Abstract — In autoregression models the design variable has traditionally been assumed to be non-stochastic and innovations are normal. In most real life situations, however, the design variable is stochastic having a non-normal distribution as the innovations. Modified maximum likelihood method is utilized to estimate unknown parameters in such situations. Closed form estimators are obtained and shown to be efficient and robust.

Keywords — Modified maximum likelihood, Nonnormality, Robustness, Stochastic design.

I. INTRODUCTION

The autoregressive model (-1 < φ < 1)

\[ y_i - \phi y_{i-1} = \gamma_0 + \gamma_1 x_i + e_i \quad (1 \leq i \leq n) \]  

has many applications in agricultural, biological, biomedical sciences, engineering and business. Also it has a special role in economic analysis since Distributed Lag Models are very common in this area and according to Koyck, adaptive expectations and partial adjustment approaches it turns to have a common autoregression form given by (1). Estimation of the parameters depends on classical assumptions: the design variable \( X \) is nonstochastic and the innovations are normally distributed [1]-[4]. In recent years, however, it has been recognized that both normality assumption of innovations and nonstochasticity of design variable are too restrictive and unrealistic: see for example [5], [6]. In this case, model (1) is called a stochastic autoregressive model. By the use of the modified maximum likelihood (MML) methodology, the model has been opened up to stochastic design variable as well as nonnormal innovations. Since the maximum likelihood (ML) estimation technique which is known to have optimal features, is very problematic in situations when the distribution of \( X \) ( marginal part) or both are nonnormal, we use MML method. Estimation by this method is carried out in three steps: (i) the maximum likelihood equations are expressed in terms of the order statistics of a sample, (ii) the non-linear functions are replaced by linear approximations so that the differences between the two converge to zero as \( n \) becomes large and (iii) the resulting equations are solved and MMLEs are obtained. They are explicit functions of sample observations and are, therefore, easy to compute. MMLEs are also considerably more efficient than the least squares (LS) estimators for all sample sizes \( n \), particularly for large \( n \). They are known to be asymptotically fully efficient (unbiased and having minimum variances) under very general regularity conditions; [7], [8] and robust to plausible deviations from the assumed distributions and to mild data anomalies (e.g. outliers). In this study, we consider the estimation of the parameters of model (1) when \( X \) has generalized logistic distribution (GL) with parameter \( b \) and \( e \) has long-tailed symmetric distribution (LTS) with parameter \( p \).

II. STOCHASTIC AUTOREGRESSIVE MODEL

Consider the reparametrized autoregressive model

\[ y_i - \phi y_{i-1} = \gamma_0 + \gamma_1 x_i + e_i, \quad u_i = (x_i - \mu_1) / \sigma_1, \quad 1 \leq i \leq n; \]  

where \( 0 \leq \phi < 1 \) and \( \mu_1 \) and \( \sigma_1 \) are the location and scale parameters in the distribution of \( X \), respectively.

The model (2) is advantageous because the MMLEs and the LSEs of the parameters are invariant to the location and scale of \( X \) [9]-[11]. Assume that \( X \) and \( e \) being mutually independent and the distributions of \( e_1 \) \( / \sigma_1 = (x_1 - \mu_1) \) \( / \sigma_1 \) and \( e_2 / \sigma = (y_1 - \phi y_{i-1} - \gamma_0 - \gamma_1 u_i) / \sigma \) are given by (3) and (4), respectively.

Generalized logistic distribution: Assume that \( e_1 \) are iid and have generalized logistic distribution (b>0) :
The expected value and variance, respectively, are
\[ E(e) = \mu, \quad V(e) = \sigma^2, \]
\[ f(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left( \frac{-x - \mu}{\sigma} \right) , \quad -\infty < x < \infty. \]

The distribution is negatively skewed for \( b<1 \) and positively skewed for \( b>1 \). For \( b=1 \), it is the well known logistic distribution.

