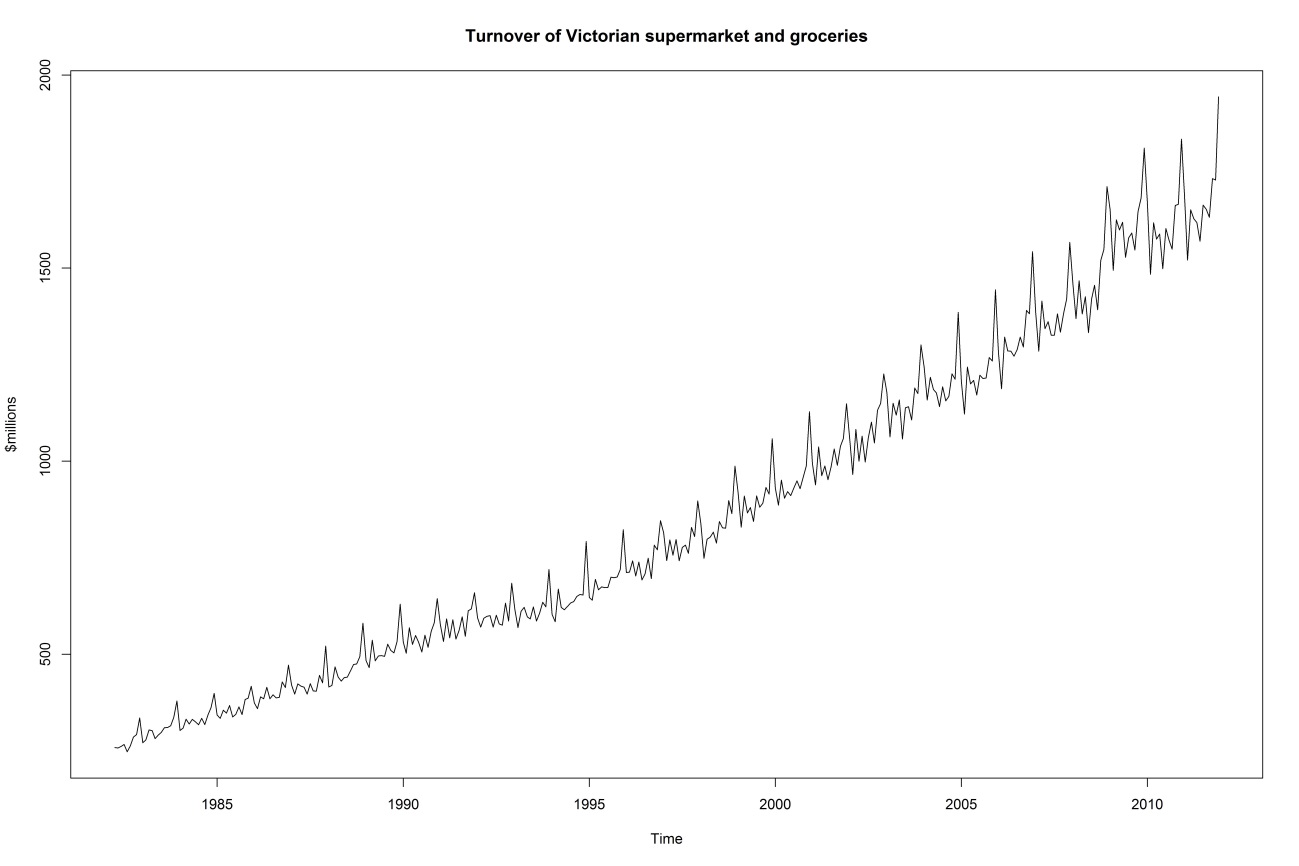
