COMPARING THE PRECISION OF ARMA MODEL ESTIMATION METHODS

JAN KODERA, QUANG VAN TRAN
University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mail: kodera@vse.cz, tran@vse.cz

Abstract

Autoregressive moving average models are well-known and widely used to capture the persistence in economic and financial time series. Due to the lack of the residual series at hand, estimation of the model parameters is often performed by several ways. The most convenient way to estimate them is to use a multistage least squares method. Then the alternatives can be the maximum likelihood method or the Kalman filter approach where unknown history is treated as nonexistent. Estimation results of these methods are often taken without any doubt. In this research we point out that estimation results of ARMA model parameters are quite far from the true ones and they should be treated with maximum precaution. We also compare precision of estimation results and make some recommendations regarding their usability.

Key words: *ARMA model estimation, least squares method, maximum likelihood method, Kalman filter, simulation.*

1. Introduction

Autoregressive moving average models are powerful and often used tool to model the persistence in economic and financial time series. As there is no residual series available at hand, the estimation of the model is far more complicated than what one has to deal with a linear model. There are three methods usable for this task: the multistage least squares method, the maximum likelihood method, and the Kalman filter approach. Although these methods have been known for some time, so far we have not seen a comprehensive assessment of their ability to solve the estimation problem of any ARMA model. To fill this gap in the literature, in this research we will evaluate the capacity of these methods for this task. To reach this purpose, first we choose a monthly series of Prague stock market index PX from January 1997 to March 2016. For this series the most suitable specification will be identified and the above mentioned methods will be used to estimate the parameters of this model. We compare the estimation results obtained by using these methods to evaluate their ability to estimate parameters of a chosen model when the true values of its coefficients are unknown. Then we generate a dataset with known parameters of the same model and use these methods to extract the coefficients from simulated data. By doing so, we can provide a comprehensive assessment how precise each method can be. The results therefore may provide some useful information to how to use and interpret the estimation results of an ARMA model from real data.

2. The ARMA Process

The autoregressive and moving average process of order p,q, ARMA(p,q), is a stationary process defined as

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t, \tag{1}$$

where $e_t \sim N(0,\sigma^2)$, ϕ and θ are real parameters. It can be also represented as

$$b(L)y_t = a(L)e_t, \tag{2}$$

where L is the so called lag (backshift) operator for which this holds:

$$LX_t = X_{t-1} \quad \text{and} \quad L^k X_t = X_{t-k},$$

and $b(L)$ and $a(L)$ are polynomials of order p, q respectively. Process y_t is said to be causal if for all $z \leq 1$ $b(z) \neq 0$ and is invertible if if for all $z \leq 1$ $a(z) \neq 0$. This condition guarantees that coefficient of a and b are identifiable through the ACF function. For more information on causality and invertibility see Brockwell and Davis (2006).

3. Methods Usable for Estimating ARMA Models

The estimation of an ARMA model has always to deal with the lack of existing residuals series. This leads to use of multistage least squares method, maximum likelihood method and the Kalman filter approach.

3.1 Linear Regression Method

The most comfortable way to estimate the ARMA model is the least squares method with two stages. At the first stage, the residuals are generated using an AR model with a long enough length of lags as any moving average model can be expressed as an AR model. There are several recommendations with respect to the length of this AR model. First, the length was recommended to be long, like $T/4$ or $\ln T$, where T is the number of observations. Later, to avoid throwing away to much observations, this length of lags was shortened. Spliid (1983) suggests that the optimal length of the AR model generating residuals could be $h = p + q$. Hannan and Kavalieris (1984) propose that the optimal lag h for this purpose is the one that minimizes the Schwartz information criterion BIC[1]. A brief summary of regression methods used for ARMA model estimation can be found in Hannan and McDougall (1988). After obtaining the residuals series, estimating the final ARMA(p,q) will be an easy task. The estimation can be performed once or recursively. Though this approach is very convenient, it is left exposed to some drawbacks. As residuals e_t are generated from y_t through an AR model, they are correlated with each other. This may deform the statistical inference of estimates or raise their inconsistency issue.

