Generalized predictive recursion maximum likelihood for robust mixture regression

Pradon Sureephong$^1$, Woraphon Yamaka$^{2,3}$ and Paravee Maneejuk$^2$

$^1$International College of Digital Innovation, Chiang Mai University, Chiang Mai, Thailand
$^2$Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
$^3$E-mail: woraphon.econ@gmail.com

Abstract In the application of econometric model, the error distribution is unknown and is not easily to specify in the likelihood function. In some situations, there might exist a mixture distribution in the errors and thus the traditional estimation method would probably yield a biased result. In this study, this mixture distribution of the error term is taken into account and the generalized semiparametric estimation is presented and applied in regression model. We also use an experiment study and the real application analysis to check the performance of this estimator in regression model. The performance of this estimation is then compared with that of conventional Least Squares method in the real data analysis.

1. Introduction

The linear regression model which has received considerable attention in many financial applications can be written as

$$Y_t = X_t \beta + \epsilon_t,$$  \hspace{1cm} (1)

where $Y_t$ is a $T \times 1$ vector of dependent variables, $X_t$ is $T \times K$ matrix of independent variables, and $\epsilon_t$ is $T \times 1$ vector of errors with density function. In the conventional estimation, one assumes that the errors have a normal density with mean zero and variance $\sigma^2$.

In the literature, the regression model could be estimated by various classical estimators such as Conditional Least Squares (CLS), Maximum Likelihood and Bayesian estimations (see [1, 2]). These classical estimators assume that the error has normal density and is free from nuisance parameters. Under these assumptions, the classical methods can provide the optimal solution for estimated parameters when the error density happens to be normal. The main reason for using a normal distribution is especially in simplicity of derivation and desirable properties of the estimated parameters which are consistent, efficient, and asymptotically normal under mild conditions, for more details (see [3, 4]). Nevertheless, if the error density is non-normal, these classical estimators will not provide a reliable and efficient result [5]. In theory, the normal distribution shows a symmetric, non-excess kurtosis and skewness; and these may not be true in the real world, especially in the financial time series data, as it often exhibits heavy tail and asymmetric distribution. Thus, when we have a non-normal error density function, alternative non-normal density estimators, such as nonparametric Maximum Likelihood or Bayesian, can be adopted to deal with these problems. However, in the most recent work of [6] emphasized that nonparametric Maximum Likelihood and Bayesian are...
complicated, leading to difficulty solving these problems. They mentioned that numerical optimization of ML is not stable and is computationally expensive, while nonparametric Bayes also consumes computational time since each marginal likelihood evaluation requires its own MCMC iterations and optimization requires several such iterations. Therefore, a general scale mixture density for the error distribution with Predictive Recursion marginal likelihood $L_{PR}(\theta)$ is developed and estimated. To do the optimization, [6] developed a hybrid Predictive Recursion-Expected Maximization (PR–EM) algorithm and applied it in the regression framework. This method is then applied to estimate the Threshold model of [7] by [8], they confirmed that this method outperforms the LS and Bayesian methods when the error has normal scale mixture distribution. However, if the error happens to have other mixture distribution such as student, skewed normal and skewed student-t scale mixture distribution, we expect that the accuracy of the PR-EM based normal mixture density is lost. This paper attempts to fill this gap in the literature by applying various combinations of mixture distribution. We propose more general form of PR-EM by applying a student-t, skewed normal and skewed student-t distributions to construct the mixture density in PR-EM. General speaking, we generalize PR-EM algorithm by replacing normal distribution with student-t, skewed normal or skewed student-t distributions, to construct the mixture density in PR-EM. Thus, the estimation becomes more flexible for solving almost all possible mixture distributions which have not been studied in earlier literature.

The main contribution of this study is to further develop and extend the PR-EM algorithm of [6] to have a mixture student-t, skewed normal or skewed -student-t. To our knowledge, this study is the first attempt to estimate regression model using PR-EM based student, skewed normal or skewed student-t method. As a consequence, the model is more flexible to overcome the difficulty arising when a general scale mixture of various parametric with unknown mixing distribution exists for the error term.

The rest of this paper proceeds as follows: Section 2 briefly presents the estimation approach to regression model. Section 3 proposes a simulation experiment to assess and illustrate the extended PR-EM algorithm as well as compare it with the conventional PR-EM. Section 4 employs the proposed method to real data while conclusions are given in Section 5.