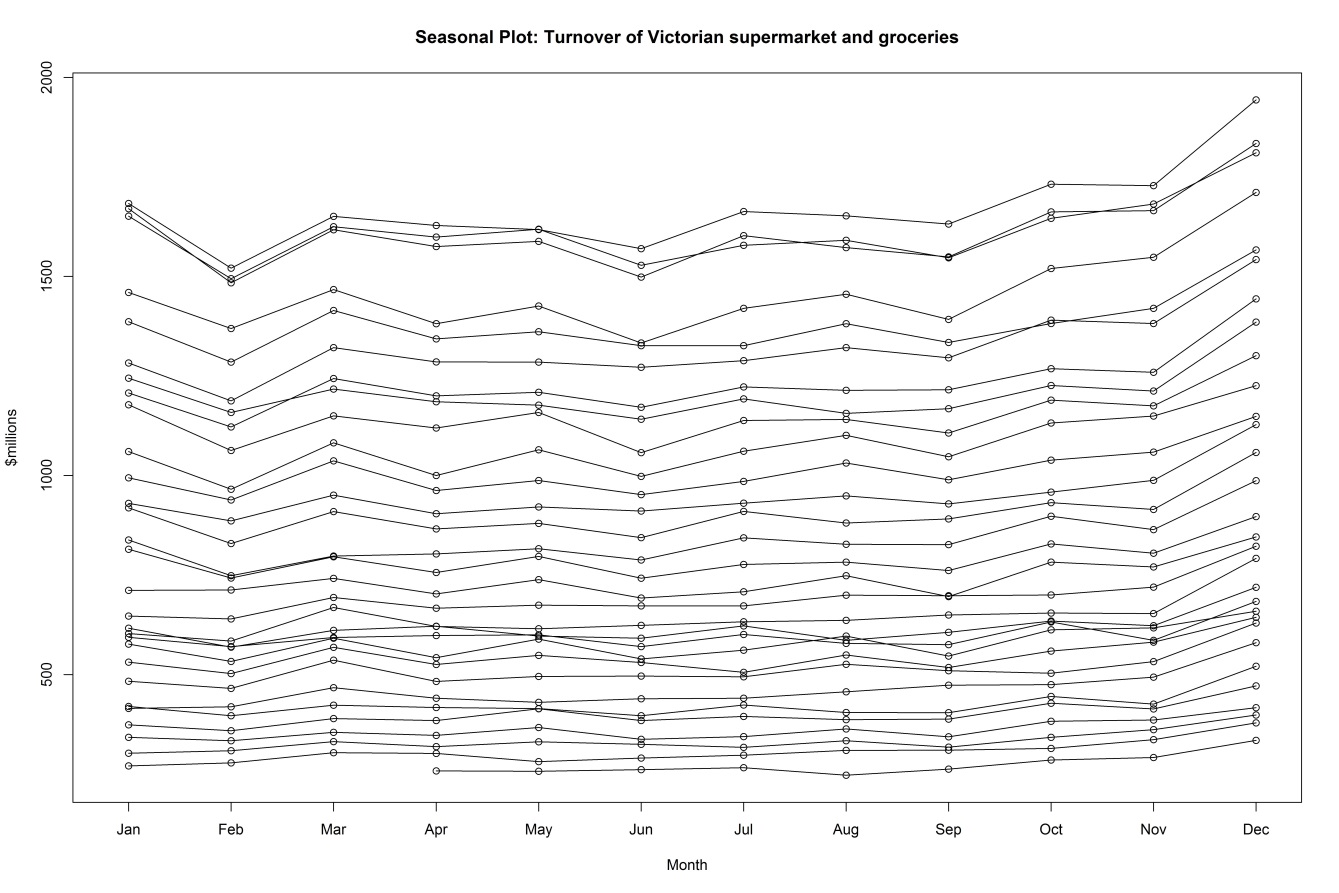
**ORIGINAL DATA**





