LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Thamerus:

Fitting a Finite Mixture Distribution to a Variable Subject to Heteroscedastic Measurement Error

Sonderforschungsbereich 386, Paper 48 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Fitting a Finite Mixture Distribution to a Variable Subject to Heteroscedastic Measurement Error

Markus Thamerus, SFB 386

Institut für Statistik, Universität München

Abstract

We consider the case where a latent variable X cannot be observed directly and instead a variable $W = X + U$ with an heteroscedastic measurement error U is observed. It is assumed that the distribution of the true variable X is a mixture of normals and a type of the EM algorithm is applied to find approximate ML estimates of the distribution parameters of X.

Keywords: Measurement error; EM algorithm; Finite mixture distribution

1 Introduction

It is well known that measurement errors in the covariates of a regression model lead to biased parameter estimates. Most likelihood based methods that adjust for this effect treat the true predictor X as a stochastic variable and require an assumption about the marginal distribution of X, see e.g. Carroll, Ruppert and Stefanski (1995). Usually an unimodal distribution is assumed and without external knowledge its parameters have to be estimated from the observed data. But if the observations suggest that the underlying statistical population of interest decomposes into several parts, the estimation problem is more flexibly addressed by the assumption of a mixture distribution. One example of specifying a mixture distribution in an errors-in-variables model can be found in Küchenhoff (1995).

Here, we assume that the distribution of the latent variable X is a finite mixture of normal distributions and that the observable variable W is related to X by an additive error U independent of X. Whereas the assumption of unbiased errors (i.e. $\mathrm{E}(U) = 0$) is useful for most applications, assuming homoscedasticity among the errors is not. As an example consider an environmental research project where data are collected through different monitoring devices each operating with its individual precision.

In such cases, a single observed data point from one device is very often taken as the mean of several measurements of the outcome variable. A heteroscedastic error structure therefore accounts also for sampling errors made by aggregation. If enough information about the measurement process is provided, we are able to determine the heteroscedastic error variances.

The methods to estimate the parameters of a mixture distribution has been the subject of a large body of literature and a very extensive survey on that topic can be found in Redner and Walker (1984). The aim of this paper is to derive a general procedure for estimating the parameters of the mixture distribution of X when the observed W is subject to heteroscedastic measurement error and the error variances are known. First, the error model and the involved types of distributions are stated. Then, we make use of the EM algorithm to find approximate ML estimates and briefly address how the information matrix associated with the parameter estimates can be derived. Finally, the results of a small simulation study are presented and in addition an empirical example is given.

2 The Error Model

Let the observed variables W_i follow a structural measurement error model with heteroscedastic error variances; that is the true variables X_i can only be observed with additive errors U_i that are assumed to be normal with mean zero and known variances σ_i^2:

$$W_i = X_i + U_i \text{ with } U_i \sim N(0, \sigma_i^2) \text{ for } i = 1, \ldots, n. \tag{1}$$

The errors U_i are mutually independent and independent of the true variables X_i for $i = 1, \ldots, n$. The distribution of the i.i.d variables $X_i, i = 1, \ldots, n$ is parametrically modeled as a mixture of normals with density

$$p(x_i \mid \Phi) = \sum_{k=1}^{m} \alpha_k \, p_k(x_i \mid \psi_k) \text{ for } k = 1, \ldots, m$$
$$\text{with } \alpha_k > 0 \text{ for } k = 1, \ldots, m \text{ and } \sum_{k=1}^{m} \alpha_k = 1. \tag{2}$$

Each component p_k itself is a normal density function with associated parameter vector $\psi_k = \binom{\mu_k}{\sigma_k^2}$. The parameter vector Φ of the mixture can therefore be denoted by $\Phi = (\alpha_1, \ldots \alpha_m, \psi_1^{'}, \ldots, \psi_m^{'})^{'}$. A natural interpretation of finite mixture densities is that the population under study is a mixture of m components with associated component densities $\{p_k\}$ and mixing proportions $\{\alpha_k\}$. Usually the observations $\{w_i\}$ are unlabeled in a sense that there is no information about their component population of origin. The objective is to find the maximum likelihood estimator for the parameter vector Φ of the mixture, we do not consider the problem of estimating the number m of components.

For estimation, the heterogeneous structure of the likelihood function of the observed sample $\{w_i\}$ has to be considered. Each contribution to this likelihood is a finite mixture of normals. Its individual parameter vector differs from Φ by adding the known error variance σ_i^2 to each component variance σ_k^2 and is given by $\Phi_i = (\alpha_1, \ldots, \alpha_m, \mu_1, \sigma_1^2+\sigma_i^2, \ldots, \mu_m, \sigma_m^2+\sigma_i^2)^{'}$. In the sequel we agree on this notation: the parameter vector to be estimated is $\Phi = (\alpha_1, \ldots, \alpha_m, \psi_1^{'}, \ldots, \psi_m^{'})^{'}$ with $\psi_k = \binom{\mu_k}{\sigma_k^2}$ and the density function for $W_i = w_i$ will be denoted by $p(w_i \mid \Phi) = \sum_{k=1}^{m} \alpha_k p_k(w_i \mid \psi_k)$ where $p_k(w_i \mid \psi_k)$ is the normal density function with parameters μ_k and $\sigma_k^2 + \sigma_i^2$. Under the constraint that the mixing proportions $\{\alpha_k\}$ sum up to one, the number of parameters to be estimated is $3m - 1$.

3 The EM algorithm

The EM algorithm is a widely used approximate method for finding maximum likelihood estimates. The proposed algorithm for a mixture of normals in the presence of heteroscedastic measurement error is an expansion of the EM algorithm as it is suggested in Redner and Walker (1984). The EM algorithm for mixture density estimation problems should, as stated by the authors above, 'best be regarded as a specialization of the general EM algorithm' formalized by Dempster, Laird and Rubin (1977).

