ARIMA Analysis in JMulTi

December 21, 2006

Markus Krätzig & Helmut Lütkepohl

Univariate ARIMA models can be specified, estimated, checked and used for forecasting in JMulTi . The relevant features will be described in the following.

1 The Basic Model

The basic ARIMA (autoregressive integrated moving average) model allowed for in JMulTi has the form

$$\Delta^d y_t = \alpha_1 \Delta^d y_{t-1} + \cdots + \alpha_p \Delta^d y_{t-p} + u_t + m_1 u_{t-1} + \cdots + m_q u_{t-q} + C D_t, \tag{1}$$

where y_t is a univariate time series variable and Δ is the differencing operator defined such that $\Delta y_t = y_t - y_{t-1}$. Thus, for example for $d = 2$, $\Delta^d y_t = \Delta \Delta y_t = y_t - 2y_{t-1} + y_{t-2}$. The vector D_t contains all deterministic terms which may consist of a constant, a linear trend, seasonal dummy variables as well as user specified other dummy variables, and u_t is an unobservable zero mean white noise process with variance σ_u^2. The α_i, m_j and the vector C are parameters.

The model (1) is usually referred to as an ARIMA(p, d, q) model. For setting up such a model the autoregressive order p, the moving average order q and the order of differencing d have to be specified. The order of differencing can be determined with the help of unit root tests in the **Initial Analysis** panel of JMulTi . The orders p and q are sometimes specified on the basis of the autocorrelations and partial autocorrelations of the time series y_t or $\Delta^d y_t$ (see (Lütkepohl, 2004)). These quantities are also available in the **Initial Analysis** panel. In JMulTi also an automatic specification possibility based on model selection criteria is offered. It is described in the specification section.

When the orders p, d, q are specified and the deterministic terms are chosen, the model can be estimated by moving to the **Estimation** panel. When the model has been estimated the **Model Checking** and **Forecasting** panels will be activated and can be used.

2 Specification of ARMA Orders

Figure 1: ARIMA Model Selection

JMulTi offers the following automatic possibility to specify the autoregressive and moving average orders of an ARIMA model. This procedure is sometimes known as **Hannan-Rissanen procedure**. It is assumed that the order of differencing, d, and the deterministic terms to include, CD, have been prespecified. Standard deterministic variables should be selected from the checkboxes, user specified deterministics can be included by denoting variables as *deterministic* and selecting them together with y in the selection component.

For simplicity the differenced variable will be denoted by y_t, that is, y_t stands for $\Delta^d y_t$ if $d > 0$. In the first stage an $AR(h)$ model with large order h is fitted by OLS to obtain residuals $\hat{u}_t(h)$. Then models of the form

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_n y_{t-n} + u_t + m_1 \hat{u}_{t-1}(h) + \cdots + m_l \hat{u}_{t-l}(h) + CD_t \qquad (2)$$

are fitted for all combinations (n, l) for which $n, l \leq p_{\max} < h$. The combination of orders minimizing

$$AIC(n, l) = \log \widetilde{\sigma}_u^2(n, l) + \frac{2}{T}(n + l),$$

3

$$HQ(n,l) = \log \widetilde{\sigma}_u^2(n,l) + \frac{2 \log \log T}{T}(n+l)$$

and

$$SC(n,l) = \log \widetilde{\sigma}_u^2(n,l) + \frac{\log T}{T}(n+l)$$

are determined and shown to the JMulTi user who can then make a choice on the basis of these recommendations. Here $\widetilde{\sigma}_u^2(n,l) = T^{-1} \sum_{t=1}^{T} \hat{u}_t(n,l)^2$, where $\hat{u}_t(n,l)$ is the residual from fitting (2) by OLS.

Here the choice of h and $p_{\max}$ may affect the estimated ARMA orders. Hannan and Rissanen (1982) suggest using h to increase slightly faster than $\log T$. The AR order h needs to be greater than $p_{\max}$ which in turn may depend on the data of interest.

3 Estimation

```
*** Thu, 21 Dec 2006 11:46:10 ***
sample range:       [1960 Q2, 1982 Q4], T = 91
Model:  ARIMA(3,1,0)


Final Results:

Iterations Until Convergence:   4

Log Likelihood:    -358.437976        Number of Residuals: 91
AIC         :     724.875952        Error Variance    : 159.894074987
SBC         :     734.919390        Standard Error    : 12.644922894

DF: 87      Adj. SSE: 14055.093721423     SSE: 13910.784523894

Dependent Variable:       cons

              Coefficients     Std. Errors      T-Ratio    Approx. Prob.
AR1          -0.01956554       0.09224872      -0.21210         0.83253
AR2           0.19444386       0.09003795       2.15958         0.03356
AR3           0.48437419       0.09358531       5.17575         0.00000
CONST        19.49700491       3.66367539       5.32171         0.00000

Total Computation Time: 0.00 (seconds)

AR Roots and Moduli:

Real :   -0.78176  -0.78176   1.16209
Imag.:    1.07954  -1.07954   0.00000
Mod. :    1.33288   1.33288   1.16209
```

Figure 2: ARIMA Estimation Results

Estimation of ARIMA models is done by Gaussian maximum likelihood (ML) assuming normal errors. The optimization of the likelihood function requires in general nonlinear optimization algorithms. In JMulTi the algorithm by Ansely (1979) is used. The maximization routine forces the AR coefficients to be invertible. The MA roots will have modulus 1 or greater. If an MA root is 1, the estimation routine will report a missing value for the MA coefficient's standard deviation, t-statistic and p-value. An MA root equal to 1 suggests that d may have been chosen too large. Starting values and convergence criteria are chosen automatically.

