)
- 5.2 Forecast Accuracy Comparison
- 5.3 Choosing the Final Model

6. Final Forecast Using All Data

- 6.1 Refitting the Final Model
- 6.2 Point Forecasts and Prediction Intervals

7. Conclusion

1. Plot the Time Series

1.1 Dataset Description and Data Preparation

The assigned dataset for this project is `aus_fertility`, which is included in the `fpp3` package. The purpose of the project is to analyze and forecast the fertility rate of women older than 40 in Australia from 1980 onward.

Since the original dataset contains fertility rates for different age groups, the data was filtered according to the assigned criteria: Region = Australia, $\text{Year} \geq 1980$, and age groups older than 40. After filtering, multiple observations remained for some years because more than one age group was included. Therefore, the fertility rates were averaged by year to obtain a single annual time series. The resulting series represents the average fertility rate of women older than 40 in Australia and was used for visualization, model fitting, model evaluation and final forecasting.

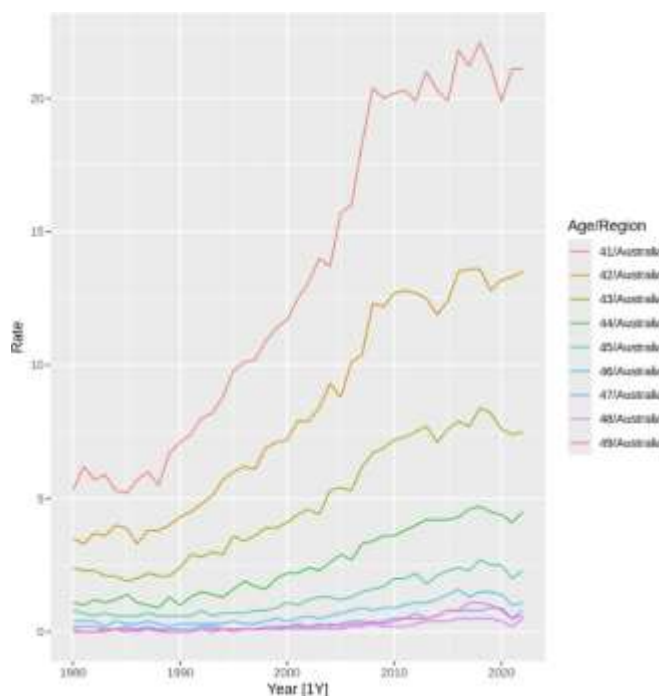


Figure 1.1 shows that after filtering the dataset for Australia, $\text{Year} \geq 1980$, and age groups older than 40, multiple observations still remain for each year because several age groups are included.

Since forecasting models such as ETS and ARIMA require a single time series, the fertility rates were averaged by year to obtain one representative annual series.

Time Series Plot

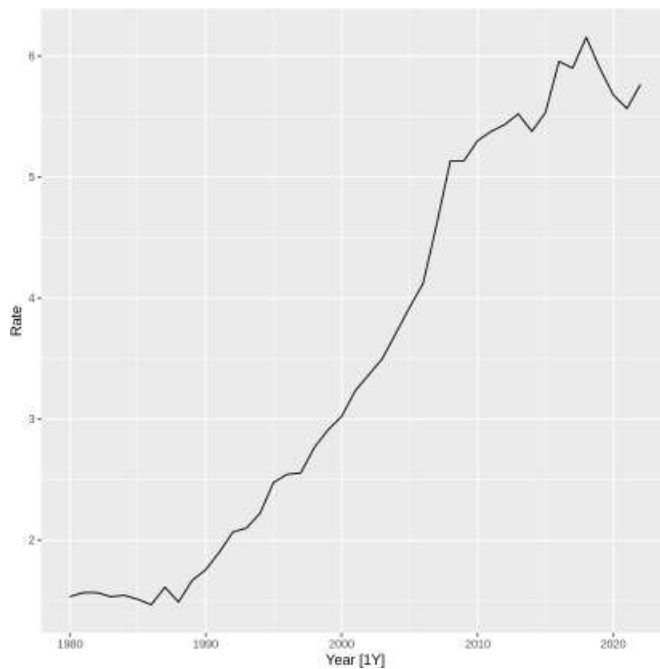


Figure 1.2 : Time Series Plot of Average Fertility Rate in Australia for Women Older Than 40

Figure 1.2 presents the time series plot of the average fertility rate of women older than 40 in Australia from 1980 onward. Since the data is annual, no regular seasonal pattern is expected. The main purpose of this plot is to examine the overall movement of the series, the presence of a possible trend, the stability of variance and whether the series appears stationary or nonstationary.

Based on the visual inspection, the series does not fluctuate around a constant mean and shows a strong and persistent upward trend over time. This indicates that the series does not maintain a stable long term level. These characteristics suggest that the original series is non stationary, mainly due to the changing mean level. Since ARIMA models require stationarity, differencing and further stationarity checks such as ACF/PACF analysis and unit root tests should be considered before model fitting.

2. Create Training and Test Sets

2.1 Training and Test Set Division

In time series forecasting, the order of observations is very important because future values depend on past values. For this reason, the dataset was divided chronologically rather than randomly. A random split could cause information leakage and reduce the reliability of model evaluation.

The observations from 1980 to 2014 were used as the training set, while the observations from 2015 onward were reserved as the test set. The training set was used for fitting ETS and ARIMA models, while the test set was used to evaluate forecasting performance by comparing predicted values with actual observations. This separation allows us to test how well the models perform on unseen future data. By preserving the natural time structure of the series, the forecasting process becomes more realistic and the final model selection becomes more reliable.