For our estimation problem, we have to incorporate an 'incomplete' data structure to make use of the algorithm. We regard our sample $\{w_i\}$ as a sample of 'incomplete'

data, where w_i has to be considered as the known part of a 'complete' observation $y_i = (w_i, z_{i1}, \ldots, z_{im})$ referring to the sample variables $Y_i = (W_i, Z_{i1}, \ldots, Z_{im})$, where

$$Z_{ik} = \begin{cases} 1 & \text{if } W_i \text{ is from } N(\mu_k, \sigma_k^2 + \sigma_i^2), \\ 0 & \text{else.} \end{cases} \tag{3}$$

The density function of the 'complete' data is therefore given by

$$f\,(y \mid \Phi) = \prod_{i=1}^{n} \prod_{k=1}^{m} \alpha_k^{z_{ik}} p_k(w_i \mid \psi_k)^{z_{ik}},$$

whereas for the 'incomplete' data it is

$$g\,(w \mid \Phi) = \prod_{i=1}^{n} \sum_{k=1}^{m} \alpha_k p_k(w_i \mid \psi_k).$$

The purpose of the EM algorithms is to maximize for a given sample S of W the 'incomplete' loglikelihood function $L(\Phi) = \log\,(g(W \mid \Phi))$ with respect to Φ. With $k\,(Y \mid W, \Phi)$ we will denote the conditional density of Y given (W, Φ) and write the loglikelihood function as

$$L(\Phi) = \log f\,(Y \mid \Phi) - \log k\,(Y \mid W, \Phi).$$

As described in Dempster et al. (1977) the loglikelihood for Φ can be decomposed for a known Φ^k into

$$\begin{aligned} L(\Phi) &= \mathrm{E}(\log f\,(Y \mid \Phi) \mid W, \Phi^k) - \mathrm{E}(\log k\,(Y \mid W, \Phi) \mid W, \Phi^k). \\ &= Q(\Phi \mid \Phi^k) - H(\Phi \mid \Phi^k) \end{aligned}$$

The EM algorithm is of an iterative nature and for a current approximation Φ^c of a maximizer of $L(\Phi)$ the next approximation Φ^n is obtained through two steps:

The E step: Determine the function $Q(\Phi \mid \Phi^c)$.

The M step: Choose the next approximation Φ^n as the set of values that maximizes $Q(\Phi \mid \Phi^c)$ with respect to Φ.

For $\Phi^c = (\alpha_1^c, \ldots, \alpha_m^c, \psi_1^c, \ldots, \psi_m^c)$ the conditional expectation of $\log f(Y \mid \Phi)$ is found by

$$
\begin{aligned}
Q(\Phi \mid \Phi^c) &= \mathrm{E}\left(\log f\left(Y \mid \Phi\right) \mid W, \Phi^c\right) \\
&= \mathrm{E}\left(\log \prod_{i=1}^{n}\prod_{k=1}^{m} \alpha_k^{Z_{ik}} p_k(W_i \mid \psi_k)^{Z_{ik}} \mid W, \Phi^c\right) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{m} \mathrm{E}\left(Z_{ik}\log\alpha_k + Z_{ik}\log p_k(W_i \mid \psi_k) \mid W, \Phi^c\right) \\
&= \sum_{k=1}^{m}\sum_{i=1}^{n} \frac{\alpha_k^c\, p_k(W_i \mid \psi_k^c)}{p(W_i \mid \Phi^c)} \log\alpha_k \\
&\quad + \sum_{k=1}^{m}\sum_{i=1}^{n} \log p_k(W_i \mid \psi_k)\, \frac{\alpha_k^c\, p_k(W_i \mid \psi_k^c)}{p(W_i \mid \Phi^c)}. \qquad (4)
\end{aligned}
$$

Note that $\mathrm{E}(Z_{ik} \mid W_i, \Phi^c) = P(Z_{ik} = 1 \mid W_i, \Phi^c) = \frac{\alpha_k^c\, p_k(W_i|\psi_k^c)}{p(W_i|\Phi^c)}$ is the posterior probability that W_i belongs to component k given W_i and the current knowledge about ψ_k. Having derived a functional form for $Q(\Phi \mid \Phi^c)$ it can be shown that the maximization problem in the M-step consists of two parts which will be treated separately. The first one involves only the proportions $\alpha_1, \ldots, \alpha_m$ and yields an unique solution. For the following we will assume that from the precedent step the algorithm provides us with a current approximation Φ^c. The maximization

$$
\sum_{k=1}^{m}\sum_{i=1}^{n} \frac{\alpha_k^c\, p_k(W_i \mid \psi_k^c)}{p(W_i \mid \Phi^c)} \cdot \log\alpha_k \longrightarrow \max_{\alpha=(\alpha_1,\ldots,\alpha_m)}
$$

has to be solved under the restriction that $\sum_{k=1}^{m} \alpha_k = 1$. This can be done easily with the help of a Lagrange multiplier and the next approximizer $\alpha^n = (\alpha_1^n, \ldots, \alpha_m^n)$ prescribed by the M-Step of the algorithm satisfies

$$
\alpha_k^n = \frac{1}{n}\sum_{i=1}^{n} \frac{\alpha_k^c\, p_k(W_i \mid \psi_k^c)}{p(W_i \mid \Phi^c)} \qquad \text{for } k = 1, \ldots, m.
$$

Notice that given Φ^c the new proportions α_k^n can be computed directly.