2. Methodology

2.1. Constructing Predictive Recursion marginal likelihood for regression model

The study applies the PR algorithm to the error to obtain the mixture density, the PR marginal likelihood for $\beta$ is written as:

$$L_{PR}(\beta) = \prod_{t=1}^{T} f_{t-1,\beta}(\epsilon_t) = \prod_{t=1}^{T} \int F(Y_t - X_t \beta | 0, u^2) \Psi(u) du$$  

(2)

where $\epsilon_t$ is the error term, $f_{t-1,\beta}(\cdot)$ is the predictive parametric density at time $t-1$. $\Psi$ is an unspecified mixing distribution in normal distribution which is supported on $u = (0, \infty)$. Thus, the density function equation (2) contains both parametric distribution function $F(\cdot)$, and other unspecified mixing distribution. Note that our study aims to propose a more general form of PR-EM by replacing normal distribution with student-t, skewed normal or skewed student-t distributions, to construct the mixture density in PR-EM, hence, the distribution $F(\cdot)$ can be assumed to be either student-t, skewed normal or skewed-student-t. To estimate the parameter $\beta$, we can use the numerical optimization method to maximize the likelihood function. When the likelihood reaches its maximum, we can obtain the optimal parameter set. Nevertheless, the study employs hybrid PR–EM algorithm of [6] to be an optimization method for estimating the model since it was proved to have an efficiency in
2.2. Hybrid PR-EM algorithm

Reference [6] proposed a hybrid PR–EM algorithm taking an advantage of the latent scale parameter structure $U_1,...,U_T$ in the mixture model. To fit this density function, we apply the PR algorithm which is a recursive estimation of mixing distribution in nonparametric mixture model to estimate the mixing density (see [10]). By taking a logarithm in equation 2, the study gets

$$\log L_{PR}(\beta) = \sum_{i=1}^{T} \log \left[ F(Y_i - X_i \beta | 0, u^2) \Psi_{t,\beta}(u) du \right]$$

$$= \sum_{i=1}^{T} \log F(Y_i - X_i \beta | 0, U_i^2) - \sum_{i=1}^{T} \log \left[ \frac{F(Y_i - X_i \beta | 0, U_i^2)}{f_{t,\beta}(Y_i - X_i \beta)} \right].$$

(3)

Then, integrate out $U_i$ with respect to the density $\Psi_{t,\beta}(u)$, thus

$$\log L_{PR}(\beta) = \sum_{i=1}^{T} \log \left[ F(Y_i - X_i \beta | 0, u^2) \right] \Psi_{t,\beta}(u) du - \sum_{i=1}^{T} \log \left[ \frac{F(Y_i - X_i \beta | 0, u^2)}{f_{t,\beta}(Y_i - X_i \beta)} \right] \Psi_{t,\beta}(u) du,$$

(4)

where

$$\Psi_{t,\beta}(u) = \frac{F(Y_i - X_i \tilde{\beta} | 0, u^2) \Psi_{t,\beta}(u) du}{f_{t,\beta}(Y_i - X_i \beta)},$$

(5)

and $\tilde{\beta}$ is the estimated parameter in the previous estimation and $\beta$ is the estimated parameter in current estimation. In practice, we can rewrite equation (4) as $Q_1(\beta | \tilde{\beta}) + Q_2(\beta | \tilde{\beta})$ which corresponds to weighted Least Squares estimation where $Q_1(\beta | \tilde{\beta})$ is computed by

$$Q_1(\beta | \tilde{\beta}) = \left[ 0.5 \sum_{i=1}^{T} w_i (e_i)^2 + \text{constant} \right],$$

(6)

where $w_i$ is a weight depending on mixing distribution for error term which is defined as

$$w_i = \int u^2 \Psi_{t,\tilde{\beta}}(u) du.$$  

(7)

In sum, this study employs PR-EM algorithm to maximize $Q_1 + Q_2$ which corresponds to weighted least squares minimization. The PR-EM algorithm is done by the following steps:

- **Given the initial value for** $\beta^0$, $U$, $\Psi^0$, and $w_1, ..., w_T$.
- **Expectation Step**: Estimate $\tilde{w}_1, ..., \tilde{w}_T$ using the ratio of $L_{PR,i}(\beta | e_i)$ and $L_{PR,i-1}(\beta | e_{i-1})$.
- **Maximization Step**: Updating the vector unknown parameters $\beta^t+1$ to maximize only $Q_1(\beta | \tilde{\beta})$, which corresponds to $\tilde{\beta} = (X^T \tilde{\Omega} X)^{-1} X^T Y$, where $\tilde{\Omega}$ is a diagonal matrix of the weights $\tilde{w}_1, ..., \tilde{w}_T$. For the $Q_2(\beta | \tilde{\beta})$, it is not needed to work in the Maximization step. The justification for the property of PR–EM, which involves some investigation into $Q_2(\beta | \tilde{\beta})$, refer to Section 3.3. in [6].
Repeat step 2 and 3 and stop when \( \| \beta^{t+1} - \beta^t \| < 0.00001 \), and thus we have the final parameter estimates \( \tilde{\beta} \) and \( \tilde{w}_1, \ldots, \tilde{w}_T \).