The original data for the turnover of supermarkets and groceries in Victoria has a strong seasonal pattern and is upwards trending. The seasonal fluctuations also gradually increase in size. The seasonal plot shows an increase in turnover in December.

**TRANSFORMATIONS/DIFFERENCING**

Performing an Augmented Dicky Fuller test and a KPSS test indicates that the original time series is non-stationary.

Due to the gradual increase in the size of seasonal fluctuations a Box Cox transformation was used to stabilise the size of the fluctuations. Using the BoxCox.lambda() function to find the optimal lambda value, lambda = 0.2310188.

To calculate the number of differences that needed to be taken, I used automatic functions ndiffs() for ordinary differencing and nsdiffs() for seasonal differencing. The output suggested that one difference should be taken for each ordinary and seasonal difference.

The original time series data set is split into a test set containing the last 24 months of given data (January 2010 to December 2011) and a training set containing all of the previous data. The training set is named retail1 and the training set is named retail2.

The training set and the test also undergo a Box Cox transformation using the same lambda as the original data did ie. lambda = 0.2310188. The new training and test sets are named retail1.bc and retail2.bc repectively.

**ARIMA**

The create a short-list of appropriate ARIMA models, the auto.arima() function was used. The function produced an ARIMA(2, 1, 1)(2, 0, 0) model with an AICc statistic of -451.99. Afterwards, variations were of this model were made to try and produce a smaller AICc statistic. The ARIMA models that were short-listed and their AICc statistics are:

|  |  |
| --- | --- |
| ARIMA | AICc |
|  | -451.99 |
|  | -518.09 |
|  | -465.89 |
|  | -492.76 |

has the lowest AICc statistic at -518.0.

These ARIMA models are then tested by fitting them to the transformed training set, forecasted 24 months and the forecasts are compared to the transformed test set using the accuracy() function.

The following table shows the models and their RMSE values.

|  |  |
| --- | --- |
| ARIMA | RMSE |
|  | 0.5065311 |
|  | 0.3414380 |
|  | 0.4334369 |
|  | 0.4039567 |

has both the lowest AICc and RMSE values and so this model will be chosen to create forecasts.

:

FORECASTS & PREDICTION INTERVALS

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

Jan 2010 19.87412 19.73416 20.01408 19.66007 20.08817

Feb 2010 19.39572 19.25196 19.53948 19.17585 19.61559

Mar 2010 19.86588 19.71213 20.01964 19.63074 20.10103

Apr 2010 19.74464 19.58292 19.90636 19.49731 19.99197

May 2010 19.79842 19.62884 19.96800 19.53907 20.05776

Jun 2010 19.60023 19.42319 19.77726 19.32948 19.87098

Jul 2010 19.72604 19.54185 19.91024 19.44434 20.00775

Aug 2010 19.76572 19.57463 19.95681 19.47347 20.05797

Sep 2010 19.67648 19.47873 19.87423 19.37405 19.97891

Oct 2010 19.93521 19.73103 20.13940 19.62294 20.24749

Nov 2010 19.97619 19.76576 20.18662 19.65437 20.29801

Dec 2010 20.48131 20.26482 20.69780 20.15022 20.81240

Jan 2011 20.01391 19.76609 20.26173 19.63491 20.39292

Feb 2011 19.64458 19.38780 19.90136 19.25187 20.03729

Mar 2011 20.08631 19.81794 20.35469 19.67587 20.49675

Apr 2011 19.87868 19.59979 20.15756 19.45216 20.30520

May 2011 19.96004 19.67091 20.24918 19.51786 20.40223

Jun 2011 19.78851 19.48951 20.08752 19.33122 20.24580

Jul 2011 19.94913 19.64056 20.25769 19.47721 20.42104

Aug 2011 20.01592 19.69808 20.33375 19.52982 20.50201

Sep 2011 19.92164 19.59479 20.24849 19.42177 20.42151

Oct 2011 20.19480 19.85918 20.53042 19.68152 20.70808

Nov 2011 20.20211 19.85795 20.54628 19.67576 20.72847

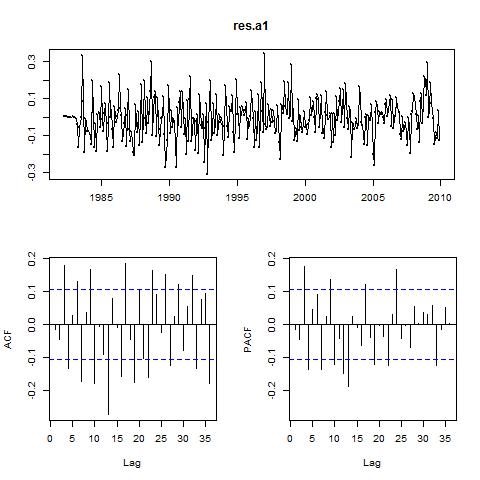
Dec 2011 20.78355 20.43105 21.13606 20.24445 21.32266