The second part of the M-Step involves the remaining parameters $\psi_1, \ldots, \psi_m$ and can be separated further into m component problems, each referring to ψ_k. We can think of this as a 'weighted' maximum likelihood estimation with sums of logarithms weighted by posterior probabilities. In fact, for each component k we want to solve

$$
\sum_{i=1}^{n} \log p_k(W_i \mid \psi_k) \cdot \frac{\alpha_k^c\, p_k(W_i \mid \psi_k^c)}{p(W_i \mid \Phi^c)} \longrightarrow max_{\psi_k}.
$$

The weights $\frac{\alpha_k^c \, p_k(W_i|\psi_k^c)}{p(W_i|\Phi^c)}$ will be denoted by $w_{i,k}^c$ in the sequel. If we write the second term of the expectation given in (4) as $Q_2(\Phi \mid \Phi^c) = \sum_{k=1}^m q_k(\mu_k, \sigma_k)$ with

$$q_k(\mu_k, \sigma_k) = \sum_{i=1}^n \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_k^2 + \sigma_i^2) - \frac{1}{2}\frac{(W_i - \mu_k)^2}{\sigma_k^2 + \sigma_i^2} \right) \cdot w_{i,k}^c \longrightarrow max_{\mu_k, \sigma_k}$$

and take the partial derivatives for μ_k and σ_k, we want to solve the equation

$$f_k = \begin{pmatrix} \frac{\partial q_k}{\partial \mu_k} \\ \\ \frac{\partial q_k}{\partial \sigma_k} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \frac{W_i - \mu_k}{\sigma_k^2 + \sigma_i^2} \cdot w_{i,k}^c \\ \\ \sum_{i=1}^n \left(-\frac{\sigma_k}{\sigma_k^2 + \sigma_i^2} + \frac{\sigma_k (W_i - \mu_k)^2}{(\sigma_k^2 + \sigma_i^2)^2} \right) \cdot w_{i,k}^c \end{pmatrix} = \begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}.$$

A closed form solution of the equation can only be found in the homogeneous case, that is $\sigma_i^2 = \sigma_u^2$ for $i = 1, \ldots, n$ and is given in Appendix A. For an heteroscedastic error variance structure we suggest to use a Newton algorithm to derive approximations for the maxima μ_k^n and σ_k^n instead. This requires an iteration within each step of the EM algorithm where in addition to f_k the Jacobian matrix $J_{f_k}(\mu_k, \sigma_k)$ of the second derivatives of $q_k(\mu_k, \sigma_k)$ has to be computed (for its elements, see also Appendix A). In the following we give a formal description of an EM algorithm for a heteroscedastic measurement error model.

Initialization:

Run an EM algorithm with the data of the sample $\{w_i\}$ under the assumption that we have no measurement error. We can use the explicit formulas of the homoscedastic error model as given in Appendix A and set $\sigma_u^2 = 0$. As a result we obtain the initial values $\alpha_k^{(0)}$ $\mu_k^{(0)}$ and $\sigma_k^{(0)}$ for each component k of the mixture.

For $r = 0, 1, 2, \ldots$ the $r+1$-th step of the algorithm is given by:

E Step:

Determine the function $Q(\Phi \mid \Phi^r)$, where Φ^r is the current parameter vector obtained from the r-th step.

M Step:

for each component k compute the k-th proportion as

$$\alpha_k^{r+1} = \frac{1}{n} \sum_{i=1}^n \frac{\alpha_k^r \, p(w_i \mid \psi_k^r)}{p(w_i \mid \psi^r)}$$

and get $\mu_k^{r+1}, \sigma_k^{r+1}$ as the final result of the Newton approximation

$$\begin{pmatrix} \mu_k^{r,i+1} \\ \\ \sigma_k^{r,i+1} \end{pmatrix} = \begin{pmatrix} \mu_k^{r,i} \\ \\ \sigma_k^{r,i} \end{pmatrix} - J_{f_k}^{-1}(\sigma_k^{r,i}, \mu_k^{r,i}) \cdot f_k^{r,i}, \tag{5}$$

where k denotes the $k-th$ component of the mixture , r is the precedent step of the EM algorithm and i is the number of the cycle of the Newton iteration. If convergence occurs for $i = i^*$ then $\mu_k^{r+1} = \mu_k^{r,i^*}$ and $\sigma_k^{r+1} = \sigma_k^{r,i^*}$.

It seems reasonable to give more weight on the overall convergence criterion of the EM algorithm than for the chosen criterion for the Newton approximation in the M-Step since the approximated values of the parameters are approximations themselves. We find it useful to apply the Newton approximation only for a predetermined number of cycles to increase the convergence rate of the EM algorithm. The performance of the algorithm strongly depends on the initial values used in its first step. Even the simplified EM algorithm, which is only used to generate starting values for the one considering the heteroscedastic error structure, requires to input some starting values for the parameters. This values can be taken more or less arbitrarily using purely descriptive methods.

4 Observed Information

Unfortunately the EM algorithm does not provide us with the mean of estimating the information matrix associated with the parameter estimates. Louis (1982) derived a procedure to compute the observed information matrix of the approximate MLE if an additional analysis is applied using the results of the algorithm. We apply this method directly to get the standard errors of the estimated distribution parameters. Therefore we will not give a detailed description of it but stress some important features of the analysis. Let $l_Y(\Phi, Y) = \log f(Y \mid \Phi)$ and $l_W(\Phi, W) = \log g(W \mid \Phi)$ denote the loglikelihood functions of the 'complete' and 'incomplete' data. $S_Y(\Phi, Y)$ and $S_W(\Phi, W)$ are the gradient vectors of l_Y and l_W and $B_Y(\Phi, Y)$ and $B_W(\Phi, W)$ denote the negatives of the associated second derivative matrices. The observed

information matrix for W is defined by $I_W(\Phi) = B_W(\Phi, W)$ with

$$\begin{aligned} B_W(\Phi, W)) &= \mathrm{E}_\Phi(B_Y(\Phi, Y) \mid W) - \mathrm{E}_\Phi(S_Y(\Phi, Y) S_Y'(\Phi, Y) \mid W) \\ &\quad + \mathrm{E}_\Phi(S_Y(\Phi, Y) \mid W)\, \mathrm{E}_\Phi'(S_Y(\Phi, Y) \mid W). \end{aligned}$$

