1 Introduction

Usually, the true data-generating process is not known when analyzing real-world data in time series analysis. Hence, to understand the process and form predictions, the parameters of a suggested theoretical process need to be estimated based on a single realization. In this regard, one estimation technique that is frequently used in econometrics is conditional maximum likelihood (ML). Accordingly, this paper examines the parameter estimation of an MA(1) process with conditional ML for both small and large sample sizes.

2 Data Generation

First, an MA(1) process needs to be simulated to generate data which can be used for the analysis. Here, an MA(1) process of the following form is considered:

$$y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t \tag{1}$$

$$\text{with } \mu = 10,\ \theta = 0.2,\ \sigma^2 = 1,\ T = 100,\ \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

The simulated realization that will be used in the analysis is displayed in Figure 1.

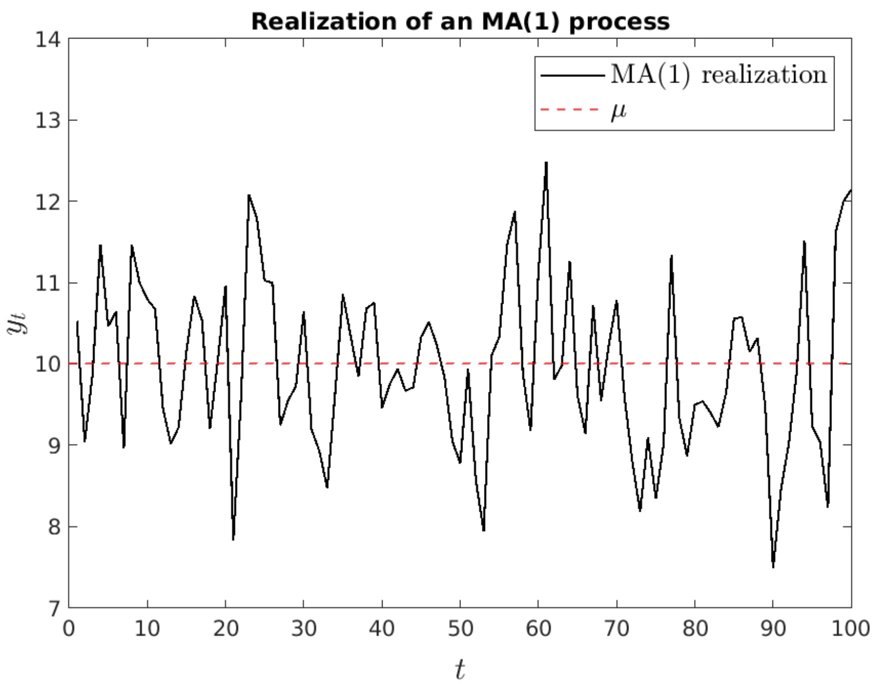

Figure 1: One simulated MA(1) realization and its expected value

3 Parameter estimation

Based on the obtained data, the parameters of the underlying process are estimated as if the true parameters are unknown. For this purpose, conditional ML is considered.

3.1 Construction of the log-likelihood function

Generally, ML estimation is based on the maximization of the likelihood function, i.e. the joint density of the random variables. This joint density can be expressed by conditional densities and one marginal density. Here, *conditional* ML is used which omits the marginal density to avoid numerical optimization.

The conditional log-likelihood function for an MA(1) process can be constructed by conditioning on ε_{t-1}. From Formula (1), it can be derived that $Y_t|\varepsilon_{t-1} \sim \mathcal{N}(\mu + \theta\varepsilon_{t-1}, \sigma^2)$. Accordingly, the conditional likelihood contribution for given parameters $\boldsymbol{\theta} = (\mu,\, \theta,\, \sigma^2)'$ can be determined.

$$f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - \mu - \theta\varepsilon_{t-1})^2}{2\sigma^2}\right] \tag{2}$$

As $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, ..., \varepsilon_T)$' is unobserved, Formula (1) needs to be rearranged to back out ε_t in an iterative approach assuming $\varepsilon_0 = 0$.

$$\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1} \tag{3}$$

Hence, Formula (2) can be reformulated using Formula (3).

$$f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(\varepsilon_t)^2}{2\sigma^2}\right] \tag{4}$$

Using Formula (4), the value of the conditional log-likelihood function is computed based on data $\boldsymbol{y}$ and parameters $\boldsymbol{\theta}$ as shown in Formula (5).

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\varepsilon_t^2}{2\sigma^2}\right] \tag{5}$$

Exemplary, Table 1 shows the log-likelihood value for two parameter sets (a) and (b). It can be seen that the parameter specifications of (a) result in a higher log-likelihood value than the ones in (b). This is reasonable as the parameters of (a) are the true parameters of the data-generating process and the parameters of (b) seem to be relatively far away from those.