DIAGNOSTICS

Box-Ljung test

data: res.a1

X-squared = 57.7509, df = 0, p-value < 2.2e-16



**EXPONENTIAL SMOOTHING**

The create a short-list of appropriate ETS models, the ets() function was used. The function produced an ETS(A, A, A) model with an AICc statistic of 4035.589. Afterwards, variations were of this model were made to try and produce a smaller AICc statistic. The ETS models that were short-listed and their AICc statistics are:

|  |  |
| --- | --- |
| ETS | AICc |
|  | 4035.589 |
|  | 4060.981 |
|  | 3943.237 |
|  | 3944.672 |

has the lowest AICc statistic at 3943.237.

These ARIMA models are then tested by fitting them to the untransformed training set, forecasted 24 months and the mean of the forecasts are compared to the untransformed test set using the accuracy() function.

The following table shows the models and their RMSE values.

|  |  |
| --- | --- |
| ETS | RMSE |
|  | 87.06782 |
|  | 70.54786 |
|  | 92.98789 |
|  | 82.678 |

The model has the lowest RMSE but also the highest AICc. Since we are comparing within models we will choose the model with the lowest AICc since this statistic is based on more data and so it is more reliable. has already been established as having the lowest AICc and so we will choose this model to forecast.

FORECASTS & PREDICTION INTERVALS

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

Jan 2010 1700.618 1642.596 1757.519 1611.704 1787.322

Feb 2010 1572.168 1515.796 1628.095 1485.042 1657.542

Mar 2010 1709.264 1644.726 1773.315 1615.370 1805.925

Apr 2010 1639.466 1575.514 1702.108 1543.975 1737.139

May 2010 1666.854 1599.872 1731.275 1563.885 1768.801

Jun 2010 1596.975 1529.231 1664.753 1495.621 1701.395

Jul 2010 1657.167 1583.654 1729.977 1547.818 1766.426

Aug 2010 1679.511 1601.669 1755.470 1564.368 1794.723

Sep 2010 1639.178 1560.590 1718.259 1519.563 1758.180

Oct 2010 1741.005 1652.926 1827.738 1606.287 1874.595

Nov 2010 1750.402 1658.569 1840.481 1611.850 1888.600

Dec 2010 1948.164 1842.451 2055.054 1790.006 2112.488

Jan 2011 1784.885 1679.528 1890.921 1631.116 1948.915

Feb 2011 1648.475 1548.105 1750.556 1501.020 1810.220

Mar 2011 1790.527 1679.650 1903.756 1617.836 1973.322

Apr 2011 1715.816 1600.358 1834.030 1538.061 1901.631

May 2011 1742.891 1622.823 1868.641 1561.985 1934.450

Jun 2011 1668.336 1547.648 1791.828 1489.583 1861.613

Jul 2011 1729.704 1597.513 1860.593 1535.533 1937.608

Aug 2011 1751.524 1615.697 1889.367 1549.363 1965.625

Sep 2011 1708.027 1573.442 1848.936 1504.270 1929.702

Oct 2011 1812.639 1661.555 1971.609 1589.458 2053.429

Nov 2011 1820.953 1664.178 1986.662 1588.264 2072.359

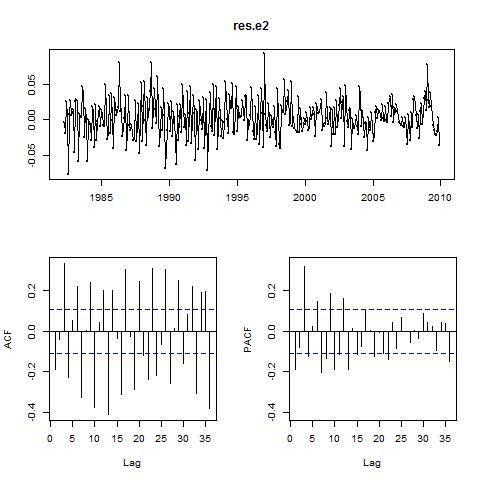
Dec 2011 2025.085 1848.450 2213.106 1755.895 2319.793

DIAGNOSTICS

Box-Ljung test

data: res.e2

X-squared = 205.0097, df = 0, p-value < 2.2e-16

****

**PREFERRED METHOD**

To compare between models the RMSE statistic from the accuracy tests against the test sets should be used. The preferred ARIMA model has a RMSE of 0.3414380 and the preferred ETS model has a RMSE of 92.98789. The ARIMA model has the lower RMSE and so this is the preferred method.

