1 Parameter estimation in ARMA models with a unit root in the MA polynomial

Let $\{X_t\}$ be a stationary causal ARMA$(p,1)$ process of the form

$$X_t - \sum_{j=1}^{p} a_j^* X_{t-j} = \eta_t - \eta_{t-1}, \tag{1}$$

where η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta^2] = \nu^*$. We use the notation $\boldsymbol{a}^* = (a_1^*, \dots, a_p^*)$ and $\boldsymbol{\theta}^* = (\boldsymbol{a}^*, \nu^*)$. We now construct a maximum likelihood estimator of the parameters $a_1^*, \dots, a_p^*$ and ω^*. The construction proceeds in two steps: We first define an infeasible likelihood function which cannot be computed in practice and then approximate it by a feasible version.

Step 1. Let $\Pi_s Z_t$ be the orthogonal projection of a general (square-integrable) random variable Z_t onto the linear space spanned by $X_1, \dots, X_s$, denoted by span$\{X_1, \dots, X_s\}$. The projection $\Pi_{t-1} X_t$ is the best linear predictor of X_t based on $X_1, \dots, X_{t-1}$. Let $\xi_t(\boldsymbol{\theta}^*) = X_t - \Pi_{t-1} X_t$ be the prediction innovations and $e_t(\boldsymbol{\theta}^*) = \mathbb{E}[\xi_t^2(\boldsymbol{\theta}^*)]$ the corresponding prediction error. Under the assumption that the innovations $\xi_t(\boldsymbol{\theta}^*)$ are i.i.d. Gaussian, the (infeasible) log-likelihood is given by

$$\mathcal{L}_T(\boldsymbol{\theta}^*) = -\frac{1}{2} \sum_{t=1}^{T} \log\big(2\pi e_t(\boldsymbol{\theta}^*)\big) - \frac{1}{2} \sum_{t=1}^{T} \frac{\xi_t^2(\boldsymbol{\theta}^*)}{e_t(\boldsymbol{\theta}^*)}.$$

The prediction innovations $\xi_t(\boldsymbol{\theta}^*)$ and the prediction errors $e_t(\boldsymbol{\theta}^*)$ can be shown to have the following representations:

$$\xi_t(\boldsymbol{\theta}^*) = V_t(\boldsymbol{a}^*) - \frac{1}{\beta(\nu^*) + t} \sum_{s=p+1}^{t-1} V_s(\boldsymbol{a}^*) + \frac{\beta(\nu^*) + p + 1}{\beta(\nu^*) + t} \Pi_p \eta_p \tag{2}$$

$$e_t(\boldsymbol{\theta}^*) = \Big(1 + \frac{1}{\beta(\nu^*) + t}\Big) \nu^* \tag{3}$$

for $t > p$, where $V_t(\boldsymbol{a}^*) = \sum_{k=p+1}^{s} (X_k - \sum_{j=1}^{p} a_j^* X_{k-j})$ and $\beta(\nu^*) = (\nu^*/\mu_p) - p - 1$ with $\mu_p = \mathbb{E}[(\eta_p - \Pi_p \eta_p)^2]$.

Step 2. For a general parameter vector $\boldsymbol{\theta} = (\boldsymbol{a}, \nu) = (a_1, \dots, a_p, \nu)$, we approximate the innovations $\xi_t(\boldsymbol{\theta})$ by

$$\widehat{\xi}_t(\boldsymbol{\theta}) = V_t(\boldsymbol{a}) - \frac{1}{t} \sum_{s=p+1}^{t-1} V_s(\boldsymbol{a})$$

and the prediction error $e_t(\boldsymbol{\theta}) = \mathbb{E}[\xi_t^2(\boldsymbol{\theta})]$ by ν. A more convenient representation of

$\widehat{\xi}_t(\boldsymbol{\theta})$ is given by

$$\widehat{\xi}_t(\boldsymbol{\theta}) = Q_{t,0} - \sum_{j=1}^{p} a_j Q_{t,j} \quad \text{with} \quad Q_{t,j} = \sum_{\ell=p+1}^{t} X_{\ell-j} - \frac{1}{t} \sum_{s=p+1}^{t-1} \sum_{\ell=p+1}^{s} X_{\ell-j}.$$

