Fitting probability distributions to economic growth: A maximum likelihood approach

Mahmood Ul Hassan∗ and Pär Stockhammar
Department of statistics, Stockholm University, Stockholm, Sweden

Appendix A. Probability distributions

The PDF of the Normal-Mixture (NM) distribution is

\[
\begin{align*}
    f_{\text{NM}}(\tilde{z}_t; \theta) &= w \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{(\tilde{z}_t - \mu_1)^2}{2\sigma_1^2} \right\} + (1 - w) \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left\{ -\frac{(\tilde{z}_t - \mu_2)^2}{2\sigma_2^2} \right\}
\end{align*}
\]

where \(0 \leq w \leq 1\) is the weight parameter and consists of the parameters \(\theta = (w, \mu_1, \mu_2, \sigma_1, \sigma_2)\).

The PDF of the Laplace (L) distribution is given by

\[
\begin{align*}
    f_{\text{L}}(\tilde{z}_t; \theta) &= \frac{1}{2\phi} \exp \left\{ -\frac{|\tilde{z}_t - \mu|}{\phi} \right\}
\end{align*}
\]

where \(\theta = (\mu, \phi)\), \(\mu \in \mathbb{R}\) is the location parameter and \(\phi > 0\) is the scale parameter.

The Normal-Laplace (NL) mixture distribution is specified by

\[
\begin{align*}
    f_{\text{NL}}(\tilde{z}_t; \theta) &= w \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{(\tilde{z}_t - \mu_1)^2}{2\sigma_1^2} \right\} + (1 - w) \frac{1}{2\phi} \exp \left\{ -\frac{|\tilde{z}_t - \mu|}{\phi} \right\}
\end{align*}
\]

for \(-\infty < \tilde{z}_t < \infty\) and for the parameters: \(-\infty < \mu < \infty\), \(0 \leq w \leq 1\), \(\phi > 0\) and \(\sigma > 0\).

The asymmetric Laplace (AL₁) distribution PDF is defined as

\[
\begin{align*}
    f_{\text{AL}_1}(\tilde{z}_t; \theta) &= \begin{cases} 
        \frac{1}{2\psi} \exp \left\{ \frac{\tilde{z}_t - \mu}{\nu} \right\} & \text{if } \tilde{z}_t \leq \mu \\
        \frac{1}{2\psi} \exp \left\{ \frac{\mu - \tilde{z}_t}{\nu} \right\} & \text{if } \tilde{z}_t > \mu
    \end{cases}
\end{align*}
\]

where again \(\mu\) is the location parameter. It consists of three parameters \(\theta = (\mu, \phi, \psi)\).

The Normal-Asymmetric Laplace-1 (NAL₁) distribution is described as

\[
\begin{align*}
    f_{\text{NAL}_1}(\tilde{z}_t; \theta) &= \frac{w}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(\tilde{z}_t - \mu)^2}{2\sigma^2} \right\} + (1 - w) \begin{cases} 
        \frac{1}{2\psi} \exp \left\{ \frac{\tilde{z}_t - \mu}{\nu} \right\} & \text{if } \tilde{z}_t \leq \mu \\
        \frac{1}{2\psi} \exp \left\{ \frac{\mu - \tilde{z}_t}{\nu} \right\} & \text{if } \tilde{z}_t > \mu
    \end{cases}
\end{align*}
\]

∗∗Corresponding author. Email: mahmood.ul-hassan@stat.su.se
where $\theta$ consists of the five parameters $(w, \mu, \sigma, \psi, \phi)$.

The PDF of Student’s t distribution is given by

$$f_{t}(z; \theta) = \frac{\nu^{-\frac{1}{2}} + \frac{1}{2} \left( \frac{1}{\nu + \frac{(z-\mu)^2}{\sigma^2}} \right)^{\frac{1+\nu}{2}}}{\text{Beta} \left[ \frac{\nu}{2}, \frac{1}{2} \right] \sigma}$$

where $\theta$ consists of the six parameters $(w, \mu, \sigma, \nu, \psi, \phi)$.

The asymmetric Laplace (AL) distribution is defined as

$$f_{AL}(z; \theta) = \frac{\nu^{-\frac{1}{2}} + \frac{1}{2} \left( \frac{1}{\nu + \frac{(z-\mu)^2}{\sigma^2}} \right)^{\frac{1+\nu}{2}}}{\text{Beta} \left[ \frac{\nu}{2}, \frac{1}{2} \right] \sigma} + (1-w) \begin{cases} \frac{1}{\nu \sigma} \exp \left\{ -\frac{z-\mu}{\nu} \right\} & \text{if } z \leq \mu \\ \frac{1}{\nu \sigma} \exp \left\{ -\frac{z-\mu}{\nu} \right\} & \text{if } z > \mu \end{cases}$$

where $\theta$ consists of the six parameters $(w, \mu, \nu, \sigma, \psi, \phi)$.