[1] $BIC(h) = T \ln \hat{\sigma}_h^2 + h \ln T$.

3.2 Maximum Likelihood Estimation Method

If a dataset $x_t, t = 1,...,T$ is i.i.d. with the marginal density $f(x_t;\mu)$, the joint density function for the whole sample will be the product of the marginal densities of each observation x_t

$$f(x;\mu) = f(x_1,...,x_T) = \prod_1^T f(x_t;\mu). \tag{3}$$

The likelihood function is then the joint density which is a function of the parameters μ for the given dataset

$$\mathfrak{L}(\mu \mid x) = \mathfrak{L}(\mu \mid x_1,...,x_T) = \prod_1^T f(x_t;\mu). \tag{4}$$

Taking the logarithm of the likelihood function, we get the so called the log-likelihood function

$$\log \mathfrak{L}(\mu \mid x) = \sum_{t=1}^{T} \log(f(x_t;\mu)). \tag{5}$$

From (1) the series of residuals can be expressed as

$$e_t = y_t - c - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j}. \tag{6}$$

They are i.i.d. by definition and lets suppose that they are normally distributed, e.i.

$$e_t \sim N(0,\sigma_e^2),$$

the log-likelihood function of residuals series is

$$\log \mathfrak{L} = \sum_{t=1}^{T} \log(f(e_t;\mu)) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\sum_{t=1}^{T} e_t^2, \tag{7}$$

where $\mu = (c,\phi,\theta,\sigma_e^2)$, $\phi = (\phi_1,...,\phi_p), \theta = (\theta_1,...,\theta_q)$. The first term of the log-likelihood function does not depend on the values of the parameters θ, hence the optimal estimates are

$$\hat{\mu}_{\text{MLE}} = \arg\max_{\mu} \left[-\frac{T}{2}\log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\sum_{t=1}^{T} e_t^2 \right]. \tag{8}$$

The vector of parameters μ of a ARMA(p,q) model is the solution to optimization problem in (8). It is fully solvable if we have the whole history of y_t and e_t, i.e. if the two vectors $(y_0,...,y_{-p+1})$ and $(e_0,...,e_{-q+1})$ are known. Then it can be reduced to the problem of minimizing $\sum_{t=1}^{T} e_t^2$, which in fact is a nonlinear least squares problem. In reality no such history exists, so $(y_0,...,y_{-p+1})$ and $(e_0,...,e_{-q+1})$ are set equal to 0 and assuming $c = 0$ we get

$$\begin{aligned} e_1(\mu) &= y_1, \\ e_2(\mu) &= y_2 - \phi_1 y_1 - \theta_1 e_1, \\ e_3(\mu) &= y_3 - \phi_1 y_2 - \phi_2 \phi_1 - \theta_1 e_2 - \theta_2 e_1, \\ &\dots \\ e_q(\mu) &= y_q - \sum_{i=1}^{q-1} \phi_i y_{q-i} - \sum_{j=1}^{q-1} \theta_j e_{q-j}, \end{aligned} \tag{9}$$

where $\phi_i = 0$ for $p < i \leq q-1$ if $p < q$. Together with $(y_1,...,y_p)$ we have a full history for all observations from $\max(p,q)$ on. For the nonlinear least squares problem $\min_{\mu} \sum_{t=1}^{T} e_t^2$, the derivative of the objective function $e_t = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j}$ is

$$D_{tj} = \frac{\partial e_t}{\partial \mu_j}. \tag{10}$$

The derivatives can be analytically calculated recursively for $t \geq \max(p,q)$. For the autoregressive terms one has:

$$\frac{\partial e_t}{\partial \phi_i} = -y_{t-i} + \sum_{k=1}^{q} \theta_k \frac{\partial e_{t-k}}{\partial \phi_i}, \ \ i = 1,...,p, \tag{11}$$

and for the moving average terms

$$\frac{\partial e_t}{\partial \theta_i} = -e_{t-i} + \sum_{k=1}^{q} \theta_k \frac{\partial e_{t-k}}{\partial \theta_i}, \ \ i = 1,...,q. \tag{12}$$

Though vector of parameters μ is estimated by a least squares approach, it still is under the framework of maximum likelihood estimation method, hence it possesses all asymptotic properties of a MLE estimate. A detailed description of this can be found in Hamilton (1992) or Enders (2004) as well as some of its partial improvement in the works of Jones (1980) and McLeod and Zhang (2008).

