Office for National Statistics

UK Centre for the Measurement of Government Activity

August 2008

From Holt-Winters to ARIMA Modelling: Measuring the Impact on Forecasting Errors for Components of Quarterly Estimates of Public Service Output

1. Introduction

The UK Centre for the Measurement of Government Activity (UKCeMGA) was launched in July 2005 in order to take forward the recommendations from the *Atkinson Review to improve the measure of Government Output and Productivity*.

The division has been established to strengthen the capability of ONS to publish authoritative and coherent measures of the output and productivity of government-funded services in the UK National Accounts. One of the key objectives of the work programme is to ensure that the measures of key government services in the UK National Accounts are fit for purpose.

UKCeMGA delivers output data to National Accounts on a quarterly basis for seven different public services, namely Adult Social Care, Children's Social Care, Civil and Family Courts, Education, the Fire and Rescue Services, Health Care and Social Security Administration.

Forecasting techniques are used extensively throughout UKCeMGA for data deliveries to National Accounts in order to extend time series beyond the most recent period for which data are available. This is often carried out because a time lag can occur in some datasets, meaning that some data are unavailable, and data that are available are sometimes incomplete. For example, data for the current financial or calendar year will often not be available until the following year. To overcome this we forecast these missing data points. Actual data content makes up just 36 per cent of the GDP(O) estimate of Government output, even three months after the end of the relevant quarter. Clearly good forecasting is essential if accurate GDP(O) estimates are to be produced.

1.1 Aim of the paper

Within UKCeMGA and throughout the rest of ONS the primary method for National Accounts forecasting is Holt-Winters. This forecasting method will soon change from Holt-Winters to ARIMA modelling as the current database system winCSDB is replaced by the new database system CORD. This change in forecasting method may affect the accuracy of UKCeMGA's output forecasts.

This paper presents the methods and findings of a UKCeMGA project which aimed to assess the extent to which UKCeMGA's forecasting accuracy may be affected by the move from Holt-Winters to ARIMA modelling. The shift in database system will involve changes to various methods associated with data deliveries to National Accounts, of which forecasting is just one. The analysis reported on in this paper hones in on the effects of changing just the method of forecasting. It does not consider the effects of any other methodological changes associated with the shift in database system. This paper is part of a wider UKCeMGA objective to investigate potential statistical errors in public service output data - see *A Framework For Identifying Sources Of Statistical Error In Estimates Of Public Service Output And Productivity* (ONS 2008).

1.2 Structure of the paper

The remainder of the paper takes the following structure:

2. Holt-Winters forecasting method
3. ARIMA modelling and forecasting method
4. Data availability
5. Methods of forecasting, evaluation and comparison
6. Annual data results
7. Quarterly data results
8. Identification of key series
9. Conclusions and next steps
10. Annexes
11. References

2. Holt-Winters Forecasting Method

Forecasting techniques can be largely classified as judgmental, univariate or multivariate. Judgmental forecasts are made by experts. Univariate forecasts involve just one explanatory variable, whilst multivariate forecasts involve more than one explanatory variable. Holt-Winters is a univariate method.

There is also a distinction between automatic and non-automatic forecasting procedures. Non-automatic procedures require user intervention whilst automatic procedures do not. Holt-Winters is generally viewed as an automatic procedure but the user can intervene if required. Holt-Winters is popular for mass produced forecasts, for example in production planning, because of its simplicity.

The Holt method uses simple exponential smoothing in order to forecast. The forecast is obtained as a weighted average of past observed values where the weights decline exponentially so that the values of recent observations contribute to the forecast more than the values of earlier observations. Most time series have three components: trend, seasonal and irregular. The irregular component is the residual after trend and seasonality have been removed. The Holt method accounts for only the trend and irregular components. The Holt-Winters method builds on this by allowing for seasonality. Forecasted values are dependent on the level, slope and seasonal components of the series being forecast.

A full algebraic description of the Holt-Winters method can be found in Annex A.

3. ARIMA Modelling and Forecasting Method

Autoregressive Integrated Moving Average models, or ARIMA models, are a class of models that can be used to produce forecasts and backcasts. The use of ARIMA models is also known as the 'Box-Jenkins' approach following the work of Box and Jenkins (Box G. and Jenkins G. 1970).

3.1 The general ARIMA model

ARIMA models have three parts, although not all parts are always necessary. The three parts are the autoregression part (AR), the integration part (I) and the moving average part (MA).

The main assumption surrounding the AR part of a time series is that the observed value depends on some linear combination of previous observed values up to a defined maximum lag (denoted p), plus a random error term ε_t. The main assumption surrounding the MA part of a time series is that the observed value is a random error term plus some linear combination of previous random error terms up to a defined maximum lag (denoted q).

To analyse a time series we require that all of the observations are independently identifiable. Hence there should be no autocorrelation in the series and the series should have zero mean. In order for these requirements to be met all of the signal (the trend and seasonal components of the series being modelled) must have been removed from the series so that we are left with only noise. Therefore it is only the irregular component of the series which is being modelled, not the trend or seasonal components. If the series has zero mean and other moments such as the variance and covariance do not depend on the passage of time, then the series is said to be stationary. In order to achieve stationarity the series must be differenced (unless it is stationary to begin with). This means taking the differences between successive observations and then analysing these differences instead of the actual observations. This process of differencing is known as integration and the order of differencing is denoted d.

The general seasonal model is denoted ARIMA(pdq)(PDQ), where p, d and q refer to the orders of the nonseasonal AR, I and MA parts of the model respectively and P, D and Q refer to the orders of the seasonal AR, I and MA parts of the model respectively.

A full algebraic description of the general ARIMA model can be found in Annex B1.

3.2 The regARIMA model

An extension of the pure ARIMA model is the regARIMA model, which allows for the inclusion of regression effects. Such regression effects could be caused by distortive features in the series such as additive outliers, level shifts, holiday effects and trading day effects, although this list is not exhaustive. The regARIMA model implies that:

i) the regression effects are subtracted from the original, untransformed series to obtain the regression error series,
ii) the regression error series is differenced to obtain a stationary series, say w_t, and finally
iii) w_t is assumed to follow a stationary ARMA process.

A full algebraic description of the regARIMA model can be found in Annex B2.

3.3 Forecasting

Forecasts can be produced once an adequate regARIMA model has been specified. Forecasting is applied using a recursive process that is similar to the Holt-Winters procedure.

Forecasts are minimum mean squared error (MMSE) linear predictions of future values based on present and past values if:
- the true ARIMA(*pdq*)(*PDQ*) orders are identified
- the true ARMA parameters are estimated
- the necessary regression variables are included in the model
- no additive outliers or level shifts will occur in the forecast period

Consider the transformed series y_t. If y_n is the most recent observation then a k-step ahead forecast is y_{n+k}, denoted $y_n(k)$. The MMSE estimator is the conditional expected value so that:

$$y_n(k) = E(y_{n+k}/y_n,\ldots, y_1)$$

where y_1 is an earlier observation.

If forecasts of more than one-step ahead are produced then earlier forecasts will be used in addition to actual data.

An algebraic description of the basic ARIMA forecasting method can be found in Annex B3.

4. Data Availability

UKCeMGA delivers output data to National Accounts on a quarterly basis for seven different public services, namely Adult Social Care, Children’s Social Care, Civil and Family Courts, Education, the Fire and Rescue Services, Health Care and Social Security Administration.

UKCeMGA receives administrative data from various Government departments across the UK. The use of administrative data carries certain hazards. In terms of the accuracy quality dimension, the data could be subject to processing errors which may cause outliers in the data. In terms of the timeliness quality dimension, not all of the data is always available in time for the relevant data delivery, hence the need for forecasting. This timeliness problem is attributable to both the scheduled production time of the data and the punctuality of release of the data.

The data are activity quantities, for example within Social Security Administration this could be the number of claims made for a particular benefit in a given time period. The periodicity of the activity data is either quarterly or annual. The periodicity of the annual data may be calendar years, financial years or, in the case of Education, academic years. These activity quantities are weighted together within each service according to their corresponding unit costs to produce a cost-weighted activity index for each service. The series are referenced to a particular year and are often chain-linked to create a period-on-period calculation of how the output is changing through time.

With the exception of the Health Care service, data for all services are forecast indirectly at the bottom level of aggregation, meaning the forecasts are appended to the original activity data before being weighted and aggregated together. Health Care forecasting, however, is carried out directly at the top level of aggregation, meaning all the sub-component activity series are weighted and aggregated together into one cost-weighted activity series, and the forecast is then appended to this series.

Where the periodicity of the activity data is annual, a spline is carried out in order to obtain the data on a quarterly basis. Forecasts are appended to the data prior to any interpolation being carried out and hence the periodicity of the forecasted data will be the same as the periodicity of the raw activity data.

For more information on the methods used by UKCeMGA, see *Documentation of UKCeMGA Methods used for National Accounts* (ONS 2007).

The availability of data, in terms of the coverage of the data and the length of the time series, varies greatly within and between the public services, hence the number of series being forecast and the date of the forecasted observation will also vary. The longest series analysed were 11 years in length, whilst the shortest were just 3 years in length. A summary of this information can be found in Annex D. In total, 278 series were analysed in the project.

5. Methods of Forecasting, Evaluation and Comparison

This section outlines how the two forecasting methods were used to produce forecasts in the project and how these forecasts were subsequently evaluated and compared.

5.1 Forecasting using the Holt-Winters method

Any time series generally consists of three components: trend, seasonal and irregular, although the seasonal component is not always present. If the seasonal component is present, it can be additive (where the size of the seasonal component is constant) or multiplicative (where the size of the seasonal component is proportionate to the level of the trend). When used automatically within the winCSDB database system, the Holt-Winters procedure will select an additive model. When used non-automatically, the user can make the choice. The user can also select a non-seasonal model for non-seasonal data. The data used in the project were checked for the presence of seasonality and decomposition type and the relevant series were adjusted accordingly when forecasts were produced using the Holt-Winters procedure. Details on the seasonality of the series can be found in Annex D. Note that this only concerns the quarterly series delivered to National Accounts as seasonality will not be present in annual series.

No automatic or manual treatment of outliers or other distortive effects were carried out on the data as such procedures are not currently carried out by the teams in UKCeMGA who process and deliver the data to National Accounts for each of the public services, therefore ensuring a like-for-like comparison.

5.2 Forecasting using ARIMA modelling

X12-ARIMA was used in the project to model the data as ARIMA and produce forecasts based on ARIMA models. X12-ARIMA is a computer package that can be used automatically - at the request of the user - in order to transform data, detect and replace outliers, identify and estimate models, seasonally adjust time series and produce forecasts and backcasts. The user can intervene in any of the above listed automatic processes as he or she deems necessary. However the programs automatic modelling capabilities were utilised in the project and these are outlined below.

5.2.1 Data transformation

Data may need to be transformed by a nonlinear transformation for several reasons, primarily for analysing a multiplicative series on an additive scale. Such a transformation can be specified by the X12-ARIMA user. Built-in transformations include those from the Box-Cox family, for example the log transformation. For this project X12-ARIMA performed an automatic test using F-adjusted Akaike's Information Criterion (AICC) to decide between a log transformation or no transformation, depending on whether the series was multiplicatively seasonal or additively seasonal (or not seasonal at all) respectively. Two models are automatically specified by the program for each series, one including a log transformation and one with no such transformation. The model that minimises the AICC measure is preferred by X12-ARIMA. There is the potential for transformation bias if the original data are transformed.

5.2.2 Regression variables

Specification of the regARIMA model requires specification of the regression variables. Although the user must have some knowledge of the series being modelled in order to decide which variables to ultimately include, X12-ARIMA can automatically identify potential outliers and distortive features and test their significance. X12-ARIMA uses dummy variables in the regression part of the model. For example, consider an additive outlier at time point t_0. Let χ_0 be a dummy variable where:

$\chi_0 = 1$ when $t = t_0$
$\chi_0 = 0$ otherwise

Adding a term $\beta\chi_0$ to the regression part of the model picks out an outlier effect of size β at time point t_0. The program then tests the significance of each type of outlier at each time point and t-statistics are produced. The critical t-values are determined by the number of observations in the span of data being tested for outliers. Significant outliers are added to the model as regression variables. When making use of the default parameters, the 'add one' (or 'forward selection') method is used, where the model is re-estimated after each single outlier is added to the model. This automatic procedure was used in the project in order to model the following regression effects.

Constant term

This allows for a nonzero mean after the series has been differenced. The program tests for this variable by default.

Easter

Easter effects arise when the level of the series is affected by the movement of the Easter holiday between different months and quarters (because Easter can fall anywhere between March 22^{nd} and April 25^{th}). The basic model assumes that the level of the series changes *E* days before Easter and remains at the new level until the day before Easter. X12-ARIMA automatically performs an AICC test for 'no Easter', *E*=1, *E*=8 and *E*=15 and includes the necessary regression variable in the regARIMA model. This automatic test was used in the project for all of the quarterly series. The annual series should not exhibit Easter effects so no such test was necessary.