The asymmetric Laplace (AL) distribution PDF is given as

$$f_{AL}(z; \theta) = \begin{cases} \frac{\alpha \beta}{\alpha + \beta} \exp \left\{ \beta z \right\} & \text{if } z \leq 0 \\ \frac{\alpha \beta}{\alpha + \beta} \exp \left\{ -\alpha z \right\} & \text{if } z > 0 \end{cases}$$

where $\alpha$ and $\beta$ describes the left and right tail of the distribution. Both $\alpha$ and $\beta$ are scale parameters with $\alpha > 0$ and $\beta > 0$.

The Normal-Asymmetric Laplace-2 (NAL) distribution is specified as

$$f_{NAL}(z; \theta) = \frac{w}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(z-\mu)^2}{2\sigma^2} \right\} + (1-w) \begin{cases} \frac{\alpha \beta}{\alpha + \beta} \exp \left\{ \beta z \right\} & \text{if } z \leq 0 \\ \frac{\alpha \beta}{\alpha + \beta} \exp \left\{ -\alpha z \right\} & \text{if } z > 0 \end{cases}$$

where $\theta$ consists of the five parameters $(w, \mu, \sigma, \alpha, \beta)$.

The Student’s t-Asymmetric Laplace-2 (TAL) distribution is defined as

$$f_{TAL}(z; \theta) = \frac{\nu^{-\frac{1}{2}} + \frac{1}{2} \left( \frac{1}{\nu + \frac{(z-\mu)^2}{\sigma^2}} \right)^{\frac{1+\nu}{2}}}{\text{Beta} \left[ \frac{\nu}{2}, \frac{1}{2} \right] \sigma} + (1-w) \begin{cases} \frac{1}{\nu \sigma} \exp \left\{ -\frac{z-\mu}{\nu} \right\} & \text{if } z \leq \mu \\ \frac{1}{\nu \sigma} \exp \left\{ -\frac{z-\mu}{\nu} \right\} & \text{if } z > \mu \end{cases}$$

where $\theta$ consists of the six parameters $(w, \mu, \nu, \sigma, \psi, \phi)$.

**Appendix B. Accuracy Measures**

- **RMSE**: $\sqrt{\frac{1}{1000} \sum_{i=1}^{1000} \left[ f_K(\tilde{z}_i) - \hat{f}(\tilde{z}_i) \right]^2}$
- **MdAPE**: $\text{median} \left( \frac{100 \times |f_K(\tilde{z}_i) - \hat{f}(\tilde{z}_i)|}{f_K(\tilde{z}_i)} \right)$
- **sMdAPE**: $\text{median} \left( \frac{200 \times |f_K(\tilde{z}_i) - \hat{f}(\tilde{z}_i)|}{f_K(\tilde{z}_i) + \hat{f}(\tilde{z}_i)} \right)$
\[ MASE = \frac{1}{1000} \frac{\sum_{i=2}^{1000} |f_K(\tilde{z}_i) - \hat{f}(\tilde{z}_i)|}{1000} \] (14)

where \( \hat{f}(\tilde{z}_i) \) is the hypothetical distribution and \( f_K(\tilde{z}_i) \) is the Kernel function of the empirical distribution.

### Appendix C. Goodness of fit tests

\[ KS = \max|\hat{F}(\tilde{z}_i) - F(\tilde{z}_i)| \] (15)

\[ CVM = \frac{1}{12n} + \sum_{i=1}^{n} \left( \frac{2i-1}{2n} - F(\tilde{z}_i) \right)^2 \] (16)

\[ U^2 = CVM - n \left( \hat{F}(\tilde{z}_i) - \frac{1}{2} \right)^2 \] (17)

\[ AD = -n - \sum_{i=1}^{n} \left[ \frac{2i-1}{n} \left( \log(1 - F(\tilde{z}_{n-i+1})) + \log(F(\tilde{z}_i)) \right) \right] \] (18)

\[ V = \max \left[ \frac{i}{n} - F(\tilde{z}_i) \right] + \max \left[ F(\tilde{z}_i) - \frac{i-1}{n} \right] \] (19)

where \( \hat{F}(\tilde{z}_i) \) is the empirical CDF of data and \( F(\tilde{z}_i) \) is the theoretical CDF of the distribution.

### Appendix D. Mathematica Code

The mixture distributions in Mathematica are defined as:

Normal-Mixture distribution
\[
\text{Normal-Mixture distribution} = \text{MixtureDistribution} \left\{ \{p, 1-p\}, \{\text{NormalDistribution}[\mu_1, \sigma_1]\}, \{\text{NormalDistribution}[\mu_2, \sigma_2]\} \right\};
\]

Normal-Asymmetric Laplace-1 distribution
\[
\text{Normal-Asymmetric Laplace-1 distribution} = \text{Piecewise}\{\{\text{Exp}[\alpha \phi], x \leq \mu\}, \{\text{Exp}[\beta \phi], x > \mu\}\};
\]

Normal-Asymmetric Laplace-2 distribution
\[
\text{Normal-Asymmetric Laplace-2 distribution} = \text{MixtureDistribution} \left\{ \{p, 1-p\}, \{\text{NormalDistribution}[\mu, \sigma]\}, \{\text{NormalDistribution}[\mu, \sigma, \alpha, \beta]\} \right\};
\]