3.3 Kalman Filter

Kalman filter is an algorithm named after one of the main author of this technique Kalman (1960). This algorithm uses measurements observed over time to obtain estimates of unknown variables (states). In mathematical term, Kalman filter is a system of equations which allows to estimate the state of a process in such a way that minimizes the mean of the squared error. The filter enables to estimate past, present, and future states, even if the precise nature of the system is unobservable. Let y_t denote a vector of m variables observed at time t. Let x_t be a vector of n possibly unobserved variables called state variables. The state space model representing the dynamics of variables included in the system is given by two systems of linear equations:

- measurement equations describing the relationship between observed and unobserved variables

$$y_t = \mathbf{A}_t x_t + \mathbf{B}_t z_t^1 + e_t, \tag{13}$$

- transition equations describing the dynamics of unobserved variables

$$x_t = \mathbf{F}_t x_{t-1} + \mathbf{H}_t z_{t-1}^2 + v_t, \tag{14}$$

where z_t^1, z_t^2 are two vectors of exogenous regressors, $\mathbf{A}, \mathbf{B}, \mathbf{F}$ and $\mathbf{H}$ are matrices of corresponding dimensions and e_t, v_t are Gaussian white noise, e.i. $e_t \sim N(0, \mathbf{Q})$ and $v_t \sim N(0, \mathbf{R})$. Kalman filter works in the following way. At the beginning of time t we use historical information up to t to predict x_t as $\hat{x}_t^-$

$$\hat{x}_t^- = \mathbf{F}_t x_{t-1} + \mathbf{H}_t z_{t-1}^2. \tag{15}$$

This prediction is connected with an apriori error $\hat{e}_t^-$

$$\hat{e}_t^- = x_t - \hat{x}_t^-. \tag{16}$$

The corresponding covariance matrix of the a priori prediction errors is

$$\mathbf{P}_t^- = \mathbb{E}(\hat{e}_t^- \hat{e}_t^{-T}). \tag{17}$$

After that we use information available at time t to compute a posteriori estimate of state variables $\hat{x}_t$ as a linear combination of an a priori estimate $\hat{x}_t^-$ and a weighted difference between an actual measurement and a measurement prediction

$$\hat{x}_t = \hat{x}_t^- + \mathbf{K}(y_t - \mathbf{A}_t \hat{x}_t^- - \mathbf{B}_t z_t^1), \tag{18}$$

where the difference between an actual measurement and a measurement prediction is called the measurement innovation and matrix $\mathbf{K}$ is the Kalman gain. This updating is connected with a posteriori estimate error

$$\hat{e}_t = x_t - \hat{x}_t, \tag{19}$$

and its covariance matrix

$$\mathbf{P}_t = \mathbb{E}(\hat{e}_t \hat{e}_t^T). \tag{20}$$

To make the understanding more straightforward, we assume there is no exogenous regressor in the observation equation, that is $B = 0$. Then the Kalman gain matrix in (18) that minimizes posteriori estimate error $\hat{e}_t$ in (19) can be of the following form

$$\mathbf{K}_t = \mathbf{P}_t^- \mathbf{A}^T (\mathbf{A}\mathbf{P}_t^- \mathbf{A}^T + \mathbf{R})^{-1}. \tag{21}$$

Intuitively, we can see that if the size of errors in the transition equations is small, then the correction in the updating part is also small (there is almost nothing to improve). Otherwise, the role of correction goes up. And if if the size of errors in the observation equations is small, then weight of innovation depends mainly on A^{-1}. To sum up the use of Kalman filter for predicting the state of unobservables with known matrices $\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{H}, \mathbf{Q}$, and $\mathbf{R}$, the procedure is as follows:

- Set initial values for $\hat{e}_0, \mathbf{P}_0$.
- Predict the a priori state estimates $\hat{x}_t^-$, the prediction errors $\hat{e}_t^-$ and the covariance matrix of the errors $\mathbf{P}_t^-$.
- Compute the Kalman gain $\mathbf{K}_t$, then calculate the posteriori state estimates $\hat{x}_t$, the posteriori errors $\hat{e}_t$ and the covariance matrix of the posteriori errors $\mathbf{P}_t$.