Year	Rate	Year	Rate
<dbl>	<dbl>	<dbl>	<dbl>
1980	1.533333	1998	2.766667
1981	1.566667	1999	2.911111
1982	1.566667	2000	3.022222
1983	1.533333	2001	3.233333
1984	1.544444	2002	3.366667
1985	1.511111	2003	3.500000
1986	1.466667	2004	3.711111
1987	1.611111	2005	3.922222
1988	1.488889	2006	4.122222
1989	1.666667	2007	4.611111
1990	1.755556	2008	5.133333
1991	1.900000	2009	5.133333
1992	2.066667	2010	5.300000
1993	2.100000	2011	5.377778
1994	2.222222	2012	5.433333
1995	2.477778	2013	5.522222
1996	2.544444	2014	5.377778
1997	2.555556		

Table 2.1 : Training Set (1980 – 2014)

Year	Rate
<dbl>	<dbl>
2015	5.533333
2016	5.955556
2017	5.900000
2018	6.155556
2019	5.900000
2020	5.677778
2021	5.566667
2022	5.766667

Table 2.2: Test Set (2015–2022)

3. Modeling the Training Data:

a. ETS models

i. Identify Candidate ETS Models

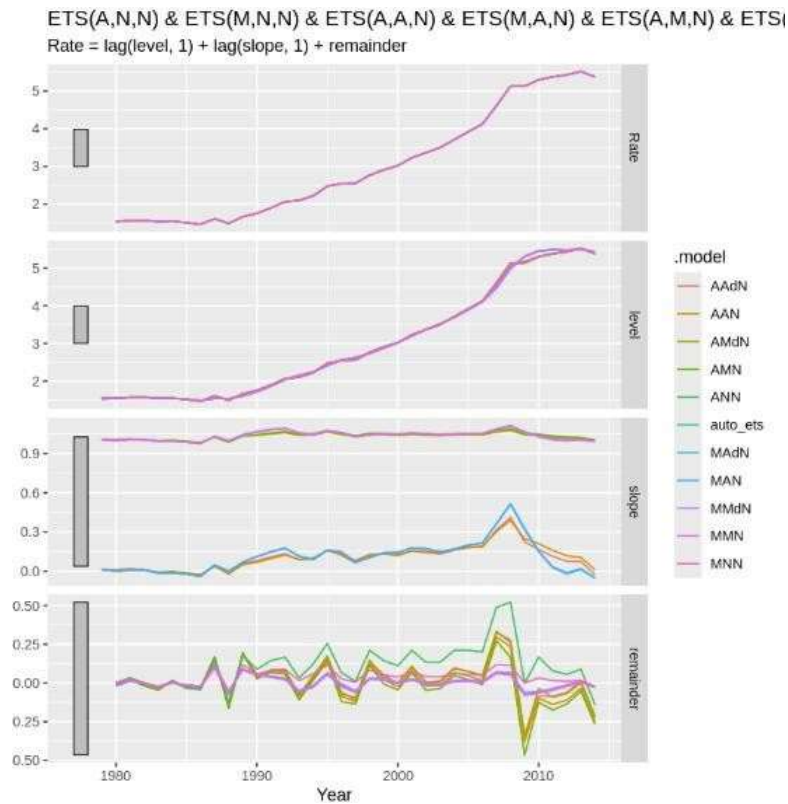


Figure 3.1 : Components of Candidate ETS Models

Exponential Smoothing (ETS) models were considered for forecasting the average fertility rate of women older than 40 in Australia. Since the time series is annual, no regular seasonal pattern is expected. Because of this all candidate models were specified without a seasonal component.

Based on the time series plot and the ACF/PACF analysis, the series shows a clear and persistent upward trend. This indicates that models including a trend component are necessary. For this reason, additive trend, multiplicative trend and damped trend models were included among the candidate ETS models.

Damped trend models were also considered because fertility rates are naturally limited by biological and demographic factors, meaning that the increasing trend cannot continue indefinitely at the same rate. Therefore, models such as ETS(A,Ad,N), ETS(M,Ad,N), ETS(A,Md,N) and ETS(M,Md,N) were included to allow the trend to gradually slow down over time.

In addition, simple exponential smoothing models such as ETS(A,N,N) and ETS(M,N,N) were also included to test whether the trend component significantly improves forecasting performance.

ii. Fitting ETS Models

.model	term	estimate	.model	term	estimate
<chr>	<chr>	<dbl>	<chr>	<chr>	<dbl>
ANN	alpha	0.999899996	AAAdN	phi	0.913695940
ANN	l[0]	1.533179647	AAAdN	l[0]	1.525221708
MNN	alpha	0.999899974	AAAdN	b[0]	0.014812406
MNN	l[0]	1.527803543	MAdN	alpha	0.516126110
AAN	alpha	0.953513032	MAdN	beta	0.516126041
AAN	beta	0.373326749	MAdN	phi	0.964786005
AAN	l[0]	1.526247937	MAdN	l[0]	1.545102135
AAN	b[0]	0.008863736	MAdN	b[0]	0.016535662
MAN	alpha	0.512052436	AMdN	alpha	0.991677156
MAN	beta	0.512052343	AMdN	beta	0.346427732
MAN	l[0]	1.548467087	AMdN	phi	0.880020241
MAN	b[0]	0.015126537	AMdN	l[0]	1.520768580
AMN	alpha	0.999899589	AMdN	b[0]	1.009827501
AMN	beta	0.373843813	MMdN	alpha	0.525178309
AMN	l[0]	1.525096494	MMdN	beta	0.525178242
AMN	b[0]	1.005395315	MMdN	phi	0.895769402
MMN	alpha	0.603376502	MMdN	l[0]	1.539435601
MMN	beta	0.456769928	MMdN	b[0]	1.011534614
MMN	l[0]	1.548787760	auto_ets	alpha	0.512052436
MMN	b[0]	1.004679529	auto_ets	beta	0.512052343
AAAdN	alpha	0.865065955	auto_ets	l[0]	1.548467087
AAAdN	beta	0.449096928	auto_ets	b[0]	0.015126537

Table 3.1 : Smoothing Parameters and Initial States of Candidate ETS Models

The candidate ETS models were fitted using the training dataset covering the years 1980–2014. For each model, the smoothing parameters and initial states were estimated automatically by the ETS procedure.

The parameter α controls the level smoothing, β controls the trend smoothing and ϕ represents the damping parameter used in damped trend models. In addition, l_0 indicates the initial level of the series, while b_0 represents the initial trend value.

Since the fertility rate series shows a strong upward trend over time, correctly estimating the trend related parameters is important for accurate forecasting. Models including a trend component are expected to perform better because the series clearly does not fluctuate around a constant mean.