**Long-tailed symmetric distribution:** In model (2) suppose that \( e_i \) are iid and have long-tailed symmetric distribution (LTS). A broad range of LTS distributions (kurtosis \( \mu_4 / \mu_2^2 > 3 \)) is obtained by minimizing the error sums of squares
\[ \sum_{i=1}^{n} (e_i - \mu_i)^2 \]
Least Squares Estimators: The least squares estimators are obtained by minimizing the error sums of squares
\[ n \sum_{i=1}^{n} (x_i - \bar{\mu})^2 \]
and
\[ n \sum_{i=1}^{n} (y_i - \phi y_i - \gamma_0 - \gamma_1 u_i)^2. \]
They are
\[ \bar{\mu}_i = \bar{x}_i - \{ \psi(b) - \psi(l) \} \tilde{\sigma}_i, \quad \tilde{\sigma}_i = s_i / \sqrt{\{ \psi(b) + \psi(l) \}}; \]
\[ s_i^2 = \sum_{i=1}^{n} (x_i - \bar{x}_i)^2 / (n-1). \]
\[ \tilde{\theta} = (\tilde{U} \tilde{U})^{-1} (\tilde{U} \tilde{Y}) \quad \text{and} \quad \tilde{\sigma} = s_e. \]

**Modified maximum likelihood estimators:**

The likelihood function \( L \) is
\[ L = L_n L_e \quad (L_e = L_f k) \]
\[ \alpha \left( \frac{b}{\sigma} \right) = \prod_{i=1}^{n} \frac{\exp(-e_i / \sigma)}{[1 + \exp(-e_i / \sigma)]^{\sigma}} \]
where \( k = 2p - 3 \).

The maximum likelihood equations expressed in terms of the variates \( u_i \) and \( z_i = (y_i - \phi y_i - \gamma_0 - \gamma_1 u_i) / \sigma \) \( (1 \leq i \leq n) \) are
\[ \frac{\partial \ln L}{\partial \mu_i} = -n \frac{1}{\sigma^2} \sum_{i=1}^{n} g_1(u_i) - \frac{2\phi y_i}{k \sigma^2} \sum_{i=1}^{n} g_2(z_i) = 0, \]
\[ \frac{\partial \ln L}{\partial \sigma_i} = -n \frac{1}{\sigma^2} \sum_{i=1}^{n} u_i - \frac{(b + 1)}{\sigma^2} \sum_{i=1}^{n} g_1(u_i) \]
\[ + \frac{2\phi y_i}{k \sigma^2} \sum_{i=1}^{n} z_i g_2(z_i) = 0, \]
\[ \frac{\partial \ln L}{\partial \gamma_0} = \frac{2p}{k \sigma^2} \sum_{i=1}^{n} g_2(z_i) = 0, \]
\[ \frac{\partial \ln L}{\partial \gamma_1} = \frac{2p}{k \sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_2(z_i) = 0, \]
\[ \frac{\partial \ln L}{\partial \phi} = \frac{2p}{k \sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_2(z_i) = 0, \]
where
\[ g_1(u) = e^{-u} / (1 + e^{-u}) \quad \text{and} \quad g_2(z) = z / (1 + (1/k)z^2). \]

The solutions of the equations (8) are the MLEs. These equations, however, have no explicit solutions. Solving so many equations by iteration is very difficult and there can be problems of convergence [13], [14]. To overcome these difficulties the modified maximum likelihood estimation (MMLE) procedure is used.

MMLE is carried out basically in three steps: (i) the maximum likelihood equations are expressed in terms of the order statistics of a sample, (ii) the non-linear functions are
replaced by linear approximations so that the differences between the two converge to zero as \( n \) becomes large and (iii) the resulting equations are solved and MMLEs are obtained.

To derive MML estimators, we first express the likelihood equations (8) in terms of the ordered variates \( z_{i0}, (y[i], y[i+1−1], u[i]) \) are the concomitants of \( z_{i0} \), i.e., the pair \( (y[i], y[i+1−1], u[i]) \) associated with the \( i^{th} \) ordered value \( z_{i0} \) obtained by arranging \( z_i \) (\( 1 \leq i \leq n \)) in increasing order of magnitude.

The second step is to replace \( g(u) \) and \( g(z) \) by linear functions

\[
g_1(u[i]) \equiv \alpha_{1i} - \beta_{1i} u[i], \quad g_2(z[i]) \equiv \alpha_{2i} + \beta_{2i} z[i] \tag{9}
\]

so that the differences between the two sides converge to zero as \( n \) becomes large. The coefficients \( \alpha_{1i}, \beta_{1i}, \alpha_{2i} \) and \( \beta_{2i} \) are obtained from the first two terms of a Taylor series expansion around the \( i^{th} \) population quantiles \( t_{i0} \) and \( t_{20} \). An IMSL (International Mathematics and Statistics Library) subroutine in FORTRAN is available to determine \( t_{i0} \) and \( t_{20} \). The resulting \( \alpha_{1i} \) and \( \beta_{1i} \) values are

\[
\alpha_{1i} = (1 + \exp( t_{1i}) + t_{1i} \exp( t_{1i})) / (1 + \exp( t_{1i})))^2, \quad \beta_{1i} = \exp( t_{1i}) / (1 + \exp( t_{1i}))^2, \tag{10}
\]

\[
\alpha_{2i} = (2/k) t_{2i}^2 / [(1 + (1/k) t_{2i})^2] \quad \text{and} \quad \beta_{2i} = (1 - (1/k) t_{2i}^2) / [(1 + (1/k) t_{2i})^2]. \tag{11}
\]