**OUT OF SAMPLE FORECASTS**

Data from the ABS website (Cat. 8501.0, Table 11) has been read into R manually and named as absretail.

ARIMA

Point Forecast Lo 80 Hi 80

Jan 2012 1776.094 1732.405 1820.626

Feb 2012 1653.539 1610.736 1697.211

Mar 2012 1786.156 1737.041 1836.331

Apr 2012 1737.150 1686.233 1789.241

May 2012 1749.783 1695.775 1805.105

Jun 2012 1697.145 1641.815 1753.898

Jul 2012 1761.530 1702.050 1822.595

Aug 2012 1781.082 1718.644 1845.249

Sep 2012 1747.204 1683.367 1812.886

Oct 2012 1843.636 1774.767 1914.543

Nov 2012 1850.425 1779.105 1923.924

Dec 2012 2057.027 1977.261 2139.245

Jan 2013 1895.571 1810.571 1983.607

Feb 2013 1771.193 1687.388 1858.163

Mar 2013 1911.602 1818.506 2008.321

Apr 2013 1844.334 1750.093 1942.429

May 2013 1873.979 1774.939 1977.214

Jun 2013 1802.393 1702.914 1906.282

Jul 2013 1863.162 1757.748 1973.371

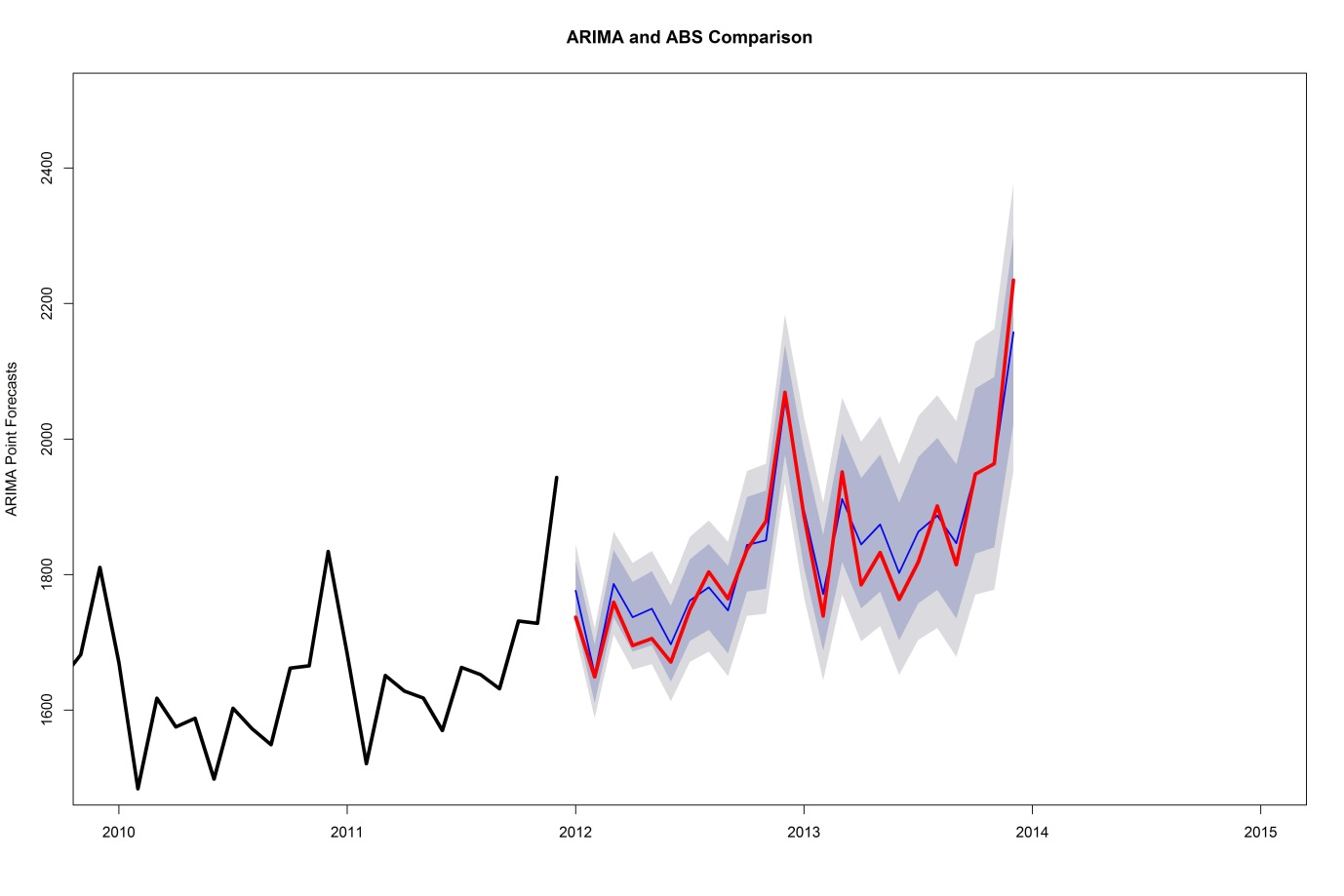
Aug 2013 1887.004 1777.277 2001.868

Sep 2013 1846.236 1735.233 1962.620

Oct 2013 1949.872 1830.917 2074.682

Nov 2013 1962.973 1840.315 2091.823

Dec 2013 2157.556 2022.347 2299.611



Forecasts are shown in blue and the ABS data is shown in red.