Replacing $\xi_t(\boldsymbol{\theta})$ and $e_t(\boldsymbol{\theta})$ by the approximations $\widehat{\xi}_t(\boldsymbol{\theta})$ and ν in in $\mathcal{L}_T(\boldsymbol{\theta})$ yields the feasible likelihood

$$L_T(\boldsymbol{\theta}) = -\frac{T-p}{2} \log(2\pi\nu) - \frac{1}{2\nu} \sum_{t=p+1}^{T} \widehat{\xi}_t^2(\boldsymbol{\theta}).$$

Estimators $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{a}}, \widehat{\nu})$ of the parameters $\boldsymbol{\theta}^* = (\boldsymbol{a}^*, \nu^*)$ are defined as

$$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{a}}, \widehat{\nu}) = \arg\max_{\theta \in \Theta} L_T(\boldsymbol{\theta}).$$

It is straightforward to solve this maximization problem and to show that

$$\begin{aligned} \widehat{\boldsymbol{a}} &= \widehat{\boldsymbol{\Gamma}}_Q^{-1} \widehat{\boldsymbol{\gamma}}_Q \\ \widehat{\nu} &= \frac{1}{T-p} \sum_{t=p+1}^{T} \Big(Q_{t,0} - \sum_{j=1}^{p} \widehat{a}_j Q_{t,j} \Big)^2, \end{aligned}$$

where $\widehat{\boldsymbol{\Gamma}}_Q = (\widehat{\gamma}_Q(i,j) : 1 \le i,j \le p)$ is a $p \times p$ matrix and $\widehat{\boldsymbol{\gamma}}_Q = (\widehat{\gamma}_Q(0,1), \dots, \widehat{\gamma}_Q(0,p))^\top$ is a vector in $\mathbb{R}^p$ with the entries $\widehat{\gamma}_Q(i,j) = \sum_{t=p+1}^{T} Q_{t,i} Q_{t,j}$.

The estimators $\widehat{\boldsymbol{a}}$ and $\widehat{\nu}$ have the following theoretical properties.

Proposition 1.1. *Suppose that the process $\{\eta_t\}$ has a finite fourth cumulant κ. Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \nu^* \Gamma^{-1} & 0 \\ 0 & 2(\nu^*)^2 + \kappa \end{pmatrix} \right),$$

where $\boldsymbol{\Gamma} = (\gamma(i-j) : 1 \le i,j \le p)$ is the autocovariance matrix of the AR(p) process $\{Y_t\}$ with $Y_t = \sum_{j=1}^{p} a_j^* Y_{t-j} + \eta_t$.

Derivation of (2) and (3). Writing $W_t(\boldsymbol{a}^*) = X_t - \sum_{j=1}^{p} a_j^* X_{t-j}$, we have

$$\boldsymbol{\xi}_t(\theta^*) = W_t(\boldsymbol{a}^*) + \Pi_{t-1}\eta_{t-1} \tag{4}$$

for $t > p$. By definition, $\Pi_t \eta_t$ belongs to the linear space spanned by $X_1, \dots, X_{t-1}$. Moreover $\Pi_t \eta_t$ is orthogonal to the space spanned by $X_1, \dots, X_{t-1}$ since $\Pi_{t-1}\Pi_t \eta_t = \Pi_{t-1}\eta_t = 0$. Noticing that $\operatorname{span}\{\xi_t(\boldsymbol{\theta}^*)\} \oplus \operatorname{span}\{X_1, \dots, X_{t-1}\} = \operatorname{span}\{X_1, \dots, X_t\}$, we can infer that

$$\Pi_t \eta_t = \frac{\mathbb{E}[\eta_t \xi_t(\boldsymbol{\theta}^*)]}{e_t(\boldsymbol{\theta}^*)} \xi_t(\boldsymbol{\theta}^*). \tag{5}$$