For \( \sigma \) to be real and positive, \( \beta_{2i} (1 \leq i \leq n) \) have to be positive. These coefficients have umbrella ordering, that is, they increase until the middle value and then decrease in a symmetric fashion. Therefore, if \( \beta_{2i} > 0 \) then all the \( \beta_{2i} \) are positive. For small \( p \), and large \( n \), however, \( \beta_{2i} \) can be negative. To rectify this situation if \( \beta_{2i} < 0 \), \( \alpha_{1i} \) is replaced by \( \alpha_{2i}^* \) and \( \beta_{2i} \) is replaced by \( \beta_{2i}^* \) [15], [16].

\[
\alpha_{2i}^* = (1/k) t_{2i}^2 / [(1 + (1/k) t_{2i})^2] \quad \text{and} \quad \beta_{2i}^* = 1 / [(1 + (1/k) t_{2i})^2]. \tag{12}
\]

The linear approximations (10)-(11) are incorporated in the likelihood equations. The solutions of the resulting equations are the MMLEs:

\[
\hat{\mu}_i = \bar{x}[i] - \frac{\Delta}{m_1} \hat{\sigma}_i; \quad \bar{x}[i] = \frac{1}{m_1} \sum_{i=1}^{n} \beta_{1i} x[i], \quad m_1 = \sum_{i=1}^{n} \beta_{1i}, \quad \Delta = \sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{n} (\alpha_{1i} - \frac{1}{b+1}), \quad \hat{\sigma}_i = (-B_1 + \sqrt{B_1^2 + 4nC_1}) / 2n, \tag{13}
\]

\[
B_1 = (b + 1) \sum_{i=1}^{n} \alpha_{1i} (x[i] - \bar{x}[i]), \quad C_1 = (b + 1) \sum_{i=1}^{n} \beta_{1i} (x[i] - \bar{x}[i])^2, \quad \hat{\gamma}_0 = \bar{v}[i] - \hat{\gamma}_1 \bar{u}[i], \quad \hat{\gamma}_1 = G + H \hat{\sigma}, \quad \hat{\phi} = K + D \hat{\sigma}
\]

\[
\hat{\sigma} = (B + \sqrt{B^2 + 4nC}) / 2n
\]

where

\[
v[i] = y[i] - \hat{\phi} y[i+1], \quad \bar{v}[i] = (1/m_2) \sum_{i=1}^{n} \beta_{2i} v[i], \quad \bar{u}[i] = (1/m_2) \sum_{i=1}^{n} \beta_{2i} u[i], \quad m_2 = \sum_{i=1}^{n} \beta_{2i}
\]

\[
G = \left[ \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2 \right], \quad H = \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i]) / \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2
\]

\[
K = \frac{1}{m_1} \left[ \sum_{i=1}^{n} \beta_{2i} y[i] w[i] \right] - \frac{1}{m_2} \sum_{i=1}^{n} \beta_{2i} y[i] \sum_{i=1}^{n} \beta_{2i} w[i], \quad D = \frac{1}{m_1} \sum_{i=1}^{n} \beta_{2i} y[i] - 1
\]

\[
\Delta_i = \left[ \sum_{i=1}^{n} \beta_{2i} y[i] - 1 - \frac{m_2}{m_2} \sum_{i=1}^{n} \beta_{2i} y[i] \right] / \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2
\]

\[
w[i] = y[i] - \hat{\gamma}_0 u[i]
\]

\[
B = \frac{2p}{k} \sum_{i=1}^{n} \beta_{2i} (v[i] - \bar{v}[i]) - \frac{\sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2 v[i] / \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2}
\]

\[
C = \frac{2p}{k} \sum_{i=1}^{n} \beta_{2i} (v[i] - \bar{v}[i]) - \frac{\sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2 v[i] / \sum_{i=1}^{n} \beta_{2i} (u[i] - \bar{u}[i])^2}
\]