To use Kalman filter we need to rewrite a ARMA(p,q) model in the state-space form. The measurement equation is

$$y_t = A_t' x_t, \tag{22}$$

whereas the transition equation is

$$x_t = \mathbf{F}_t x_{t-1} + \vartheta' v_t, \tag{23}$$

where x_t is a $r \times 1$ vector of state variables, $\mathbf{F}_t$ is a matrix of size $r \times r$, A_t, ϑ are $r \times 1$ vectors, $r = \max(p, q+1)$. They are defined as follows

$$\mathbf{F}_t = \begin{bmatrix} \phi_1 & 1 & 0 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & 0 & \cdots & 0 \\ \phi_3 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \ddots & 0 \\ \phi_{r-1} & 0 & 0 & 0 & \cdots & 1 \\ \phi_r & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}; \quad \vartheta = \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{r-1} \\ \theta_r \end{bmatrix}; \quad A_t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

The state vector x_t is

$$x_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \cdots + \phi_r y_{t-r+1} + \theta_1 e_t + \cdots + \theta_{r-1} e_{t-r+2} \\ \vdots \\ \phi_{r-1} y_{t-1} + \phi_r y_{t-2} + \theta_{r-2} e_t + \theta_{r-1} e_{t-1} \\ \phi_r y_{t-1} + \theta_{r-1} e_t \end{bmatrix}.$$

Let us set the initial value of $\hat{x}_1^- = 0$. Then Kalman filter allows us to recursively compute the prediction of state variables $\hat{x}_t^- = \mathbb{E}(x_t \mid y_{t-1}, ..., y_0; \hat{x}_0)$ as well as the covariance matrix of the a priori prediction errors

$$\mathbf{P}_t^- = \mathbb{E}((x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)^T).$$

Having the estimate $\hat{x}_t^-$, we can calculate the innovations using the observation equation as

$$e_t = y_t - \mathbb{E}(y_t \mid y_0, ..., y_{t-1}; x_0) = y_t - A' \hat{x}_t^-,$$

and their covariance matrix

$$\begin{aligned} \omega_t &= \mathbb{E}\left((y_t - A'\hat{x}_t^-)(y_t - A'\hat{x}_t^-)^T\right) \\ &= \mathbb{E}\left((A'x_t - A'\hat{x}_t^-)(A'x_t - A'\hat{x}_t^-)^T\right) \\ &= A'\mathbf{P}_t^- A. \end{aligned}$$

Using Kalman filter equations we can derive the following evolution of the matrices $\mathbf{P}_{t+1}^-$

$$\mathbf{P}_{t+1}^- = \mathbf{F}\left[\mathbf{P}_t^- - \mathbf{P}_t^- A A' \mathbf{P}_t^-\right] \mathbf{F}' + \vartheta \vartheta' \sigma^2. \tag{24}$$

For the initial value $\hat{x}_1^- = 0$ we have the corresponding matrix $\hat{\mathbf{P}}_1^- = 0$ and the likelihood function of an observation vector $(y_0, y_1, ..., y_T)$ is

$$\mathfrak{L} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\omega_t}} \exp\left(-\frac{e_t^2}{2\omega_t}\right), \tag{25}$$

and taking logarithm of it, we get the log-likelihood function as

$$\log \mathfrak{L} = -\frac{T}{2}\log(2\pi) - \sum_{t=1}^{T} \log(\omega_t) - \sum_{t=1}^{T} \frac{e_t^2}{2\omega_t}. \tag{26}$$

