arfimasoc — Obtain lag-order selection statistics for ARFIMAs

**Description**

arfimasoc reports Akaike’s information criterion (AIC), Schwarz’s Bayesian information criterion (BIC), and the Hannan and Quinn information criterion (HQIC) for a series of autoregressive fractionally integrated moving-average (ARFIMA) models. These criteria are used to select the number of autoregressive (AR) and moving-average (MA) lags to be used in the ARFIMA model.

**Quick start**

Compute AIC, BIC, and HQIC for ARFIMA models of y with up to 2 AR and 2 MA lags
arfimasoc y

Same as above, but compare ARFIMA models of y with up to 7 AR lags and 1 MA lag
arfimasoc y, maxar(7) maxma(1)

Limit the number of iterations in arfima estimation to 50
arfimasoc y, arfimaopts(iterate(50))

Compute AIC, BIC, and HQIC for ARFIMA models of y with exogenous regressor x and with up to 2 AR lags and 2 MA lags
arfimasoc y x

Compute the information criteria for ARFIMA models of the first difference of y
arfimasoc d.y

**Menu**

Statistics > Time series > ARFIMA > Lag-order selection
Syntax

arfimasoc depvar [ indepvars ] [ if ] [ in ] [ , options ]

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxar(#)</td>
<td>set maximum AR order to #; default is maxar(2)</td>
</tr>
<tr>
<td>maxma(#)</td>
<td>set maximum MA order to #; default is maxma(2)</td>
</tr>
<tr>
<td>n(#)</td>
<td>use $N$ when calculating BIC and HQIC</td>
</tr>
<tr>
<td>arfimaopts( opts )</td>
<td>specify options of arfima for model estimation</td>
</tr>
</tbody>
</table>

You must tsset your data before using arfimasoc; see [TS] tsset.
depvar and indepvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.

Options

**Main**

maxar(#) specifies the maximum AR lag order for which the information criteria are to be calculated. The default is maxar(2).

maxma(#) specifies the maximum MA lag order for which the information criteria are to be calculated. The default is maxma(2).

n(#) sets $N$ to be used when calculating BIC and HQIC; see [R] IC note.

arfimaopts( opts ) specifies options of arfima to include in the ARFIMA fit by arfimasoc. opts may be noconstant, smemory, difficult, technique(), iterate(), tolerance(), ltolerance(), nrtolerance(), gtolerance(), nonrtolerance, and collinear. See [TS] arfima for a description of these options.

Remarks and examples

Many statistics have been developed to assist researchers in fitting an ARFIMA model of the correct order. The arfimasoc command computes three information criteria (AIC, BIC, and HQIC) that help researchers determine the best number of AR and MA lags to be included in an ARFIMA model. arfimasoc calculates these criteria for the ARFIMA models with up to $p$ AR lags and $q$ MA lags, where $p$ and $q$ are predetermined numbers. arfimasoc keeps the sample and option specifications the same in the estimation of all the different ARFIMA models.

In general, the value of the information criterion decreases with the model’s goodness of fit, as assessed by the likelihood function, and increases with the number of parameters. Therefore, the selected model is the one that minimizes the information criterion, or equivalently, the model that best fits the data while using the least number of parameters possible. However, different information criteria may choose different models.

Among the three different information criteria available, BIC and HQIC have the advantage that they are consistent. This means that as the sample size grows, they select the correct number of lags with probability approaching one. However, there is a positive probability that AIC will select more lags than necessary, even with an infinite sample size; see Brockwell and Davis (2016, 149–151).
Example 1: Basic example

Economists often debate whether inflation behaves like a long-memory or short-memory process. To investigate this question, we use arfimasoc on U.S. macro data to fit several ARFIMA models of the inflation rate.

use https://www.stata-press.com/data/r18/usmacro (Federal Reserve Economic Data - St. Louis Fed)
arfimasoc inflation
Fitting models (9): ........... done
Lag-order selection criteria
Sample: 1955q3 thru 2010q4 Number of obs = 222

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARFIMA(0,0)</td>
<td>-334.9385</td>
<td>3</td>
<td>675.877</td>
<td>686.0851</td>
<td>679.9984</td>
</tr>
<tr>
<td>ARFIMA(0,1)</td>
<td>-236.0778</td>
<td>4</td>
<td>480.1557</td>
<td>493.7664</td>
<td>485.6508</td>
</tr>
<tr>
<td>ARFIMA(0,2)</td>
<td>-231.3761</td>
<td>5</td>
<td>472.7522</td>
<td>489.7656</td>
<td>479.6212</td>
</tr>
<tr>
<td>ARFIMA(1,0)</td>
<td>-226.7425</td>
<td>4</td>
<td>461.485</td>
<td>475.0957</td>
<td>466.9801</td>
</tr>
<tr>
<td>ARFIMA(1,1)</td>
<td>-225.7163</td>
<td>5</td>
<td>461.4327</td>
<td>478.4461</td>
<td>468.3016</td>
</tr>
<tr>
<td>ARFIMA(1,2)</td>
<td>-210.8595</td>
<td>6</td>
<td>433.7189</td>
<td>454.135</td>
<td>441.9617</td>
</tr>
<tr>
<td>ARFIMA(2,0)</td>
<td>-225.6939</td>
<td>5</td>
<td>461.3878</td>
<td>478.4012</td>
<td>468.2568</td>
</tr>
<tr>
<td>ARFIMA(2,1)</td>
<td>-202.8705</td>
<td>6</td>
<td>417.7411</td>
<td>438.1571</td>
<td>425.9838</td>
</tr>
<tr>
<td>ARFIMA(2,2)</td>
<td>-217.4265</td>
<td>7</td>
<td>448.8531</td>
<td>472.6718</td>
<td>458.4696</td>
</tr>
</tbody>
</table>

Selected (max) LL: ARFIMA(2,1)
Selected (min) AIC: ARFIMA(2,1)
Selected (min) BIC: ARFIMA(2,1)
Selected (min) HQIC: ARFIMA(2,1)

The default maximum AR lag $p$ and MA lag $q$ are both 2. The table provides results for each AR and MA combination, beginning with a constant-only model ARFIMA(0,0). The column LL reports the log likelihood, and the column df reports the number of estimated parameters. In this example, the log likelihood is maximized with the ARFIMA(2,1) model. All information criteria select the ARFIMA(2,1) model as well. Although they agree here, model selection criteria can disagree because they put different penalties on the complexity of the model, as measured by the number of parameters estimated.