3. Simulation study
To assess the accuracy of our generalized PR-EM, we conduct an experiment study. Also, the purpose of this experiment is to compare the performance of our estimation under a variety of mixture error distributions as well as the conventional PR-EM of [6] and traditional least squares (LS) methods. The simulation is now considered the simple linear regression. The study generates simulated data from the following model specification:

\[
Y_i = 1 + 2X_i + \varepsilon_i, \quad \varepsilon_i \sim \text{distribution}.
\]

where the error term \( \varepsilon_i \) has been generated independently from the following distributions

- Normal scale mixture with respect to a uniform distribution \( \varepsilon \sim \mathcal{N}(0,1) - \text{U}(0,1) \)
- Skewed normal scale mixture with respect to a uniform distribution \( \varepsilon \sim \text{sN}(0,1) - \text{U}(0,1) \)
- Student-t scale mixture with respect to a uniform distribution \( \varepsilon \sim \mathcal{T}(0,1) - \text{U}(0,1) \)
- Skewed student-t scale mixture with respect to a uniform distribution \( \varepsilon \sim \text{sT}(0,1) - \text{U}(0,1) \)

Table 1. MSE results of the Monte Carlo experiment.

| Methods                  | \( T = 100 \)       | \( T = 200 \)       |
|--------------------------|---------------------|---------------------|
| \( \varepsilon \sim \mathcal{N}(0,1) - \text{U}(0,1) \) |                     |                     |
| PREM-normal mix          | \textbf{0.1202} 0.0032 | \textbf{0.1028} 0.0026 |
| PREM-skewed normal mix   | 0.2441 0.0032       | 0.2383 0.0022       |
| PREM - student-t mix     | 0.1931 0.0037       | 0.1834 0.0025       |
| PREM - skewed student-t mix | 0.2523 0.0035 | 0.2417 0.0026       |
| LS                       | 0.1225 0.0057       | 0.1203 0.0039       |
| \( \varepsilon \sim \text{sN}(0,1) - \text{U}(0,1) \) |                     |                     |
| PREM-normal mix          | 0.2021 0.0026       | 0.1972 0.0008       |
| PREM-skewed normal mix   | \textbf{0.1291} 0.0022 | \textbf{0.1234} 0.0008 |
| PREM - student-t mix     | 0.1875 0.0026       | 0.1792 0.0011       |
| PREM - skewed student-t mix | 0.2450 0.0027 | 0.2405 0.0010       |
| LS                       | 0.1324 0.0039       | 0.1244 0.0026       |
| \( \varepsilon \sim \mathcal{T}(0,\sigma = 1, \nu = 4) - \text{U}(0,1) \) |                     |                     |
| PREM-normal mix          | 0.2032 0.0026       | 0.1981 0.0009       |
| PREM-skewed normal mix   | 0.2376 0.0028       | 0.2334 0.0010       |
| PREM - student-t mix     | \textbf{0.1299} 0.0027 | \textbf{0.1818} 0.0009 |
| PREM - skewed student-t mix | 0.2446 0.0030 | 0.2395 0.0017       |
| LS                       | 0.1318 0.0051       | 0.1961 0.0016       |
| \( \varepsilon \sim \text{sT}(0,\sigma = 1, \nu = 4, \gamma = 1.5) - \text{U}(0,1) \) |                     |                     |
| PREM-normal mix          | 0.1864 0.0031       | 0.1846 0.0015       |
| PREM-skewed normal mix   | 0.2246 0.0026       | 0.2223 0.0017       |
| PREM - student-t mix     | 0.1727 0.0033       | 0.1717 0.0015       |
| PREM - skewed student-t mix | 0.1276 0.0028 | \textbf{0.1659} 0.0013 |
| LS                       | 0.1280 0.0042       | 0.2205 0.0019       |
We consider different sample sizes, say, $T = 100, 200$. For each sample size, the study generates $m = 100$ datasets. For all scenarios, the Mean Square Error (MSE) and absolute Bias of $m$ replications with each of the method are obtained to make the comparison. The computing language is written in R software and run on a desktop computer with 1.7GHz IntelCore i3 with 4GB RAM capacity. According to table 1-2, the overall pattern is similar in all scenarios. We can observe that if the estimator is correctly specified, the PR-EM based correct distribution always outperforms excellently and provides the best fit estimation. In addition, we can see a pattern of convergence to zero of the MSE and bias when $T$ increases. As a general rule, we can say that bias and MSE tend to approach zero when the sample size increases, indicating that the estimates based on the generalized PR-EM have good asymptotic properties. In the computation aspect, among all methods for estimating the unknown parameter, the computational cost of PR-EM is higher when compared to the LS, however, it can greatly enhance the parameter estimation and thereby yielding more accurate and reliable results.

