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*]

options	Description
Main	
`maxar(#)`	set maximum AR order to #; default is `maxar(2)`
`maxma(#)`	set maximum MA order to #; default is `maxma(2)`
`n(#)`	use N when calculating BIC and HQIC
`arfimaopts(`*opts*`)`	specify options of `arfima` for model estimation

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 ARFIMAs 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 US macro data to fit several ARFIMA models of the inflation rate.

```
. use https://www.stata-press.com/data/r19/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
```

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

```
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
```

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

```
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(hqic_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(hqic_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

$$\text{AIC} = -2\ln L + 2k$$

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

$$\text{BIC} = -2\ln L + k\ln N$$

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

$$\text{HQIC} = -2\ln L + 2k\ln\ln N$$

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Brockwell, P. J., and R. A. Davis. 2016. *Introduction to Time Series and Forecasting*. 3rd ed. Cham, Switzerland: Springer.

Hannan, E. J., and B. G. Quinn. 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, B ser., 41: 190–195. https://doi.org/10.1111/j.2517-6161.1979.tb01072.x.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464. https://doi.org/10.1214/aos/1176344136.

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 VAR and VEC models