The estimation output shows the number of iterations needed for convergence. It also shows some other statistics and the parameter estimates with standard errors, t-statistics and tail probabilities. The latter quantities cannot be computed in the usual way if one of the moving average roots is on the unit circle in which case the ARMA process is not invertible and the usual asymptotic theory does not apply. Such an outcome of the estimation procedure can also be a result of an overspecified model. This should be checked carefully.

Also the AR roots, that is, the roots of the estimated polynomial

$$\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$$

5

and the MA roots, that is, the roots of the polynomial

$$m(z) = 1 - m_1 z - \cdots - m_q z^q$$

are shown with their moduli. It should be noted that SBC stands for the Schwarz criterion here.

4   Model Checking
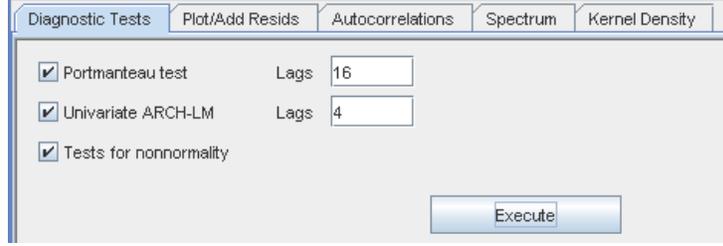
Figure 3: ARIMA Residual Analysis

JMulTi offers a range of tools for checking the adequacy of an ARIMA model. Tests for residual autocorrelation and nonnormality are offered under **Residual Analysis → Diagnostic Tests** and for plots of residuals, autocorrelations and desity estimates see the **Initial Analysis** help.

4.1   Tests for Residual Autocorrelation

The portmanteau test checks the pair of hypotheses

$$H_0 : \rho_{u,1} = \cdots = \rho_{u,h} = 0 \text{ versus } H_1 : \rho_{u,i} \neq 0 \text{ for at least one } i = 1, \ldots, h,$$

where $\rho_{u,i} = \mathsf{Corr}(u_t, u_{t-i})$ denotes the autocorrelation coefficients of the residual series. If the $\hat{u}_t$ are residuals from an estimated $\text{ARMA}(p,q)$ model, the portmanteau test statistic is

$$Q_h = T \sum_{j=1}^{h} \hat{\rho}_{u,j}^2,$$

where $\hat{\rho}_{u,j} = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t^s \hat{u}_{t-j}^s$ and $\hat{u}_t^s = \hat{u}_t / \hat{\sigma}_u$ are the standardized estimation residuals. The test statistic has an approximate $\chi^2(h - p - q)$-distribution if the null hypothesis holds. An adjusted version with potentially better small sample properties was proposed by Ljung and Box (1978). In JMulTi the following verison is available:

$$Q_h^* = T^2 \sum_{j=1}^{h} \frac{1}{T - j} \hat{\rho}_{u,j}^2 \approx \chi^2(h - p - q).$$

4.2   Lomnicki-Jarque-Bera Test for Nonnormality

This test for nonnormality based on the third and fourth moments or, in other words, on the skewness and kurtosis of a distribution. Denoting by u_t^s the standardized true model residuals, i.e., $u_t^s = u_t / \sigma_u$, the test checks the pair of hypotheses

$$H_0 : E(u_t^s)^3 = 0 \text{ and } E(u_t^s)^4 = 3 \quad \text{versus} \quad H_1 : E(u_t^s)^3 \neq 0 \text{ or } E(u_t^s)^4 \neq 3,$$

7

that is, it checks whether the third and fourth moments of the standardized residuals are consistent with a standard normal distribution. Denoting the standardized estimation residuals by $\hat{u}_t^s$, the test statistic is

$$LJB = \frac{T}{6}\left[T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^3\right]^2 + \frac{T}{24}\left[T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^4 - 3\right]^2,$$

where $T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^3$ is a measure for the skewness of the distribution and $T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^4$ measures the kurtosis. The test statistic has an asymptotic $\chi^2(2)$-distribution if the null hypothesis is correct and the null hypothesis is rejected if LJB is large.

ARCH-LM Test

In JMulTi the test for neglected conditional heteroskedasticity (ARCH) is based on fitting an ARCH(q) model to the estimation residuals,

$$\hat{u}_t^2 = \beta_0 + \beta_1\hat{u}_{t-1}^2 + \cdots + \beta_q\hat{u}_{t-q}^2 + error_t, \tag{3}$$

and checking the null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_q = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0 \text{ or } \ldots \text{ or } \beta_q \neq 0.$$

Under normality assumptions the LM test statistic is obtained from the coefficient of determination, R^2, of the regression (3):

$$ARCH_{LM}(q) = TR^2.$$

It has an asymptotic $\chi^2(q)$ distribution if the null hypothesis of no conditional heteroskedasticity holds (Engle (1982)).