Additive outliers

Additive outliers, or point outliers, only affect single observations and occur when an observation is unseasonally high or low. If detected (through dummy variables and significance testing as described above) then the value of the observation is replaced so that it is in line with what the program would 'expect'. Automatic additive outlier detection and replacement was used in the project.

Level shifts

A level shift occurs when there is a sudden increase or decrease in the level of the series by some constant amount after a certain time point. This is tested automatically by X12-ARIMA by using dummy variables in the regression part of the regARIMA model such that the dummy regression variable is equal to 0 before the level shift and 1 after the level shift. The effects are then tested for their significance and any effects whose t-statistic exceeds the critical value are included as variables in the regression model. The values of the observations before the level shift are adjusted upwards or downwards by some constant amount that is equal to the size of the effect of the level shift. Automatic level shift outlier detection and replacement was used in the project.

5.2.3 Model selection

Selection of the most appropriate ARIMA model is crucial for the purposes of accurate forecasting. If the selected model fits the current and previous observations well then it is hoped that it will be capable of predicting future observations successfully.

The ARIMA model can be selected automatically using a procedure called automdl, which is based closely on the TRAMO (Time series Regression with ARIMA noise, Missing values and Outliers) modelling procedure (Gomez V. and Maravall A. 1997). It is this procedure that was used to model the quarterly data in the project. The automdl procedure consists of 5 stages which are summarised in Annex C.

5.2.4 Length of series and periodicity of data

To maximise the benefit of modelling and hence forecasting with X12-ARIMA, the series being modelled should be at least 12 years long. The program has some minimum requirements in terms of the length of the series for certain functions to work. The absolute minimum length of any series for the program to perform any kind of modelling operation is 3 full years. 11 of the series modelled in the project were shorter than 3 full years and hence these series were excluded from the project. For series that are longer than 3 full years but shorter than 5 years and 1 quarter, automatic modelling cannot be performed. For the quarterly series in the project that were affected by this requirement the ARIMA model was fixed to (0 1 1)(0 1 1). The three numbers in the first set of brackets indicate the orders of the nonseasonal AR, I and MA parts of the ARIMA model respectively, whilst the three numbers in the second set of brackets indicate the seasonal orders of the same model.

For the annual series modelled in the project the ARIMA model was fixed to (0 2 2). The frequency of the observations in annual data is just one per year, compared to four observations per year for quarterly data and twelve observations per year for monthly data. Consider a series that is 12 years in length. If the periodicity of the data is quarterly the program shall specify a model based on 48 observations. If the periodicity of the data is monthly the program shall specify a model based on 144 observations. For annual data however the program can only attempt to specify a model based on just 12 observations. With so few observations being modelled, the automatic modelling procedure is unlikely to be able to fit a suitable model to the data. A simpler method, such as Holt-Winters, is likely to produce more robust forecasts for annual data than those produced by X12-ARIMA. The nonseasonal Holt-Winters method is a special case of the (0 2 2) ARIMA model and hence this model should be used for analysing annual data with X12-ARIMA. The results produced should be similar to those produced by Holt-Winters.

5.2.5 Forecasting

If a series has been transformed then forecasts are produced in the transformed scale and are obtained on the original scale by inverse transformation. Forecasts are MMSE predictions for the transformed series but not necessarily the original series.

95 per cent forecast confidence intervals (or 'prediction intervals') are constructed around point forecasts of the transformed data. Intervals on the original scale are obtained by inverse transformation.

Series that exhibit seasonality can be automatically seasonally adjusted by X12-ARIMA. When forecasts are produced for these series, the forecasts are appended to the original series and the seasonal adjustment is applied to the forecast rather than the forecast being appended to the seasonally-adjusted series.

5.3 Methods of evaluating and comparing forecasts

The project considered the Holt-Winters and ARIMA procedures for forecasting the next value of a time series from its own current and past values. That is, given a series of equally spaced observations Y_t, $t = 1,2,\ldots,n$ on some quantity Y, we required to forecast Y_{n+1}.
For the purpose of comparing the two methods it was necessary to calculate the error of the forecasts, meaning the forecasted value of each series was compared against the actual observed value in the relevant period. Therefore, depending on the date of the last observation for each individual series, the most recent data point available was excluded from the time series and was then forecast using both the Holt-Winters method and the ARIMA method. Both forecasted values were then compared against the actual observed value that was excluded from the series. This procedure represents a simulated out-of-sample forecasting comparison between the two

methods. So from above, instead of forecasting Y_{n+1} we required to forecast Y_n based on the observations Y_t, t = 1,2,….,n-1. One-step-ahead forecasts were produced as this is in line with current practice in terms of forecasting for the majority of UKCeMGA's data in the National Accounts. The timeliness of UKCeMGA's data varies both within and between services and hence the forecasting horizon for some series is more than one-step-ahead. In some cases the forecasting horizon is as much as two-years-ahead, which corresponds to eight-quarters-ahead. Despite this variation in forecasting horizons, the project analysed only one-step-ahead forecast comparisons for both the annual and the quarterly data.

Note that the analysis only considered one one-step-ahead forecasting comparison for each series. The rationale behind this method was that there is a trade-off between the number of observations in the out-of-sample forecast period and the number of observations in the series being modelled. Approximately 25 per cent of the series in the analysis would fall below 5 full years in length if the forecasting period was enlarged to beyond just one observation and this would result in losing the function of automatic modelling in X12-ARIMA, which is undesirable. Three series in the analysis would fall below 3 full years in length if the forecasting period was enlarged to beyond just one observation and these series would need to be omitted from the analysis as X12-ARIMA would not be capable of modelling them at all. In order to achieve a consistent method of comparison across all series, it was necessary to forecast only one one-step-ahead observation for every series in the analysis.

The error could then be calculated as the absolute difference between the actual value and the forecasted value, i.e. $|A_t - F_t|$ where t is the time period of the most recent data point available.

These errors were calculated for the two different procedures for all of the individual series within each of the public services. They were used to calculate the number - and hence percentage - of times that one method outperformed the other in terms of the number of series within each of the public services. Mean Absolute Percentage Error (MAPE) was also used as an evaluative measure of predictive performance for each method within each service. This measure is detailed below.

5.3.1 Mean Absolute Percentage Error

MAPE is calculated by dividing the difference between actual value A_{it} and forecasted value F_{itm} (known as the forecasting error) by the actual value A_{it}, where i is the series, t is the forecast period and m is the forecasting method. The resulting value is multiplied by 100 to obtain the forecasting percentage error. The absolute value of this calculation was summed accross every series within each public service and divided by the number of series n in that service in order to obtain the MAPE value.

$$MAPE_m = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{A_{it} - F_{itm}}{A_{it}}\right)\times 100\right|$$

The forecasting method that minimised the MAPE within each public service was considered the preferred method for forecasting the activity data in that service.

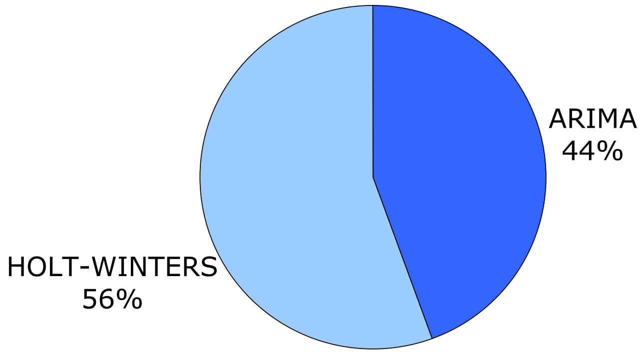
6. Annual Data Results

6.1 Adult Social Care (36 series)

Table 6.1 - Adult Social Care results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	15.5	56
ARIMA	15.1	44

Chart 6.1 - Percentage of times that one method outperforms the other for Adult Social



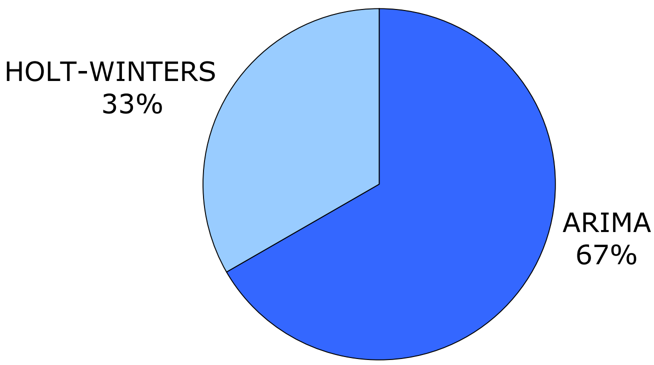
Holt-Winters outperforms ARIMA more frequently than ARIMA outperforms Holt-Winters, although ARIMA is favoured by the MAPE. The differences in accuracy are on average greater for the series where ARIMA is preferred than for the series where Holt-Winters is preferred.

6.2 Children's Social Care (6 series)

Table 6.2 - Children's Social Care results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	9.2	33
ARIMA	9.8	67

Chart 6.2 - Percentage of times that one method outperforms the other for Children's



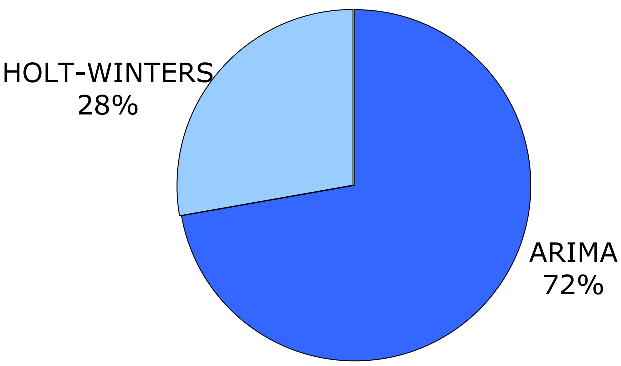
ARIMA outperforms Holt-Winters more frequently than Holt-Winters outperforms ARIMA, although Holt-Winters is favoured by the MAPE. The differences in accuracy are on average greater for the series where Holt-Winters is preferred than for the series where ARIMA is preferred.

6.3 Civil and Family Courts (18 series)

Table 6.3 - Civil and Family Courts (annual data) results

	MAPE(%)	Percentage of times that one method outperforms the other
Holt-Winters	19.8	28
ARIMA	10.0	72

Chart 6.3 - Percentage of times that one method outperforms the other for Civil and Family Courts (annual data)



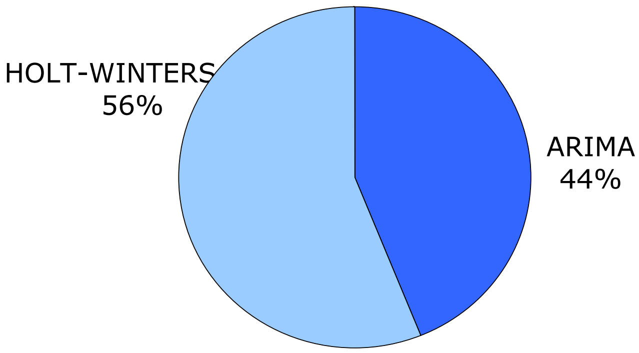
ARIMA is the preferred forecasting method according to both evaluative measures.

6.4 Education (75 series)

Table 6.4 - Education results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	6.5	56
ARIMA	7.3	44

Chart 6.4 - Percentage of times that one method outperforms the other for Education



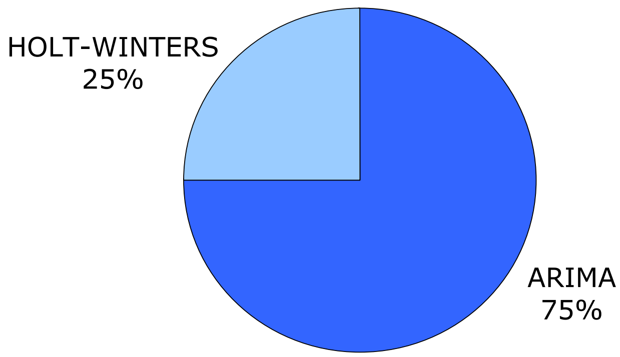
Holt-Winters is the preferred forecasting method according to both evaluative measures.