	μ	θ	σ^2	$\ln \mathcal{L}$
(a)	10	0.2	1	-139.431
(b)	5	0.7	2	-369.776

Table 1: Log-likelihood values for two MA(1) parameter specifications

3.2 Estimation of parameters

Using Formula (5), the parameters $\boldsymbol{\theta}$ of the MA(1) process can be estimated by maximizing the conditional log-likelihood function, i.e. $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta})$. For this purpose, three optimization functions supplied by MATLAB can be used (1: `fminsearch`, 2: `fminunc`, 3: `patternsearch`). As these are minimization algorithms, Formula (5) is slightly adapted by multiplying it with (-1).
Using the `fminunc` algorithm, the parameters of the MA(1) realization displayed in Figure 1 are estimated with the true parameters values as the starting values for the optimization. The results along with the respective true parameter values are presented in Table 2.

	μ	θ	σ^2
True value	10.000	0.200	1.000
Estimated value	9.979	0.473	0.888

Table 2: Parameter values for an MA(1) process using conditional ML

It can be seen that $\hat{\mu}$ and $\hat{\sigma}^2$ are already quite close to their true values given the small sample size $T = 100$. In particular, it is apparently easier to estimate the constant parameter μ than the MA(1) parameter θ. Notice that the estimated parameter values deviate from the true values due to the small number of observations and the randomness in the process simulation and estimation.

3.3 Computation of standard errors

To evaluate estimation precision, standard errors of the parameter estimates are computed using the Hessian-based estimate for the variance-covariance matrix.

$$\widehat{Var}_{Hessian}(\hat{\boldsymbol{\theta}}) = -\left[\frac{\partial^2 \mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]^{-1} \tag{6}$$

Subsequently, the standard errors for $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\theta}, \hat{\sigma}^2)$ can be obtained by taking the square root of the k^{th} row, k^{th} column element of $\widehat{Var}(\hat{\boldsymbol{\theta}})$. The respective standard errors are reported in Table 3.

	$\hat{\mu}$	$\hat{\theta}$	$\hat{\sigma}^2$
Standard error	0.138	0.093	0.126

Table 3: Hessian-based standard errors of the conditional ML estimates

3.4 Hypothesis testing

Based on the parameter estimates and their standard errors, hypothesis tests can be conducted. Considering $H_0 : \theta = 0.4$, the following t-statistic is computed.

$$t = \frac{\hat{\theta} - \bar{\theta}}{s.e.(\hat{\theta})} = \frac{0.473 - 0.4}{0.093} = 0.788^1 \tag{7}$$

Considering $t \overset{a}{\sim} \mathcal{N}(0,1)$ under H_0, a two-sided test with $\alpha = 0.05$ leads to the critical values $\Phi^{-1}(0.025) = -1.96$ and $\Phi^{-1}(0.975) = 1.96$. As $-1.96 < 0.788 < 1.96$, $H_0 : \theta = 0.4$ cannot be rejected.

3.5 Corresponding p-values

To compute the two-sided p-value of the t-statistic from Section 3.4, the probability mass outside the interval $\pm|t|$ of the standard normal distribution is computed.

$$p\text{-value} = P(X \geq t = 0.788) \times 2 = 0.431$$

Since the resulting two-sided p-value exceeds the pre-specified significance level of $\alpha = 0.05$ considerably (i.e. $0.431 > 0.05$), $H_0 : \theta = 0.4$ cannot be rejected as already stated above.

3.6 Confidence Intervals

Furthermore, the 95% confidence intervals for $\hat{\mu}$ and $\hat{\theta}$ are derived using the estimate, the confidence level from the standard normal distribution and the standard error, resulting in:

$$\mu\text{: } [9.708,\ 10.249] \text{ with } \hat{\mu} = 9.979$$
$$\theta\text{: } [0.291,\ 0.655] \text{ with } \hat{\theta} = 0.473$$

Conducting a t-test with any value inside the confidence interval as the hypothesized value would lead to a non-rejection of H_0 for $\alpha = 0.05$. In a more technical interpretation, the true parameters are captured by the stochastic confidence interval on

[1] Due to rounding issues, the numbers displayed here might not always add up. The reported results are the exact outcomes of the computations in MATLAB.

average in 95% of the estimations given multiple realizations of the process are observed. However, usually only one realization is observed when analyzing real-world data and therefore, this interpretation generally does not offer much value.
Since the true data generating process is known in the present case, it can be seen that θ_0 does not lay within the confidence interval. As stated above, this happens with a probability of 5%. To evaluate this, confidence intervals for 1,000 different realizations and estimations of the process are computed. There, θ_0 is captured by the confidence interval in 92.3% of the estimations which supports this statement.