Table 3.1 presents the estimated smoothing parameters and initial states for all candidate ETS models. These fitted models were later compared using information criteria and forecast accuracy measures in order to determine the best performing ETS model.

b. ARIMA models

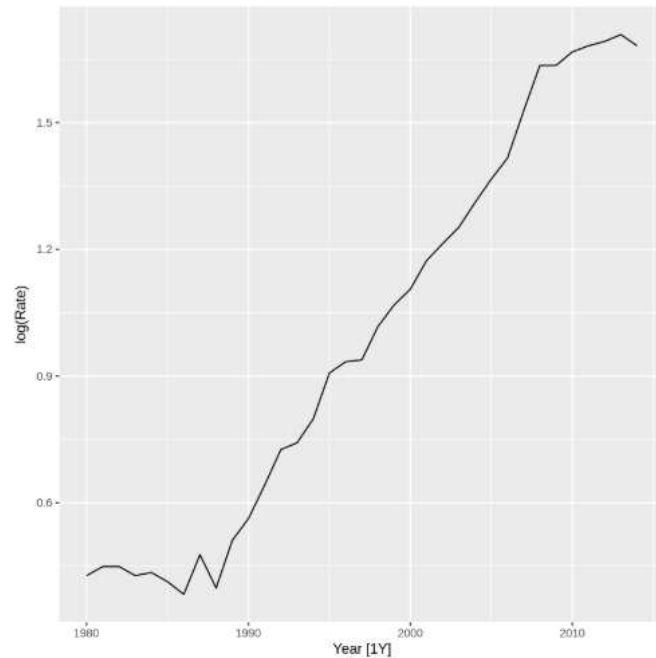


Figure 3.4 : Log-Transformed Fertility Rate Series

After observing the original time series, it was noticed that the variability of the fertility rate appears to increase gradually over time. Since ARIMA models perform better when the variance is more stable, a logarithmic transformation was applied to the series.

Figure 3.4 presents the log-transformed fertility rate series for women older than 40 in Australia. The purpose of this transformation is to stabilize the variance and reduce the effect of increasing fluctuations observed in the original series.

After the logarithmic transformation, the series becomes more suitable for ARIMA modeling because the effect of large fluctuations is reduced and the overall pattern becomes smoother and easier to model. However, the upward trend still remains, which indicates that first order and possibly second order differencing are still necessary to achieve stationarity. Therefore, after variance stabilization through log transformation, differencing procedures were applied in the next step to remove the trend component and obtain a stationary series.

i. Transformation

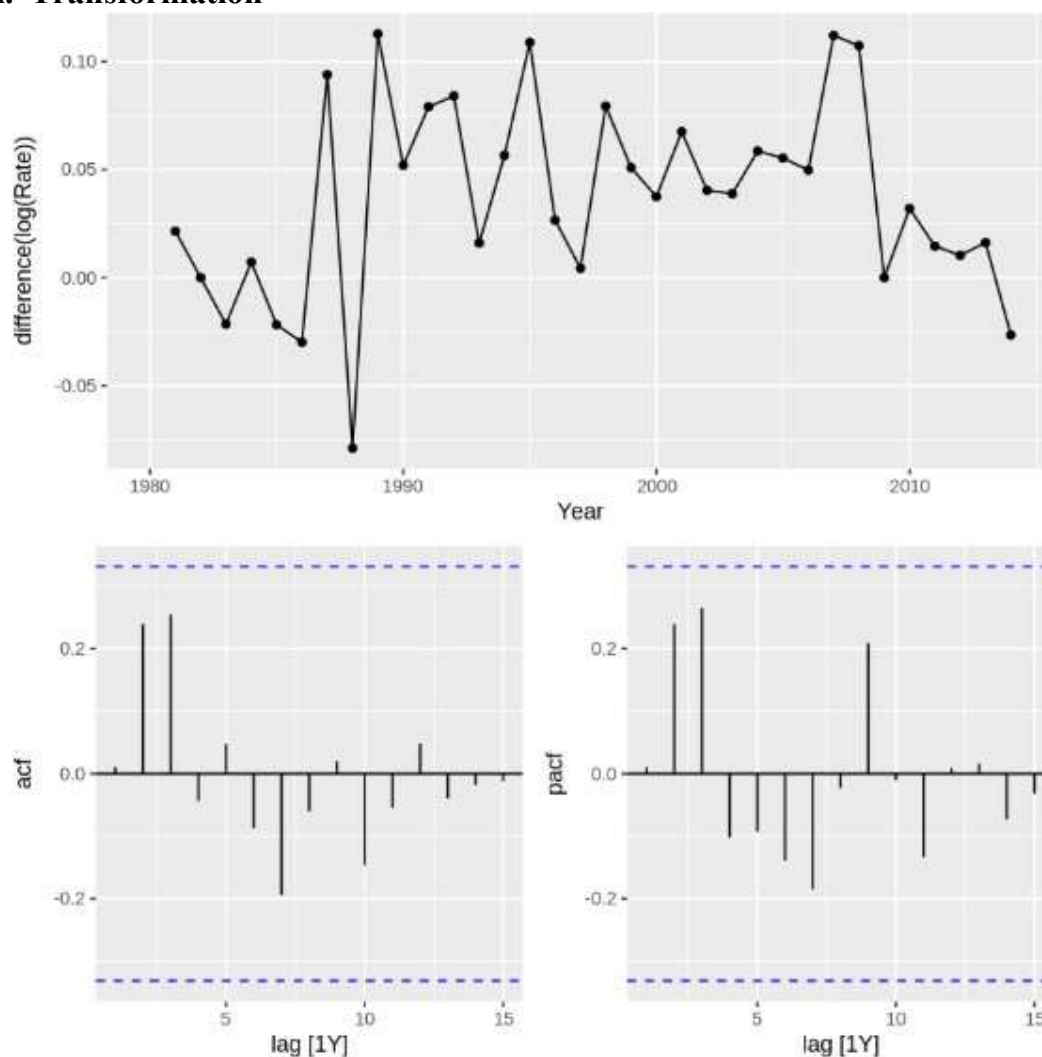


Figure 3.5 : ACF and PACF of First-Differenced Log Series

Figure 3.5 presents the ACF and PACF plots of the first differenced log-transformed series. Although the trend is reduced after first differencing, the series still does not fully fluctuate around a constant mean and some autocorrelations remain significant. This indicates that first differencing alone is not sufficient to achieve full stationarity.