6.5 Fire and Rescue Services (4 series)

Table 6.5 - Fire and Rescue Services (annual data) results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	6.9	25
ARIMA	4.9	75

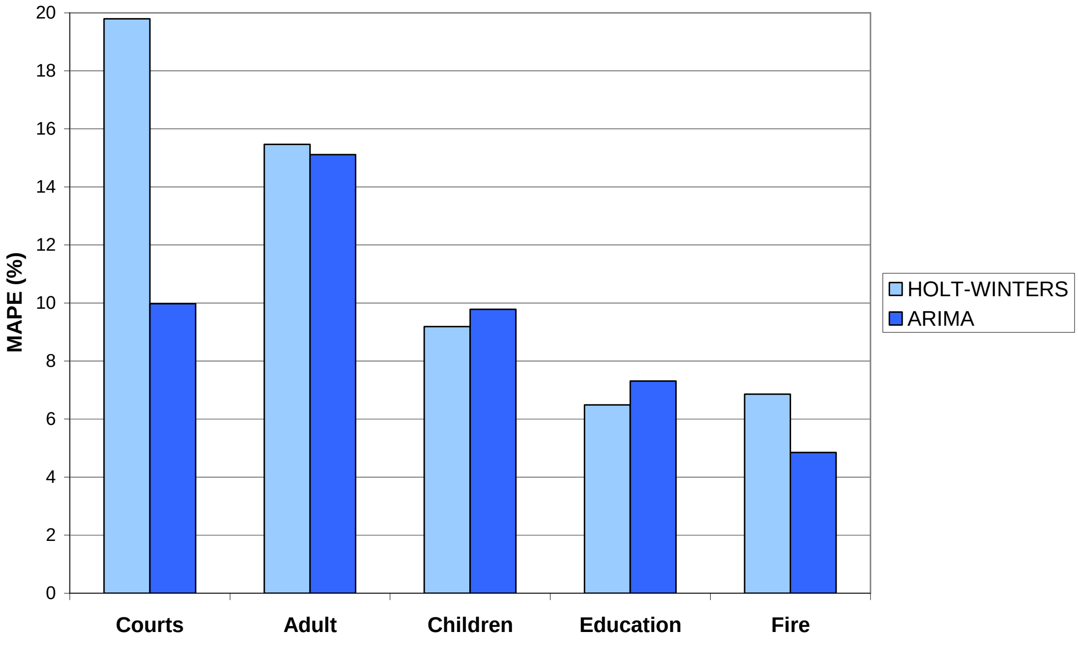
Chart 6.5 - Percentage of times that one method outperforms the other for the Fire and Rescue Services (annual data)



ARIMA is the preferred forecasting method according to both evaluative measures.

6.6 MAPE analysis

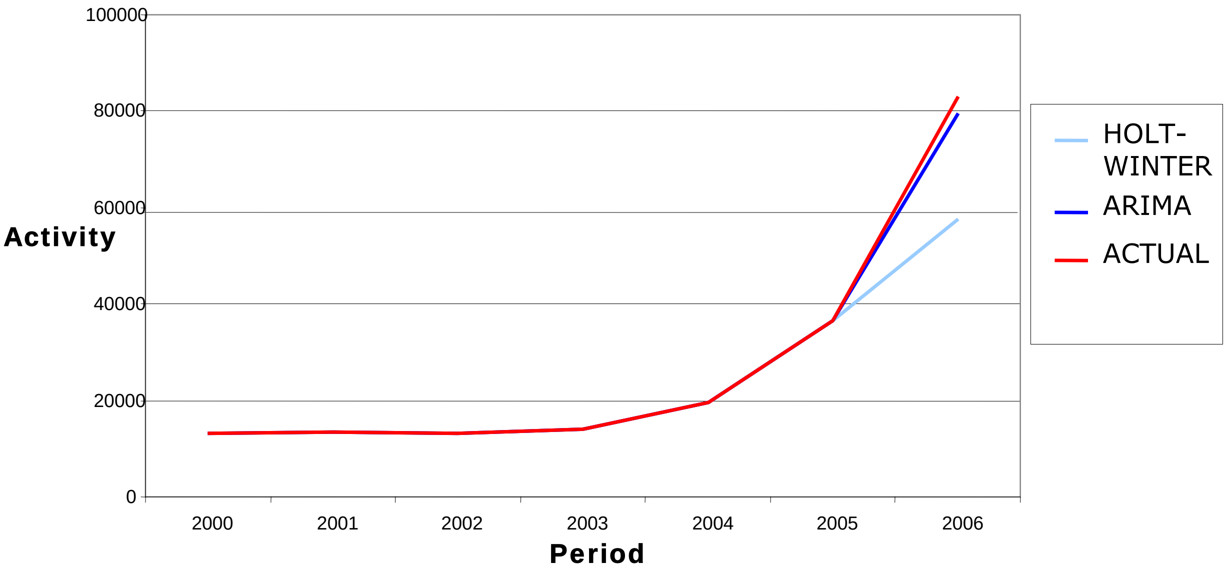
Chart 6.6.1 - Annual data results for each service and each method



The graph above indicates that, according to the MAPE, ARIMA modelling performs better than Holt-Winters in three out of the five services that deliver annual data to National Accounts, namely Civil and Family Courts, Adult Social Care and the Fire and Rescue Services. The biggest difference is in the Civil and Family Courts result, where the difference in MAPE values is 9.8 percentage points. Forecasts appear to be marginally better using the Holt-Winters procedure for Children's Social Care and Education.

The annual data was forecast in ARIMA using a (0 2 2) model, which should produce similar forecasting results to those produced by Holt-Winters. This is apparent in the results with the exception of Civil and Family Courts. ARIMA forecasting outperforms Holt-Winters by a substantial margin in the majority of the Civil and Family Courts annual sub-component series and the large difference in Mean Absolute Percentage Error between the two methods cannot be attributed to simply a few extreme results.

Chart 6.6.2 - Example of a Civil and Family Courts annual sub-component series



The example above is typical of the majority of the Civil and Family Courts annual sub-component series. The out-of-sample forecast period here is 2006 and the difference in forecasting accuracy of the two methods is apparent. The absolute percentage forecasting error produced by Holt-Winters for this series is over 30 per cent, whilst the same error produced by ARIMA modelling is just over 4 per cent.

Some of the series in each of the public services have very high absolute percentage errors for both the Holt-Winters method and the ARIMA method, caused by outliers in the forecast periods. As the forecast period is generally the most recent observation and therefore the final data point in the series, this outlier could be either a level shift or an additive outlier. The two types cannot be distinguished as the level of the series after the discontinuity is unknown. For the purposes of this analysis, we shall assume that the outliers in the forecast periods of the affected series are additive outliers. For a forecast to be a minimum mean squared error (MMSE) prediction of some future value based on present and past values, a key assumption is that there are no additive outliers or level shifts in the forecast period. If this assumption is violated then the forecast cannot be a MMSE prediction. This is the case for several of the series analysed in this project. The information below describes the relevant MAPE values if these series were to be omitted from the analysis. The plotted activity data for the relevant series can be found in Annex E and this can be used to highlight the additive outliers in the out-of-sample forecast periods.

- Within Adult Social Care the series entitled 'Referrals and assessments - learning disabilities' has very high absolute percentage errors for both methods (109 per cent and 122 per cent for Holt-Winters and ARIMA respectively). If this series is omitted from the analysis the MAPE values for Holt-Winters and ARIMA are 12.8 per cent and 12.1 per cent respectively as opposed to the existing values of 15.5 per cent and 15.1 per cent respectively.
- 'Secure accommodation' within Children's Social Care has an error of 27 per cent for Holt-Winters and 39 per cent for ARIMA. Omitting this series from the analysis would produce new MAPE's for Holt-

Winters and ARIMA of 5.6 per cent and 3.9 per cent respectively. As their previous values were 9.2 per cent and 9.8 per cent respectively, this omission results in ARIMA being the preferred method.

- 'Welsh health education – degrees' within Education has absolute percentage errors of 61 per cent in Holt-Winters and 63 per cent in ARIMA. Omitting this series from the analysis would reduce the MAPE's for Holt-Winters and ARIMA to 5.8 per cent and 6.6 per cent respectively (compared to 6.5 per cent for Holt-Winters and 7.3 per cent for ARIMA previously).

6.7 Prediction interval analysis

The X12-ARIMA program automatically constructs 95 per cent prediction intervals around point forecasts. It was therefore possible to calculate the percentage of Holt-Winters forecasts that lie within the upper and lower bounds of the prediction intervals constructed around the ARIMA forecasts. The results of the analysis are shown below for each public service.

Table 6.7 - Percentage of Holt-Winters forecasts that lie within the upper and lower bounds of the prediction intervals constructed around the ARIMA forecasts

Service	Percentage of Holt-Winters forecasts within X12-ARIMA prediction intervals
Adult Social Care	78
Children's Social Care	83
Education	81
Civil and Family Courts	61
Fire & Rescue Services	75

The results show that the majority of the forecasts produced by the Holt-Winters procedure lie within the prediction intervals constructed around the forecasts produced by ARIMA modelling. Note that eleven series were excluded from the analysis in Adult Social Care as these series were too short for prediction intervals to be constructed by X12-ARIMA. The results suggest that the forecasts produced by the two forecasting methods are similar. The reader should consider that prediction intervals were only constructed around ARIMA forecasts as such intervals were not provided as standard output of the system used to produce the Holt-Winters forecasts.

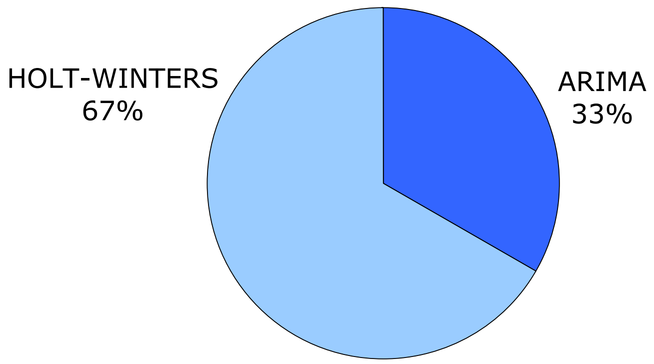
7. Quarterly Data Results

7.1 Civil and Family Courts (62 series)

Table 7.1 - Civil and Family Courts (quarterly data) results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	13.5	67
ARIMA	14.9	33

Chart 7.1 - Percentage of times that one method outperforms the other for Civil and Family Courts (quarterly data)



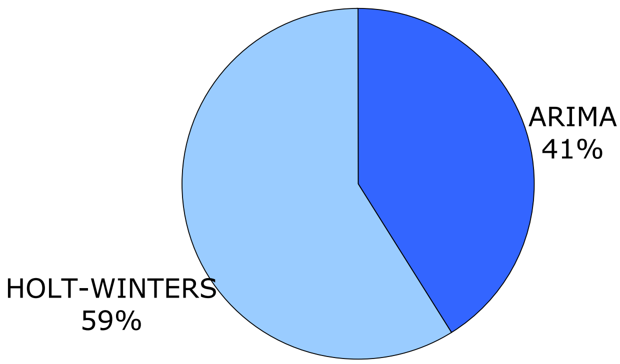
Holt-Winters is the preferred forecasting method according to both evaluative measures.

7.2 Fire and Rescue Services (51 series)

Table 7.2 - Fire and Rescue Services (quarterly data) results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	21.6	59
ARIMA	29.1	41

Chart 7.2 - Percentage of times that one method outperforms the other for the Fire and Rescue Services (quarterly data)



Holt-Winters is the preferred forecasting method according to both evaluative measures.

7.3 Health Care (1 series)

Currently, Health Care data is only forecast directly at the top level of aggregation, meaning only one cost-weighted activity series was forecast using the two procedures in the project. It was therefore not necessary to consider the MAPE or the percentage of times that one method outperforms the other. Instead, the forecasting error generated by each of the methods for the one single Health Care series should be considered. The forecasting percentage errors are displayed in table 7.3 below:

Table 7.3 - Health Care results

ABSOLUTE PERCENTAGE ERROR (%)	
Holt-Winters	ARIMA
0.2	0.5

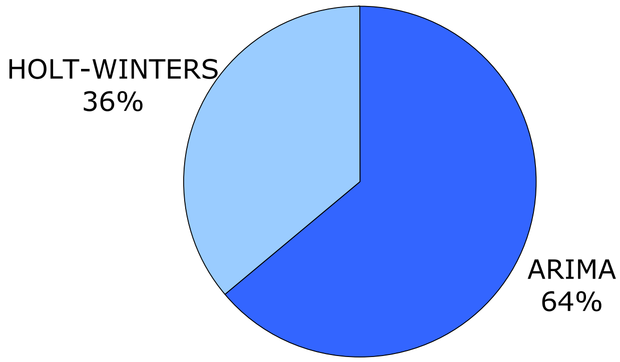
The table above indicates that when forecasting the cost-weighted Health Care activity series, ARIMA modelling is slightly outperformed by the Holt-Winters procedure. When considering these results, it should be remembered that the errors were produced from one single forecast appended to one single series for each method, rather than an average measure of multiple forecasting errors.