Accuracy test

ME RMSE MAE

Training set 2.500917 21.92478 16.59903

Test set 61.261048 77.36985 64.19955

MPE MAPE MASE

Training set 0.2471592 2.100711 0.3312126

Test set 3.2224935 3.392793 1.2810204

ACF1 Theil's U

Training set -0.01651142 NA

Test set 0.43329690 0.669133

ETS

Point Forecast Lo 80 Hi 80

Jan 2012 1759.837 1699.474 1821.458

Feb 2012 1614.781 1556.781 1672.819

Mar 2012 1758.299 1692.569 1824.232

Apr 2012 1696.465 1628.385 1763.318

May 2012 1716.960 1645.478 1787.412

Jun 2012 1645.449 1573.638 1716.189

Jul 2012 1717.590 1636.700 1796.081

Aug 2012 1726.855 1638.386 1811.587

Sep 2012 1689.106 1600.910 1774.361

Oct 2012 1793.819 1695.843 1891.137

Nov 2012 1798.266 1695.789 1901.642

Dec 2012 2000.081 1877.034 2121.437

Jan 2013 1823.995 1706.263 1944.627

Feb 2013 1671.970 1558.747 1786.299

Mar 2013 1818.795 1687.382 1949.838

Apr 2013 1753.169 1623.173 1883.142