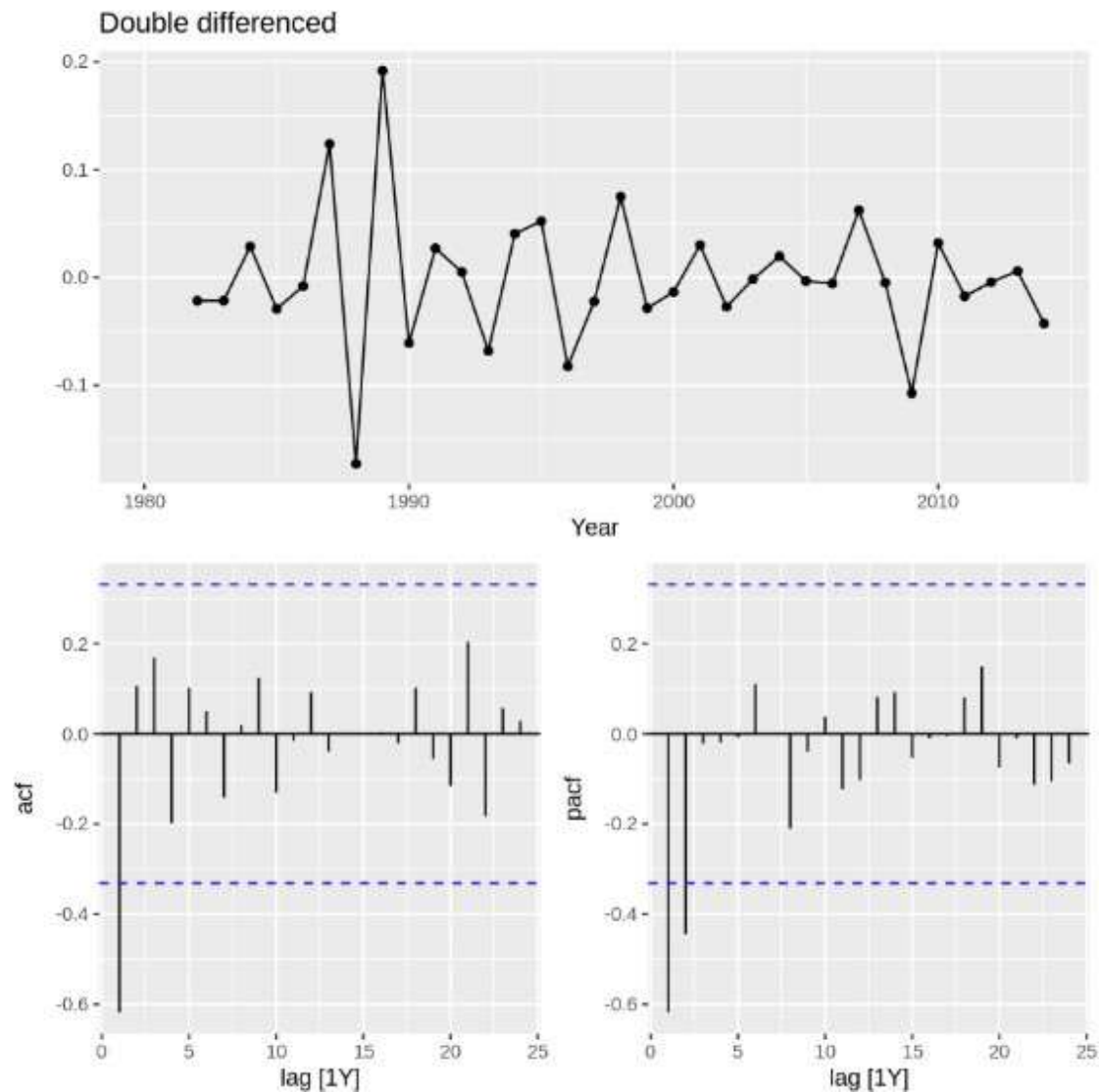


Figure 3.6 : ACF and PACF of Double-Differenced Log Series

Second differencing was applied. Figure 3.6 presents the ACF and PACF plots of the double-differenced log series. After second differencing, the series appears more stable around a constant mean and most autocorrelation values fall within the confidence bounds.

The ACF plot shows a significant negative spike at lag 1, while the PACF plot shows significant spikes at the first 2 lags. This pattern suggests that the transformed series is now closer to stationarity and supports the consideration of ARIMA models with differencing order $d = 2$. Based on these results, second order differencing was selected for the ARIMA modeling process.

ii. Stationarity Check

Although the forecast may not appear intuitive, as similarly observed in the class examples, the model is considered valid since the residuals follow a white noise process ($p > 0.05$), confirming that $d = 2$ was sufficient to achieve stationarity and that no significant autocorrelation structure remains in the residuals.

iii. Identify Candidate ARIMA Models

After achieving stationarity through second order differencing ($d = 2$), the ACF and PACF plots were examined to determine appropriate candidate ARIMA models.

In the ACF plot, a single strong negative spike is observed at lag 1, while the remaining lags mostly remain within the confidence bounds. This pattern suggests the presence of a Moving Average component of order 1, indicating a possible MA(1) structure. In the PACF plot, significant spikes are observed at lag 1 and lag 2, while higher order lags are relatively weak. This indicates a possible Autoregressive component of order 2, suggesting an AR(2) structure.

Based on these observations, two manual candidate ARIMA models were selected:

- ARIMA(0,2,1)
- ARIMA(2,2,0)

In addition to these manually identified models, automatic model selection methods such as stepwise and search procedures were also applied to compare forecasting performance and information criteria.

iv. Fitting ARIMA Models

Model Type	Model	AICc	Selection Method
Manual	ARIMA(2,2,0)	-105.0595	ACF/PACF based
Manual	ARIMA(0,2,1)	-103.7371	ACF/PACF based
Automatic	ARIMA(0,2,2)	-104.2641	Stepwise
Automatic	ARIMA(2,2,0)	-105.0595	Search

Table 3.2: AICc Comparison of Candidate ARIMA Models

The manually selected ARIMA models and the automatically selected models were fitted using the training dataset. The manual models were chosen based on the ACF and PACF patterns of the double-differenced log series. In addition, stepwise and exhaustive search methods were applied to compare the manually selected models with automatically selected alternatives.

Table 3.2 presents the AICc values of the candidate ARIMA models. AICc was used as the

main comparison criterion, where a lower AICc value indicates a better model fit.