7.4 Social Security Administration (25 series)

Table 7.4 - Social Security Administration results

	MAPE (%)	Percentage of times that one method outperforms the other
Holt-Winters	3.5	36
ARIMA	3.4	64

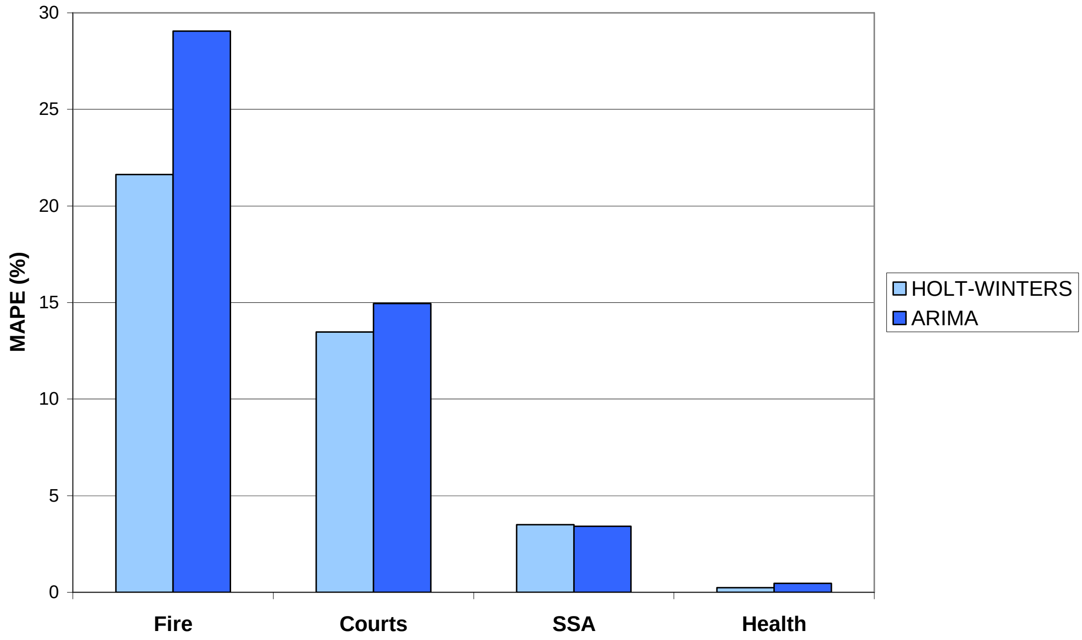
Chart 7.4 - Percentage of times that one method outperforms the other for Social Security Administration



ARIMA is the preferred forecasting method according to both evaluative measures.

7.5. MAPE analysis

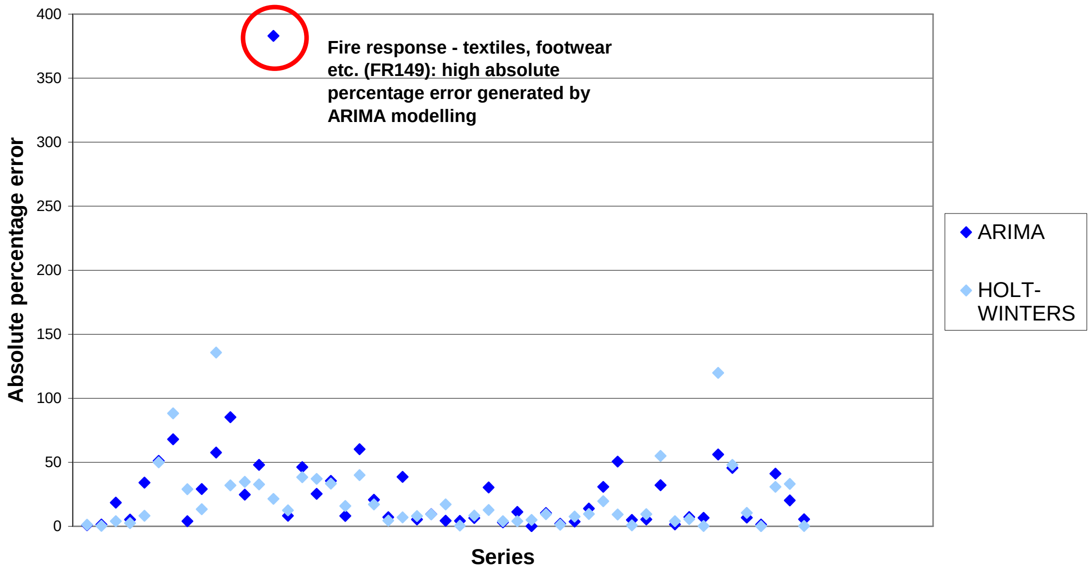
Chart 7.5.1 - Quarterly data results for each service and each method



In contrast to the annual series, the forecasts for the quarterly series for both the Fire and Rescue Services and Civil and Family Courts are better when using the Holt-Winters procedure according to the MAPE, although the difference between the two methods for the Civil and Family Courts data is marginal. This is also the case for the cost-weighted Health Care activity series. The MAPE's generated in Social Security Administration (SSA) are very similar for both methods, although slightly better when using ARIMA modelling.

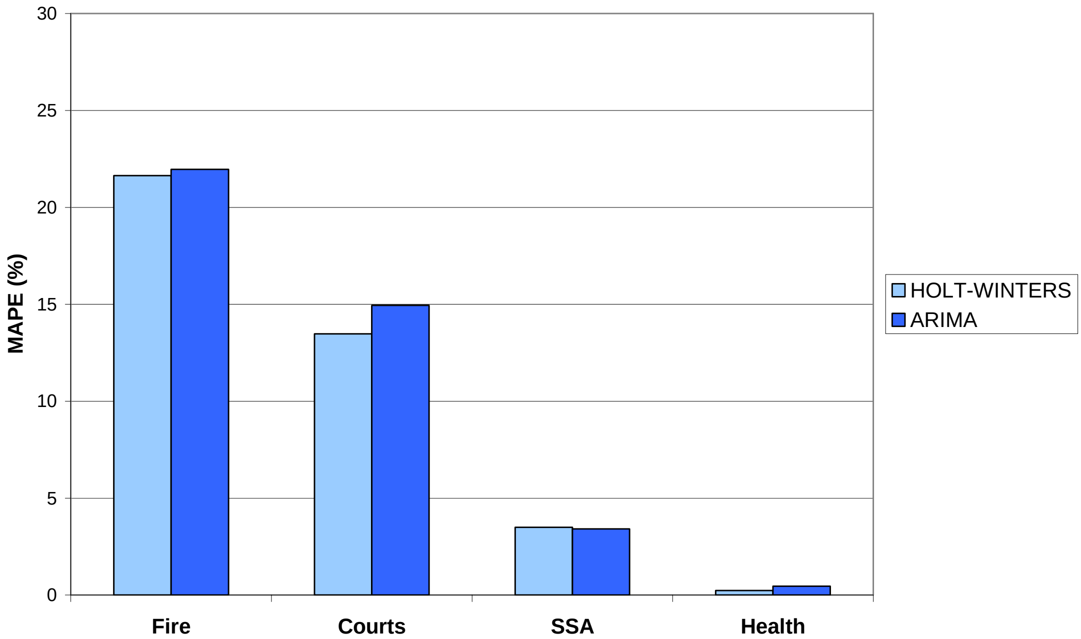
The results for the Fire and Rescue Services data show that Holt-Winters has outperformed ARIMA modelling by 7.5 percentage points. This difference is substantial when compared to the other results for the quarterly data. The high absolute forecasting error produced by ARIMA modelling is due to an extreme result for one series - 'Fire response - textiles, footwear etc'. The error produced by ARIMA for this series is 383 per cent, whilst the corresponding error produced by Holt-Winters is just over 21 per cent. This extreme result is circled in chart 7.6 below.

Chart 7.5.2 - Absolute percentage forecast errors for all quarterly series within Fire and Rescue Services



When this series is omitted from the analysis, the MAPE value for Holt-Winters does not change and remains at 21.6 per cent. The MAPE value for ARIMA modelling decreases from 29.1 per cent to 22.0 per cent, a fall of 7.1 percentage points. Although Holt-Winters is still the preferred forecasting method in terms of the MAPE after the omission, the results are now much more marginal. These adjusted results are shown graphically in chart 7.7 below.

Chart 7.5.3 - Adjusted quarterly data results for each service and each method



The series that contain an additive outlier in the forecast period will generate large absolute percentage errors for both of the forecasting methods. These series are considered below and an indication is given of what the MAPE values would be should these series be excluded from the analysis. The plotted activity data for the relevant series can be found in Annex E and this can be used to highlight the additive outliers in the out-of-sample forecast periods.

- 'Land repossession - initial hearing' within Civil and Family Courts has absolute percentage errors of 264 per cent in Holt-Winters and 343 per cent in ARIMA. Omitting this series would generate new MAPE's of 9.4 per cent in Holt-Winters and 9.6 per cent in ARIMA (compared with the previous MAPE's of 13.5 per cent and 14.9 per cent respectively).
- Within Social Security Administration, 'Social fund grants and loans' generates absolute percentage errors of 8.9 per cent and 11.2 per cent in Holt-Winters and ARIMA respectively. Omitting this series would reduce the MAPE's to 3.3 per cent for Holt-Winters and 3.1 per cent for ARIMA (compared with the previous MAPE's of 3.5 per cent and 3.4 per cent respectively).

7.6 Prediction interval analysis

The results of the prediction interval analysis are shown below for each public service.

Table 7.6 - Percentage of Holt-Winters forecasts that lie within the upper and lower bounds of the prediction intervals constructed around the ARIMA forecasts

Service	Percentage of Holt-Winters forecasts within X12-ARIMA prediction intervals
Health Care	100
Civil and Family Courts	90
Fire & Rescue Services	98
Social Security Administration	96

The results show that the majority of the forecasts produced by the Holt-Winters procedure lie within the prediction intervals constructed around the forecasts produced by ARIMA modelling. The results suggest that the forecasts produced by the two forecasting methods are similar. The percentages for the quarterly data are markedly higher than the percentages for the annual data. The reader should consider that prediction intervals were only constructed around ARIMA forecasts as such intervals were not provided as standard output of the system used to produce the Holt-Winters forecasts. The reader should also remember that the Holt-Winters and ARIMA forecasts for the Health Care data were appended to one single series.

8. Identification of Key Series

In a mass production environment such as UKCeMGA, automated processes are vital as they allow for large time savings when compared to fully manual processes. The automatic features built-in to the X12-ARIMA program will be utilised if and when UKCeMGA changes its forecasting method from Holt-Winters to ARIMA modelling. The analysis outlined in this paper has indicated the potential impact on public service activity forecasts of using such methodology.

The X12-ARIMA user may intervene in any of the automatic processes as he or she deems necessary. The automatic features will generally produce sufficient modelling and forecasting results and in many cases manual intervention is undesirable as such intervention may only produce inferior results to those produced by the fully automatic procedure. However, sometimes manual intervention in some or all parts of the modelling process can improve the ability of the model to predict the values of future observations. As UKCeMGA is a mass production environment, attempting to improve the modelling and forecasting results produced by the fully automatic X12-ARIMA procedure for every series delivered to National Accounts would not be feasible. Therefore part of the project outlined in this paper was devoted to the identification of 'key' public service activity series that produced high ARIMA forecasting errors and carry a relatively large proportion of the weight within each of their respective public services.

For a series to have been considered a 'key' series, it must have met two criteria:

1. an absolute ARIMA forecasting error greater than 10 per cent and
2. a weight value that exceeds some critical weight value for the relevant public service.

The project considered out-of-sample forecasts for each series and each method, where the forecast Yn was based on the observations Yt, t = 1,2,….,n-1. The weight of each series is the expenditure on that series as a proportion of total expenditure for the relevant public service. The expenditure on each series is simply the quantity of activity for each series in time period n-1 multiplied by each activity's associated unit cost in the same time period. The total expenditure for the relevant public service is simply an aggregation of the expenditure values for each of the sub-component series within that service in time period n-1. Therefore the weight of each series in a particular service can be expressed algebraically as:

$$w_i = \frac{c_i}{C}$$

where:

w_i = the weight of series i
c_i = the expenditure on series i
C = the total expenditure on the relevant service

Establishing a method to derive the critical weight value for each service that could be applied consistently to each service's data was not a straight-forward task as the number of sub-component series within each service and the distribution of the weight values within each service vary greatly between each of the services. It was therefore not possible to simply choose, for example, the five series in each service that had the largest expenditure weights.

The critical weight value was derived as the expected weight value multiplied by 1.5 where the expected weight value was one (or 100 per cent) divided by the number of sub-component series within the relevant service. Therefore the weight of a specific series was greater than the critical weight if:

$$w_i \geq \left(\frac{1}{s} \times 1.5 \right)$$

where *s* is the number of sub-component series within the relevant service.

After applying this somewhat arbitrary methodology to all of the series analysed in the project, the following series were identified as 'key' series.