If $\hat{\Phi}$ is the MLE found by the EM algorithm and if we further assume that the $Y_1, \ldots, Y_n$ are independent, the observed information can be computed as

$$\begin{aligned} \hat{I}_W(\hat{\Phi}) &= \sum_{i=1}^{n} \mathrm{E}_{\hat{\Phi}}(B_{Y_i}(\hat{\Phi}, Y_i) \mid W_i) - \sum_{i=1}^{n} \mathrm{E}_{\hat{\Phi}} \left(S_{Y_i}(\hat{\Phi}, Y_i) S_{Y_i}'(\hat{\Phi}, Y_i) \mid W_i \right) \\ &\quad - \sum_{\substack{i,j \\ i \neq j}} \mathrm{E}_{\hat{\Phi}}(S_{Y_i}(\hat{\Phi}, Y_i) \mid W_i)\, \mathrm{E}_{\hat{\Phi}}'(S_{Y_j}(\hat{\Phi}, Y_j) \mid W_j). \end{aligned}$$

All these conditional expectations can be computed after the last cycle of the algorithm and require lengthy but straightforward differentiations. If we notice that $Z_{ik} Z_{ij} = 0$ for $k \neq j$ and that the expectations $\mathrm{E}_{\hat{\Phi}}(Z_{ik} \mid W_i) = \frac{\hat{\alpha}_k p_k(W_i, \hat{\psi}_k)}{\sum_{j=1}^{m} \hat{\alpha}_j p_j(W_i, \hat{\psi}_j)}$ equal the estimated weights $\hat{w}_{ik}$, the programming of $\hat{I}_W(\hat{\Phi})$ can be done easily.

5 Simulation and Example

Simulation was carried out for a two and for a three components mixture model. In both cases observations of a random variable X following a normal mixture distribution were drawn. Then independently simulated heteroscedastic measurement errors U_i were added to get the observations of the sample variables W_i. The U_i's were each drawn from a normal distribution with zero mean and variance σ_i^2, where σ_i^2 itself was uniformly distributed over the interval $[0, c]$. Only the sample $\{w_i\}$ was used for estimation. For different sample sizes and different values of c 1000 replications of each experiment were run. In tables 1 and 2 we present the results of both models for a measurement error of medium size, which is given for the simulated mixtures at a value for $c = 0.3$. The results of small ($c = 0.1$) and large ($c = 0.5$) measurement errors are given in the Appendix B.

For each parameter of the mixture distribution we calculated the average of the parameter estimates over the number of replications (avg. est.). In each experiment the

observed information matrix of the parameter estimates was calculated and finally the mean of the estimated variances of the parameter estimates (avg. $\hat{\sigma}^2_{est.}$) was taken. This value can be compared to the sample variance of the estimates (S^2_{est}). With the help of the estimates and its estimated standard errors, we constructed confidence intervals for $1-\alpha = 0.5, 0.9$ and 0.95 and computed the frequency how often the true parameter values fell into this intervals.

For the two components mixture model the true parameters are $\alpha = 0.7, \mu_1 = 0, \mu_2 = 5$, and $\sigma_1 = \sigma_2 = 1$. For all parameters the average estimates show satisfactory results and their precision increases with the sample size, but it is obvious that the estimates of the standard deviations σ_k do not perform as well as the parameter estimates for the means and proportions. As expected the mean estimated variance of the parameters are getting closer to its sample variance if n increases. This also holds for the three components model, where we have $\alpha_1 = 0.7, \alpha_2 = 0.3, \alpha_3 = 0.1, \mu_1 = 0, \mu_2 = 5, \mu_3 = 10$ and $\sigma_1 = \sigma_2 = \sigma_3 = 1$. As in the two components model the $\hat{\sigma}_k$'s do not show the same satisfactory results as the other parameter estimates, which is also reflected in the produced coverage rates for their confidence intervals. They clearly show deviations from the expected rates and if we would test for the unknown rate on a 95 % confidence level we would have to reject the null hypothesis for almost all of them in the case of a medium sample size of $n = 100$.

The convergence of the algorithm depends on the structure of the data. If there are clearly distinct components, the algorithm performs well even for small sample sizes. In other cases where the data show almost an unimodal structure, difficulties arise due to the disability of the algorithm to identify different components of a mixture and it seems not worthwhile to further investigate such ill conditioned problems.

The main purpose of this simulation study was to see if the EM algorithm can be used to handle this sort of data, where, in addition to the task of estimating distribution parameters of a finite mixture, the data can only be observed with an individual measurement error. The results obtained are promising and we will finally give an empirical example, where all these difficulties can be found.