Basically, the estimates of (ϕ, θ, σ^2) are the ones that maximize (26) with respect to (ϕ, θ, σ^2). We can simplify the computation by initialize the filter with covariance matrix $\bar{\mathbf{P}}_t^- = \sigma^2 \mathbf{P}_t^-$. After dropping the constant term, the log-likelihood function then becomes

$$\log \mathfrak{L} = -\sum_{t=1}^{T}\left[\log(\sigma^2\omega_t) + \frac{e_t^2}{2\sigma^2\omega_t}\right].$$

It can be shown that optimal value of σ^2 can be directly obtained as

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{e_t^2}{\omega_t}, \tag{27}$$

and the log-likelihood function after some rearrangement will be of the form

$$\log \mathfrak{L} = -\left[T\log\sum_{t=1}^{T}\frac{e_t^2}{\omega_t} + \sum_{t=1}^{T}\log(\omega_t)\right], \tag{28}$$

which will be optimized with respect only to $[\phi,\theta]$. More about the Kalman filter can be found in the already mentioned book of Hamilton (1992) or in the work of Harvey (1989).

4. Empirical Analysis and Results

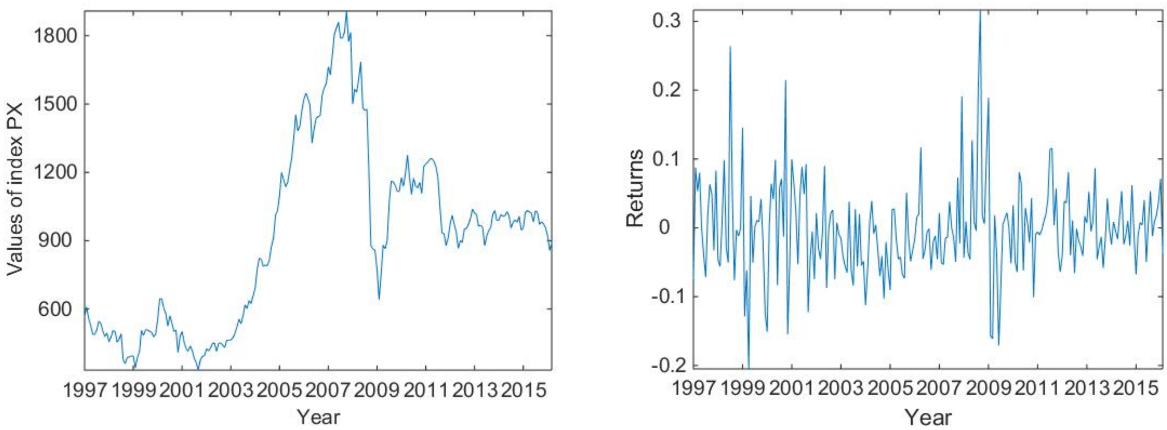
To evaluate the ability of three estimation methods described above to capture the persistence present in a time series, we choose the monthly series of Prague stock market index PX from January of 1997 to March of 2016. The original series of index PX values is transform into a logarithmic returns series and their developments over this period of time are shown at Figure 1. The corresponding descriptive statistics are in Table 1. Figure 1 shows that the examined returns series on the right panel is stationary at least in the mean and might not be stationary in variance. Therefore the stationarity of this series is also checked by the ADF test and the null hypothesis of having a unit root is strongly rejected as the test statistic is very high.

Table 1: The descriptive stattistics of data – Prague stock market index PX

PX	Mean	Median	Maximum	Minimum	Std. dev.	Skewness	Kurtosis
Level	909.01	938.70	1908.30	331.90	399.24	0.479	2.4072
Returns	0.0020	0.0053	0.2050	-0.3165	0.0688	-0.8283	6.1330

Source: the authors.