Table 8.1 - Series identified as 'key' series for all public services

Service	Series	Absolute Error (%)	Expected Weight Value	Critical Weight Value	Actual Weight
Adult Social Care	Older people aged 65+ - residential care provided by others	**19**	0.028	0.042	**0.141**
	Adults aged 18-64 - residential care provided by others	**12**	0.028	0.042	**0.096**
Civil and Family Courts	Specified money - small claims hearing	**11**	0.013	0.019	**0.068**
	Specified money - general order	**10**	0.013	0.019	**0.087**
	Land repossession - initial hearing	**343**	0.013	0.019	**0.021**
	Land repossession - warrant	**20**	0.013	0.019	**0.020**
Fire & Rescue Services	Fire response - grassland etc	**56**	0.018	0.027	**0.065**
Social Security Administration	Social fund grants and loans	**11**	0.038	0.058	**0.082**

As detailed in section 7.5, 'Land repossession - initial hearing' within Civil and Family Courts and 'Social fund grants and loans' within Social Security Administration have large forecasting errors which are attributable to outliers in the out-of-sample forecast periods. If these series are excluded from this part of the analysis then the remaining series in table 8.1 are the 'key' series that have been identified according to the methodology detailed above. These series have a large ARIMA forecasting error and carry a relatively large proportion of the weight within each of their respective public services.

The ‘key’ series identified above should ideally be reviewed under the guidance of the Time Series Analysis Branch within the ONS. Improvements in the ARIMA forecasting errors for these series are important as these errors could potentially have the largest impact on the final quarterly estimates of public service output if and when UKCeMGA changes its forecasting method to ARIMA modelling. Improvements may be achieved through manual intervention in some or all of the automatic modelling and forecasting procedure built-in to the X12-ARIMA program. Such manual intervention could include identification of the seasonal and nonseasonal ARMA model orders, identification of the seasonal and nonseasonal differencing orders and estimation of the seasonal and nonseasonal ARMA model parameters. In addition, the user can choose to exclude regression variables automatically included in the regARIMA model or include regression variables not automatically included in the regARIMA model.

The main hazard regarding the use of ARIMA modelling for forecasting is that an automatically selected model may not necessarily be a good model. It may happen to fit the observed data well but it may also model the series badly, and this could lead to unreliable forecasts being produced. This may be the case for the ‘key’ series identified above. However, it may also be the case that the models selected for these series do model the series well and they just so happen to produce poor forecasts in the out-of-sample forecast periods used in the analysis. Certain diagnostics produced by X12-ARIMA can be used as quality measures in order to evaluate models used for forecasting. They can help the user to determine whether or not the automatically selected model models the series well and hence whether or not manual intervention in the automatic modelling and forecasting procedure is necessary. If the automatically selected model does not model the series well then manual intervention may be necessary. If the automatically selected model does model the series well then manual intervention may be detrimental and unnecessarily time consuming.

The quality measures discussed above are listed and detailed in Annex F.

9. Conclusions and Next Steps

9.1 Conclusions

This paper presented the methods and findings of a UKCeMGA project which aimed to assess the extent to which UKCeMGA's forecasting accuracy may be affected by the move from Holt-Winters to ARIMA modelling.

The results for the annual data showed that forecasting will be improved by using the ARIMA method for the Adult Social Care, Civil and Family Courts and Fire and Rescue Services data according to the MAPE, although Holt-Winters outperformed ARIMA more frequently than ARIMA outperformed Holt-Winters for the Adult Social Care data. The biggest gains could potentially be realised in Civil and Family Courts and to a lesser extent the Fire and Rescue Services. The Holt-Winters method was marginally superior for the Children’s Social Care and Education data according to the MAPE, although ARIMA outperformed Holt-Winters more frequently than Holt-Winters outperformed ARIMA when forecasting the Children’s Social Care data.

The forecasts for the quarterly data for both the Fire and Rescue Services and Civil and Family Courts were better when using the Holt-Winters procedure according to the MAPE, although the difference between the two methods for the Civil and Family Courts data was marginal. This was also the case for the cost-weighted Health Care activity series. The MAPE’s generated in Social Security Administration were very similar for both methods, although slightly lower when using ARIMA modelling. The large difference between the MAPE values in the Fire and Rescue Services results was caused by an extreme result for one series forecast using the ARIMA method. When this series was omitted from the analysis the Holt-Winters method was still preferred although the difference in performance between the two methods was much more marginal.

Several of the series analysed produced MAPE values that were markedly large for both forecasting methods. After reviewing the plotted data it was apparent that these large errors were the result of outliers in the out-of-sample forecast periods. When these series were omitted from the analysis the preferred method for each service pre-omission was the same as the preferred method for each service post-omission with the exception of Children’s Social Care. For this service, Holt-Winters was the preferred method prior to the omission whilst ARIMA modelling was the preferred method after the omission.

The prediction interval analysis showed that the majority of the forecasts produced by the Holt-Winters procedure fell within the prediction intervals constructed around the forecasts produced by ARIMA modelling for both the annual and the quarterly data. The reader should consider that prediction intervals were only constructed around the ARIMA forecasts as such intervals were not provided as standard output of the system used to produce the Holt-Winters forecasts. Had prediction intervals been constructed around the Holt-Winters forecasts, it would have been possible to calculate the percentage of series within each service where the interval attached to one method overlapped with the interval attached to the other. Such an analysis would be more thorough than the analysis presented in this paper.

Much of the analysis in this paper has used the MAPE as the primary measure of predictive performance and hence the conclusions presented in this section reflect this. However the results are less clear-cut than the analysis suggests and the conclusions may have been quite different had an alternative measure, such as the Mean Squared Error (MSE), been used as the primary tool for evaluation and comparison.

When interpreting the results the reader should keep in mind that the analysis only considered one one-step-ahead forecasting comparison for each series for reasons discussed in section 5.3. This is especially relevant when interpreting the Health Care results, which were obtained from one one-step-ahead forecasting comparison for just one series, as this is in line with current practice in terms of forecasting.

The overriding conclusion of this paper is that ultimately there is little difference in predictive performance between the Holt-Winters and ARIMA methods when producing one-step-ahead forecasts of components of quarterly estimates of public service output. Hence this paper recommends that UKCeMGA should continue to use the Holt-Winters method for forecasting data in the National Accounts. UKCeMGA should only change its forecasting method if and when such a change in method becomes mandatory.

9.2 Next steps

The findings of the project will be utilised by UKCeMGA in order to assess the extent to which UKCeMGA's forecasting errors may be affected by the move from Holt-Winters to ARIMA modelling.

The ‘key’ series identified in the project should ideally be reviewed under the guidance of the Time Series Analysis Branch within the ONS. Improvements in the ARIMA forecasting errors for these series are important as these errors could potentially have the largest impact on the final quarterly estimates of public service output if and when UKCeMGA changes its forecasting method to ARIMA modelling. Improvements may be achieved through manual intervention in some or all of the automatic modelling and forecasting procedure built-in to the X12-ARIMA program. This manual intervention can be replicated for the relevant series when the transition in method to ARIMA modelling begins.

Work on a related project - a comparison of forecasting performances achieved by direct and indirect forecasts of the Health Care data - is currently in progress.

The analysis reported on in this paper hones in on the effects of changing just the method of forecasting. It does not consider the effects of any other methodological changes associated with the shift from the winCSDB database system to CORD. Future work could potentially evaluate other such methodological changes, or even all of the methodological changes, associated with the shift in database system and assess the effects of these changes on public service output estimates.

10. Annexes

Annex A: Holt-Winters forecasting method

In general, the forecast function for the time series $y_1, \quad y_2, \quad \ldots, \quad y_n$ is denoted by $\hat{y}(n,l)$,

where the integer l represents the forecasting horizon. Assuming the presence of a multiplicative seasonal component, the Holt-Winters method takes the following form:

$$\hat{y}_{n+l|n} = (m_n + lb_n)c_{n-s+l} \quad [1]$$

where m_n is the local level of the series, b_n is the component of the trend, and c_{n-s+l} is the relevant seasonal component, with s signifying the seasonal period (4 for quarterly data and 12 for monthly data.)

Therefore if a monthly time series is considered, the one-step-ahead forecast is given by:

$$\hat{y}_{n+1|n} = (m_n + b_n)c_{n-11} \quad [2]$$

The updating formulae for the three components will each require a smoothing constant. If once again α_0 is used as the parameter for the local level, α_1 for the trend, and a third constant α_2 is added as the smoothing constant for the seasonal factor, the updating equations will be:

$$m_t = \alpha_0 \frac{y_t}{c_{t-s}} + (1-\alpha_0)(m_{t-1} + b_{t-1}) \quad [3]$$

$$b_t = \alpha_1 (m_t - m_{t-1}) + (1-\alpha_1) b_{t-1} \quad [4]$$

$$c_t = \alpha_2 \frac{y_t}{m_t} + (1-\alpha_2) c_{t-s} \quad [5]$$

where α_0, α_1, and α_2 all lie between zero and one. If the aforementioned additive version of Holt-Winters was used, the seasonal factor is simply added as opposed to multiplied into the one-step-ahead forecast function, thus:

$$\hat{y}_{n+1|n} = m_n + b_n + c_{n-11} \quad [6]$$

and the level and seasonal updating equations involve differences as opposed to ratios:

$$m_t = \alpha_0 (y_t - c_{t-s}) + (1-\alpha_0)(m_{t-1} + b_{t-1}) \quad [7]$$

$$c_t = \alpha_2 (y_t - m_t) + (1-\alpha_2) c_{t-s} \quad [8]$$

The trend component, b_t, remains unchanged.

The choice of starting values of the series components and of the smoothing parameters is of some importance. For starting values, it seems sensible to set the local level component m_0, equal to the average observation in the first year, i.e.

$$m_0 = \sum_{t=1}^{s} y_t / s \qquad [9]$$

where s is the number of seasons. The starting value for the trend component can be taken from the average difference per time period between the first and second year averages. That is:

$$b_0 = \frac{\left\{ \sum_{t=1}^{s} y_t / s \right\} - \left\{ \sum_{t=s+1}^{2s} y_t / s \right\}}{s} \qquad [10]$$

Finally, the seasonal index starting value can be calculated after allowing for a trend adjustment, as follows:

$$c_0 = \frac{\{ y_k - (k-1) b_0 / 2 \}}{m_0} \qquad \text{(multiplicative)} \qquad [11]$$

$$c_0 = y_k - \{ m_0 + (k-1) b_0 / 2 \} \qquad \text{(additive)} \qquad [12]$$

where k =1, 2,..., s. Obviously this will lead to s separate values for c_0, which is what is required to gain the initial seasonal pattern.

Annex B: ARIMA modelling and forecasting method

B1: The general ARIMA model

The main assumption surrounding the AR part of a time series Y_t is that the observed value y_t depends on some linear combination of previous observed values up to a defined maximum lag (denoted p), plus a random error term ε_t. This is the basis of the AR(p) model and this can be expressed algebraically as:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots + \varphi_p y_{t-p} + \varepsilon_t \qquad [13]$$

where the parameters φ_t are constants.

The main assumption surrounding the MA part of a time series Y_t is that the observed value y_t is a random error term plus some linear combination of previous random error terms up to a defined maximum lag (denoted q). This is the basis of the MA(q) model and this can be expressed algebraically as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_q \varepsilon_{t-q} \qquad [14]$$

where the parameters θ_t are constants.

When these models are combined this leads to a mixed ARMA(pq) model. This can be expressed algebraically as:

$$y_t = \varphi_1 y_{t-1} + \ldots\ldots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots + \theta_q \varepsilon_{t-q} \qquad [15]$$

Rearranging equation [15]:

$$y_t - \varphi_1 y_{t-1} - \ldots\ldots - \varphi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots + \theta_q \varepsilon_{t-q} \qquad [16]$$

The backshift operator B (where $By_t = y_{t-1}$, $B^2 y_t = y_{t-2}$, and so on) can be incorporated in to equation [16] such that:

$$y_t - \varphi_1 B y_t - \varphi_2 B^2 y_t - \ldots - \varphi_p B^p y_t = \varepsilon_t + \theta_1 B \varepsilon_t + \theta_2 B^2 \varepsilon_t + \ldots + \theta_q B^q \varepsilon_t \qquad [17]$$

Equation [17] can be re-written as:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \ldots\ldots - \varphi_p B^p) y_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_p B^p) \varepsilon_t \qquad [18]$$

Equation [18] can be re-written as:

$$\varphi_p(B) y_t = \theta_q(B) \varepsilon_t \qquad [19]$$

where $\varphi_p(B)$ is the nonseasonal AR operator and $\theta_q(B)$ is the nonseasonal MA operator. This is known as the general nonseasonal ARMA(pq) model.