4.3   ARCH Analysis of Residuals

The ARCH analysis of the residuals is accessible by choosing **Model Checking → ARCH Analysis of Residuals**. It offers the same features as the ARCH Analysis accessible in the main menu. Help details can be found in the **Univariate ARCH and GARCH models** chapter.
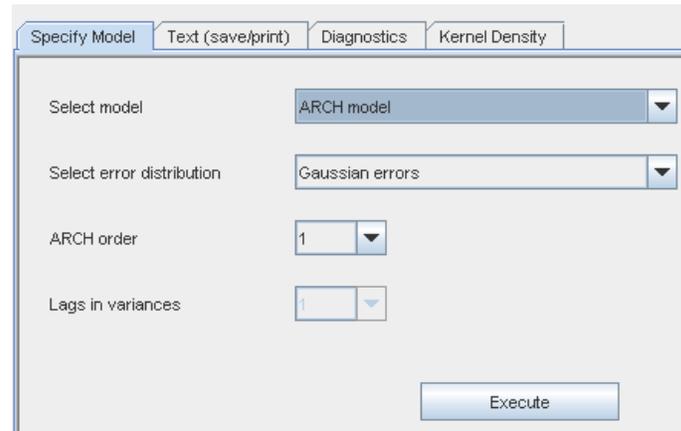
Figure 4: ARCH Analysis of Residuals

5 Forecasting
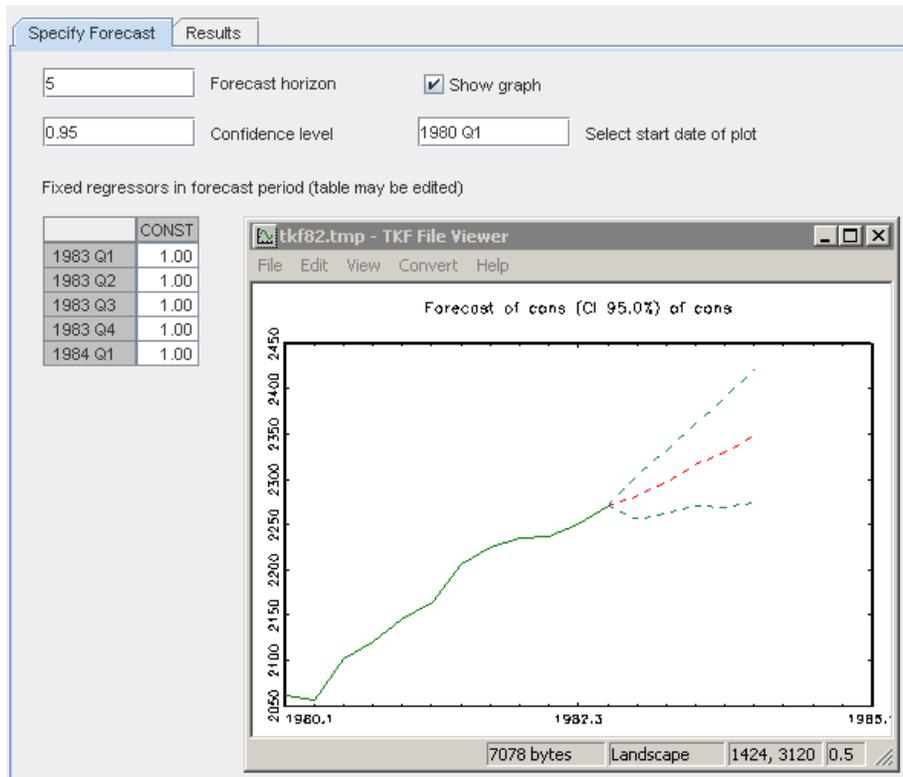
Figure 5: ARIMA Forecasting

ARIMA forecasts in JMulTi are based on Granger and Newbold (1986). The procedure forecasts the levels of y by using the estimated AR and MA coefficients in a recursive procedure. Confidence intervals are based on the assumption of normal errors.

Users should select the forecast horizon and the confidence level. For convenience, the start date of the forecast plot may be adjusted. By default, the underlying level's series y is plotted for all periods where it is available. If there are sample values available also during the forecast period, then the forecasts may be compared with the actual values.

It should be noted that the forecasting tool automatically extrapolates deterministic series to the forecasting period. If user specified deterministic variables have been specified, the routine first checks, whether sample values are available. If not, the last value is carried over for all periods. All values to be used for the forecast may be changed by manually editing the table where the determinics are shown.

References

Ansely, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process, *Biometrika* **66**: 59–65.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdoms inflations, *Econometrica* **50**: 987–1007.

Granger, C. and Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd edn, San Diego: Academic Press.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed atoregressive-moving average order, *Biometrika* **69**: 81–94.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time-series models, *Biometrika* **65**: 297–303.

Lütkepohl, H. (2004). Univariate time series analysis, *in* H. Lütkepohl and M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 8–85.