**Computations:** Computation of the MMLEs \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) is straightforward. Using the LSEs \( \hat{\gamma}_0, \hat{\gamma}_1 \) and \( \hat{\phi} \), the estimated residuals

\[
\tilde{c}_i = y_i - \hat{\phi} y_{i+1} - \hat{\gamma}_0 - \hat{\gamma}_1 \hat{u}_i, \quad \hat{u}_i = (x_i - \hat{\mu}_i) / \hat{\sigma}_i (1 \leq i \leq n)
\]

are obtained. The concomitants \( (y[i], y[i+1], \hat{u}[i]) \) correspond to the \( i^{th} \) ordered value \( \tilde{c}_i \) (\( 1 \leq i \leq n \)) are used to calculate MMLEs from Equations (13). The LS estimators \( \hat{\gamma}_0, \hat{\gamma}_1 \) and \( \hat{\phi} \) are then replaced by \( \hat{\gamma}_0, \hat{\gamma}_1 \) and \( \hat{\phi} \), respectively and new concomitants are obtained. The revised MMLEs are computed.
from these new concomitants. The process is repeated one more time. Thus, the MMLEs are computed in two iterations besides computing the LSEs initially. Not more than two iterations are needed for the estimates to stabilize sufficiently.

The reason is that only the relative magnitudes (not necessarily the true values) of \( e_i \) \((1 \leq i \leq n)\) are needed to identify the concomitants. See also [9], [15], [16].

### III. RELATIVE EFFICIENCIES OF THE LSEs

The LS estimators are widely used irrespective of the nature of the underlying distribution and this can result in loss of efficiency. To evaluate the relative efficiencies of the LSEs

\[
\text{RE} = 100(\text{variance of MMLE})/(\text{variance of LSE}) \quad (14) 
\]

we have used simulations based on \([100,000/n]\) Monte Carlo runs. The MMLEs are known to be asymptotically equivalent to the MLEs.

**TABLE I**

VALUES OF (1) SIMULATED VARIANCE OF THE MMLE AND RELATIVE EFFICIENCY (RE) OF THE LSE; \(n=50\).

<table>
<thead>
<tr>
<th>( \mu_1 )</th>
<th>( \sigma_1 )</th>
<th>( \gamma_0 )</th>
<th>( \gamma_1 )</th>
<th>( \phi )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0.5, p=2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.103</td>
<td>0.015</td>
<td>0.317</td>
<td>0.038</td>
<td>0.0004</td>
</tr>
<tr>
<td>RE</td>
<td>92</td>
<td>74</td>
<td>93</td>
<td>76</td>
<td>99</td>
</tr>
<tr>
<td>b = 2.0, p=4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.046</td>
<td>0.013</td>
<td>0.272</td>
<td>0.035</td>
<td>0.0003</td>
</tr>
<tr>
<td>RE</td>
<td>98</td>
<td>81</td>
<td>84</td>
<td>81</td>
<td>76</td>
</tr>
<tr>
<td>b = 2.0, p=4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0.043</td>
<td>0.013</td>
<td>0.327</td>
<td>0.038</td>
<td>0.0004</td>
</tr>
<tr>
<td>RE</td>
<td>98</td>
<td>84</td>
<td>96</td>
<td>85</td>
<td>96</td>
</tr>
</tbody>
</table>

\[
\text{RE}(1) = 0.053 \quad 
\text{RE}(2) = 0.012 \quad 
\text{RE}(3) = 0.004 \quad 
\text{RE}(4) = 0.0002 \quad 
\text{RE}(5) = 0.002 \quad 
\text{RE}(6) = 0.0012 
\]

Given in Table 1 are the simulated values of the variances of the MMLEs and the relative efficiencies of the corresponding LSEs. Here, \( \gamma_0 = 1.0 \), \( \gamma_1 = 1.5 \) and \( \sigma_1 \) and \( \sigma \) are taken to be 1 without loss of generality. Simulated means are not given because the bias in all the estimators turned out to be negligibly small. We give values only for \( n=50 \), for conciseness. From the values of Table I, it can be seen that the MMLEs of \( \mu_1, \sigma_1, \gamma_0, \gamma_1, \phi \) and \( \sigma \) are more efficient than the corresponding LSEs.

### IV. ROBUSTNESS

In statistical applications, the objective is to obtain estimators which have certain optimal properties with respect to an assumed distribution. In practice, however, the deviations from an assumed distribution is very common. Thus the issue of robustness becomes important. An estimator is called robust if it is fully efficient (or nearly so) for an assumed distribution but maintains high efficiency for plausible alternatives or when a sample contains mild data anomalies (e.g., outliers); see for example, [5], [17].