To analyse a time series and fit an ARMA(pq) model to the observations we require that all of the observations are independently identifiable. Hence there should be no autocorrelation in the series and the series should have zero mean. In order for these requirements to be met all of the signal must have been removed from the series so that we are left with only noise. If the series has zero mean and other moments such as the variance and covariance do not depend on the passage of time, then the series is said to be stationary. In order to achieve stationarity the series must be differenced (unless it is stationary to begin with). This means taking the differences between successive observations and then analysing these differences instead of the actual observations. This can be expressed algebraically as:

$$\Delta^1 y_t = y_t - y_{t-1} \quad [20]$$

or:

$$\Delta^1 y_t = (1-B)y_t \quad [21]$$

The expressions above describe a first difference. A second difference would involve taking differences of the observations and then taking differences of the differences and analysing these second differences. This process of differencing removes the integrated, nonstationary component from the series being modelled. Incorporating this in to the ARMA(*pq*) model leads to the following expression:

$$\varphi_p(B)\Delta^d y_t = \theta_q(B)\varepsilon_t \quad [22]$$

where Δ^d is the nonseasonal difference operator and d is the order of differencing required to produce a stationary series. Equation [22] is known as the general, or pure, nonseasonal ARIMA(*pdq*) model. We can allow for the presence of seasonality in a series by applying the above concepts and techniques to same-period observations across different years. This leads to the general seasonal ARIMA(*pdq*)(*PDQ*) model, where P, D and Q refer to the orders of the seasonal AR, I and MA parts of the model respectively. This can be expressed algebraically as:

$$\varphi_p(B)\Phi_P(B^s)\Delta^d\Delta_s^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad [23]$$

where $\Phi_P(B^s)$ is the seasonal AR operator, Δ_s^D is the seasonal I operator, $\Theta_Q(B^s)$ is the seasonal MA operator and s is the seasonal period (for example 4 for quarterly data or 12 for monthly data).

B2: The regARIMA model

Suppose we have a linear regression equation such that:

$$z_t = \sum_{i=0}^{n} \beta_i \chi_{it} + y_t \quad [24]$$

where z_t is the time series being modelled, χ_{it} are the regression variables, β_i are the regression parameters which measure the size of the regression effects and y_t is the time series of regression errors due to:

$$y_t = z_t - \sum_{i=0}^{n} \beta_i \chi_{it} \quad [25]$$

The time series of regression errors can be incorporated in to the general pure ARIMA model such that:

$$\varphi_p(B)\Phi_P(B^s)\Delta^d\Delta_s^D\left(z_t - \sum_{i=0}^{n} \beta_i \chi_{it}\right) = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad [26]$$

Equation [26] implies that:

i) the regression effects are subtracted from the original, untransformed series to obtain the regression error series,
ii) the regression error series is differenced to obtain a stationary series, say w_t, and finally

iii) w_t is assumed to follow a stationary ARMA process such that:

$$\varphi_p(B)\Phi_P(B^s)w_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad [27]$$

Equation [27] describes the general regARIMA model.

B3: Forecasting

Consider the transformed series y_t. If y_n is the most recent observation then a k-step ahead forecast is y_{n+k}, denoted $y_n(k)$. The MMSE estimator is the conditional expected value so that:

$$y_n(k) = E(y_{n+k}/y_n,\ldots, y_1) \qquad [28]$$

where y_1 is an earlier observation. Equation [28] implies that the forecast is an expected value given some number of successive previous observed vales and this number will depend on the specified ARIMA model. For example, consider a simple (1 0 0) model:

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t \qquad [29]$$

When $t = n$:

$$y_n = \varphi_1 y_{n-1} + \varepsilon_n \qquad [30]$$

The one-step ahead forecast $y_n(1)$ can therefore be expressed algebraically as:

$$y_n(1) = \varphi_1 y_n + \varepsilon_{n+1} \qquad [31]$$

Given that $E(\varepsilon_{n+1}) = 0$:

$$y_n(1) = \varphi_1 y_n \qquad [32]$$

Equation [32] implies that a one-step ahead forecast for a (1 0 0) model will depend on only the most recent observation y_n and the corresponding AR parameter φ_1.

Now consider a (2 0 0) model:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t \qquad [33]$$

If the process used for the (1 0 0) model is applied to the above (2 0 0) model then:

$$y_n(1) = \varphi_1 y_n + \varphi_2 y_{n-1} \qquad [34]$$

So for a (2 0 0) model the one-step ahead forecast will depend on the previous two successive observations and their associated AR parameters.

Annex C: The automdl procedure

Step 1 - Default model estimation

The default model is the (0 1 1)(0 1 1) model which is known as the ‘airline’ model after it was successfully fitted to international airline passenger data. The three numbers in the first set of brackets indicate the orders of the non seasonal AR, I and MA parts of the ARIMA model respectively. The three numbers in the second set of brackets indicate the seasonal orders of the same model. This model is appropriate for 70 - 80 per cent of time series used in official statistics. The AR and MA coefficients are estimated for the model by Iterative Generalised Least Squares (IGLS) where an iterative least squares process is carried out until convergence is reached according to some convergence criteria. Automatic outlier and regression effect identification and replacement is carried out and hence a full regARIMA model is specified and a number of diagnostics are produced. These diagnostics will be compared later to those of the model selected by the automatic model identification procedure. If the automatically selected model shows no improvement over the default model in terms of these diagnostics then the default model shall ultimately be selected.

Step 2 - Identification of the differencing orders

The data used at this stage is not the original data: it is the data after the automatic outlier and regression effect identification and replacement has been carried out in the first stage. The data is differenced up to some order which is determined by a series of automatic tests.

Step 3 - Identification of the ARMA model orders

The program uses a variant of the BIC model selection criterion to determine the orders of the AR and MA parts of the model. The maximum order for the non seasonal parts is 3 and the maximum order for the seasonal parts is 2. The AR and MA coefficients are then estimated by IGLS. If the identified model is different to the default model then automatic outlier and regression effect identification and replacement is carried out again.

Step 4 - Comparison of the identified model with the default model

The diagnostics produced by the two models are compared and the preferred model is chosen.

Step 5 - Final model checks

Once a model is selected a set of final tests are completed to check that this model is adequate. These tests include an assessment of whether or not a constant term variable is needed in the regression model and a test for insignificant ARMA parameters in an attempt to simplify the model.

Annex D: Data availability

Table D1 - Adult Social Care data availability

Coverage	England and Scotland only
Sources	England - DH (Department of Health) Scotland - Scottish Government
No. of Series	37 (36 excluding Scotland – Community Care Assessments which is not forecast as series is too short (less than 3 full years of data))
Frequency	Annual (Financial Year)
Seasonality	N/A
Forecast Period	2005/6

Table D2 - Children's Social Care data availability

Coverage	England only
Sources	DH: PSSEX1 Return Ofsted: Performance Assessment Framework Indicators
No. of Series	6
Frequency	Annual (Financial Year)
Seasonality	N/A
Forecast Period	2005/6

Table D3 - Civil and Family Courts data availability

Coverage	England & Wales only
Sources	MoJ (Ministry of Justice): Family courts via the Family-man database, Civil courts via the Business Management System database
No. of Series	106 (although 26 are not forecast as these data are not included in National Accounts deliveries)
Frequency	Quarterly Except Insolvency Cases and Insolvency Hearings which are Annual (Calendar Year)
Seasonality	No Seasonality
Forecast Period	Insolvency Cases - 2005 Insolvency Hearings - 2006 Quarterly - 2007 Q3 Except Land Repossession - Initial Land Hearing, Adoption Application, Adoption - Further Directions and Adoption - Final Hearing which are forecast at 2007 Q1

Table D4 - Education data availability

Coverage	United Kingdom
Sources	England - DCSF (Department for Children, Schools and Families): Annual Schools Census Scotland, Wales & N.I. - Devolved Administrations
No. of Series	84 (75 excluding all Health Education Scotland components which are not forecast as series are too short (less than 3 full years of data))
Frequency	Annual (Academic Year)
Seasonality	N/A
Forecast Period	2005 Except three NPISH (Non-Profit Institutes Serving Households) components which are forecast at 2004

Table D5 - Fire and Rescue Services data availability

Coverage	Fire Response - United Kingdom Fire Prevention - England & Wales only Fire Special Services - England, Wales & N.I.
Sources	CLG (Communities and Local Government)
No. of Series	55
Frequency	Fire Response – Quarterly Fire Prevention and Fire Special Service – Annual (Financial Year)
Seasonality	Fire Response – Some components are multiplicative, the rest have no seasonality
Forecast Period	Quarterly series - 2006 Q4 Annual series - 2005/6

Table D6 - Health Care data availability

Coverage	England, limited amount from N.I.
Sources	England - DH: Hospital Episode Statistics, KH03a return, QMAE return, Prescription Pricing Authority N.I. - Devolved Administration
No. of Series	1
Frequency	Quarterly
Seasonality	Additive
Forecast Period	2006 Q4

Table D7 - Social Security Administration data availability

Coverage	UK
Sources	Majority from DWP (Department for Work and Pensions). Local Authorities and HMRC (Her Majesty's Revenue and Customs) provide the rest
No. of Series	26 (25 excluding Child Benefit Claims which is not forecast as series is too short (less than 3 full years of data))
Frequency	Quarterly
Seasonality	Most series are multiplicative although some have no seasonality
Forecast Period	2006 Q4 Except Pension Credit Claims and Pension Credit Load which are forecast at 2007 Q3

Annex E: Graphical representation of outliers in the forecast periods

Several of the series analysed had very high absolute percentage errors for both the Holt-Winters method and the ARIMA method, caused by an outlier in the forecast period. The charts below highlight these outliers.

E1: Adult Social Care - Referrals and assessments - learning disabilities (A113)

Chart E1.1 - Plotted activity data for series A113

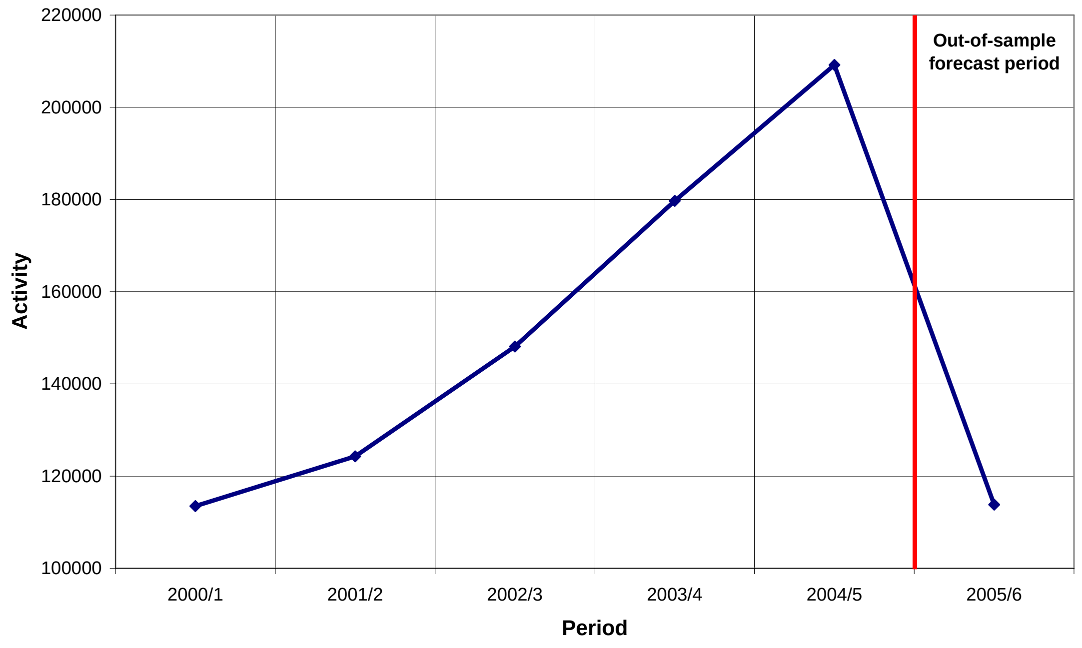
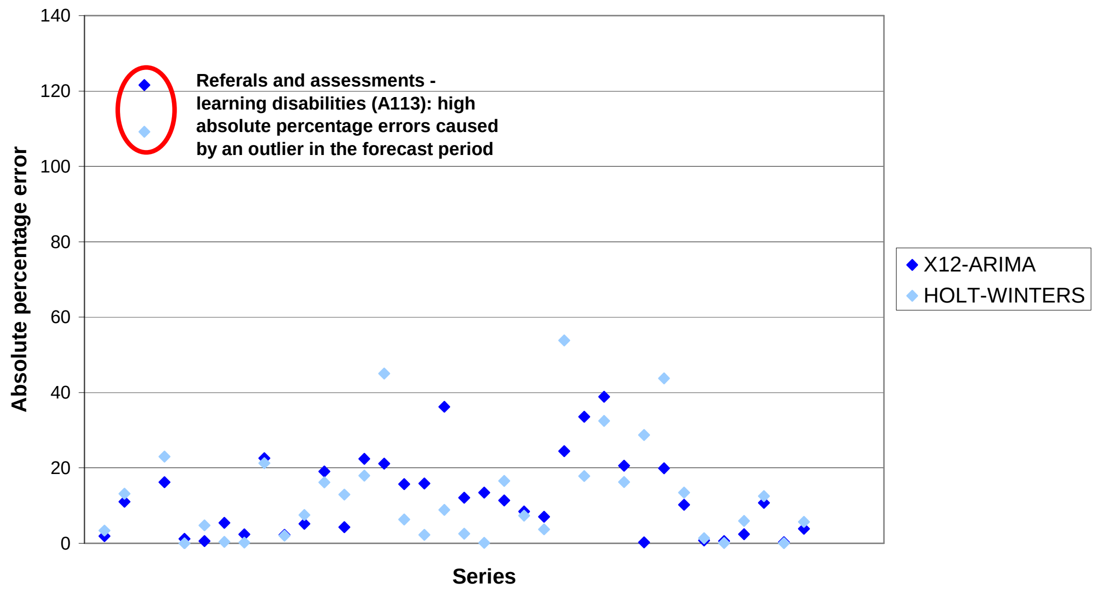