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha = 0.7$	$n = 50$	0.6947	0.0049	0.0044	0.542	0.899	0.953
	$n = 100$	0.6974	0.0023	0.0023	0.506	0.887	0.937
	$n = 500$	0.6997	0.0005	0.0004	0.524	0.894	0.947
$\mu_1 = 0$	$n = 50$	0.1216	0.0391	0.0220	0.543	0.897	0.945
	$n = 100$	-0.0046	0.0188	0.0193	0.488	0.896	0.939
	$n = 500$	0.0002	0.0037	0.0038	0.509	0.900	0.949
$\mu_2 = 5$	$n = 50$	5.0053	0.1225	0.1157	0.466	0.876	0.933
	$n = 100$	4.9894	0.0525	0.0489	0.506	0.891	0.932
	$n = 500$	4.9991	0.0096	0.0093	0.490	0.908	0.953
$\sigma_1 = 1$	$n = 50$	0.9621	0.0263	0.0246	0.492	0.861	0.912
	$n = 100$	0.9833	0.0126	0.0135	0.494	0.877	0.920
	$n = 500$	0.9963	0.0024	0.0027	0.504	0.879	0.934
$\sigma_2 = 1$	$n = 50$	0.9460	0.0775	0.0750	0.484	0.825	0.879
	$n = 100$	0.9782	0.0356	0.0364	0.502	0.864	0.913
	$n = 500$	0.9938	0.0065	0.0066	0.484	0.892	0.943

Table 1. Simulation results for the two components mixture model with heteroscedastic measurement error of medium size (c=0.3). For each sample size n 1000 replications of the experiment were conducted.

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha_1 = 0.7$	$n = 100$	0.6941	0.0024	0.0024	0.489	0.894	0.956
	$n = 500$	0.6991	0.0004	0.0004	0.537	0.923	0.960
$\alpha_2 = 0.2$	$n = 100$	0.2024	0.0021	0.0020	0.493	0.886	0.932
	$n = 500$	0.1991	0.0004	0.0004	0.517	0.890	0.949
$\alpha_3 = 0.1$	$n = 100$	0.1034	0.0011	0.0010	0.501	0.889	0.947
	$n = 500$	0.1018	0.0002	0.0002	0.502	0.902	0.957
$\mu_1 = 0$	$n = 100$	0.1045	0.0181	0.0099	0.467	0.870	0.924
	$n = 500$	0.0696	0.0035	0.0016	0.241	0.777	0.883
$\mu_2 = 5$	$n = 100$	4.9879	0.0990	0.1109	0.447	0.837	0.902
	$n = 500$	4.9978	0.0159	0.0151	0.494	0.910	0.954
$\mu_3 = 10$	$n = 100$	9.9863	0.2153	0.2031	0.457	0.858	0.908
	$n = 500$	10.0106	0.0320	0.0340	0.476	0.874	0.928
$\sigma_1 = 1$	$n = 100$	0.9632	0.0120	0.0129	0.461	0.852	0.918
	$n = 500$	0.9814	0.0023	0.0021	0.490	0.884	0.943
$\sigma_2 = 1$	$n = 100$	0.9462	0.0918	0.1016	0.430	0.812	0.855
	$n = 500$	0.9797	0.0150	0.0159	0.469	0.878	0.933
$\sigma_3 = 1$	$n = 100$	0.9570	0.1265	0.1245	0.455	0.823	0.868
	$n = 500$	1.0171	0.0216	0.0213	0.490	0.899	0.947

Table 2. Simulation results for the three components mixture model with heteroscedastic measurement error of medium size (c=0.3). For each sample size n 1000 replications of the experiment were conducted.

The Radon Problem

In 1992 a Swiss study on the effect of radon on the occurrence of lung cancer cases was carried out and some of the results can be found in Minder and Völkle (1995). We will apply our model to the radon data of this study to give an example of the use of an heteroscedastic measurement error model.

Most researchers who want to obtain reliable data on radon, will agree that this is a difficult matter. This arises partly from the nature of radon itself and partly from the various environmental sources of influences on the measurement process. First of all, the amount of radon strongly depends on local geological conditions and second, once in the air, it decomposes into other substances like polonium, lead and wismut. Indoor measurements are affected by the building structure and the constructing material of the place as well as by the amount of ventilation.

In this study radon averages from 46 different Swiss regions were observed. In each region n_i measurements of radon were taken from different locations to obtain the regional averages $W_i = n_i^{-1} \sum_j W_{ij}$. Due to all the difficulties described above, the single radon measurements observed in region i follow the error model

$$W_{ij} = X_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_i}), \text{ with } i = 1, \ldots, 46 \text{ and } j = 1, \ldots n_i,$$

where W_{ij} is the j-th observed value in region i and X_i is the true regional mean. Therefore the observed means are deviations from a existing true mean, that is

$$W_i = X_i + U_i, \quad U_i \sim N(0, \sigma_i^2) \text{ with } i = 1, \ldots, 46.$$

The heteroscedastic error variances are given by $\sigma_i^2 = \sigma^2_{\epsilon_i}/n_i$ and even if the $\sigma^2_{\epsilon_i}$'s are equal for all regions, heteroscedasticity in the errors U_i is caused by the different number of observations. In the study, n_i varies from 16 to 511 and in addition the sample variances S_i^2 of the n_i measurements are given for each region. The estimated error variances $\hat{\sigma}_i^2 = S_i^2/n_i$ range from 0.769 to 426.983 and those values will serve as the error variances σ_i^2, which we earlier assumed to be known.

Figure 1 shows a histogram of the 46 mean radon measurements. A kernel estimator is drawn into the picture to illustrate that the assumption of a mixture distribution for the true average seems reasonable.