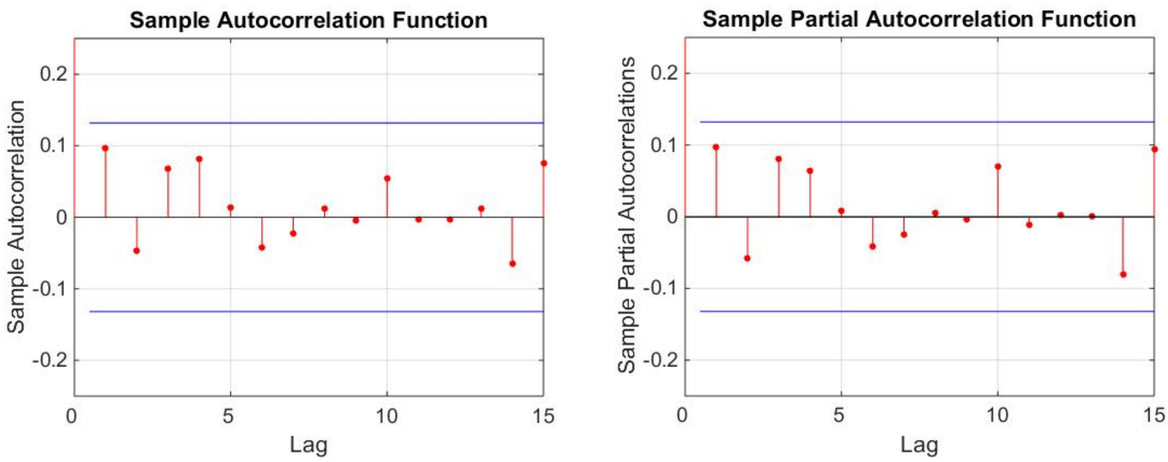
Figure 1: The development of stock market index PX over the last 20 years



Source: the authors.

After that we use Box - Jenkins methodology to preliminarily identify the possible specification for the ARMA model used to model the monthly return series of index PX. In order to proceed further, we first carry out autocorrelation analysis. The results of this analysis is shown in Figure 2. Both ACF and PACF functions display a similar pattern with several outstanding peaks of the first lags. Then the rest of them of both functions seems to decay as undampened sine-waves. This course of ACF and PACF functions leads us to think that our series can be modeled by an ARMA model and the initial specification can be an ARMA(4,4), where the maximum order for both AR and MA terms is 4. The whole identification process is proceeded in accordance with suggestions proposed by Choi (1992). However, for the identification purpose of an ARMA(*p,q*) model this initial step is not of great importance as real data rarely exhibit simple patterns. The correct specification can be achieved at a later stage with the use of diagnostic testing.

Figure 2: The ACF and PACF of returns of stock market index PX



Source: the authors.

We estimate coefficients of the ARMA(4,4) model by all three methods and use the F-test or LR test to ex post select the most suitable specification for the PX return series. It turns out that the best specification is an ARMA(2,3) model. The estimation is conduct in Matlab with our own programs written for this purpose. More information on numerical methods is available in Brandimarte (2006). The estimation results are reported in Tables 2,3, and 4 for the least squares method, the maximum likelihood method and Kalman filter method respectively.

Table 2: The estimation results using the LS method

Term	Coefficient	Standard Error	t - Stat	p - Value
C	-0.0028	0.0049	-0.5630	0.5740
AR(1)	0.6018	0.4459	1.3497	0.1786
AR(2)	-0.4860	0.4249	-1.1438	0.2540
MA(1)	-0.5039	0.4513	-1.1165	0.2654
MA(2)	0.3833	0.4218	0.9088	0.3645
MA(3)	0.1781	0.0878	2.0300	0.0436

Source: the authors.

Table 3: The estimation results using the MLE method

Term	Coefficient	Standard Error	z - Stat	p - Value
C	-0.0021	0.0044	-0.4769	0.6335
AR(1)	0.8120	0.0053	13.4193	0.0000
AR(2)	-0.8800	0.0053	-14.8738	0.0000
MA(1)	-0.7221	0.0053	-8.4221	0.0000
MA(2)	0.7802	0.0053	9.2493	0.0000
MA(3)	0.1800	0.0053	2.5658	0.0103

Source: the authors.