If we were now to fit the selected ARFIMA(2,1) model, we would find that the confidence interval for $d$ is $[-0.05, -0.50]$. This provides evidence that inflation seems to follow a long-memory process.

arfima inflation, ar(1/2) ma(1)
Example 2: Adding exogenous variables

We use `arfimasoc` to fit several models of the federal funds rate, allowing for inflation and the output gap as covariates in estimation.

```
.arfimasoc fedfunds inflation ogap
Fitting models (9): .......... done
```

Lag-order selection criteria

Sample: 1955q3 thru 2010q4 Number of obs = 222

<table>
<thead>
<tr>
<th>Model</th>
<th>LL</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARFIMA(0,0)</td>
<td>-319.1342</td>
<td>5</td>
<td>648.2684</td>
<td>665.2818</td>
<td>655.1374</td>
</tr>
<tr>
<td>ARFIMA(0,1)</td>
<td>-279.6027</td>
<td>6</td>
<td>571.2054</td>
<td>591.6215</td>
<td>579.4481</td>
</tr>
<tr>
<td>ARFIMA(0,2)</td>
<td>-276.9722</td>
<td>7</td>
<td>567.9444</td>
<td>591.7632</td>
<td>577.5609</td>
</tr>
<tr>
<td>ARFIMA(1,0)</td>
<td>-274.6944</td>
<td>6</td>
<td>561.3889</td>
<td>581.8049</td>
<td>569.6316</td>
</tr>
<tr>
<td>ARFIMA(1,1)</td>
<td>-274.9036</td>
<td>7</td>
<td>563.8072</td>
<td>587.626</td>
<td>573.4238</td>
</tr>
<tr>
<td>ARFIMA(1,2)</td>
<td>-268.6171</td>
<td>8</td>
<td>553.2342</td>
<td>580.4556</td>
<td>564.2245</td>
</tr>
<tr>
<td>ARFIMA(2,0)</td>
<td>-272.6333</td>
<td>7</td>
<td>559.2666</td>
<td>583.0854</td>
<td>568.8832</td>
</tr>
<tr>
<td>ARFIMA(2,1)</td>
<td>-268.9673</td>
<td>8</td>
<td>553.9347</td>
<td>581.1561</td>
<td>564.925</td>
</tr>
<tr>
<td>ARFIMA(2,2)</td>
<td>-268.6169</td>
<td>9</td>
<td>555.2339</td>
<td>585.858</td>
<td>567.598</td>
</tr>
</tbody>
</table>

Selected (max) LL: ARFIMA(2,2)
Selected (min) AIC: ARFIMA(1,2)
Selected (min) BIC: ARFIMA(1,2)
Selected (min) HQIC: ARFIMA(1,2)

Here `fedfunds` is the dependent variable, `inflation` and `ogap` are independent variables included in every estimation, and `arfimasoc` is searching across all AR and MA combinations up to a maximum of two lags each.

All three information criteria select a model with one AR lag and two MA lags. Because these selected models contain the maximum number of MA terms, it may be worthwhile to search over even more MA terms to ensure that a true minimum of the information criteria has been reached.

Stored results

`arfimasoc` stores the following in `r()`:

Scalars
- `r(N)` number of observations
- `r(ar_max)` maximum AR lag order
- `r(ma_max)` maximum MA lag order

Macros
- `r(depvar)` name of endogenous variable
- `r(covariates)` names of exogenous variables
- `r(aic_sel)` selected ARFIMA model by AIC
- `r(bic_sel)` selected ARFIMA model by BIC
- `r(hbic_sel)` selected ARFIMA model by HQIC
- `r(ll_sel)` selected ARFIMA model by LL
- `r(aic_cmd)` selected ARFIMA command by AIC
- `r(bic_cmd)` selected ARFIMA command by BIC
- `r(hbic_cmd)` selected ARFIMA command by HQIC
- `r(ll_cmd)` selected ARFIMA command by LL

Matrices
- `r(table)` table of results
- `r(converged)` 1 if converged, 0 otherwise
Methods and formulas

Akaike’s (1974) information criterion is defined as

$$AIC = -2 \ln L + 2k$$

where $\ln L$ is the maximized log likelihood of the model and $k$ is the number of parameters estimated. Some authors define AIC as the expression above divided by the sample size.

Schwarz’s (1978) Bayesian information criterion is another measure of fit. It is defined as

$$BIC = -2 \ln L + k \ln N$$

where $N$ is the sample size. See [R] IC note for additional information on calculating and interpreting BIC.

The Hannan and Quinn (1979) information criterion is another measure of fit. It is defined as

$$HQIC = -2 \ln L + 2k \ln \ln N$$

References


Also see

[TS] **arfima** — Autoregressive fractionally integrated moving-average models

[TS] **arimasoc** — Obtain lag-order selection statistics for ARMAs

[TS] **varsoc** — Obtain lag-order selection statistics for VARs and VECMs