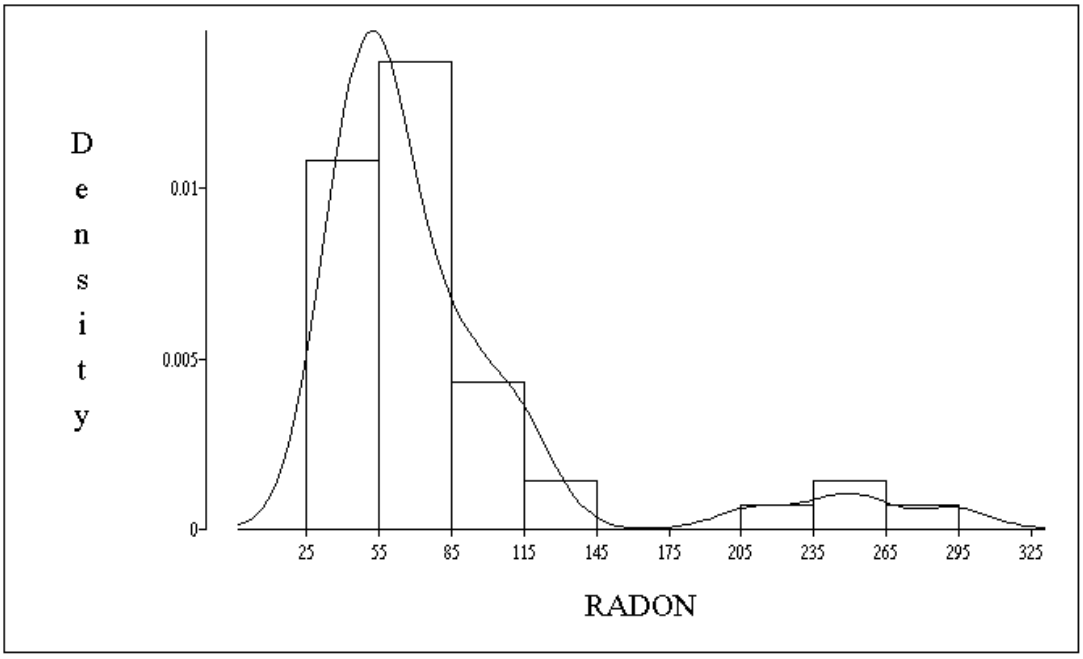

Figure 1. Histogram of the mean Radon measurements of 46 different regions in Switzerland. The solid line represents the estimated Gaussian kernel density.

Usually radon measurements are assumed to be lognormally distributed, but this mainly holds for data coming from a homogeneous stratum. As our data are collected from all over Switzerland, a mixture distribution makes more allowance for regional differences for the occurrence of radon. In Figure 2 we plotted the observed radon averages against their standard errors so that the heteroscedastic structure of the errors can be seen. We fixed the number of components to be three, well aware that this will cause large standard errors for the third component, which will only be identified by four data points. But three of those are neighboring regions, so their means are coherent and can be regarded as a cluster. Parameter estimation was carried out via application of the EM algorithm and its results are given in Table 3. In view of the descriptive plot in Figure 1 the obtained estimation results are not surprising. Their large standard errors are mainly due to the fact that only 46 observations

were available for estimation. It would be interesting to have a larger data base to test the model.

Swiss Radon data						
Estimate for	Component 1		Component 2		Component 3	
proportions α_k	0.6225	(0.0815)	0.2905	(0.0776)	0.0870	(0.0416)
means μ_k	49.7573	(1.8943)	94.2082	(6.4696)	254.5958	(13.9715)
stand. dev. σ_k	8.4732	(1.4605)	15.5282	(5.3460)	25.3031	(11.4102)

Table 3. Estimation results for Swiss radon data fitting a mixture of three normal distributions to the observed mean values. The standard errors of the parameter estimates are given in brackets.

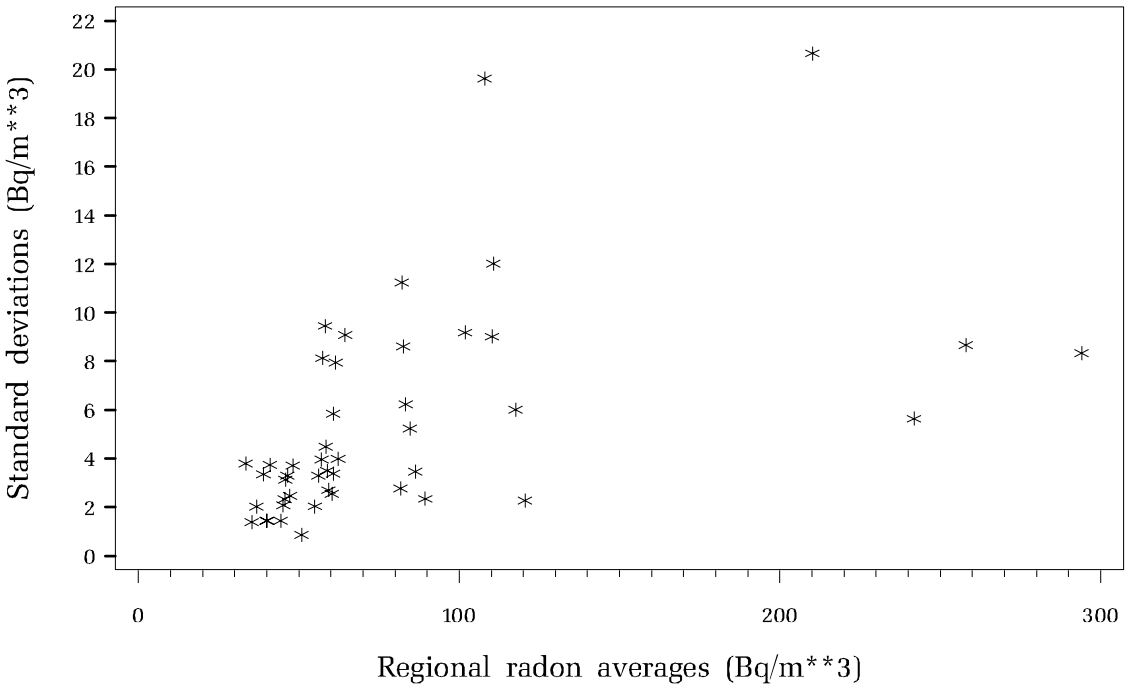

Figure 2. Scatterplot of the regional radon measurements against the estimated standard deviations of their measurement errors.