Table 4: The estimation results using Kalman filter

Term	Coefficient	Standard Error	z - Stat	p - Value
C	-0.0021	0.0188	-0.1115	0.9113
AR(1)	0.8228	0.0452	18.2211	0.0000
AR(2)	-0.8874	0.0315	-28.2970	0.0000
MA(1)	-0.7420	0.0628	-11.8050	0.0000
MA(2)	0.8053	0.0566	14.2291	0.0000
MA(3)	0.1842	0.0512	3.5968	3.2e-4

Source: the authors.

The estimation results show that all estimated coefficients of the ARMA(2,3) model except the one for term MA(3) are statistically insignificant at 5%. On the other hand, all estimates obtained by using the other two methods except the constant are strongly statistically significant. The insignificance of coefficients estimated by a least squares method may come from the mutual dependence of regressors. When using this method, first we have to generate a residuals series by an AR model. The residuals then depend on the AR terms. After that we estimate a model where the lagged observation are dependent on these residuals. Otherwise, the maximum likelihood and Kalman filter methods provide us with similar results. All methods give us estimate with the same sign for our series.

We also measure the quality of three methods by another criterion which is the sum of squared errors (SSE) of the ARMA(2,3) model estimated by these methods. In this aspect, the best one is the maximum likelihood method which induces the SSE = 1.0283 for our series. The second best is the method with Kalman filter approach when SSE generated from coefficients obtained by this method is 1.0406. The worst one is the least squares method with SSE = 1.0538. As we do not know the true values of the coefficients of our model, we proceed with another experiment in which we simulate a thousand runs of an ARMA(2,3) model with $c = 0$, $\phi = [0.75 - 0.85]$, and $\theta = [-0.7\ 0.8\ 0.2]$. These values are similar to those we obtain from PX returns series. Then we use all three methods to estimate back the known coefficients from these series to examine whether they can fully recover these values from simulated data. In Tables 5,6, and 7 some descriptive statistics of coefficients estimated by each method. Here the standard deviations are calculated from the estimation results. In Table 8 the results when the series of residuals is also delivered is shown for comparison.

Table 5: The simulation results with the LS method

Coefficient	Mean	Median	Maximum	Minimum	Std. dev.
C	0.0026	0.0014	0.3520	-0.4200	0.0913
AR(1)	0.2909	0.3101	1.0541	-0.7796	0.2760
AR(2)	-0.3649	-0.3811	0.6124	-1.1397	0.2438
MA(1)	-0.2207	-0.2352	0.9112	-1.0202	0.2821
MA(2)	0.2948	0.3044	1.1181	-0.7550	0.2375
MA(3)	0.1212	0.1217	0.3960	-0.1541	0.0865

Source: the authors.

Table 6: The simulation results with the MLE method

Coefficient	Mean	Median	Maximum	Minimum	Std. dev.
C	-0,0025	-2.2e-4	0.7549	-1.0720	0.1609
AR(1)	0,3720	0.6415	1.9059	-1.9490	0.6376
AR(2)	-0,6505	-0.8206	0.9902	-1.0741	0.4356
MA(1)	-0,2825	-0.5386	2.3194	-1.7990	0.6593
MA(2)	0,5720	0.6982	1.7410	-1.0518	0.4559
MA(3)	0,1234	0.1436	0.4472	-0.2946	0.1274

Source: the authors.

Table 7: The simulation results with Kalman filter

Coefficient	Mean	Median	Maximum	Minimum	Std. dev.
C	-	-	-	-	-
AR(1)	0.3992	0.6793	1.9274	-1.7577	0.6206
AR(2)	-0.6641	-0.8435	0.9935	-1.0146	0.4490
MA(1)	-0.3157	-0.5910	1.9844	-1.8916	0.6464
MA(2)	0.5966	0.7506	1.1412	-1.0212	0.4996
MA(3)	0.1348	0.1637	0.4281	-0.2938	0.1324

Source: the authors.

Table 8: The simulation results with the LS method with known residuals series

Coefficient	Mean	Median	Maximum	Minimum	Std. dev.
C	6.7e-4	-0.0019	0.2666	-0.2677	0.0697
AR(1)	0.7388	0.7436	1.4688	-0.0810	0.2072
AR(2)	-0.8842	-0.8590	-0.3526	-1.7512	0.1982
MA(1)	-0.6971	-0.7016	0.1681	-1.4476	0.2214
MA(2)	0.8325	0.8088	1.7005	0.3140	0.2068
MA(3)	0.2012	0.2024	0.3899	0.0027	0.0697

Source: the authors.