Since $\xi_t(\boldsymbol{\theta}^*) = \eta_t + (\Pi_{t-1}\eta_{t-1} - \eta_{t-1})$, it holds that $\mathbb{E}[\eta_t \xi_t(\boldsymbol{\theta}^*)] = \nu^*$ and $e_t(\boldsymbol{\theta}^*) = \nu^* + \mu_{t-1}$ with $\mu_t = \mathbb{E}[(\eta_t - \Pi_t \eta_t)^2]$. Plugging this into (5) yields

$$\xi_t(\boldsymbol{\theta}^*) = W_t(\boldsymbol{a}^*) + \frac{\nu^*}{\nu^* + \mu_{t-2}} \xi_{t-1}(\boldsymbol{\theta}^*). \tag{6}$$

The term μ_t can be rewritten as

$$\mu_t = \mathbb{E}[(\eta_t - \Pi_t \eta_t)^2] = \nu^* - \mathbb{E}(\Pi_t \eta_t)^2 = \nu^* - \frac{(\nu^*)^2}{\nu^* + \mu_{t-1}} = \frac{\nu^* \mu_{t-1}}{\nu^* + \mu_{t-1}}.$$

This yields the recurrence equation $1/\mu_t = 1/\nu^* + 1/\mu_{t-1}$, which can be recursively applied to obtain that $1/\mu_t = (t-p)/\nu^* + 1/\mu_p$ for $t > p$. Using this in (6) gives that

$$\frac{\nu^*}{\nu^* + \mu_{t-2}} = \frac{\nu^*/\mu_{t-2}}{1 + \nu^*/\mu_{t-2}} = \frac{\nu^*/\mu_p + t - 2 - p}{\nu^*/\mu_p + t - 1 - p}$$

and thus

$$\xi_t(\boldsymbol{\theta}^*) = W_t(\boldsymbol{a}^*) + \frac{\beta(\nu^*) + t - 1}{\beta(\nu^*) + t} \xi_{t-1}(\boldsymbol{\theta}^*) \tag{7}$$

with $\beta(\nu^*) = \nu^*/\mu_p - p - 1$ for $t > p+1$. By iteratively applying (7), we arrive at

$$\xi_t(\boldsymbol{\theta}^*) = \sum_{s=p+1}^{t} \frac{\beta(\nu^*) + s}{\beta(\nu^*) + t} W_s(\boldsymbol{a}^*) + \frac{\beta(\nu^*) + p + 1}{\beta(\nu^*) + t} \Pi_p \eta_p,$$

which can be equivalently written as

$$\xi_t(\boldsymbol{\theta}^*) = V_t(\boldsymbol{a}^*) - \frac{1}{\beta(\nu^*) + t} \sum_{s=p+1}^{t-1} V_s(\boldsymbol{a}^*) + \frac{\beta(\nu^*) + p + 1}{\beta(\nu^*) + t} \Pi_p \eta_p.$$

Moreover, using the representation $e_t(\boldsymbol{\theta}^*) = \nu^* + \mu_{t-1}$ and the formulas on μ_t from above, it is easily seen that

$$e_t(\boldsymbol{\theta}^*) = \Big(1 + \frac{1}{\beta(\nu^*) + t}\Big) \nu^*.$$

Proof of Proposition 1.1. Let the process $\{Y_t\}$ be defined by the equations $Y_t = \sum_{j=1}^{p} a_j^* Y_{t-j} + \eta_t$. Since $X_t = Y_t - Y_{t-1}$, we obtain that

$$V_t(\boldsymbol{a}) = \sum_{k=p+1}^{t} \Big(X_k - \sum_{j=1}^{p} a_j X_{k-j}\Big) \tag{8}$$

$$= \{Y_t - Y_p\} - \sum_{j=1}^{p} a_j \{Y_{t-j} - Y_{p-j}\}. \tag{9}$$

From (8), it immediately follows that

$$\widehat{\xi}_t(\boldsymbol{\theta}) = \eta_t(\boldsymbol{a}) - \frac{1}{t}\sum_{k=p+1}^{t-1}\eta_k(\boldsymbol{a}) - \frac{p+1}{t}\eta_p(\boldsymbol{a}) \quad \text{with} \quad \eta_t(\boldsymbol{a}) = Y_t - \sum_{j=1}^{p} a_j Y_{t-j}, \tag{10}$$

where $\eta_t(\boldsymbol{a})$ equals η_t for $\boldsymbol{a} = \boldsymbol{a}^*$, that is, $\eta_t(\boldsymbol{a}^*) = \eta_t$. With the help of (9), we can further write

$$\widehat{\xi}_t(\boldsymbol{\theta}) = U_{t,0} - \sum_{j=1}^{p} a_j U_{t,j} \quad \text{with} \quad U_{t,j} = Y_{t-j} - \frac{1}{t}\sum_{k=p+1}^{t-1} Y_{k-j} - \frac{p+1}{t}Y_{p-j}. \tag{11}$$