Student’s t-Asymmetric Laplace-1 distribution
\[
\text{Student’s t-Asymmetric Laplace-1 distribution} = \text{MixtureDistribution} \left\{ \{p, 1-p\}, \{\text{StudentTDistribution}[\mu, \sigma, \nu]\}, \{\text{NormalDistribution}[\mu, \sigma]\} \right\};
\]

Student’s t-Asymmetric Laplace-2 distribution
\[
\text{Student’s t-Asymmetric Laplace-2 distribution} = \text{MixtureDistribution} \left\{ \{p, 1-p\}, \{\text{StudentTDistribution}[\mu, \sigma, \nu]\}, \{\text{NormalDistribution}[\mu, \sigma]\} \right\};
\]
\( \text{ALF}_2 = \text{Piecewise}\{\{\alpha \beta \exp[\beta x]/(\alpha + \beta), x \leq 0\}, \alpha \beta \exp[-\alpha x]/(\alpha + \beta)\}; \)
\( \text{AL}_2 = \text{ProbabilityDistribution}\{\text{ALF}_2, \{x, -\infty, \infty\}, \text{Assumptions} \to \{\alpha > 0, \beta > 0\} \}; \)
\( \text{TAL}_2 = \text{MixtureDistribution}\{\{p, 1 - p\}, \{\text{StudentTDistribution}[\mu, \sigma, \nu], \text{AL}_2\}\}; \)

Import the data in Mathematica
\(\text{us} := \text{Import["GDPZ.xlsx", \{"xlsx", "Data", 1\}};\)

For optimization, the \text{NMaximize} function is used. In \text{NMaximize}, we need to give the ML function of the distributions, parameter constrains and starting value. We have used the MLE estimate of the parameters of each distribution in mixture as initial value in \text{NMaximize}. The code for the accuracy measures given below is only Normal-Mixture distribution. Other mixtures are treated similarly and is available upon request

\( \text{result[data, w]} := \)
\( \text{Module}\{n = \text{Length[data]}, \text{sdata} = \text{Sort[data]}, m = \text{Mean[data]}, s2 = \text{Variance[data]}, \}
\( \text{optimalMixture} = \text{Quiet}\{\text{Check}[\text{NMaximize}\{\text{Total}[\log[\text{PDF}[\text{NN}, \text{sdata}]], \}
\( \text{Join}\{\{\sigma_1 > 0\}, \{\sigma_2 > 0\}, \{0 < p < 1\}\}, \text{Join}\{\{\{\sigma_1, \text{Sqrt}[s2] - .01, \text{Sqrt}[s2] + .01\}, \{\mu_1, \text{Mean[data]} - .1, \text{Mean[data]} + .1\}, \}
\( \{\mu_2, \text{Mean[data]} - .1, \text{Mean[data]} + .1\}\}, \{\{p, w - .1, w + .1\}\}], \text{None}\}\}, \}
\( \text{quantilesMixture} := \text{Quiet}\{\text{Table}[x / \text{FindRoot}[\text{CDF}[\text{NN}, x]/\text{optimalMixture}[2]] == i/(n + 1.), \{x, \text{sdata}[i], \text{sdata}[i] + .05\}, \{i, n\}]\}; \)
\( \text{RMSE1} := \text{Row}\{\text{RMSE1} = \text{Chop}\[\text{RootMeanSquare}[\text{quantilesMixture} - \text{sdata}], 10^{-6}]\}; \)
\( G = \text{SmoothKernelDistribution[data]}; \)
\( \text{div} = \text{FindDivisions}[\{\text{Min[data]}, \text{Max[data]}\}, 1000]; \)
\( a = \text{PDF}[G, \text{div}]; b = \text{PDF}[\text{NN}, \text{div}]/\text{optimalMixture}[2]; \)
\( \text{RMSE} = \text{Sqrt}[\text{Total}[(a - b)^2]/1000]]; \)
\( \text{MDAPE} = \text{Median}[100 \ast \text{Abs}[a - b]]/(a)\];
\( \text{sMDAPE} = \text{Median}[(200 \ast \text{Abs}[a - b])/(a + b)]; \)
\( \text{MASE} = (1/1000) \ast \text{Total}[\text{Abs}[a - b]]/(1/999) \ast \text{Total}[\text{Abs}[\text{Differences}[a]]] \);
\( \text{rr} := \{\text{optimalMixture}, \text{RMSE1}, \text{RMSE}, \text{MDAPE}, \text{sMDAPE}, \text{MASE}\}; \)
\( \text{rr} \)

For the goodness of fit tests, we need to specify the distribution with estimated parameters obtain from above optimization code in the \text{DistributionFitTest} function in Mathematica

\( H = \text{DistributionFitTest[data, \text{NN}, \{\text{HypothesisTestData}\}}; \)
\( H[\{\text{TestDataTable}, \text{All}\}] \)

Reference

H. Ruskeep and M.A. Ghorbani, \textit{Maximum Likelihood Estimation of Ordinary and Finite Mixture Distributions}. From the Wolfram Demonstrations Project. Retrieved from http://demonstrations.wolfram.com/MaximumLikelihoodEstimationOfOrdinaryAndFiniteMixtureDistrib/