4. Real data analysis
Consider a simple linear regression problem, we study the ways Chinese domestic factors (such as Gross Domestic Product (GDP), monetary based (M1), and domestic credit (DC) affected their

| Methods                  | $\beta_0$  | $\beta_1$  | $\beta_0$  | $\beta_1$  |
|--------------------------|------------|------------|------------|------------|
| $\varepsilon \sim \mathcal{N}(0,1) - U(0,1)$            |            |            |            |            |
| PREM-normal mix          | 0.3430     | 0.0011     | 0.3418     | 0.0011     |
| PREM-skewed normal mix   | 0.4866     | 0.0024     | 0.3486     | 0.0012     |
| PREM - student-t mix     | 0.4302     | 0.0063     | 0.4264     | 0.0019     |
| PREM - skewed student-t mix | 0.4954 | 0.0026     | 0.4918     | 0.0016     |
| LS                       | 0.3431     | 0.0175     | 0.3421     | 0.0047     |
| $\varepsilon \sim s\mathcal{N}(0,1) - U(0,1)$            |            |            |            |            |
| PREM-normal mix          | 0.4454     | 0.0113     | 0.4425     | 0.0024     |
| PREM-skewed normal mix   | 0.3462     | 0.0088     | 0.2832     | 0.0023     |
| PREM - student-t mix     | 0.4284     | 0.0098     | 0.4214     | 0.0042     |
| PREM - skewed student-t mix | 0.4915 | 0.0103     | 0.4884     | 0.0022     |
| LS                       | 0.3579     | 0.0047     | 0.3459     | 0.0034     |
| $\varepsilon \sim T(0, \sigma = 1, \nu = 4) - U(0,1)$    |            |            |            |            |
| PREM - normal mix        | 0.4477     | 0.0043     | 0.4437     | 0.0019     |
| PREM-skewed normal mix   | 0.4847     | 0.0057     | 0.4817     | 0.0020     |
| PREM - student-t mix     | 0.3324     | 0.0042     | 0.3247     | 0.0014     |
| PREM - skewed student-t mix | 0.4917 | 0.0053     | 0.4880     | 0.0028     |
| LS                       | 0.3561     | 0.0068     | 0.3516     | 0.0022     |
| $\varepsilon \sim sT(0, \sigma = 1, \nu = 4, \gamma = 1.5) - U(0,1)$ | | | | |
| PREM - normal mix        | 0.4279     | 0.0122     | 0.4284     | 0.0093     |
| PREM-skewed normal mix   | 0.4709     | 0.0132     | 0.4709     | 0.0098     |
| PREM - student-t mix     | 0.4114     | 0.0121     | 0.4056     | 0.0059     |
| PREM - skewed student-t mix | 0.3348 | 0.0120     | 0.4017     | 0.0058     |
| LS                       | 0.3497     | 0.0182     | 0.4504     | 0.0093     |

We consider different sample sizes, say, $T = 100, 200$. For each sample size, the study generates $m = 100$ datasets. For all scenarios, the Mean Square Error (MSE) and absolute Bias of $m$ replications with each of the method are obtained to make the comparison. The computing language is written in R software and run on a desktop computer with 1.7GHz IntelCore i3 with 4GB RAM capacity.
exchange market pressure (EMP). We collect quarterly data from 1999 Q1 through 2017 Q1. The data is collected from the Federal Reserve Economic Data (FRED) provided by the Federal Reserve Bank of St. Louis. We have considered the following regression model for exchange market pressure:

\[ EMP_t = \beta_0 + \beta_1 GDP_t + \beta_2 M1_t + \beta_3 DC_t \]  