3.7 Variance estimation

Besides the Hessian-based approach shown in Section 3.3, two alternative techniques to estimate the variance-covariance matrix and derive the standard errors are considered. Formula (8) shows an estimation using the outer product of the gradient (OPG) and Formula (9) a Quasi-Maximum-Likelihood (QML) approach. Notice that Formula (9) is a combination of Formulas (6) and (8).

$$\widehat{Var}_{OPG}(\hat{\boldsymbol{\theta}}) = \left[\frac{\partial\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} \times \frac{\partial\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right]^{-1} \tag{8}$$

$$\widehat{Var}_{QML}(\hat{\boldsymbol{\theta}}) = -\left[\frac{\partial^2\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]^{-1} \times \left[\frac{\partial\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} \times \frac{\partial\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right] \times -\left[\frac{\partial^2\mathcal{L}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]^{-1} \tag{9}$$

Table 4 shows the OPG- and QML-based standard errors compared to the Hessian-based ones from Section 3.3.

	$\hat{\mu}$	$\hat{\theta}$	$\hat{\sigma}^2$
Hessian-based	0.138	0.093	0.126
OPG-based	0.139	0.085	0.133
QML-based	0.138	0.104	0.121

Table 4: Comparison of standard errors

Overall, the standard errors do not differ considerably. The small differences between the Hessian- and the OPG-estimates indicate a correct specification of the conditional density, as stated by the Information Matrix Equality (further explained in Section 3.8). Since the true data generating process is known and the conditional density is correctly specified, the similarity of the estimates are in line with the expectations. Furthermore, the QML-based standard error of $\hat{\theta}$ is slightly larger.

This is the case because a misspecification of the conditional density is accepted in the QML setting. This increases the level of uncertainty about the estimate and therefore results in a larger standard error.

3.8 Large sample estimation

So far, a realization of length $T = 100$ was considered. To move towards limit results, a realization of length $T = 50,000$ is now simulated. The data generating process remains the same as defined in Section 2.
The estimation is conducted as explained above and its results are shown in Table 5. As consistent estimators for the parameters and a stationary process are used, the new estimates are expected to improve with increasing sample size.

	μ	θ	σ^2
True value	10.000	0.200	1.000
Estimated value	9.999	0.196	0.995
Standard error (Hessian-based)	0.005	0.004	0.006

Table 5: Parameter estimates and standard errors with $T = 50,000$

As expected, the parameter estimates are considerably closer to the true values and the standard errors are close to 0. Conducting a t-test with $H_0 : \theta = 0.4$ results in a test statistic $t = -46.747$ with a corresponding two-sided p-value of 0 for $\alpha = 0.05$. Therefore, H_0 can be rejected.
To evaluate which range of hypothesized values would lead to a non-rejection of H_0 for $\hat{\mu}$ and $\hat{\theta}$, 95% confidence intervals are constructed.

$$\mu\text{: } [9.989,\ 10.010] \text{ with } \hat{\mu} = 9.999$$
$$\theta\text{: } [0.187,\ 0.204] \text{ with } \hat{\theta} = 0.196$$

With increasing sample size, the confidence intervals center closer around the true parameters.
Using the alternative estimates for the variance-covariance matrix introduced in Section 3.7 results in the following standard errors.

	$\hat{\mu}$	$\hat{\theta}$	$\hat{\sigma}^2$
Hessian-based	0.005	0.004	0.006
OPG-based	0.005	0.004	0.006
QML-based	0.005	0.004	0.006

Table 6: Comparison of standard errors with $T = 50,000$

Now, all methods result in the same rounded standard errors. The Information Matrix Equality states that the expected value of the Hessian-based covariance matrix (see Formula (6)) equals the expected value of the OPG-based covariance matrix (see Formula (8)) if the conditional density is correctly specified. In this case, it is known that the conditional density is correctly specified as the true data generating process is known. Therefore, the Information Matrix Equality should hold and the Hessian- and OPG-approach should converge to the same standard errors with increasing sample size (law of large numbers). This is supported by Table 6.

The QML-based standard errors are calculated by a "sandwich" formula as shown in Formula (9), consisting of both OPG- and Hessian-based covariance calculation. As the Information Matrix Equality should hold in this case, the first two terms in Formula (9) cancel out leaving only the third term, the Hessian-based covariance matrix (compare Formula (6)). Hence, the QML- and the Hessian-based approach should also result in the same values, as shown in Table 6.