4. Model Selection on Training Data

4.1 Selecting the Best ETS Model

.model	sigma2	log_lik	AIC	AICc	BIC	MSE	AMSE	MAE
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
ANN	0.032723804	-1.344963	8.6899267	9.4641203	13.355971	0.03085387	0.11585697	0.13144315
MNN	0.003782347	2.894413	0.2111742	0.9853678	4.877218	0.03085475	0.11586500	0.04761105
AAN	0.018822620	9.427403	-8.8548061	-6.7858406	-1.078066	0.01667146	0.04912659	0.09470972
MAN	0.002009758	13.897306	-17.7946115	-15.7256460	-10.017871	0.01905582	0.05883867	0.03269099
AMN	0.021154589	7.383446	-4.7668924	-2.6979269	3.009848	0.01873692	0.06499316	0.09796585
MMN	0.002124392	12.778340	-15.5566790	-13.4877135	-7.779939	0.02153948	0.07845120	0.03355118
AAdN	0.018761016	10.058594	-8.1171885	-5.1171885	1.214900	0.01608087	0.04678915	0.09229241
MAdN	0.002074937	13.989624	-15.9792474	-12.9792474	-6.647159	0.01819878	0.05487017	0.03277360
AMdN	0.019142579	9.706248	-7.4124963	-4.4124963	1.919592	0.01640793	0.04892009	0.09238326
MMdN	0.002124290	13.601297	-15.2025942	-12.2025942	-5.870506	0.01895801	0.06215525	0.03339025
auto_ets	0.002009758	13.897306	-17.7946115	-15.7256460	-10.017871	0.01905582	0.05883867	0.03269099

Table 4.1: AICc Comparison of Candidate ETS Models

Table 4.1 presents the information criteria values of the candidate ETS models fitted using the training dataset.

Among all candidate models, ETS(M,A,N) produced the lowest AICc value of -15.7256460. This result is also confirmed by the automatic model selection (auto_ets), which selected the same model.

ETS(M,A,N) includes multiplicative error, additive trend, and no seasonal component. Since the fertility data is annual and no regular seasonal pattern is observed, excluding the seasonal component is appropriate. In addition, the additive trend structure allows the model to capture the long-term upward movement of the fertility rate series. Therefore, ETS(M,A,N) was selected as the best ETS model for further residual diagnostics and forecasting.

4.2 Residual Diagnostics of ETS(M,A,N)

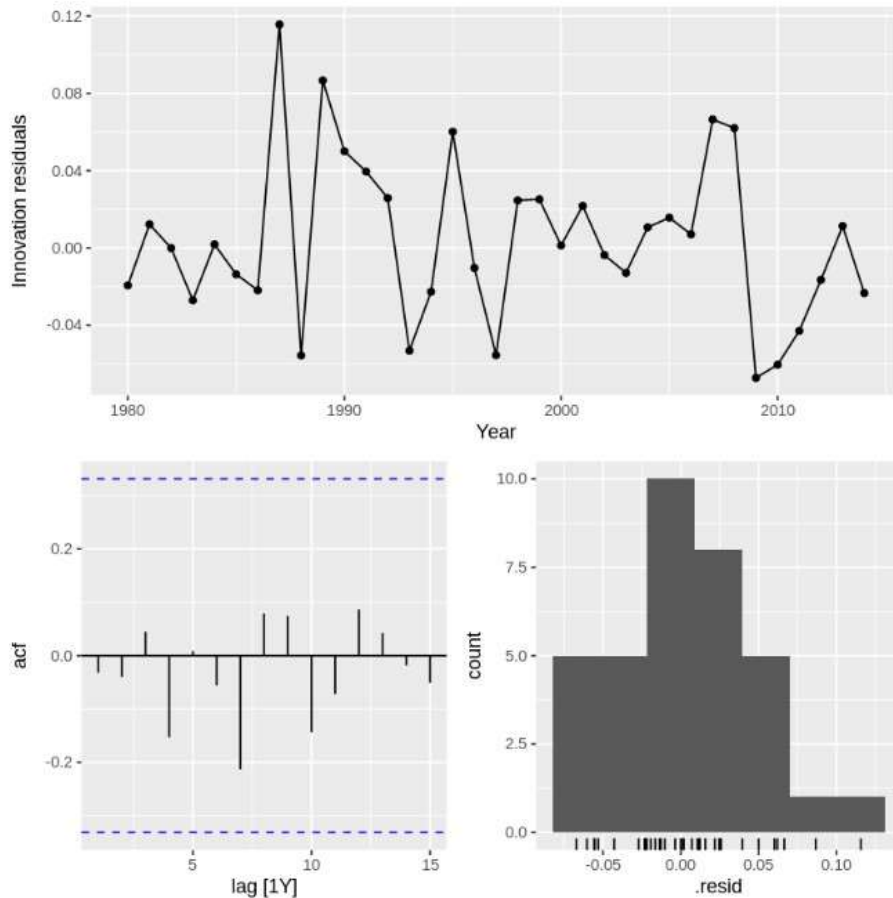


Figure 4.1: Residual Diagnostics of ETS(M,A,N)

The residual time plot is used to check whether the residuals fluctuate randomly around zero. The ACF plot helps determine whether any significant autocorrelation remains after fitting the model, while the histogram is used to visually examine the distribution of residuals.

Most residual autocorrelations remain within the confidence bounds, indicating that there is no significant remaining autocorrelation in the residuals. This suggests that the ETS(M,A,N) model captures the main structure of the fertility rate series successfully.

In addition, the histogram appears approximately symmetric and close to a normal distribution, which supports the adequacy of the model assumptions. However, a Ljung-Box test is also required to statistically confirm the independence of residuals.

4.3 Ljung-Box Test for ETS(M,A,N)

.model	lb_stat	lb_pvalue
<chr>	<dbl>	<dbl>
MAN	6.45491	0.5964133

Table 4.2: Ljung-Box Test Result for ETS(M,A,N)

The Ljung-Box test was applied to statistically check whether the residuals of the selected ETS(M,A,N) model contain any remaining autocorrelation.