(9)

These data are transformed to be growth rate in order to avoid the nonstationary data and spurious regression. We then fit the model using PR-EM based normal, student-t and skewed student-t mixture distributions as was described in Section 2. To compare them, we consider the Bayesian information criterion (BIC), Mean Square Error (MSE), and the value of the estimated log-likelihood function. Table 2 presents the obtained results for making a comparison. According to table 3, we can observe that the skewed student-t mixture provides the lowest MSE and BIC, indicating the best mixture distribution is the skewed student-t. In addition, we compare the best fit PR-EM method with the least squares (LS) method to confirm the performance of the PR-EM over the conventional method and the result reveals the superiority of PR-EM based skewed student-t over LS in terms of the lower MSE.

Figure 1. Scatterplot with fitted regression lines under various methods.

Under the skewed student-t based method, table 4 presents the parameter estimates for linear regression model. The result shows that \( M1 \) and \( DC \) have statistical significant effect on EMP. The fitted regression line of each method is plotted in figure 1.
Table 3. Model comparison.

|          | Normal mixture | Skewed normal mixture | Student-t mixture | Skewed student-t mixture | LS     |
|----------|----------------|-----------------------|-------------------|--------------------------|--------|
| BIC      | 469.1234       | 475.7575              | 476.2146          | 462.0504                 |        |
| MSE      | 19.7800        | 20.4269               | 19.7918           | 19.6619                  | 19.800 |
| LL       | -225.9808      | -227.1526             | -227.3812         | -218.1538                |        |

Table 4. Parameter estimation under PR-EM with mixture skewed student-t.

| Parameter | Estimate | Std.Error | p-value |
|-----------|----------|-----------|---------|
| $\beta_0$ | -3.3845  | 0.2862    | 0.0000  |
| $\beta_1$ | -0.9995  | 0.9762    | 0.3094  |
| $\beta_2$ | 22.0362  | 8.4825    | 0.0115  |
| $\beta_3$ | 54.9696  | 4.3143    | 0.0000  |

5. Conclusions

Our proposed method generalizes the recent work of [6]. We generalize PR-EM algorithm by introducing various distribution, namely student-t, skewed normal and skewed student-t distributions, to construct the mixture density in PR-EM. Thus, the estimation becomes more flexible of capturing almost all possible mixture which have not been studied in earlier literature. This robust estimation also simultaneously accommodates asymmetry and heavy tails and thereby allowing researcher to analyze data in a flexible way from different areas.

We conduct a simulation and experiment study and the assessment of the PR-EM was based on the average of the MSE of parameters in the framework of a linear regression. The numerical study shows that the PR-EM algorithm exhibits the best performance and has good finite sample performance when it estimates the correct model. Last but not least, we apply this approach to study the effect of Chinese domestic factors such as Gross Domestic Product (GDP), monetary based (M1), and domestic credit (DC) on exchange market pressure (EMP) and find that PR-EM algorithm based on skewed student-t is perform better when compared with other mixing distributions including LS.

References

[1] Gelman A, Carlin J B, Stern H S, Dunson D B, Rubin D B and Vehtari A 2004 Bayesian data analysis
[2] Van de Geer S 1990 Estimating a regression function The Annals of Statistics 907-924
[3] Amemiya T 1985 Advanced econometrics (Harvard university press)
[4] Andrews D F and Mallows C L 1974 Scale mixtures of normal distributions Journal of the Royal Statistical Society Series B (Methodological) 99-102
[5] Pipitpojanakarn V Mancejuk, P, Yamaka W, Sriboonchitta S 2017 How does economic growth affect the well-being in Asia? International Journal of Economic Research 14(6)
[6] Martin R and Han Z 2016 A semiparametric scale-mixture regression model and predictive recursion maximum likelihood Computational Statistics & Data Analysis 94 75-85
[7] Tong H 1983 Threshold Models in Non-linear Time Series Analysis Lecture Notes in Statistics (New York Springer)
[8] Pastipipatkul P, Yamaka W & Sriboonchitta S 2017 Predictive Recursion Maximum Likelihood of Threshold Autoregressive Model In Robustness in Econometrics (Springer Cham) pp 349-362
[9] Martin R and Tokdar S T 2011 Semiparametric inference in mixture models with predictive recursion marginal likelihood Biometrika 98(3) 567-582
[10] Newton M A 2002 On a nonparametric recursive estimator of the mixing distribution Sankhya Ser. A 64(2) 306-322