Chart E1.2 - Adult Social Care: Absolute percentage errors for all series



E2: Children's Social Care - Secure accommodation (C22)

Chart E2.1 - Plotted activity data for series C22

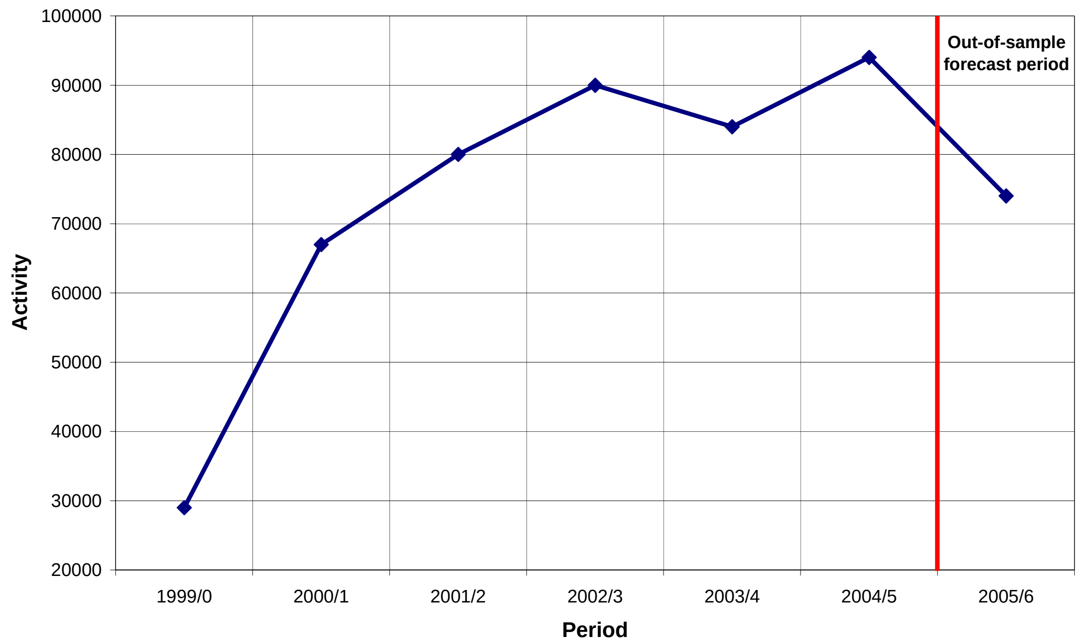
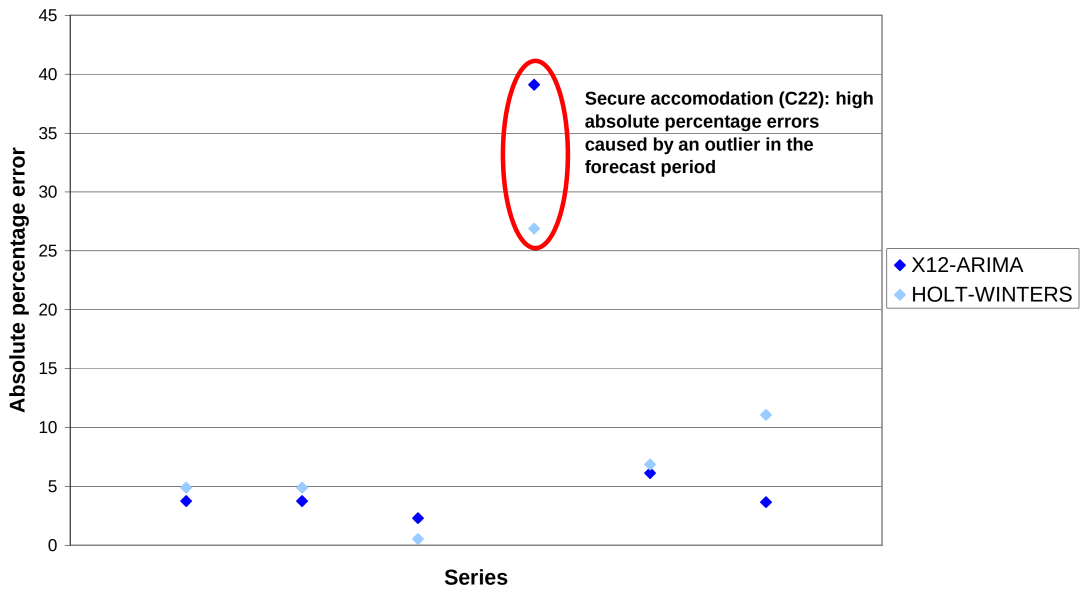


Chart E2.2 - Children's Social Care: Absolute percentage errors for all series



E3: Civil and Family Courts - Land repossession - initial hearing (V23)

Chart E3.1 - Plotted activity data for series V23

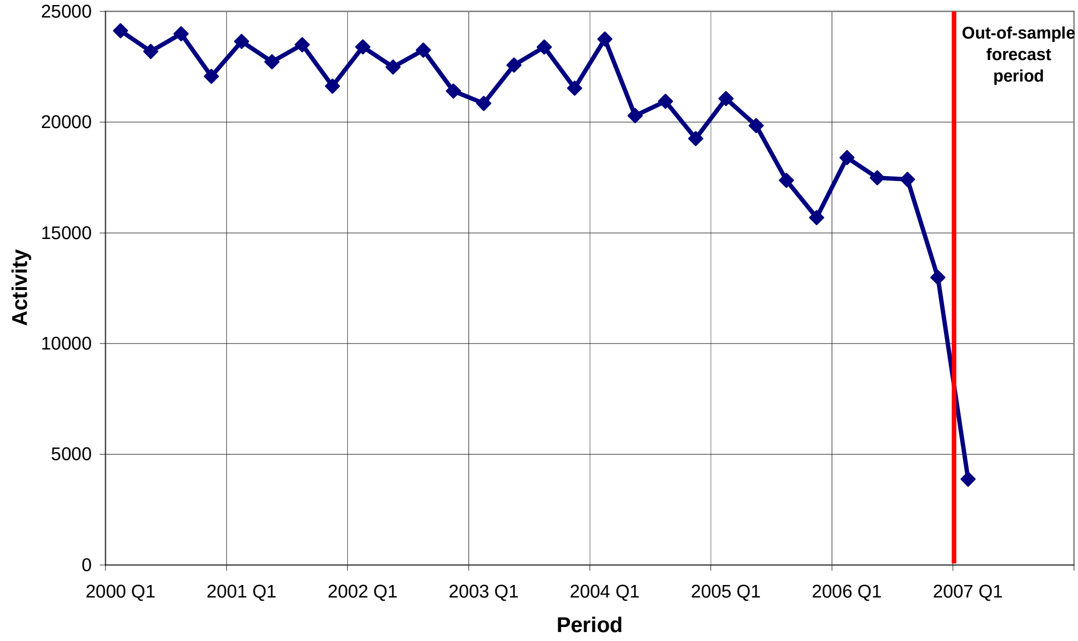
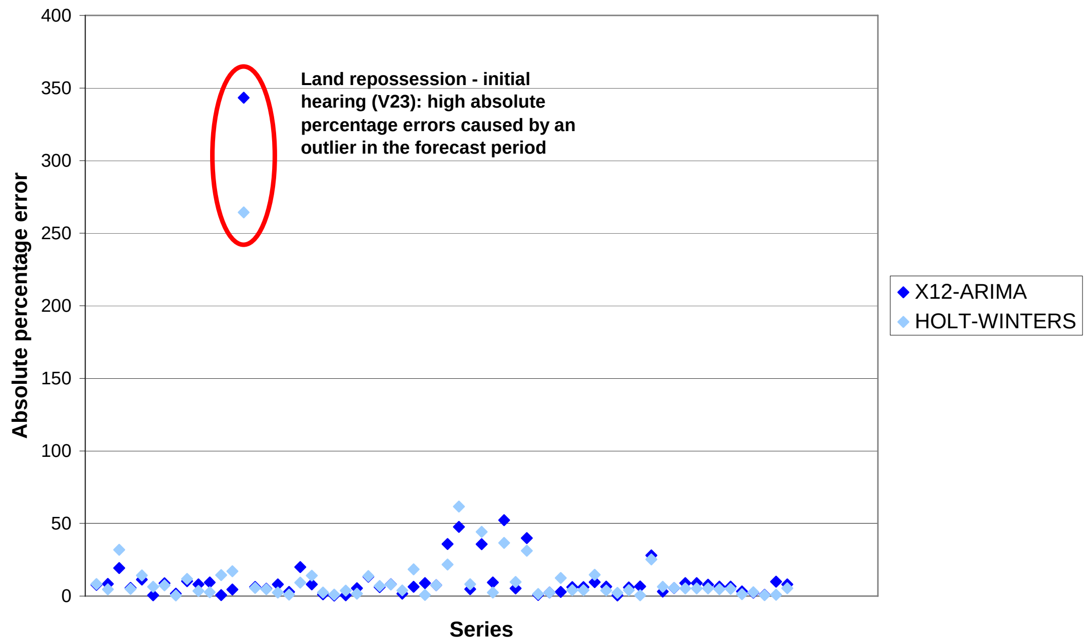


Chart E3.2 - Civil and Family Courts: Absolute percentage errors for all series



E4: Education - Welsh health education - degrees (E1362)

Chart E4.1 - Plotted activity data for series E1362

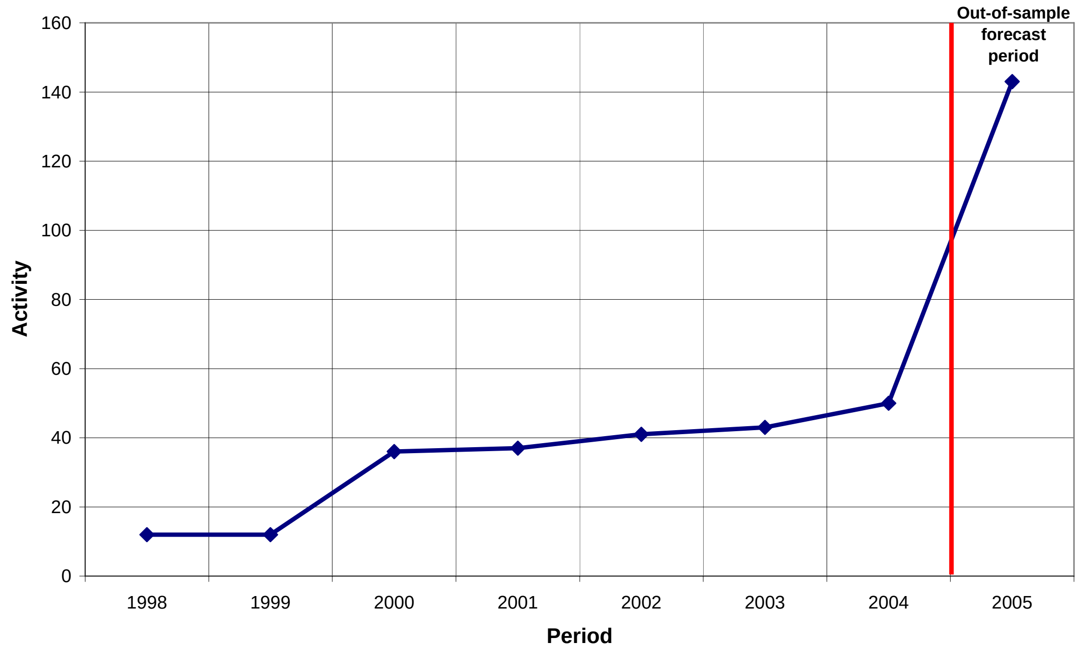
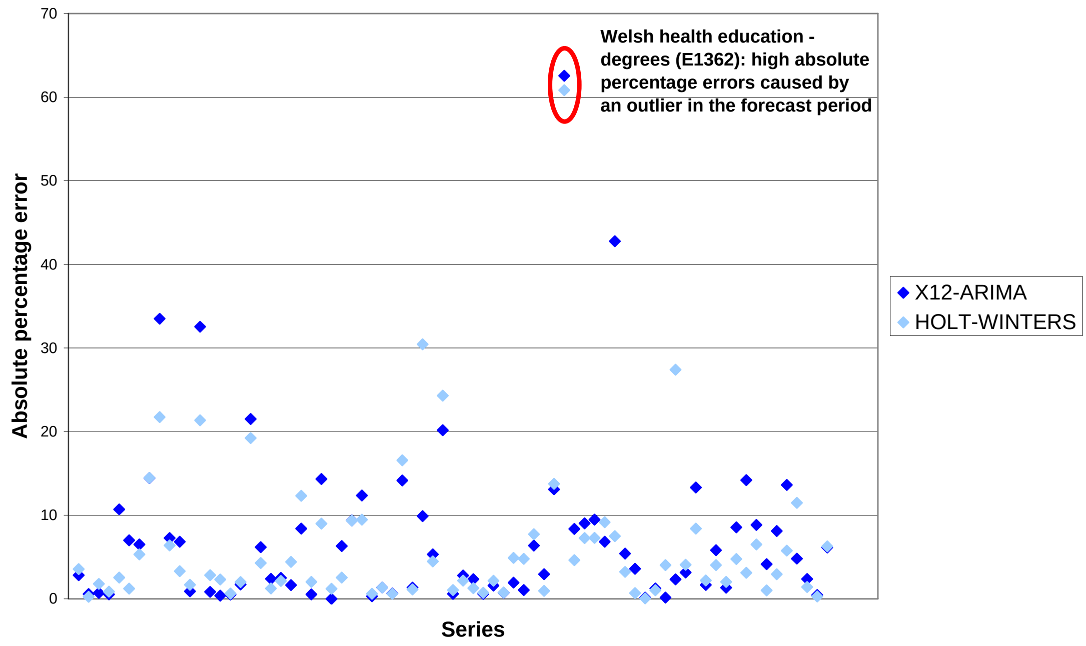