The null hypothesis of the Ljung-Box test states that the residuals are independently distributed and behave like white noise. As shown in Table 4.2, the p-value is 0.5964, which is greater than 0.05. Therefore, the null hypothesis cannot be rejected. This indicates that there is no significant remaining autocorrelation in the residuals and confirms that the ETS(M,A,N) model is statistically adequate and well fitted for forecasting. This result also supports the residual diagnostics shown in Figure 4.1 and strengthens the reliability of the selected ETS model.

4.4 Selecting the Best ARIMA Model

Model Type	Model	AICc	Selection Method
Manual	ARIMA(2,2,0)	-105.0595	ACF/PACF based
Manual	ARIMA(0,2,1)	-103.7371	ACF/PACF based
Automatic	ARIMA(0,2,2)	-104.2641	Stepwise
Automatic	ARIMA(2,2,0)	-105.0595	Search

Table 4.3: AICc Comparison of Candidate ARIMA Models

Among the candidate ARIMA models, ARIMA(2,2,0) produced the lowest AICc value of -105.0595. The same model was also selected by the exhaustive search method, which supports the manual model selection based on the ACF and PACF analysis.

The stepwise method selected ARIMA(0,2,2), but its AICc value was slightly higher at -104.2641. Similarly, the manually selected ARIMA(0,2,1) model had a higher AICc value of -103.7371.

Since lower AICc values indicate better model fit, ARIMA(2,2,0) was selected as the best ARIMA model for further residual diagnostics and forecasting. This consistency between manual selection and automatic search increases the reliability of the model selection process.

4.5 Residual Diagnostics of ARIMA(2,2,0)

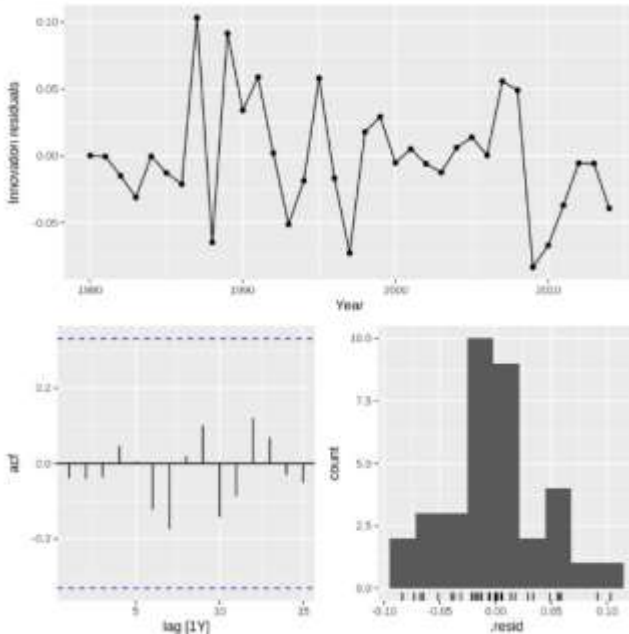


Figure 4.2: Residual Diagnostics of ARIMA(2,2,0)

The residual time plot is used to check whether the residuals fluctuate randomly around zero. The ACF plot of the residuals is examined to determine whether any significant autocorrelation remains after fitting the model. The histogram is used to visually inspect the distribution of residuals.

All residual autocorrelations remain within the confidence bounds, suggesting that the ARIMA(2,2,0) model captures the main dependence structure of the transformed series. The histogram also appears approximately symmetric and close to normal distribution, which supports the adequacy of the model assumptions. Therefore, the residual diagnostics indicate that the model is considered adequate for forecasting. However, a Ljung-Box test is also required to statistically confirm the independence of residuals.

4.6 Ljung-Box Test for ARIMA(2,2,0)

<code>.model</code>	<code>lb_stat</code>	<code>lb_pvalue</code>
<code><chr></code>	<code><dbl></code>	<code><dbl></code>
arima220	3.934293	0.8630026

Table 4.4: Ljung-Box Test Result for ARIMA(2,2,0)

The Ljung-Box test was applied to statistically check whether the residuals of the selected ARIMA(2,2,0) model contain remaining autocorrelation. The null hypothesis of the Ljung-Box test states that the residuals are independently distributed and behave like white noise. As shown in Table 4.4, the p-value is 0.8630, which is greater than 0.05. Therefore, the null hypothesis cannot be rejected. This indicates that there is no significant remaining autocorrelation in the residuals. Therefore, ARIMA(2,2,0) is statistically adequate, well-fitted, and suitable for forecasting.

4.7 Selected Candidate Models

After model selection on the training data, two candidate models were retained for forecasting the test set: ETS(M,A,N) and ARIMA(2,2,0). ETS(M,A,N) was selected as the best ETS model, while ARIMA(2,2,0) was selected as the best ARIMA model. These two models were then used to generate forecasts for the test set and compare their out of sample forecast accuracy.