6 Discussion

Within the framework of the errors-in-variables models, likelihood based approaches to fit regression models are very attractive. As soon as a regressor variable is assumed to be stochastic, its distribution plays an important role in the analysis of such models. A first step is to specify a model for the distribution of the incorrectly observed variable. Finite mixture distributions can add considerable information when the observed variables do not come from a homogeneous population and therefore we should be able to estimate its parameters even in the presence of a measurement error. The main task is a computational one, but as long as the means of the components in relation to their variances, are not too close to each other, we made good experiences with the proposed algorithm. [1]

Acknowledgement

I want to thank Dr.Ch.E. Minder for discussion and providing the data and Ekkehard Kessner for his help with the programming.

[1] A macro of the algorithm written in the GAUSS language can be obtained by the author.

Appendix A:

For a homoscedastic measurement error model, that is $U_i \sim N(0, \sigma_u^2)$ for $i = 1, \ldots, n$, the M-Step of the algorithm $l_k(\mu_k, \sigma_k) \longrightarrow max_{\mu_k, \sigma_k}$ yields unique solutions for each component k and the updated parameters for the next step are given by

$$\begin{aligned} \mu_k &= \frac{\sum_{i=1}^n W_i \cdot w_{i,k}^c}{\sum_{i=1}^n w_{i,k}^c} \quad \text{and} \\ \sigma_k^2 + \sigma_u^2 &= \frac{\sum_{i=1}^n (W_i - \mu_k^n)^2 \cdot w_{i,k}^c}{\sum_{i=1}^n w_{i,k}^c}, \end{aligned}$$

where the weights $w_{i,k}^c$ are given as stated above. It is worth noting that in this case the algorithm provides us in each step with an approximate estimation of the sum of the component variance and the error variance.

The Jacobian matrix $J_{f_k}(\mu_k, \sigma_k)$ of the second derivatives of $q_k(\mu_k, \sigma_k)$ used in the Newton approximation of the M Step of the algorithm is given by

$$J_{f_k}(\mu_k, \sigma_k) = \begin{pmatrix} \frac{\partial^2 q_k}{\partial \mu_k \partial \mu_k} & \frac{\partial^2 q_k}{\partial \sigma_k \partial \mu_k} \\ \frac{\partial^2 q_k}{\partial \mu_k \partial \sigma_k} & \frac{\partial^2 q_k}{\partial \sigma_k \partial \sigma_k} \end{pmatrix}.$$

Its elements are computed as

$$\begin{aligned} \frac{\partial^2 q_k}{\partial \mu_k \partial \mu_k} &= -\sum_{i=1}^n \frac{w_{i,k}^c}{\sigma_k^2 + \sigma_i^2}, \\ \frac{\partial^2 q_k}{\partial \sigma_k \partial \mu_k} &= -2 \sum_{i=1}^n \frac{\sigma_k (W_i - \mu_k)}{(\sigma_k^2 + \sigma_i^2)^2} \cdot w_{i,k}^c, \\ \frac{\partial^2 q_k}{\partial \sigma_k \partial \sigma_k} &= \sum_{i=1}^n \frac{w_{i,k}^c}{(\sigma_k^2 + \sigma_i^2)^3} \left(\sigma_k^4 - \sigma_i^4 + (\sigma_i^2 - 3\sigma_k^2)(W_i - \mu_k)^2 \right). \end{aligned}$$

Appendix B:

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha = 0.7$	$n = 50$	0.6953	0.0047	0.0044	0.530	0.897	0.940
	$n = 100$	0.6990	0.0022	0.0023	0.486	0.885	0.942
	$n = 500$	0.6989	0.0004	0.0004	0.496	0.889	0.950
$\mu_1 = 0$	$n = 50$	0.1170	0.0334	0.0177	0.535	0.888	0.934
	$n = 100$	-0.0005	0.0167	0.0173	0.484	0.890	0.942
	$n = 500$	-0.0013	0.0033	0.0035	0.478	0.886	0.945
$\mu_2 = 5$	$n = 50$	4.9870	0.1024	0.1146	0.465	0.854	0.912
	$n = 100$	4.9933	0.0437	0.0444	0.487	0.858	0.931
	$n = 500$	4.9956	0.0083	0.0089	0.507	0.883	0.945
$\sigma_1 = 1$	$n = 50$	0.9589	0.0203	0.0208	0.477	0.843	0.901
	$n = 100$	0.9910	0.0101	0.0104	0.486	0.884	0.938
	$n = 500$	0.9983	0.0020	0.0019	0.524	0.909	0.957
$\sigma_2 = 1$	$n = 50$	0.9604	0.0604	0.0646	0.464	0.815	0.871
	$n = 100$	0.9684	0.0273	0.0288	0.453	0.856	0.896
	$n = 500$	0.9961	0.0052	0.0052	0.493	0.889	0.939

Table 4. Simulation results for the two components mixture model with heteroscedastic measurement error of small size (c=0.1). For each sample size n 1000 replications of the experiment were conducted.