The obtained results show that without the knowledge of the residuals series, none of the used methods can fully extract the coefficients of the model from simulated series. Further, the estimation results are widely distributed around the true values regardless of which method is used for the estimation purpose. On the other hand, on average, all three methods can provide estimates with correct signs of the true values, all are lower in absolute values. This means that methods used to estimate coefficients of an ARMA model tend to underestimate the actual values. In terms of how far they can provide their estimates from the true values, we measure this distance by the euclidean distance from the average estimates and the true values. The results are as follows. The worst method measured by this distance is the least squares method with $d = 0.9657$, then the maximum likelihood method with $d = 0.6441$ and the best one is Kalman filter with $d = 0.5912$. Otherwise if we disposed the residuals series this distance would be $d = 0.0486$.

5. Conclusion

In this paper we compare the ability methods most often used to estimate the coefficients of an ARMA model. First we apply them to estimate an ARMA(2,3) model for index PX returns series. Our finding is that for a model with unknown parameters the least squares method is the least suitable method for this purpose both in terms of statistical significant and modeling ability measured by the sum of squared errors, which is the goodness of fit measure. Measured by the goodness of fit, the best method is MLE while Kalman filter can provide similar results. Kalman filter based estimation does not require the calculation of initial residuals up to q. Then using simulated data we have found that none of these methods can fully extract the true coefficients from simulated series and they tend to underestimate them. In our opinion, the least squares method the least reliable method for estimation objective and the two remaining methods can give similar results and they should be considered as the first option. Furthermore, since the estimates are far from their true values, any interpretation of the estimation results should be cautious.

Acknowledgements

The support from the Czech Science Foundation under Grant P402/12/G097 is gratefully acknowledged.

References

[1] BRANDIMARTE, P. 2006. Numerical methods in finance and economics : A MATLAB-based introduction. 2nd ed. New York : Wiley, 2006. ISBN 978-0-471-74503-7.

[2] BROCKWELL, P. J., DAVIS, R. A. 2006. Time Series : Theory and methods. 2nd ed. London : Springer, 2006. ISBN 3-540-97429-6.

[3] CHOI, B. S. 1992. ARMA model identification. Berlin : Springer, 1992. ISBN 978-1-4613-9745-8.

[4] ENDERS, W. 2004. Applied econometric time series. 3rd ed. New York : Wiley, 2009. ISBN 978-0470505397.

[5] HAMILTON, J. D. 1992. Time series analysis. Princeton : Princeton University Press, 1994. ISBN 9780691042893.

[6] HANNAN, E. J., KAVALIERIS, L. 1984. A Method for autoregressive-moving average estimation. In Biometrika, 1984, vol. 71, iss. 2, pp. 273-280.

[7] HANNAN, E. J., McDOUGALL, A. J. 1988. Regression procedures for ARMA estimation. In Journal of the American Statistical Association, 1988, vol. 83, iss. 409, pp. 490-498.

[8] HARVEY, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge : Cambridge University Press, 1989. ISBN 0521405734.

[9] JONES, R: H. 1980. Canonical structure of linear dynamical systems. In Technometrics, 1980, vol. 22, iss. 3, pp. 389-395.

[10] KALMAN, R. E. 1962. A new approach to linear filtering and prediction problems. In Proceedings of the National Academy of Sciences of the United States of America, 1962, vol. 48, iss. 4, pp. 596-600.

[11] MCLEOD, A. I., ZHANG, Y. 2008. Faster ARMA maximum likelihood estimation. In Journal Computational Statistics and Data Analysis, 2008, vol. 52, iss. 4, pp. 2166-2176.

[12] SPLIID, H. 1983. A fast estimation method for the vector autoregressive moving average model with exogenous variables. In Journal of the American Statistical Association, 1983, vol. 78, pp. 843-849.