5. Forecasting the Test Set

5.1 Generating Forecasts for ETS(M,A,N) and ARIMA(2,2,0)

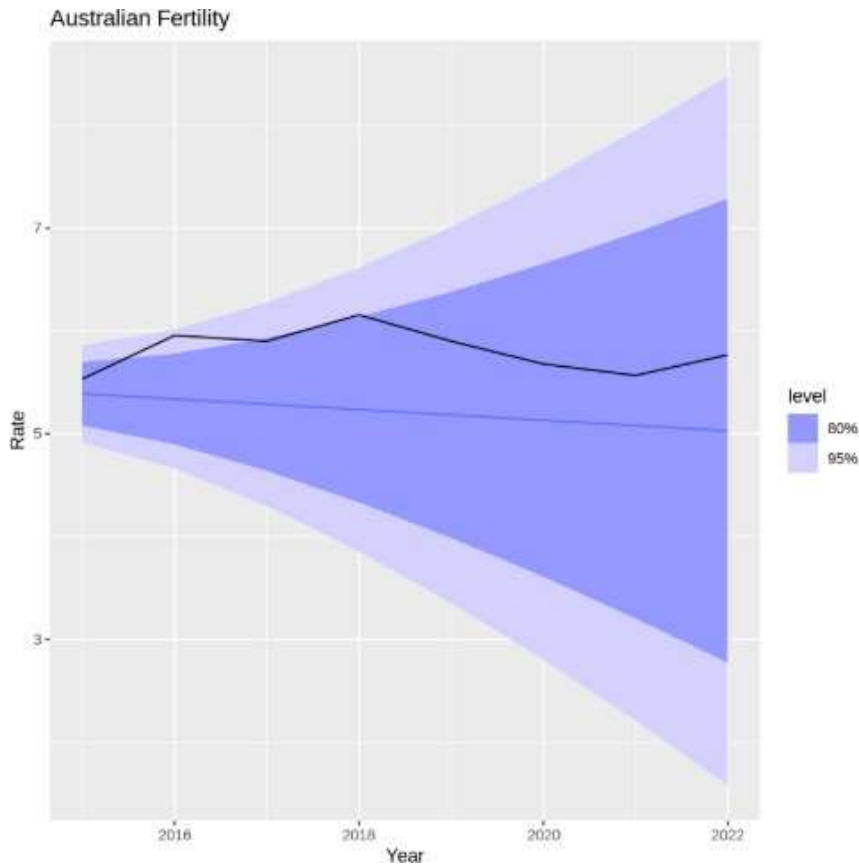


Figure 5.1: ETS(M,A,N) Forecasts for the Test Set

The ETS(M,A,N) model was used to generate forecasts for the test period from 2015 to 2022.

Figure 5.1 shows the forecasted values together with the actual observations in the test set.

The model captures the general long term trend of the fertility rate series, although some deviations between the forecasted and actual values can be observed. The prediction intervals widen over time, reflecting increasing uncertainty for longer forecast horizons. This behavior is expected in time series forecasting and indicates the uncertainty associated with future predictions.

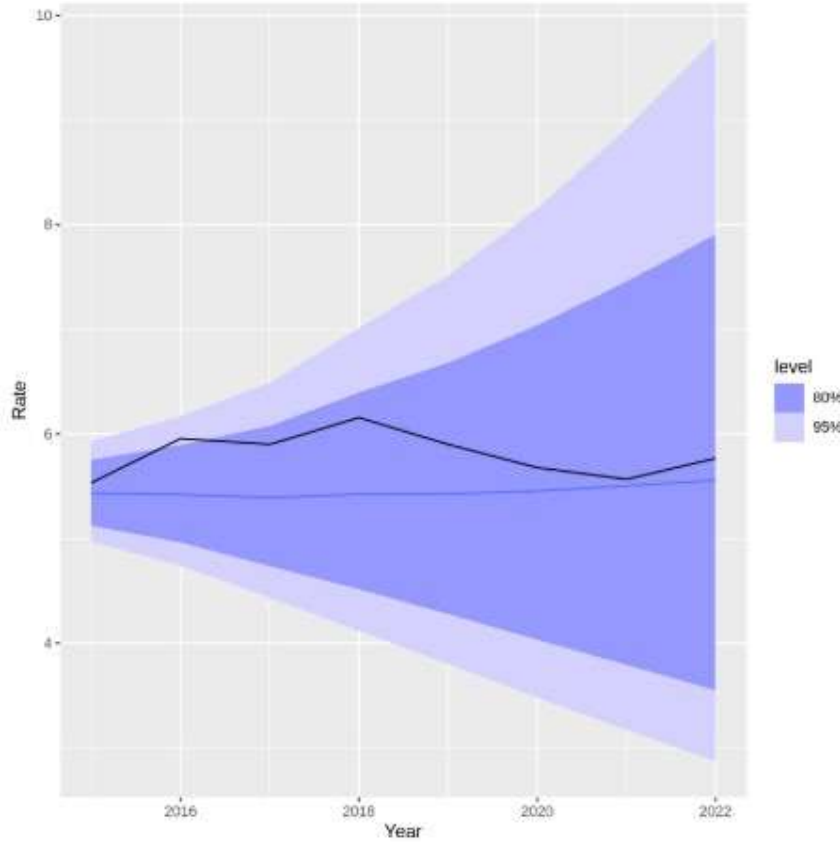


Figure 5.2: ARIMA(2,2,0) Forecasts for the Test Set

The ARIMA(2,2,0) model was also used to generate forecasts for the same test period.

Figure 5.2 presents the forecasted values and actual observations for the test set.

The forecast pattern is similar to the ETS model, but the ARIMA model shows slightly different behavior in capturing the fluctuations of the series. The widening prediction intervals again indicate increasing uncertainty over time.

Both models appear reasonable visually, but forecast accuracy measures are required to determine which model performs better.

5.2 Forecast Accuracy Comparison

Model	RMSE	MAE	MAPE
ETS(M,A,N)	0.384728	0.598431	10.21274
ARIMA(2,2,0)	0.420262	0.355657	6.003898

Table 5.1: Forecast Accuracy Comparison of ETS and ARIMA Models on the Test Set

Table 5.1 presents the out of sample forecast accuracy measures of the two selected candidate models based on the test dataset.

The main comparison criterion was RMSE (Root Mean Squared Error), since it is commonly used to evaluate forecasting performance and penalizes large forecast errors more strongly. Although ARIMA(2,2,0) produced lower MAE and MAPE values, ETS(M,A,N) achieved the lower RMSE value of 0.384728 compared to 0.420262 for ARIMA(2,2,0). Since lower RMSE indicates better overall forecasting performance, ETS(M,A,N) was selected as the final model. This result shows that ETS provides slightly better overall forecast accuracy for this dataset.

5.3 Choosing the Final Model

Based on the forecast accuracy comparison, ETS(M,A,N) was selected as the final forecasting model because it produced the lowest RMSE value among the selected candidate models. Although ARIMA(2,2,0) had slightly lower MAE and MAPE values, RMSE was selected as the main criterion because it penalizes large forecast errors more heavily and provides a better measure of overall forecasting performance in this study. Therefore, ETS(M,A,N) was considered to provide the best overall forecasting performance. The final selected model is ETS(M,A,N), which represents an exponential smoothing model with multiplicative error, additive trend and no seasonal component.