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha = 0.7$	$n = 50$	0.6974	0.0052	0.0049	0.536	0.876	0.938
	$n = 100$	0.6963	0.0025	0.0024	0.514	0.902	0.946
	$n = 500$	0.6987	0.0005	0.0004	0.507	0.906	0.957
$\mu_1 = 0$	$n = 50$	0.1310	0.0428	0.0226	0.516	0.906	0.947
	$n = 100$	0.1000	0.0210	0.0124	0.486	0.876	0.939
	$n = 500$	-0.0029	0.0041	0.0043	0.484	0.890	0.947
$\mu_2 = 5$	$n = 50$	5.0042	0.1590	0.1380	0.481	0.867	0.919
	$n = 100$	4.9964	0.0688	0.0736	0.488	0.885	0.940
	$n = 500$	5.0017	0.0109	0.0115	0.475	0.884	0.939
$\sigma_1 = 1$	$n = 50$	0.9466	0.0303	0.0301	0.488	0.854	0.907
	$n = 100$	0.9648	0.0151	0.0144	0.498	0.877	0.921
	$n = 500$	0.9973	0.0029	0.0029	0.508	0.897	0.945
$\sigma_2 = 1$	$n = 50$	0.9694	0.1014	0.0973	0.462	0.847	0.898
	$n = 100$	1.0001	0.0476	0.0527	0.467	0.879	0.928
	$n = 500$	0.9972	0.0079	0.0081	0.523	0.894	0.943

Table 5. Simulation results for the two components mixture model with heteroscedastic measurement error of large size (c=0.5). For each sample size n 1000 replications of the experiment were conducted.

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha_1 = 0.7$	$n = 100$	0.6962	0.0024	0.0022	0.487	0.901	0.962
	$n = 500$	0.6990	0.0004	0.0004	0.484	0.902	0.945
$\alpha_2 = 0.2$	$n = 100$	0.2001	0.0020	0.0018	0.489	0.895	0.947
	$n = 500$	0.1999	0.0003	0.0003	0.499	0.901	0.947
$\alpha_3 = 0.1$	$n = 100$	0.1037	0.0010	0.0010	0.476	0.882	0.938
	$n = 500$	0.1012	0.0002	0.0002	0.491	0.897	0.949
$\mu_1 = 0$	$n = 100$	0.0961	0.0167	0.0080	0.495	0.902	0.947
	$n = 500$	0.0639	0.0032	0.0015	0.256	0.799	0.892
$\mu_2 = 5$	$n = 100$	4.9911	0.1030	0.0951	0.457	0.849	0.907
	$n = 500$	5.0009	0.0134	0.0126	0.502	0.909	0.950
$\mu_3 = 10$	$n = 100$	9.9961	0.1671	0.2322	0.448	0.833	0.894
	$n = 500$	10.0083	0.0260	0.0261	0.484	0.897	0.943
$\sigma_1 = 1$	$n = 100$	0.9797	0.0100	0.0097	0.512	0.864	0.918
	$n = 500$	0.9878	0.0019	0.0020	0.478	0.871	0.915
$\sigma_2 = 1$	$n = 100$	0.9626	0.0806	0.0861	0.441	0.783	0.836
	$n = 500$	0.9830	0.0109	0.0108	0.492	0.871	0.921
$\sigma_3 = 1$	$n = 100$	0.9461	0.0936	0.1070	0.465	0.791	0.840
	$n = 500$	0.9998	0.0162	0.0157	0.500	0.904	0.944

Table 6. Simulation results for the three components mixture model with heteroscedastic measurement error of small size (c=0.1). For each sample size n 1000 replications of the experiment were conducted.

parameter	sample	avg. est.	avg. $\hat{\sigma}^2_{est}$	S^2_{est}	$KI_{0.5}$	$KI_{0.9}$	$KI_{0.95}$
$\alpha_1 = 0.7$	$n = 100$	0.6955	0.0026	0.0025	0.500	0.894	0.950
	$n = 500$	0.6990	0.0005	0.0004	0.508	0.911	0.961
$\alpha_2 = 0.2$	$n = 100$	0.2010	0.0023	0.0023	0.489	0.877	0.943
	$n = 500$	0.2002	0.0004	0.0004	0.494	0.907	0.949
$\alpha_3 = 0.1$	$n = 100$	0.1035	0.0011	0.0010	0.522	0.901	0.942
	$n = 500$	0.1008	0.0002	0.0002	0.476	0.907	0.954
$\mu_1 = 0$	$n = 100$	0.1116	0.0202	0.0098	0.454	0.879	0.936
	$n = 500$	0.0774	0.0040	0.0019	0.209	0.751	0.854
$\mu_2 = 5$	$n = 100$	4.9638	0.1366	0.1296	0.447	0.845	0.900
	$n = 500$	4.9929	0.0187	0.0184	0.503	0.899	0.946
$\mu_3 = 10$	$n = 100$	10.0038	0.2436	0.2583	0.440	0.808	0.871
	$n = 500$	10.0107	0.0392	0.0391	0.465	0.889	0.944
$\sigma_1 = 1$	$n = 100$	0.9507	0.0143	0.0135	0.484	0.858	0.911
	$n = 500$	0.9807	0.0028	0.0027	0.456	0.881	0.934
$\sigma_2 = 1$	$n = 100$	0.9400	0.1221	0.1387	0.444	0.767	0.838
	$n = 500$	0.9771	0.0202	0.0223	0.483	0.882	0.920
$\sigma_3 = 1$	$n = 100$	0.9571	0.1516	0.1395	0.472	0.838	0.893
	$n = 500$	1.0302	0.0278	0.0253	0.503	0.928	0.963

Table 7. Simulation results for the three components mixture model with heteroscedastic measurement error of large size (c=0.5). For each sample size n 1000 replications of the experiment were conducted.

7 References

Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society.* **B 39**, 1-38.

Küchenhoff, H. (1995). *Schätzmethoden in mehrphasigen Regressionsmodellen.* Habilitationsschrift. Institute of Statistics, University of Munich.

Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society.* **B 44**, 226-233.

Minder, Ch. E. and Völkle, H. (1995). Radon und Lungenkrebssterblichkeit in der Schweiz. *324. Bericht der mathematisch-statistischen Sektion der Forschungsgesellschaft Johanneum*, 115-124.

Redner, A.R. and Walker, H.F. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review.* Vol.26, 195-240.