May 2013 1772.712 1632.518 1914.074

Jun 2013 1697.357 1560.799 1838.124

Jul 2013 1770.232 1617.267 1927.346

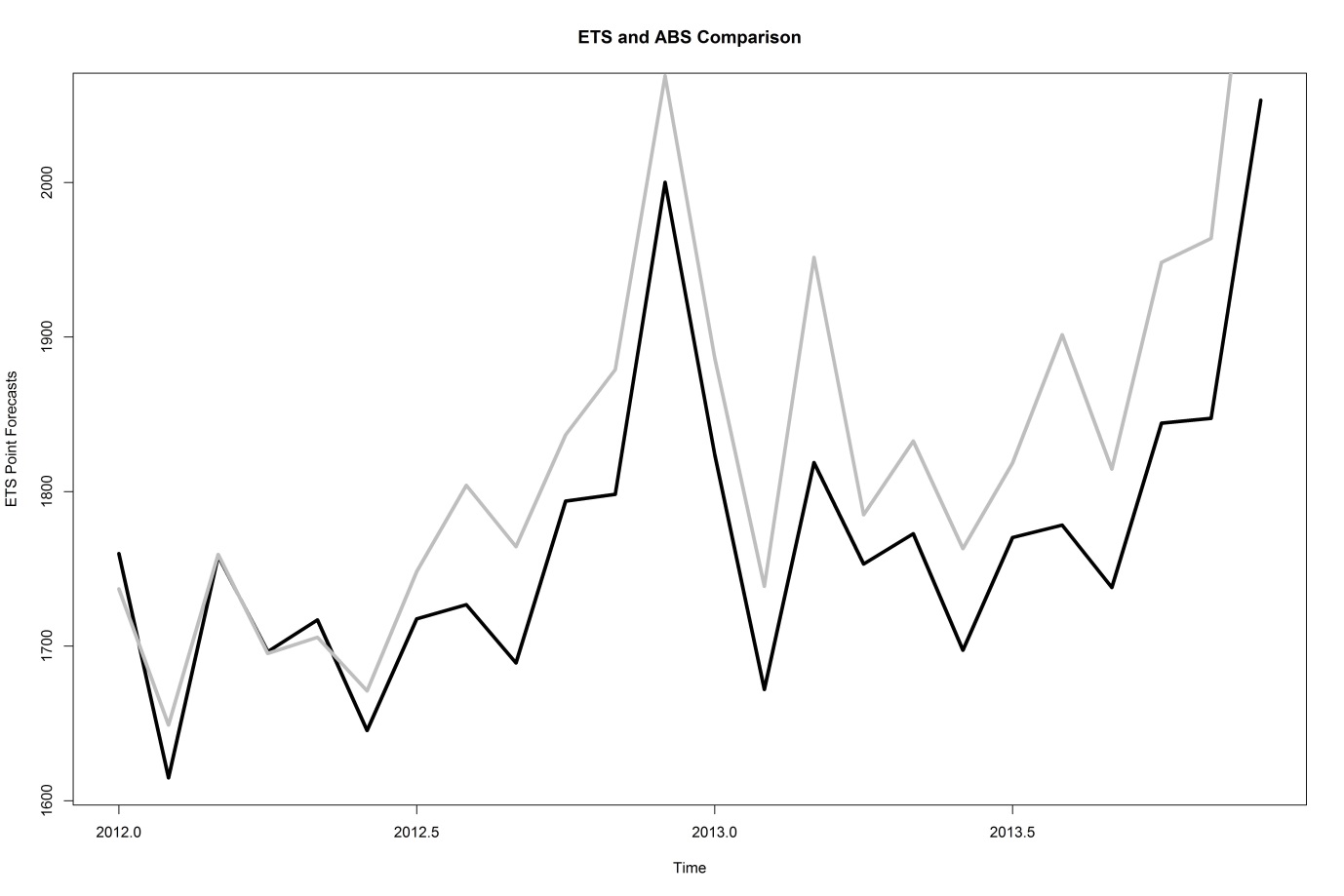
Aug 2013 1778.274 1622.219 1941.416

Sep 2013 1737.969 1581.026 1899.544

Oct 2013 1844.236 1670.393 2025.536

Nov 2013 1847.371 1666.349 2039.959

Dec 2013 2053.145 1847.596 2272.355



Forecasts are shown in black and the ABS data is shown in grey.

Accuracy test

ME RMSE MAE

Training set 2.500917 21.92478 16.59903

Test set 61.261048 77.36985 64.19955

MPE MAPE MASE

Training set 0.2471592 2.100711 0.3312126

Test set 3.2224935 3.392793 1.2810204

ACF1 Theil's U

Training set -0.01651142 NA

Test set 0.43329690 0.669133

ABS

Jan Feb Mar Apr May Jun

2012 1737.0 1649.0 1759.3 1695.3 1705.7 1671.0

2013 1886.5 1738.8 1951.5 1785.0 1832.6 1763.1

Jul Aug Sep Oct Nov Dec

2012 1748.2 1804.0 1764.4 1836.7 1878.9 2068.9

2013 1818.5 1901.4 1814.6 1948.2 1963.8 2234.6