6. Final Forecast Using All Data

6.1 Refitting the Final Model

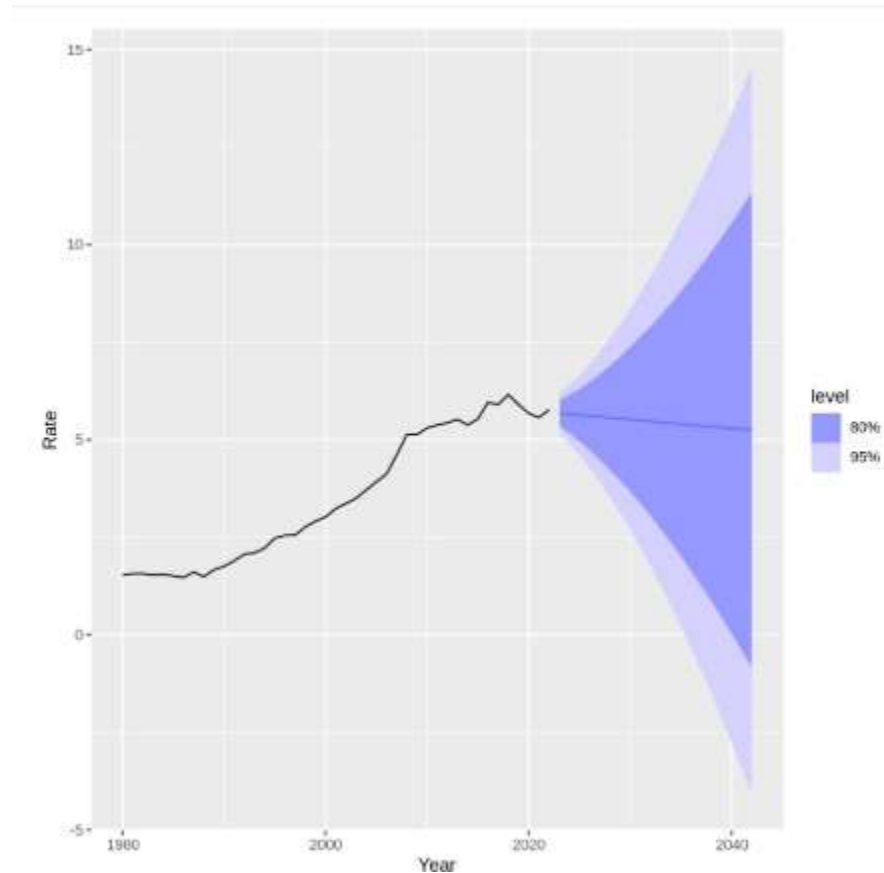


Figure 6.1: Final 20 Year Forecast Using ETS(M,A,N)

After selecting ETS(M,A,N) as the final forecasting model, the model was refitted using the full data set (1980–2022) in order to generate the final long term forecasts. Using the complete dataset improves model reliability since all available information is included in the estimation process.

Figure 6.1 presents the 20 year forecast horizon together with the historical fertility rate series.

The forecast shows a slight downward trend in future fertility rates, while the prediction intervals widen significantly over time. This widening uncertainty is expected in long term forecasting and reflects the increasing difficulty of predicting distant future observations.

An 80% prediction interval was included to show the increasing uncertainty of long term forecasts.

6.2 Point Forecasts and Prediction Intervals

.model	Year	.mean	80%
<chr>	<dbl>	<dbl>	<hilo>
finalMAN	2023	5.678668	[5.34330747, 6.014028]80
finalMAN	2024	5.655936	[5.16689759, 6.144975]80
finalMAN	2025	5.633205	[4.96422255, 6.302187]80
finalMAN	2026	5.610473	[4.73997826, 6.480969]80
finalMAN	2027	5.587742	[4.49684508, 6.678639]80
finalMAN	2028	5.565011	[4.23656365, 6.893458]80
finalMAN	2029	5.542279	[3.96036345, 7.124195]80
finalMAN	2030	5.519548	[3.66915779, 7.369938]80
finalMAN	2031	5.496816	[3.36364416, 7.629989]80
finalMAN	2032	5.474085	[3.04436119, 7.903809]80
finalMAN	2033	5.451354	[2.71172366, 8.190984]80
finalMAN	2034	5.428622	[2.36604530, 8.491199]80
finalMAN	2035	5.405891	[2.00755441, 8.804227]80
finalMAN	2036	5.383159	[1.63640490, 9.129914]80
finalMAN	2037	5.360428	[1.25268431, 9.468172]80
finalMAN	2038	5.337697	[0.85641988, 9.818974]80
finalMAN	2039	5.314965	[0.44758301, 10.182348]80
finalMAN	2040	5.292234	[0.02609281, 10.558375]80
finalMAN	2041	5.269503	[-0.40818128, 10.947186]80
finalMAN	2042	5.246771	[-0.85541737, 11.348960]80

Table 6.1: 20 Year Point Forecasts and 80% Prediction Intervals

Table 6.1 presents the point forecasts and 80% prediction intervals for the next 20 years obtained from the final ETS(M,A,N) model.

The point forecasts show a gradual decline in fertility rates over the forecast horizon, while the prediction intervals become considerably wider over time. For example, the forecasted mean value for 2023 is approximately 5.68, while by 2042 it decreases to nearly 5.25. At the same time, the lower and upper bounds of the prediction interval expand significantly, indicating higher uncertainty for long term forecasts. This behavior is common in time series forecasting, since uncertainty accumulates as the forecast horizon increases. Therefore, long term forecasts should be interpreted carefully, especially when prediction intervals become very wide.

7. Conclusion

In this project, the fertility rate of women older than 40 in Australia was analyzed using the `aus_fertility` dataset from the `fpp3` package. After filtering the data for Australia, years after 1980, and age groups above 40, yearly average fertility rates were calculated to obtain a single annual time series. Since the series showed a strong upward trend and no seasonal pattern, ETS and ARIMA models were considered for forecasting. For ETS modeling, ETS(M,A,N) was selected as the best model based on the lowest AICc value and successful residual diagnostics. The Ljung-Box test also confirmed that the residuals behaved like white noise. For ARIMA modeling, ARIMA(2,2,0) was selected after evaluating manual models, stepwise selection and exhaustive search methods. Residual diagnostics also supported the adequacy of this model. Although both ETS and ARIMA performed reasonably well, forecast accuracy comparison showed that ETS(M,A,N) achieved the lowest RMSE value and therefore provided the best overall forecasting performance. As a result, ETS(M,A,N) was selected as the final forecasting model and used to generate the final 20 year forecast. The final forecast suggests a slight decline in fertility rates in future years, while prediction intervals widen significantly over time, reflecting increasing uncertainty in long term forecasting. Overall, the project shows that ETS(M,A,N) provides the most reliable forecasting structure for this dataset.