Chart E4.2 - Education: Absolute percentage errors for all series



E5: Social Security Administration - Social fund grants and loans (S5)

Chart E5.1 - Plotted activity data for series S5

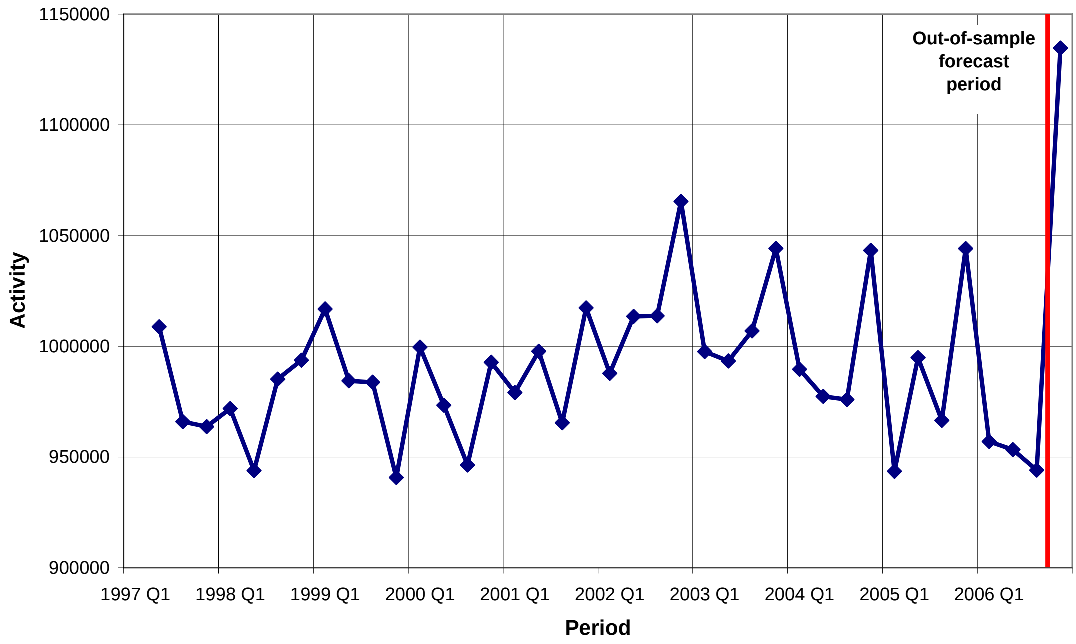
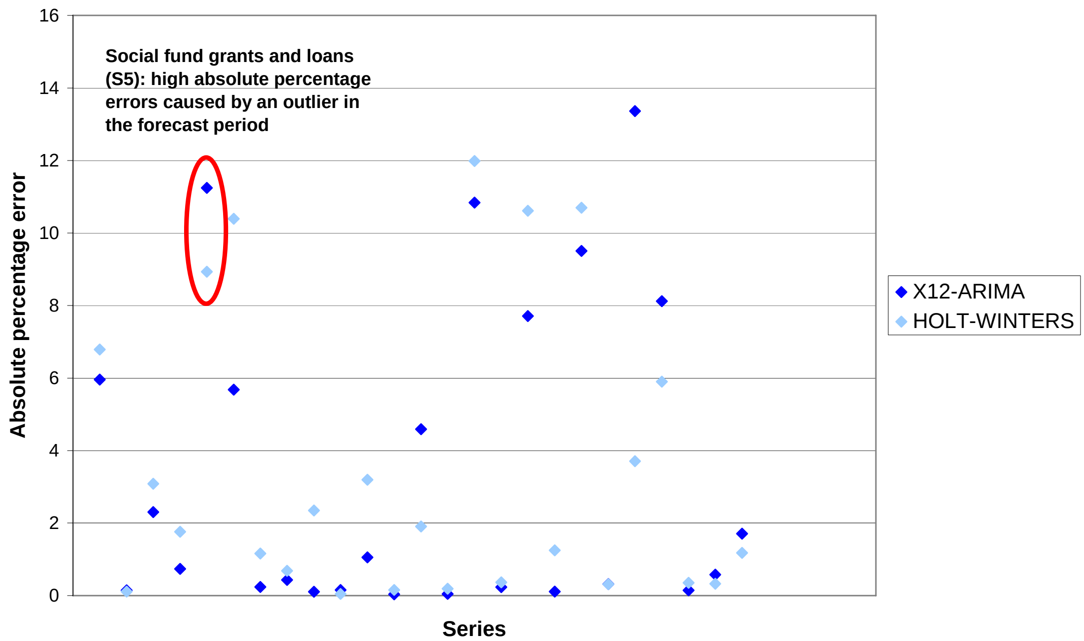


Chart E5.2 - Social Security Administration: Absolute percentage errors for all series



Annex F: Quality measures for evaluating ARIMA models used for forecasting

F1: P-values

The Ljung-Box Q-statistic is used to test simultaneously whether a set of autocorrelations are all zero. This is desirable as the presence of autocorrelation could indicate that the series being modelled is non-stationary and hence the observed residuals may not be independently identifiable (a key assumption of time series analysis). A small P-value at any lag (but especially the first 12 for quarterly data) indicates a large Q-statistic value which is a sign of potential autocorrelation. Small P-values at every fourth lag could indicate residual seasonality for quarterly data. Customarily the critical P-value is 0.05. If a total of *n* ARMA parameters are estimated in the model then no P-values will be calculated for the first *n* lags as *n* degrees of freedom will be automatically lost.

F2: ARMA parameter values

The general condition for a stationary AR process is that the root of the model lies outside the unit circle. For example, consider a (1 0 0) model. The variance can be expressed as:

$$V(y_t) = \frac{\sigma^2}{1-\varphi_1^2} \qquad [35]$$

where σ^2 is the variance of the ε_t terms, which are independently identifiable with mean zero.

The variance can only be finite if $|\varphi_1| < 1$.

If we write the (1 0 0) model in difference form we have:

$$\left(1-\varphi_1 B\right) y_t = \varepsilon_t \qquad [36]$$

When this is set equal to zero, the root of the equation must be:

$$B = \left|\varphi_1^{-1}\right| > 1 \qquad [37]$$

Therefore the root of the equation lies outside the unit circle. This is a necessary and sufficient condition of stationarity. If the absolute value of a root is greater than one then the absolute value of the corresponding AR parameter must be less than one. This diagnostic is automatically produced by X12-ARIMA.

Similarly, the absolute value of each of the roots of the MA part of the ARMA model must be greater than one. If this is the case, these roots are said to lie outside the unit circle. This is known as invertibility of the MA operators. If the absolute value of a root is greater than one then the absolute value of the corresponding MA parameter must be less than one. This diagnostic is automatically produced by X12-ARIMA.

ARMA parameter values greater than absolute 0.99 should be regarded as border cases. This could indicate non-stationarity of the AR operators or non-invertibility of the MA operators. Either of these situations could potentially produce meaningless modelling and forecasting results.

F3: Within-sample error

For each series, X12-ARIMA automatically calculates the mean absolute percentage error (MAPE) of within-sample errors for the final 3 years of each of the series. Values of 15% or greater should be regarded as unsuitable. This is an indication of a model's ability to forecast.

F4: Prediction intervals

A 95% prediction interval is automatically constructed around each point forecast produced by X12-ARIMA. The value of the forecast in relation to the upper and lower limits of the interval is an important indication of a model's ability to forecast. The width of the interval relative to the value of the forecast should also be considered.

F5: Normality test

The assumption of normality is used in the construction of the prediction intervals attached to each forecast. If this assumption is rejected then the intervals may be distorted. A significant value of either Geary's ɑ statistic or the Sample Kurtosis b_2 statistic indicates that the standardised residuals do not follow a normal distribution and hence the reported confidence intervals might not represent the actual ones. The formulae for these two statistics are given below:

$$\alpha = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \overline{X}\right|}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}} \qquad [38]$$

$$b_2 = \frac{n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)^2} \qquad [39]$$

where n is the number of residuals in the series being modelled, X_i is the value of residual i and $\overline{X}$ is the mean value of the residuals.

References

Atkinson T. (2005) *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*, available at:
http://www.statistics.gov.uk

Box G. and Jenkins G. (1970) *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Chamberlain G. (2006) 'Fitting Trends to Time Series Data', *Economic Trends*, number 633, pp64-70, available at:
http://www.statistics.gov.uk

Chatfield C. (1977) 'Some Recent Developments in Time-Series Analysis', *Journal of the Royal Statistical Society*, volume 140, number 4, pp492-510.

Chatfield C. (1978) 'The Holt-Winters Forecasting Procedure', *Applied Statistics*, volume 27, number 3, pp264-279.

Chatfield C. (1997) 'Forecasting in the 1990's', *The Statistician*, volume 46, number 4, pp461-473.

Chatfield C. and Yar M. (1988) 'Holt-Winters Forecasting: Some Practical Issues', *The Statistician*, volume 37, number 2, pp129-140.

Findley D. and Hood C. (2000) *X-12-ARIMA and its Application to Some Italian Indicator Series*, U.S. Census Bureau, available at:
http://www.census.gov

Findley D. Monsell B. Bell W. Otto M. and Chen B. C. (1998) *New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program*, U.S. Census Bureau, available at:
http://www.census.gov

Gomez V. and Maravall A. (1997) *Programs TRAMO and SEATS: Instructions for the User*, Beta Version, available at:
http://www.istat.it

Hussain F. McLaren C. H. and Stuttard N. (2007) *The Use of Seasonal Adjustment Software Within the Office for National Statistics*, paper presented at the fifth International Conference of the Association for Survey Computing, 'The Challenges of a Changing World', Southampton University, 12th-14th September 2007.

Makridakis S. Hibon M. and Moser C. (1979) 'Accuracy of Forecasting: an Empirical Investigation', *Journal of the Royal Statistical Society*, volume 142, pp97-145.

Newbold P. and Granger W. J. (1974) 'Experience with Forecasting Univariate Time Series and the Combination of Forecasts', *Journal of the Royal Statistical Society*, volume 137, number 2, pp131-165.

ONS (Draft March 2007) *Guide to Seasonal Adjustment with X-12-ARIMA*.

ONS (2008) *A Framework For Identifying Sources Of Statistical Error In Estimates Of Public Service Output And Productivity*, available at:
http://www.statistics.gov.uk

ONS (2008) *Documentation of UKCeMGA Methods used for National Accounts*, available at:
http://www.statistics.gov.uk

Skipper H. (2005) 'Early Estimates of GDP: Information Content and Forecasting Methods', *Economic Trends*, number 617, pp26-35, available at:
http://www.statistics.gov.uk

U.S. Census Bureau (2007) *X-12-ARIMA Reference Manual*, Version 0.3, available at:
http://www.census.gov

Wood J. and Elliott D. (2007) 'Methods Explained: Forecasting', *Economic and Labour Market Review*, volume 1, number 12, pp55-58, available at:
http://www.statistics.gov.uk