Using (11) and taking the first derivatives of the likelihood $L_t(\boldsymbol{\theta})$, we obtain the first-order conditions

$$\frac{\partial L_T(\boldsymbol{\theta})}{\partial a_k} = \frac{1}{\nu}\sum_{p+1}^{T}\Big(U_{t,0} - \sum_{j=1}^{p} a_j U_{t,j}\Big)U_{t,k} \stackrel{!}{=} 0 \quad \text{for } 1 \le k \le p \tag{12}$$

$$\frac{\partial L_T(\boldsymbol{\theta})}{\partial \nu} = -\frac{T-p}{2\nu} + \frac{1}{2\nu^2}\sum_{t=p+1}^{T}\widehat{\xi}_t^2(\boldsymbol{\theta}) \stackrel{!}{=} 0. \tag{13}$$

From (12) together with some straightforward calculations, we get that

$$\sum_{j=1}^{p}\Big(\frac{1}{T-p}\sum_{t=p+1}^{T} U_{t,j}U_{t,k}\Big)\big(\widehat{a}_j - a_j^*\big) = \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\xi}_t(\theta^*)U_{t,k} \tag{14}$$

for $1 \le k \le p$, or equivalently,

$$\widehat{\boldsymbol{\Gamma}}_U(\widehat{a} - a^*) = \widehat{\boldsymbol{\rho}}_U, \tag{15}$$

where $\widehat{\boldsymbol{\rho}}_U = (\widehat{\rho}_U(1), \dots, \widehat{\rho}_U(p))^\top$ with $\widehat{\rho}_U(k) = (T-p)^{-1}\sum_{t=p+1}^{T}\widehat{\xi}_t(\theta^*)U_{t,k}$ and

$$\widehat{\boldsymbol{\Gamma}}_U = \begin{pmatrix} \widehat{\gamma}_U(1,1) & \dots & \widehat{\gamma}_U(p,1) \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}_U(1,p) & \dots & \widehat{\gamma}_U(p,p) \end{pmatrix}$$

with $\widehat{\gamma}_U(j,k) = (T-p)^{-1}\sum_{t=p+1}^{T} U_{t,j}U_{t,k}$. From (13) and (14), it further follows that

$$\widehat{\nu} = \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\xi}_t^2(\boldsymbol{\theta}^*) - \sum_{j=1}^{p}(\widehat{a}_j - a_j^*)\Big(\frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\xi}_t(\boldsymbol{\theta}^*)U_{t,j}\Big). \tag{16}$$

Noting that $\partial\widehat{\xi}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (-U_{t,1}, \dots, -U_{t,p})$, Lemmas 5 and 6 in Pham-Dinh (1978, Estimation of parameters in the ARMA model when the characteristic polynomial of the MA operator has a unit zero, AOS) yield that $\widehat{\mathbf{\Gamma}}_U = \mathbf{\Gamma} + o_p(1)$ and

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\rho}}_U \\ \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\xi}_t^2(\boldsymbol{\theta}^*) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \nu^*\mathbf{\Gamma} & 0 \\ 0 & 2(\nu^*)^2 + \kappa \end{pmatrix}\right). \tag{17}$$

(To prove (17), one uses that $U_{t,j} = Y_{t-j} - t^{-1}\sum_{k=p+1}^{t-1} Y_{k-j} - \{(p+1)/t\}Y_{p-j}$ with $\{Y_t\}$ being a stationary, causal AR(p) process and $\widehat{\xi}_t(\boldsymbol{\theta}^*) = \eta_t - t^{-1}\sum_{k=p+1}^{t-1}\eta_k - \{(p+1)/t\}\eta_p$ with η_t being i.i.d. variables.) Proposition 1.1 follows upon applying these results to (15) and (16). □