3.9 Log-Likelihood estimation for varying θ

To examine the behaviour of the conditional log-likelihood function with respect to θ, a realization with $T = 200$ is simulated. The corresponding conditional log-likelihood function is computed for varying $\theta \in [-0.6, 0.8]$ in steps of 0.05 while μ and σ^2 are fixed at their true values ($\mu_0 = 10$ and $\sigma_0^2 = 1$). Figure 2 displays the resulting likelihood values along with the true and estimated parameter value.

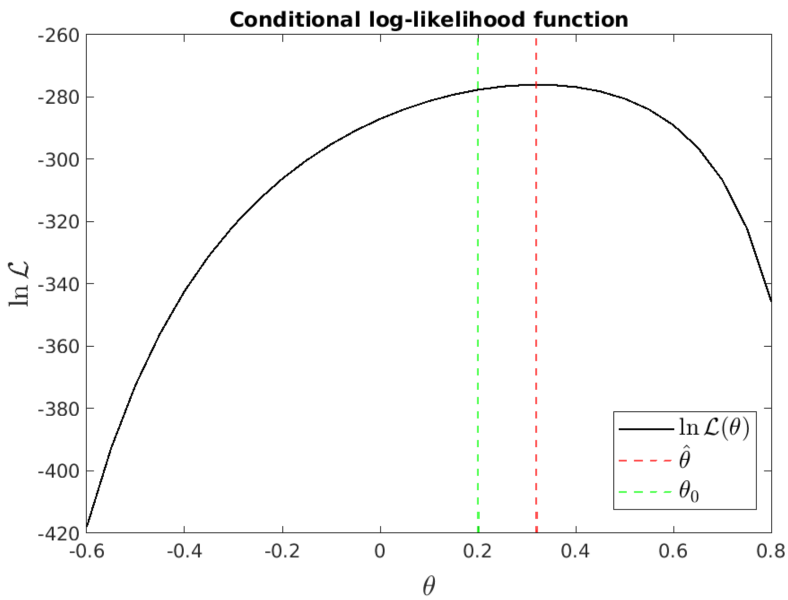

Figure 2: Conditional log-likelihood function for varying θ

A rather peaked log-likelihood function is observed, indicating a rather high estimation precision. It is maximized around the parameter estimate ($\hat{\theta} = 0.321$) which is slightly higher than the true parameter value ($\theta_0 = 0.2$). Again, this might be due to the small number of observations, the randomness, and the step size of θ.

3.10 Parameter estimation for a non-invertible MA(1)

Lastly, a simulated MA(1) with $\theta = 5, \mu = 10, \sigma^2 = 1$ and $T = 200$ is analyzed and its parameters $\boldsymbol{\theta}$ are estimated using conditional ML and `fminunc`. Theoretically, the estimation of MA(1)-processes causes trouble when the MA(1) parameter of the data generating process $|\theta| \geq 1$. This is the case because the process needs to be inverted to back out ε_t (see Formula (3)). Because θ enters in powers in the recursion, the innovations "explode" with $|\theta| \geq 1$ for increasing T and are not useful anymore for estimation purposes. Luckily, those non-invertible processes have a "good twin", another MA(1) process with different parameter values which is observationally equivalent. Its parameters $\tilde{\boldsymbol{\theta}}$ can be theoretically derived from the original ones as shown in Table 7. Accordingly, the MA(1) parameter $\tilde{\theta}$ of this twin process turns out to be $|\tilde{\theta}| < 1$ if $|\theta_0| \geq 1$.

True Parameter	Twin Parameter	True Value	Twin Value	Estimated Value
μ	$\tilde{\mu} = \mu$	10	10	9.597
θ	$\tilde{\theta} = \frac{1}{\theta}$	5	$\frac{1}{5} = 0.2$	0.190
σ^2	$\widetilde{\sigma^2} = \theta^2 * \sigma^2$	1	$5^2 \times 1^2 = 25$	27.767

Table 7: Expected and estimated parameters for a non-invertible MA(1)

The conditional ML estimation leads to parameter estimates close to the theoretical ones of the "good twin". As the log-likelihood function has maxima at both parameter specifications, the optimization algorithm needs to stay on the side of the "good twin" to obtain $|\hat{\theta}| < 1$. Since the starting value ($\theta = 0.6$) is rather far away from $\theta_0 = 5$, conditional ML estimates the values of the "good twin".

4 Conclusion

It is shown that conditional ML estimation is suitable for MA(1) parameter estimation based on one realization. Furthermore, estimation precision improves with increasing sample size. Also, a larger sample size supports the Information Matrix Equality. Hence, the Hessian-, OPG- and QML-based covariance matrices result in the same standard errors when using a correctly specified conditional density.