To illustrate the robustness of the MML estimators we consider the distributions of \( X \) and \( e \) are GL(b=0.5, \( \sigma_1 \)) and LTS(p=2.5, \( \sigma \)), respectively (called as population models). As plausible alternatives, we consider the following which we will call sample models.

Misspecification of the distribution:

i) GL(b=1, \( \sigma_1 \)),

ii) LTS(p=3.5, \( \sigma \))

Outlier model:

iii) \((n-r) X_i \) come from GL(b=0.5, \( \sigma_1 \)) and independently \( r \) (we do not know which) come from GL(b=0.5, 3 \( \sigma_1 \)); \( r \in [0.5 \times 0.1 \text{n}] \).

iv) \((n-r) e_i \) come from LTS(p=2.5, \( \sigma \)) and independently \( r \) (we do not know which)
v) If the two samples \((n-r) X_i\) and \(e_i\) come from \(GL(b=0.5, \sigma_1)\) and \(GL(p=2.5, \sigma)\) and independently \(r X_i\) and \(e_i\) come from \(GL(b=0.5, \sigma_1)\) and \(GL(p=2.5, 3\sigma)\), respectively: \(r=[0.5+0.1n]\).

The random numbers generated were divided by appropriate constants to make their variances the same as the assumed distributions. The variances of MMLEs and the relative efficiencies of LSEs are given in Table II, bias in all the estimators being negligible is not reported.

It is seen that data anomalies have devastating effect on the LSEs. Clearly, the MMLEs are robust. The results are the same for numerous other alternatives, e.g., mixture and contamination models.

<table>
<thead>
<tr>
<th>Model (i)</th>
<th>Model (ii)</th>
<th>Model (iii)</th>
<th>Model (iv)</th>
<th>Model (v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_1)</td>
<td>(\sigma_1)</td>
<td>(\gamma_0)</td>
<td>(\gamma_l)</td>
<td>(\phi)</td>
</tr>
<tr>
<td>Var.</td>
<td>0.118</td>
<td>0.033</td>
<td>0.285</td>
<td>0.080</td>
</tr>
<tr>
<td>RE.</td>
<td>56</td>
<td>41</td>
<td>56</td>
<td>40</td>
</tr>
<tr>
<td>Model (i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.103</td>
<td>0.05</td>
<td>0.261</td>
<td>0.038</td>
</tr>
<tr>
<td>RE.</td>
<td>90</td>
<td>76</td>
<td>89</td>
<td>76</td>
</tr>
<tr>
<td>Model (ii)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.046</td>
<td>0.011</td>
<td>0.126</td>
<td>0.027</td>
</tr>
<tr>
<td>RE.</td>
<td>62</td>
<td>29</td>
<td>64</td>
<td>30</td>
</tr>
<tr>
<td>Model (iii)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.101</td>
<td>0.016</td>
<td>0.243</td>
<td>0.037</td>
</tr>
<tr>
<td>RE.</td>
<td>92</td>
<td>75</td>
<td>86</td>
<td>74</td>
</tr>
<tr>
<td>Model (iv)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.049</td>
<td>0.01</td>
<td>0.126</td>
<td>0.027</td>
</tr>
<tr>
<td>RE.</td>
<td>60</td>
<td>28</td>
<td>60</td>
<td>29</td>
</tr>
<tr>
<td>Model (v)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi = 0.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.123</td>
<td>0.034</td>
<td>0.335</td>
<td>0.080</td>
</tr>
<tr>
<td>RE.</td>
<td>55</td>
<td>40</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Model (i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.106</td>
<td>0.015</td>
<td>0.309</td>
<td>0.037</td>
</tr>
<tr>
<td>RE.</td>
<td>93</td>
<td>75</td>
<td>90</td>
<td>74</td>
</tr>
<tr>
<td>Model (ii)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.050</td>
<td>0.011</td>
<td>0.170</td>
<td>0.027</td>
</tr>
<tr>
<td>RE.</td>
<td>61</td>
<td>32</td>
<td>64</td>
<td>33</td>
</tr>
<tr>
<td>Model (iii)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.105</td>
<td>0.015</td>
<td>0.312</td>
<td>0.036</td>
</tr>
<tr>
<td>RE.</td>
<td>92</td>
<td>76</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>Model (iv)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var.</td>
<td>0.045</td>
<td>0.010</td>
<td>0.152</td>
<td>0.026</td>
</tr>
<tr>
<td>RE.</td>
<td>61</td>
<td>29</td>
<td>57</td>
<td>29</td>
</tr>
</tbody>
</table>

REFERENCES